DISCUSSION PAPER SERIES

IZA DP No. 6480

The Effects of Prize Spread and Noise in Elimination Tournaments: A Natural Field Experiment

Josse Delfgaauw
Robert Dur
Arjan Non
Willem Verbeke

April 2012

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

The Effects of Prize Spread and Noise in Elimination Tournaments: A Natural Field Experiment

Josse Delfgaauw
Erasmus University Rotterdam and Tinbergen Institute

Robert Dur
Erasmus University Rotterdam, Tinbergen Institute, CESifo and IZA

Arjan Non
Erasmus University Rotterdam and Tinbergen Institute

Willem Verbeke
Erasmus University Rotterdam and ERIM

Discussion Paper No. 6480
April 2012

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 6480
April 2012

ABSTRACT

The Effects of Prize Spread and Noise in Elimination Tournaments: A Natural Field Experiment*

We conduct a natural field experiment in a large retail chain to test basic predictions of tournament theory regarding prize spread and noise. A random subset of the 208 stores participates in two-stage elimination tournaments. Tournaments differ in the distribution of prize money across winners of the first and second round of the tournament. As predicted by theory, we find that a more convex prize spread increases performance in the second round at the expense of first-round performance, although the magnitude of these effects is small. Moreover, the treatment effect is significantly larger for stores that historically have relatively stable performance as compared to stores with more noisy performance.

JEL Classification: C93, M51, M52

Keywords: elimination tournaments, incentives, prize spread, performance measurement, field experiment

Corresponding author:

Robert Dur
Erasmus University
Erasmus School of Economics/H 8-15
P.O. Box 1738
3000 DR Rotterdam
The Netherlands
E-mail: dur@ese.eur.nl

* We gratefully acknowledge comments and suggestions by Steffen Altmann, Jordi Blanes-i-Vidal, Jan Brinkhuis, Bas Donkers, Alex Koning, Hessel Oosterbeek, John van Reenen, Matthias Wibral, seminar attendants at the London School of Economics and participants to the 2011 Advances with Field Experiments conference at the University of Chicago.

1 Introduction

Tournament theory is a cornerstone of incentive theory in organizations. Pioneered by Lazear and Rosen (1981), Green and Stokey (1983), Nalebuff and Stiglitz (1983), and Rosen (1986), tournament theory can explain many prominent organizational features. Examples include large wage increases upon promotions (as found by e.g. Murphy 1985, Baker et al. 1994, McCue 1996), a convex wage structure across the levels of the hierarchy (Murphy 1985, Baker et al. 1994, Gibbs 1995), and a positive relation between the prize spread and the number of people competing for a promotion (Eriksson 1999, Bognanno 2001). Waldman (2008) provides an extensive discussion of empirical evidence on tournament theory. Crucially, predictions from tournament theory for organizational architecture follow from employees' presumed responses to tournament incentives.

In this paper, we report the results of a natural field experiment we conducted in a privately-held company. We design elimination tournaments with two rounds that allow us to test several basic hypotheses on employees' behavior as derived from standard tournament theory. First, we vary the distribution of total prize money over the two rounds of the elimination tournament. Theory predicts that a more convex prize structure while keeping total prize money constant (i.e. simultaneously decreasing the prize for winning the first round and increasing the prize for winning the second round) leads to better second-round performance at the expense of first-round performance. Second, we investigate whether the level of noise in contestants' performance affects their performance in the tournament. In theory, noise dilutes incentives to perform, as it reduces the marginal effect of effort on the probability of winning.[1]

To test these hypotheses, we run an elimination tournament among a randomly chosen subset of the 208 stores of a retail chain in the Netherlands selling music, movies, and video games. Both rounds of the tournament last four weeks. Performance in the tournament is measured by the Average number of Products per Customer (APC), a relatively stable

[1] Our design allows for a clean test of the effects of prize spread and noise on employees' incentives to perform well. Tournament theory also generates predictions on the effects of participant heterogeneity (Lazear and Rosen 1981, Rosen 1986), the incentives to sabotage (Lazear 1989), the choice of low-risk versus high-risk strategies (Knoeber and Thurman 1994), and self-selection into tournaments (Lazear and Rosen 1981). See Charness and Kuhn (2010) and Lazear and Oyer (2009) for recent overviews.

and well-known performance measure in this company. In the first round, the 144 participating stores are assigned to groups of four stores that are comparable in terms of historical performance. After the first round, the two worst-performing stores of each group are eliminated, whereas the two best-performing stores of each group win a prize and proceed to the second round. In the second round, stores are once more assigned to groups of four comparable stores. The two best-performing stores of each group again win a prize.

To investigate the relation between prize structure and the incentive effects of the tournament, participating stores are assigned to two different treatments. The treatments differ by the prize spread only, we keep the total amount of prize money constant. In the low-spread treatment, prizes are identical in the two rounds, whereas in the high-spread treatment the second-round prize is four times as large as the first-round prize. For employees in the participating stores, the ex ante expected earnings are about 2 percent of monthly earnings, with prize money ranging from 1.2 percent to 6 percent of monthly earnings.

Our findings are by and large in line with theoretical predictions. First, we find an average treatment effect of the tournament on APC of approximately 1.5 percent. This effect is statistically significant. Second, we find that second-round performance is 1 percentage point higher in the high-spread treatment as compared to the low-spread treatment, while first-round performance is 0.8 percentage point lower. These differences are qualitatively in line with theory, but they are not statistically significant. Third, in the high-spread treatment, the estimated second-round treatment effect is significantly higher than the first-round treatment effect, as predicted by theory. Fourth, while theory predicts a higher first-round treatment effect as compared to the second-round treatment effect in the low-spread treatment, we find the reverse, albeit insignificantly so. As a result, most of the average treatment effect is concentrated in the second round of the tournament.

To test for the effect of noise in measured performance on the effect of the tournament, we use the variance in performance prior to the experimental period as our measure of noise. In the assignment of stores to groups, we take their level of noise into account, so that stores with relatively low (high) noise in performance are matched to other stores with relatively low (high) noise. As predicted by theory, noise has a negative effect on the response to the tournament. This effect is mainly concentrated in the second period. The impact is substantial relative to the average treatment effect: while the stores with least noise experience an estimated treatment effect of about 2.4 percent, the estimated treatment effect is zero for the quartile of stores with highest noise in performance.

The remainder of this paper is organized as follows. The next section discusses related empirical work. The design of the experiment is discussed in Section 3. In Section 4, we analyse a simple elimination tournament model and derive five testable hypotheses regarding the influence of prize structure and noise on performance in the tournament. Section 5 provides summary statistics and Section 6 describes our estimation strategy. In Section 7, we present and discuss our findings as well as several robustness checks, in particular with respect to differences in heterogeneity in stores' historical performance across groups. Section 8 concludes.

2 Previous Studies

Two studies use non-experimental field data to test similar hypotheses from tournament theory. Audas et al. (2004) use the administrative records of a British financial firm to investigate the effects of prize spread and noise in promotion decisions on absenteeism of employees. They find that larger prize spreads, defined as the difference in average earnings between two adjacent layers in the firm's hierarchy, reduce absenteeism. More unexplained variation in promotion decisions increases absenteeism. Based on data from a cross-section of firms, DeVaro (2006) estimates a structural model treating prize spread, performance, and promotions as endogenous. He finds a positive effect of prize spread on workers' performance ratings, a negative effect of noise on performance, and a positive effect of noise on the prize spread. We see our methodology as complementary. By conducting a field experiment rather than analysing actual career paths, we generate exogenous variation in prize spread and obtain a simple measure of noise in performance. This allows for an easy identification of the effects of prize spread and noise on performance in tournaments within an organization.[2]

[2] Several studies test elements of tournament theory in a non-organizational setting. Ehrenberg and Bognanno (1990) show that golf players' performance increases in the effect of improvements in relative positions on prize money. Becker and Huselid (1992) find that race car drivers' performance increases in prize spread. Knoeber and Thurman

Field experiments in organizations are scarce. To our knowledge, this is the first field experiment that studies the effects of an elimination tournament. Field experiments with one-stage tournaments have been conducted by Erev et al. (1993) and Bandiera et al. (2009) among fruit pickers and by Casas-Arce and Martinez-Jerez (2009) and Delfgaauw et al. (2009) among retailers. These studies find a positive effect of tournament incentives on performance, but none of them varies the prize spread. Lim et al. (2009) vary both the number and the distribution of prizes in contests among fundraisers, keeping total prize money constant. They find that contests with multiple identical prizes elicit higher effort as compared to single-prize contests, but differentiating prizes by rank has no further effect on effort. None of these studies looks at the effect of noise in the performance measure.

In terms of design, our paper is closely related to a recent laboratory experiment by Altmann et al. (2012). In a stated-effort setting, they find that subjects choose significantly higher effort in the first stage of a two-stage elimination tournament as compared to a strategically equivalent one-stage tournament. A more convex prize spread in the elimination tournament, obtained by decreasing the prize for winning the first round, does not affect effort in either stage of the tournament, in contrast to theory.[3] Sheremeta (2010) and Höchtl et al. (2011) compare a single-stage contest with a multi-stage contest for the case of a single prize, using contest success functions to determine the winner. Whereas parameters are set such that total effort should be equivalent in the two treatments, both studies find substantial overprovision of effort in the elimination tournament relative to the one-shot contest. Sheremeta (2010) shows that the amount of overprovision relates to the level of effort subjects chose in a contest without a monetary prize, suggesting that some people experience non-monetary utility from winning.

The effects of noise on performance in tournaments are rarely studied in experiments. An exception is Bull et al. (1987), who find in a laboratory experiment with stated-effort that a simultaneous change in noise and marginal cost of effort such that equilibrium effort is unaffected indeed leads to similar levels of effort as chosen by subjects. Given the prevalence of rela-

(1994) find a similar result in competitions among broiler producers.

[3] Several other lab experiments have analysed the effects of prize spread in a single-stage tournament setting, see e.g. Bull et al. (1987), Harbring and Ihrlenbusch (2003), and Freeman and Gelber (2010). Charness and Kuhn (2010) provide a recent overview of these studies.

tive performance incentives and noisy performance measures in real-world settings, our study provides an important test of this part of tournament theory.

3 Experimental Design

The experiment took place in the period September - November 2010 in a retail chain in the Netherlands that sells computer games, music, and movies. In September 2010, the retail chain owned 208 geographically dispersed stores, operating under two different brands. Each store employs on average 5 employees, including a store manager. All strategically important decisions are made by the company's top management. The company's management decides on the range of products sold, pricing, as well as advertisement. New products arrive in stores complete with instructions on how to sell them. Store managers have limited discretion: they are responsible for day-to-day operations. They can primarily boost sales by effective use of the sales force, and by encouraging customers to buy product complements or otherwise related products. Employees receive rather weak incentive pay on top of their base salary. Payments are based on yearly growth in the average number of products and revenues per transaction, the average number of transactions per hour, and a subjective performance evaluation. In addition, store managers have the opportunity to earn a yearly bonus based on reductions in wage costs as a percentage of revenues and on reductions in waste. These incentive schemes remained in place during the experiment.

We set up an elimination tournament for a randomly selected set of stores. The performance measure in the tournament is the Average number of Products sold per Customer (APC). This performance measure can be directly computed from the company's operational database, which records the number of products sold per store per week, and the number of customers (i.e. transactions) per store per week. Everyone in the company is familiar with APC as performance measure. It is part of employees' standard incentive scheme, and stores receive a weekly report on their performance including APC. The reason that APC was chosen as a performance measure and not, for instance, sales is twofold. First, it makes unequivocally clear how stores can enhance performance: through an increase in cross-selling. Second, there is relatively little variation in this measure over time. Figure

1 shows the average APC per week for the period of week 30 in 2009 until week 45 in 2010.

The elimination tournament consisted of 2 rounds, both lasting four weeks with a two-week break in between. The first round ran from week 36 to week 39 in 2010, the second round from week 42 to week 45. In the first round, the participating stores were assigned to groups of four stores. All employees of the two best-performing stores per group, i.e. those with highest APC cumulative over the four weeks in round 1, received a bonus. Moreover, these stores qualified for the second round of the tournament, while all other stores (the bottom-two stores of each group) were eliminated. In the second round, qualified stores were again assigned to groups of four. As before, all employees of the top-two performing stores per group in the second round received a bonus. After round 2, the tournament ended.

We scheduled a two-week gap between the end of the first round and the start of the second round. This period was used to communicate the results of the first round to all treatment stores and to inform the winners of the first round of their second-round assignment. This two-week period is not included in the estimations below, as otherwise a possible response to winning or losing would affect the estimates of the store-fixed effects.

In February 2010, we ran another experiment at the same retail chain, aimed at studying dynamic incentive effects of relative performance pay (the results are reported in Delfgaauw et al. 2010). At that time, a randomly selected set of stores could earn an additional bonus, while the remaining stores were promised a similar opportunity later in the year. Hence, all stores that did not participate in the first experiment (115 stores) do participate in the current elimination tournament. Furthermore, to check whether assignment to treatment or control in the first experiment affects performance in the elimination tournament, we randomly select an additional 29 stores from the stores that comprised the treatment group in the first experiment to participate also in the current tournament. Below, we check whether these 29 stores respond differently to the current treatment as compared to the stores that were part of the control group in the February 2010 experiment. As we find no differences, we can be assured that there are no spill-over effects of the February 2010 experiment. In that experiment, a total of 15 stores were (non-randomly) not allowed to participate in the first experiment, for a variety of reasons. One of these stores was closed during 2010.

Furthermore, 6 new stores were opened during the year. These 20 stores all participate in the current tournament, but since they were non-randomly assigned, they are left out of the analysis. Furthermore, two stores were not allowed to participate in the current experiment and, hence, are also left out of the analysis. This leaves us with 186 stores in the analysis. Of these stores, 62 comprise the control group, while the remaining 124 comprise the treatment group (see also Figure 2).

To study the effect of prize spread on the incentive effect of the tournament, we assign the participating stores to two different treatment groups. The only difference between the two treatments is the prize spread. Thus, we keep total prize money identical across the treatments. In the first round of the low-spread treatment, the bonus for being one of the two best-performing stores in the group is 35 euro gross for a full-time employee. In the second round, the bonus is again 35 euro gross. Hence, per eight stores, employees of two stores win in total 70 euro, employees of two other stores win 35 euro, while the employees of four stores win nothing. In the high-spread treatment, the first-round bonus is 17.50 euro gross. The bonus in the second round is 70 euro. Hence, per eight stores, employees of two stores in the high-spread treatment earn 87.50, employees of two other stores earn 17.50, and four stores receive nothing. Comparing the two treatments, the expected monetary bonus per employee at the start of the tournament is identical in both treatments (26.25 euro), while the prize spread is higher in the high-spread treatment than in the low-spread treatment.[4] All bonuses earned were paid after the tournament ended (in December 2010).

We also examine the effect of noise in performance on the incentive effect of the tournament. We take stores' standard deviation in the performance measure APC over the period August 2009 to August 2010 as our measure of noise. Note that this period does not include the experimental period, so that our measure of noise is not affected by the response to the tournament incentives. Furthermore, store's assignment to groups was partially based on this performance measure, as described below, so that high-noise (low-noise) stores competed against other high-noise (low-noise) stores.

The assignment of stores in the treatment group to the different treatment conditions (low prize spread and high prize spread) went as follows.

[4] For full-time employees, a bonus of 35 euro is about 2.5 percent of monthly gross earnings. Parttimers receive an amount proportional to their contract size.

First, we stratified the stores by their noise in the performance measure. We divided them in two equally large groups, one group with the stores with the highest standard deviation in APC and one group with stores with the lowest standard deviation. Subsequently, we randomly assigned half of the stores in each noise-group to the low-spread treatment and the other half to the high-spread treatment. By doing so, we created four categories of equal size (31 stores) that differ in two dimensions: high noise stores versus low noise stores, and low-spread treatment versus high-spread treatment. A similar procedure was used to assign the 20 non-randomly selected stores to these four treatment-noise categories, so that each category contains 36 stores. In the first round of the tournament, stores compete against three other stores from the same category. The assignment to groups of competing stores is based on historical performance, so as to create a level playing field. Per treatment-noise category, we rank stores on average performance (APC) in the period August 2009 to August 2010. The best-scoring four stores are placed together in a group, as well as numbers 5 to 8, and so on. This creates in total 36 groups, with 9 groups for each treatment-noise category.

In the second round, we again assigned stores to groups on the basis of average performance (APC) in the period August 2009 to August 2010. Assignment was not based on performance in the first round, so as to avoid ratchet effects. In both treatments, we kept the stores in the high-noise category and the low-noise category separate, with one exception: in both treatments, the two stores with the lowest APC in each of the two noise-categories were placed together in a group.[5] Hence, in the second stage of the tournament, we have in total 72 stores divided over 18 groups, with 4 groups per treatment-noise category plus 1 group per treatment with stores from both the low-noise category and the high-noise category. Out of the 20 non-randomly assigned stores participating in the tournament, 9 made it to the second round. Hence, we can use 63 participating stores in the analysis of the second-round treatment effects, which are almost equally divided over the four treatment-noise categories.

[5] As it turns out, seven of these 8 stores were among the 20 stores that were non-randomly assigned to the treatment group and are therefore left out of the analysis. The remaining store belongs to the high-spread, high-noise category. We treat this store the same as all other stores in this category. Leaving the store out of the analysis does not affect the results.

All communication about the elimination tournament to the stores went through the company's internal communication channel. Stores were not aware of our involvement, so that our experiment qualifies as a natural field experiment (Harrison and List 2006). A week before the first round started, all stores of the retail chain learned that a new incentive event would take place. A letter explained that all stores who did not participate in the February 2010 incentive event would participate in the current incentive event, as well as a randomly selected number of stores that did participate in February. A few days later, all participating stores received a message with the rules of the elimination tournament. Stores in the high-spread and low-spread treatment received identical messages, except for the amounts of prize money mentioned for winning the first and second round. Stores were informed that some other stores, randomly selected, faced a different division of prize money, to reduce confusion and suspicion that might arise when employees learn during the tournament that other stores were entitled to different prizes. It was also explained that assignment to groups would be based on the average APC over the period of August 2009 up to August 2010. It was mentioned explicitly that assignment to groups in the second round would not be based on performance during the first round. Just before the start of each round, the stores (still) participating in the tournament received the assignment to groups for all stores, with for each store the average APC over the period of August 2009 up to August 2010. Hence, the stores could verify that they were matched to stores with similar historical performance. During the tournament, each store received weekly feedback on the ranking of stores in its group in the form of a poster with APC-figures for all stores in the group. These posters could be attached to a larger poster, which store managers were instructed to hang in a prominent place (typically the store's canteen).

4 Deriving Hypotheses

In this section, we develop and analyse a simple model to derive the hypotheses that our experiment allows us to test. For a general treatment of the effects of prize spread and noise in tournaments, see Lazear and Rosen (1981) and Gibbs (1996). A general model of incentive effects of elimination tournaments can be found in Rosen (1986).[6]

Consider four identical agents that participate in a two-stage elimination tournament. In the first stage, the agents compete pairwise. The winners of the first stage receive prize $B_1 \geq 0$ and go on to the second stage of the tournament. The first-stage losers are eliminated from the tournament and receive nothing. In the second stage, the two first-stage winners compete against each other for one prize with value $B_2 > 0$.[7]

Let $Q_{i,t}$ be agent i's performance in stage t. Performance depends on effort $e_{i,t}$ and on idiosyncratic noise $\mu_{i,t}$:

$$Q_{i,t} = q(e_{i,t}) + \mu_{i,t},$$

where $q(\cdot)$ is concave. Effort and noise are not observable, performance is verifiable. Agent i's utility in stage t depends on income $w_{i,t}$ and effort cost:

$$U_{i,t} = w_{i,t} - c(e_{i,t}),$$

where $c(\cdot)$ is strictly convex: $c' > 0$, $c'' > 0$. We neglect discounting across stages of the tournament and assume an interior solution for optimal effort. We aim to derive a symmetric subgame-perfect Nash equilibrium.

In the contest between agents i and j in stage t, let $\Delta\mu_{ij,t} = \mu_{i,t} - \mu_{j,t}$ be the noise difference. We assume that $\Delta\mu_{ij,t}$ is distributed according to density function $f(\cdot)$ which is unimodal and symmetric around zero and twice continuously differentiable, with cumulative density function $F(\cdot)$. Across stages, draws of $\Delta\mu_{ij,t}$ are independent. Given effort $\hat{e}$ of contender j in a given stage, agent i's probability of winning that stage is given by $p[q(e_{i,t}) - q(\hat{e}) > -\Delta\mu_{ij,t}] = 1 - F[q(\hat{e}) - q(e_{i,t})]$. Hence, the marginal effect of effort on the winning probability is given by $f[q(\hat{e}) - q(e_{i,t})]q'(e_{i,t})$.

First, consider agents' behaviour in the second stage of the elimination

[6] Recent theoretical advances on elimination tournaments include endogenizing the number of stages and the prize structure (Fu and Lu 2009) and optimal seeding when participants are heterogeneous (Groh et al. 2010).

[7] In the experiment, we have competition between teams rather than between individuals. Also, rather than competition in groups of 2, we have 4 contestants per group competing for two prizes per group. This does not qualitatively affect the theoretical predictions regarding the efects of prize spread and noise. Gibbs (1996) shows that the effect of changing the value of winning is independent of the number of participants, while an increase in the variance of the error term unambiguously lowers the marginal effect of effort on the winning probability whenever the ratio of prizes to participants is $\frac{1}{2}$, as is the case in our experiment.

tournament. In a symmetric equilibrium, both agents optimally exert the same level of effort, as implicitly given by first-order condition

$$f(0)q'(e^*_{t=2})B_2 - c'(e^*_{t=2}) = 0. \tag{1}$$

In the symmetric equilibrium, the probability of winning the second stage is equal to $F(0) = \frac{1}{2}$, so that second-stage expected utility (conditional on winning the first stage) equals $U_{t=2} = \frac{1}{2}B_2 - c(e^*_{t=2})$. As a result, the expected value of winning the contest in the first stage is given by $B_1 + \frac{1}{2}B_2 - c(e^*_{t=2})$. Maximising first-stage utility yields the following first-order condition for optimal effort

$$f(0)q'(e^*_{t=1})\left[B_1 + \frac{1}{2}B_2 - c(e^*_{t=2})\right] - c'(e^*_{t=1}) = 0. \tag{2}$$

By applying the implicit function theorem to first-order conditions (2) and (1), we derive the following predictions regarding the effects of noise in the performance measure and of the prize structure on performance in the elimination tournament. Proposition 1 describes the effect of noise.[8]

Proposition 1 *A larger variance of the noise distribution $f(\cdot)$, so that mass is shifted from the mode to the tails, leads to lower performance in both stages of the tournament.*

Proof. *Higher variance reduces $f(0)$. Totally differentiating (1) gives*

$$\frac{\partial e^*_{t=2}}{\partial f(0)} = -\frac{q'(e^*_{t=2})B_2}{f(0)q''(e^*_{t=2})B_2 - c''(e^*_{t=2})} > 0.$$

The effect on first-round effort is similar. ■

Next, we derive the effects of increasing the convexity of the prize spread. Consider two tournaments with identical total prize money, but different prize spreads. Using superscript L (H) to refer to the tournament with low (high) prize spread, we have $B_1^L > B_1^H$, $B_2^L < B_2^H$, and $2B_1^L + B_2^L =$

[8] In estimating the effects of noise, we use the variance in individual stores' performance μ_i rather than the variance of the difference in the error terms $\Delta\mu_{i,j}$ as in Proposition 1. This has qualitatively no effect on the hypothesis of the effect of noise. The distribution of the difference between two i.i.d. random variables with density $f(\cdot)$ is unimodal with a maximum at zero when $f(\cdot)$ is unimodal (Vogt 1983). By Bienaymé's formula, the variance of the difference of two i.i.d. random variables is the sum of the variance of the two variables. Hence, the variance of $\Delta\mu_{i,j}$ is increasing in the variance of μ_i.

$2B_1^H + B_2^H$. This yields the following predictions regarding the effect of prize spread on performance in the second and first round, respectively.

Proposition 2 *Second-stage performance in the high-spread tournament is better than second-stage performance in the low-spread tournament.*
Proof. *Totally differentiating (1) shows that $e^*_{t=2}$ increases in B_2:*

$$\frac{\partial e^*_{t=2}}{\partial B_2} = -\frac{f(0)q'(e^*_{t=2})}{f(0)q''(e^*_{t=2})B_2 - c''(e^*_{t=2})} > 0.$$

$B_2^H > B_2^L$ *implies that* $e^H_{t=2} > e^L_{t=2}$. ■

Proposition 3 *First-stage performance in the low-spread tournament is better than first-stage performance in the high-spread tournament.*
Proof. *By Proposition 2 and $B_2^L < B_2^H$, second-stage effort is higher in the high-spread treatment, so that $c(e^L_{t=2}) < c(e^H_{t=2})$. As total prize money is identical, we have $B_1^L + \frac{1}{2}B_2^L = B_1^H + \frac{1}{2}B_2^H$, so that $B_1^L + \frac{1}{2}B_2^L - c(e^L_{t=2}) > B_1^H + \frac{1}{2}B_2^H - c(e^H_{t=2})$. By (2), this implies that $e^L_{t=1} > e^H_{t=1}$.* ■

Propositions 2 and 3 together imply that a higher prize spread increases second-round performance at the expense of first-round performance. A higher second-stage bonus induces agents to exert more effort in the second round, which reduces the expected value of winning the first round.

Lastly, given a certain prize structure, the model provides predictions on first-round performance in the tournament relative to second-round performance.

Proposition 4 *If $B_1 \geq B_2$, performance in the first stage is better than performance in the second stage.*
Proof. *Second-stage utility $\frac{1}{2}B_2 - c(e^*_{t=2}) > 0$. Hence, if $B_1 \geq B_2$, $B_1 + \frac{1}{2}B_2 - c(e^*_{t=2}) > B_2$. Comparing (1) and (2), it follows that $e^*_{t=1} > e^*_{t=2}$.* ■

Proposition 5 *If $B_1 \leq \frac{1}{2}B_2$, performance in the first stage is worse than performance in the second stage.*
Proof. *Second-stage effort cost $c(e^*_{t=2}) > 0$. Hence, if $B_1 \leq \frac{1}{2}B_2$, $B_1 + \frac{1}{2}B_2 - c(e^*_{t=2}) < B_2$. Comparing (1) and (2), it follows that $e^*_{t=1} < e^*_{t=2}$.* ■

In the experiment, we have two treatments with identical total prize money but different prize structures. The first treatment has a relatively low prize spread, with equal prizes in both rounds: $B_1^L = B_2^L$. The second treatment has a relatively high prize spread, with $B_1^H = \frac{1}{4}B_2^H$. Hence, Propositions 2 and 3 predict that stores in the low-spread treatment show better first-round performance than stores in the high-spread treatment, but lower second-round performance. Furthermore, Proposition 4 predicts that in the low-spread treatment the first-round treatment effect should be higher than the second-round treatment effect, while Proposition 5 predicts the reverse for the high-spread treatment. Lastly, in both treatments we divide the stores in two groups depending on the historical variance of the performance measure. Proposition 1 predicts that, for a given prize spread, we should observe a lower treatment effect among stores with noisy performance compared to stores with relatively stable performance.

Our assignment of stores to groups made competing stores as homogenous as possible in terms of performance before the tournament. Still, some differences remain, and some groups contain a more heterogenous set of stores than others. Differences in ability across competitors reduce the incentive effect of a given tournament scheme (Lazear and Rosen 1981, O'Keefe et al. 1984). In elimination tournaments, the most able and responsive competitors are more likely to survive the early stages, which might yield more homogenous contests in later stages (Rosen 1986). Our experiment is not designed to test for the effects of contestant heterogeneity on performance in the tournament. However, we should make sure that our findings on the effects of the prize spread on performance in the tournament are not due to differences in competing stores' heterogeneity. Therefore, we perform several robustness checks in Section 7.4.

5 Summary Statistics

In our estimations, performance is given by the Average number of Products per Customer per week (APC). Table 1 shows that on average, a customer buys 1.82 products per transaction. During the experimental period, the average APC-score is somewhat lower than in the year preceding the tournament. Comparing the stores in the control group with the stores in the high and low prize spread group, we find no differences in historical performance. APC is a relatively stable performance measure. Averaged across stores, the within-store standard deviation over the period August 2009 to

August 2010 is 0.15. There is some variation in this measure of noise across stores, as it ranges from a minimum of 0.07 to a maximum of 0.54, with a median of 0.13. Figure 3 shows that the distribution of noise is very similar across the control group and the high-spread and low-spread treatment groups.[9] In other observable store characteristics, we find no statistically significant differences except for the share of female employees: stores in the control group have relatively few female employees.

Grouping the treatment stores by noise group, we find that treatment stores with a large standard deviation in APC show a higher average APC, which is an indication of heteroscedasticity. The difference in noisiness of the performance measure between the low-noise and the high-noise group is substantial. The standard deviation of APC in the high-noise group is about 50% larger than in the low-noise group. Proposition 1 states that the treatment effect should be decreasing in the variance of the performance measure, provided that the density at the mode of the error distribution is smaller for high-noise stores than for low-noise stores. Figure 4 suggests that this holds in the data, by showing kernel densities of the residuals of a regression of APC on store-fixed effects using all observations before the tournament starts. The peak of the kernel density is clearly lower for stores in the high-noise group than for stores in the low-noise group. In both groups, the peak lies marginally to the left of zero. Other store characteristics show no differences between the high and low-noise stores.

6 Estimation

We assess the effects of the tournament incentives using OLS with week fixed effects and store fixed effects. Let $y_{i,w}$ be the performance of store i in week w. Let T_i (C_i) be a dummy variable that takes value 1 for treatment (control) stores. Furthermore, based on the results of the first round of the tournament we create a dummy W_i that takes value 1 for stores that have won in the first round (and, hence, take part in the second round of the tournament) and a dummy E_i that takes value 1 for the stores that are eliminated from the tournament after the first round. Lastly, R_1 and R_2 are two dummy variables indicating the weeks in which the first and second

[9]Figure 3 shows that there are a few stores with unusually large standard deviations in APC. None of the results in this paper are affected by removing these stores from the analysis.

round of the tournament took place, respectively. We estimate the average treatment effect by

$$y_{i,w} = \alpha_i + \tau_w + \beta T_i \left[R_1 + W_i R_2\right] + \delta E_i R_2 + \varepsilon_{i,w} \tag{3}$$

where α_i and τ_w are store and week fixed effects, respectively, and $\varepsilon_{i,w}$ is an error term.[10] Coefficient β gives the average treatment effect of stores in competition versus the stores in the control group. The stores that lost in the first round are non-randomly selected and may respond to losing. Hence, these stores cannot be used as control stores in the second round. We therefore include $E_i R_2$ as a control variable. The estimate of β may be biased because of non-random selection into the second round of the tournament. We deal extensively with several possible selection biases in the next section. It is straightforward to adjust (3) to separate the first and second round average treatment effect, by replacing the term $\beta T_i \left[R_1 + W_i R_2\right]$ by $\beta_1 T_i R_1 + \beta_2 T_i W_i R_2$.

To estimate how the level of noise in a store's performance measure affects the response to the tournament incentives, we use the standard deviation in the performance measure APC over the period August 2009 to August 2010 as a measure of noise. By interacting the treatment dummy with the standard deviation sd_i, we can assess whether the treatment effect is heterogenous in noise, as predicted by Proposition 1. This implies estimating

$$y_{i,w} = \alpha_i + \tau_w + \beta T_i \left[R_1 + W_i R_2\right] + \nu T_i \left[R_1 + W_i R_2\right] sd_i + \mu \left[R_1 + \left(W_i + C_i\right) R_2\right] sd_i + \delta E_i R_2 + \varepsilon_{i,w} \tag{4}$$

where ν measures how sensitive the treatment effect is to noise, and the term $\mu \left[R_1 + \left(W_i + C_i\right) R_2\right] sd_i$ measures how performance during the experimental period relates to the standard deviation in APC. The latter term is necessary to control for any time-specific effects of noise, which might otherwise be picked up by ν. Note that we continue to leave out the first-round losers from the estimation of the second-round effects. In estimating (4), we take up our measure of noise sd_i in deviation from its mean.

To estimate the effect of prize spread, we split dummy T_i into two treat-

[10] In our estimations we cluster standard errors at the store level to correct for serial correlation within stores and heteroscedasticity across stores, as suggested by Bertrand et al. (2004).

ment group dummies. Variable T_L (T_H) takes value 1 when store i is assigned to the low-spread (high-spread) treatment. Replacing T_i in (3) by the two treatment group dummies gives

$$y_{i,w} = \alpha_i + \tau_w + [\beta_L T_L + \beta_H T_H] [R_1 + W_i R_2] + \delta E_i R_2 + \varepsilon_{i,w} \qquad (5)$$

Again, this expression is easily adjusted to estimate the treatment effects in the two tournament rounds separately.

7 Results

7.1 Average treatment effect

The first column of Table 2 gives the results of estimating (3). We find a statistically significant effect of the tournament on performance. The average treatment effect is 0.028 extra products per customer. This corresponds to an increase of 1.5% of the mean score on Average number of Products sold per Client (APC) and to 20% of within-store standard deviation of APC. The second column of Table 2 separates the estimated average treatment effect by tournament round. On average, the first-round effect is positive but statistically insignificant. In the second round, the treatment effect is about 2.5 percent extra products per customer, which differs significantly from zero (p-value < 0.01). The difference between the estimates for the first and second-round treatment effect is statistically significant with a p-value of 0.056. Both estimations show that the stores that lost in the first round perform about as well as the stores in the control group during the second-round period. Hence, two weeks after their elimination, these stores seem to have returned to business-as-usual performance.[11]

The estimated treatment effect in column 1 as well as the estimated treatment effect of second round of the tournament in column 2 may be biased if stores differ in time trends. If some stores experience an upward trend while others experience a downward trend, then relatively many stores on a positive time trend will be selected into the second round, resulting in

[11] If we include the first-round losers as treated stores in the second round, instead of taking them out as in Table 2, the estimated overall treatment effect is a 1 percent increase in performance which differs significanty from zero (p-value $= 0.08$). All p-values are based on two-sided tests.

an upward bias in the estimated treatment effect. To analyse this, we run a pseudo-tournament among the stores in the control group. First, we group the control stores into groups in a similar way as the assignment of the treatment stores. We create 13 groups of 4 stores and two groups of 5 stores with similar average performance over the period August 2009 to August 2010. We identify for each of the groups the two stores with the highest performance during the first round of the experiment. Next, we compare the performance of the 'winners' and 'losers' of this pseudo-competition during the second round of the experiment with the performance of the real winners and losers from the treatment group. Figure 5 shows for each of these four groups the kernel densities of performance during the second round of the experiment. The performance distributions of the winners and losers of the pseudo-competition are very similar. Hence, in the control group, the stores that perform relatively well during the first round of the tournament do not show better or worse performance during the second round as compared to stores that performed poorly during the first round. Furthermore, the performance distribution of the first-round losers of the real tournament is similar to the performance distributions of the control stores. This again suggests that treatment stores not making it to the second round return to regular performance within two weeks of their elimination. In contrast, the second-round performance distribution of first-round treatment group winners is shifted to the right and has more mass between 2 and 2.4 as compared to the other groups. Hence, the second-round treatment effect is not caused by a selection of stores that experience a positive trend in performance.[12]

Next, we analyse whether there are carry-over effects from the earlier experiment we did in this company. As described in Section 3, all stores comprising the control group in the earlier experiment participated in the current tournament, as well as 29 randomly selected stores from the treatment group of the earlier experiment. Columns 1 and 2 of Table 3 show that the response of stores that did participate in the earlier experiment is not significantly different from the response of the other stores, neither in the first round nor in the second round. Hence, our current results are not

[12] Another possible source of selection bias is a difference between stores in responsiveness to competitive incentives. In Section 7.3 we will show that there is no evidence for such a difference.

affected by the earlier experiment.

7.2 Noise

The first hypothesis that we test concerns the effect of noise on the responsiveness of stores' performance to the tournament. As described by Proposition 1, we hypothesize that stores experiencing more noisy performance respond less to tournament incentives. Column 1 in Table 4 reports the results of estimating (4). We find that noisiness of the performance measure has a negative effect on the response to the tournament. This negative effect is close to being significantly different from zero at conventional levels (p-value = 0.11).[13] An increase in the level of noise by one standard deviation reduces the treatment effect by 1 percentage point. As we have taken up the variable noise in deviation from its mean, the first coefficient in column 1 gives the estimated treatment effect at the mean level of noise. This effect is about 1.3 percent and statistically different from zero. Wald tests show that the estimated treatment effect is significantly different from zero for stores with a standard deviation in APC up to 0.15 (i.e. for 70 percent of the stores). Note also that higher noise relates to weaker performance during the experimental period, underlining the importance of controlling for this time-specific effect.[14]

In column 2 of Table 4, the estimation of (4) is separated by tournament round. We find a small and statistically insignificant effect of noise in the first round. In the second round, however, noise has a strongly significant, negative effect on performance in the tournament. Wald tests show that the second-round treatment effect is statistically different from zero for stores with a standard deviation below 0.16. Hence, we find support for Proposition 1, in particular in the second round of the tournament.

Recall that noise in the performance measure is a pre-existing store characteristic, not randomly assigned. Hence, it is possible that noise is partially caused by or correlated with other store-specific characteristics. When these other store-specific characteristics affect stores' responsiveness to the tournament incentives, the effect of noise found in Table 4 might be spurious or

[13] Our measure of noise is the within-store standard deviation of APC. This includes both idiosyncratic and common shocks (i.e. time-fixed effects in APC). If we exclude the common shocks from our measure of noise, the estimated effect of noise on the treatment effect has the same magnitude, but is more precisely estimated (p-value = 0.08).

[14] A quadratic specification of the effect of sd_i does not improve the estimation.

estimated with bias. Insofar as these store characteristics are unobservable (at least for us), we cannot rule out this possibility. However, for observable store characteristics, this can be assessed. First, we run an OLS regression of our measure of noise on all available store characteristics (regression output can be found in the Appendix, Table A1). The observable store characteristics explain about 25 percent of the variation in noise across stores. Most explanatory power comes from the average level of performance APC and regional variation. Next, we take the residuals from this cross-section regression of noise, and use these residuals instead of the standard deviation of APC in estimating (4). Hence, we use the variation in noise that is not correlated with observable store characteristics. The results of this estimation are presented in Columns 1 and 2 in Table 5. We find that the estimates of the effect of residual noise on the response to competition are very similar to the estimates when using our standard measure of noise. This rules out that the negative effect of noise on the response to competition is caused by one or more of the observable store characteristics.

7.3 Prize spread

To analyse the effects of prize spread, we estimate the effects of the two treatments separately. The first column of Table 6 shows the estimated average treatment effects over both rounds for the low-spread and high-spread treatments separately, as given by (5). Both treatments have a similar effect on performance, of around 1.5 percent in magnitude. Both estimates are significantly different from zero with a p-value of about 0.05.

Column 2 of Table 6 differentiates these estimates by tournament round. This estimation allows us to test the hypotheses that follow from Proposition 2 to 5. First, we focus on comparing the low-spread and the high-spread treatment. Propositions 2 and 3 predict better second-round performance in the high-spread treatment and better first-round performance in the low-spread treatment, respectively. We find that first-round performance in the low-spread treatment is indeed 0.8 percentage point better than in the high-spread treatment, but a Wald test reveals that the difference is not statistically significant. In the second round, the treatment effect is 1 percentage point higher in the high-spread treatment, but again the difference is not statistically significant. Hence, both effects have the sign as predicted by theory, but the effects are small.

Next, we compare first and second-round performance within a treatment. Proposition 4 predicts that in the low-spread treatment, the first-round treatment effect is higher than the second-round treatment effect. The estimation results in column 2 of Table 6 show that we actually find higher second-round performance, although the 1 percentage point difference is not statistically significant. Proposition 5 predicts that in the high-spread treatment, first-round performance should be lower than second-round performance. This is clearly borne out in column 2 of Table 6, where the difference between first and second-round performance is estimated at 2.7 percent, which is significant at a p-value of 0.02.

Overall, our interpretation of these findings is that they are, by and large, in line with theoretical predictions. Including the effect of noise as discussed in the previous subsection, four out of five estimated effects have the sign predicted by theory, although only two of these effects are statistically significant. Regarding the convexity of the prize spread, we cannot exclude the possibility that the treatment effects by tournament round are similar in the low spread treatment and the high spread treatment. Still, the pattern is suggestive: an increase in the convexity of the prize spread increases second-round performance at the expense of first-round performance.

Three extensions of the basic model presented in Section 4 might explain why the second-round treatment effect is higher than the first-round treatment effect in the low-spread treatment, in contrast to the model's prediction. First, if stores differ in their ability to perform, second-round groups could contain a more homogenous set of stores than first-round groups. This would lead to fiercer competition in the second round compared to the first round. We examine the effects of heterogeneity among stores extensively in the next subsection and find no evidence that our estimations are affected by differences in group's heterogeneity across tournament rounds.

Second, another selection effect could arise when stores differ in responsiveness to competition. In that case, relatively responsive stores are selected into the second round. If so, we should compare the first and second-round treatment effect of the stores that won the first round. However, the stochastic nature of performance implies that we cannot simply compare the first and second-round performance of the first-round winners. Given that a store won the first round, its expected value of the stochastic component in APC during the first round is positive, yielding an upward bias in the estimate of the first-round effect. Here, we can use the pseudo-competition we conducted in the control group, as described in Section 7.1, to assess the magnitude of this bias, as follows. The pseudo-competition gives us winners and losers of a competition purely determined by the stochastic component, in the same period as the first round of the tournament. We can compare the difference in performance between the winners and losers of the pseudo-competition to the difference in performance between winners and losers of the first round of the actual tournament. If stores are homogenous, the theory as described in Section 4 predicts that, while on average winners (losers) in the real tournament perform better than the winners (losers) in the pseudo-competition, the difference between winners and losers is similar across the treatment and control groups. If, on the other hand, stores differ in responsiveness to competition, the difference between winners and losers should be larger in the real tournament than in the pseudo-tournament.

Column 3 in Table 6 examines whether stores are heterogenous in responsiveness to competition. The first five coefficients give the estimated performance during the first round of the experiment for five groups of stores, all relative to the performance of the stores in the control group that 'lost' the pseudo-competition. First, the 'winners' of the pseudo-competition perform about 4 percentage points better than the 'losers'. Comparing the difference in performance between the first-round winners and losers in the treatment groups, we see that in the low-spread treatment this difference is marginally higher at 4.3 percentage points, while in the high-spread treatment it is even smaller at 3.3 percentage points. These differences are nowhere close to being statistically significant. Hence, we find no evidence for differences in responsiveness across stores.

A third explanation for the uniformly higher treatment effect in the second round is that winning a competition may provide employees with non-monetary benefits such as status, social recognition, or simply the joy of winning (Auriol and Renault 2008, Besley and Ghatak 2008, Frey and Neckermann 2008, Moldovanu et al. 2007). Several recent empirical studies suggest that these non-monetary benefits are substantial, by showing that people respond to competition even when there is no money at stake (i.e. when only relative performance information and/or symbolic awards are provided), see Azmat and Iriberri (2010), Blanes i Vidal and Nossol (2011), Delfgaauw et al. (2009), Kosfeld and Neckermann (2011), and Sheremeta

(2010). Accepting the presence of non-monetary utility of winning a competition, the result that the second-round treatment effect is higher than the first-round treatment effect in the low-spread treatment would suggest that winning the second round yields higher non-monetary utility than winning the first round. Note that the addition of a non-monetary benefit of winning the second round of the tournament to the basic model in Section 4 does not change the predictions of the effects of changes in the prize spread. In particular, the difference between first-round and second-round performance in the high-spread tournament should be larger than this difference in the low-spread tournament. Computed from the estimates in the second column of Table 6, the magnitude of this difference-in-differences is about 1.8 percentage points, but a Wald test reveals that this difference is not statistically significant (p-value = 0.30).

7.4 Robustness check: heterogeneity

In this subsection, we check whether the results on prize spread are robust to the inclusion of controls for heterogeneity in ability. The theoretical model in Section 4 assumed that contestants were homogenous in ability. In assigning stores to groups, we have made the groups as homogenous as possible, by grouping stores with similar historical performance together. Still, some differences in historical performance between stores in a group remain. Theory predicts that differences in ability between contestants reduce the incentive effect of a tournament (Lazear and Rosen 1981, O'Keefe et al. 1984, Rosen 1986). Low-ability contestants realize that their efforts are less likely to result in victory, while high-ability contestants realize that high effort does not improve their winning probability much over modest effort. This implies that if the groups in some treatment-round are more heterogenous than in others, then the estimations on the effect of prize spread may partially reflect these differences in heterogeneity.

Differences in groups' heterogeneity seem more likely when comparing across rounds of the tournament rather than across treatments. Comparing treatments, in expectation there are no differences in the heterogeneity of groups, as the procedure of assigning stores to groups was identical in the two treatments. However, comparing across rounds, groups might differ in heterogeneity, as the set of stores that compete in the second round is non-randomly selected. This could go either way. Second-round groups may have been more heterogenous than first-round groups, as for each store, two of the most similar stores in terms of historical performance have been eliminated in the first round. On the other hand, if the most responsive stores make it to the second round, second-round groups may have been more homogenous in terms of performance under competition than first-round groups. The latter effect we addressed in the previous subsection (Table 6, column 3), where we found that the difference in performance during the first round between winners and losers in the tournament was not significantly different from the difference in performance between the 'winners' and 'losers' in the artificial pseudo-competition in the control group. Hence, we find no indication that differences in responsiveness to competition between stores make second-round groups more homogenous than first-round groups, as discussed by Rosen (1986).

Here, we check whether heterogeneity in historical performance across stores affects our estimates of the effects of prize spread. Note that our experiment is not designed to *test* theoretical predictions regarding heterogeneity. Our objective here is to analyse whether our earlier findings are robust to controlling for differences in heterogeneity across stores competing in a group. Bull et al. (1987), Schotter and Weigelt (1992), Van Dijk et al. (2001), and Harbring and Lunser (2008) study heterogenous contests in the lab, typically finding that heterogeneity depresses average performance, although disadvantaged players tend to slack off less than predicted.[15]

We create two measures of heterogeneity within groups. The first measure is the standard deviation of stores' mean historical performance (APC) per group. As described in Section 3, we used stores' mean historical performance in assigning stores to groups and stores were informed about mean historical performance of their competitors. Figure 6 shows the distribution of groups' standard deviation of store's mean historical performance,

[15] Our focus here is on differences in ex ante performance, possibly reflecting differences in ability. Differences in performance can also arise in the course of the competition, as a tournament round lasts four weeks and stores receive weekly performance rankings of their group. This implies that in some groups, large differences in performance in the tournament may arise early in a round, while in other groups stores may perform close together. This may create ex post differences in competitiveness across groups. However, such differences arise in both treatments and in both rounds. Comparing across treatments and rounds, there is no reason why groups would develop on average more or less competitively in some treatment or tournament round. These dynamic incentive effects of relative performance pay are studied by Casas-Arce and Martinez-Jerez (2009), Frank and Obloj (2011), and Delfgaauw et al. (2010).

separated by rounds. In many groups, stores are nearly identical in terms of historical performance. Across rounds, these distributions are not very different, but it is clear that groups in the second round tend to be more heterogenous.[16]

Let $\eta_{i,R}$ be the standard deviation of mean historical APC in store i's group in round $R \in [1, 2]$. We modify (5) by including the interaction of our first measure of heterogeneity $\eta_{i,R}$ with a treatment dummy, and estimate

$$\begin{aligned} y_{i,w} \quad = \quad & \alpha_i + \tau_w + \beta_{L,1} T_L R_1 + \beta_{H,1} T_H R_1 + \beta_{L,2} W_i T_L R_2 + \beta_{H,2} W_i T_H R_2 + \\ & + \beta_\eta \left[T_L + T_H\right] \left[R_1 \eta_{i,1} + W_i R_2 \eta_{i,2}\right] + \delta E_i R_2 + \varepsilon_{i,w} \end{aligned} \qquad (6)$$

Hence, we estimate the average treatment effect separated by treatment and by round, as in Subsection 7.3, but now controlling for the effect of differences in groups' standard deviation of stores' historical performance. In estimating (6), we take up our measure of heterogeneity in deviation from its mean.

The second column in Table 7 gives the results of estimating (6), while the first column repeats the estimation results without controlling for heterogeneity (copied from Table 6, column 2). First, we find that larger differences in store's historical performance within a group leads to a lower response to competition, as predicted by tournament theory. Moreover, controlling for heterogeneity hardly affects the estimates of the treatment effect by prize spread and round. Most importantly, the differences in the estimated effects between treatments and rounds remain the same. Hence, we conclude that the estimated differences in the response to competition between treatments and rounds, which form the basis for our tests of Propositions 2 to 5, are not caused by differences in groups' heterogeneity in historical performance across treatments and rounds.

Our second measure of heterogeneity is the difference between a store's mean historical performance (APC over the period August 2009 to August 2010) and the average of the mean historical performance of the stores ranked second and third in the group. This gives a measure of the distance

[16] One participating store has an exceptionally large historical APC of 2.56. While matched to other stores with relatively high historical APC, the first- and second-round groups of this store still have a relatively high standard deviation of store's historical APC. We have redone our whole analysis excluding this one store as well as excluding these two groups (6 stores). Our results remained qualitatively unaffected.

to the winning positions, from an ex ante perspective. While this measure does not capture the full heterogeneity within the group, it does provide a relevant measure of heterogeneity that varies at store level.

Figure 7 gives the distribution of this measure, separated by rounds. We find that many stores are close to the (ex ante) winning threshold in their group. This is another way of showing that in many groups, stores with very similar historical performance compete against each other. Comparing the first and second round distributions, the second round distribution has somewhat wider tails. This again suggests that ex ante differences between stores in a group are larger in the second round, which could affect our estimates of the difference in response to competition across rounds.

In controlling for the effect of the distance to winning positions, we allow the effect to be different for those who are ahead (i.e. the stores ranked first and second on the basis of mean historical performance) as compared to those who are behind (those ranked third and fourth). For stores who are behind, a higher value implies being closer to the threshold, which should increase the incentive effect from the tournament. For stores who are ahead, however, a higher value implies being further away from the threshold, which should decrease the incentive effect. Therefore, we estimate

$$\begin{aligned} y_{i,w} = {} & \alpha_i + \tau_w + \beta_{L,1} T_L R_1 + \beta_{H,1} T_H R_1 + \beta_{L,2} W_i T_L R_2 + \beta_{H,2} W_i T_H R_2 + \\ & + \gamma \left[T_L + T_H\right] \left[R_1 \Delta^+_{i,1} + W_i R_2 \Delta^+_{i,2}\right] + \varphi \left[T_L + T_H\right] \left[R_1 \Delta^-_{i,1} + W_i R_2 \Delta^-_{i,2}\right] + \\ & + \delta E_i R_2 + \varepsilon_{i,w} \end{aligned} \tag{7}$$

where $\Delta^+_{i,R}$ is the distance to the threshold for store i that is ahead in round $R \in [1, 2]$ (taking the value zero for stores that lie behind) and $\Delta^-_{i,R}$ is the distance to the threshold for store i that lies behind in round $R \in [1, 2]$ (taking the value zero for stores that are ahead).

The results of estimating (7) are presented in column 3 in Table 7. We find that for stores that lie ahead, a larger distance to the threshold indeed significantly reduces the incentive effect of the competition. For stores that are behind, there is an insignificant negative effect of being closer to the threshold. This is in line with findings in lab experiments, where disadvantaged contestants tend to invest more than predicted by theory (Bull et al. 1987, Schotter and Weigelt 1992, Van Dijk et al. 2001, Harbring and Lunser

2008). Again, we find that controlling for heterogeneity does not affect the estimated average effects of the tournament on performance by treatment and by round. The differences in coefficients between the two treatments and the two rounds remain similar in magnitude. This implies that differences in homogeneity in historical performance across the low spread and high spread treatments and across tournament rounds do not affect the estimated differences in response to competition between treatments and rounds. Hence, our tests of Propositions 2 to 5 are not affected by differences in groups' heterogeneity across treatments and tournament rounds.[17]

This subsection has shown that controlling for differences in groups' heterogeneity among competing stores in terms of historical performance does not affect our results on the effects of the convexity of the prize structure. In the previous subsection, we found that the first round did not select a particularly responsive set of stores into the second round (Table 6, column 3). Hence, we are confident that our conclusions regarding the effects of prize spread are not influenced by differences in heterogeneity of competing stores across treatments and tournament rounds.

8 Concluding Remarks

Examining whether workers respond as predicted to tournament incentives in their natural working environment is important for linking tournament theory to organizational policies regarding wages and promotions. We have designed a natural field experiment in a private company to test several predictions on the effects of prize structure and noise in a two-stage elimination tournament. As predicted, we find that increased convexity of the prize spread increases second-round performance at the expense of first-round performance, although some differences between the estimated effects by round are not statistically significant. Furthermore, workers with relatively volatile performance hardly respond to tournament incentives, while workers whose performance measure is more stable increase performance significantly.

The magnitude of the effects we find should be considered in the light of the strength of the incentive. On average, employees earned a bonus equal to 2 percent of their monthly earnings in the experiment, which took place

[17] Allowing the effects of heterogeneity to vary by treatment or by tournament round does not affect the estimations of the effect of prize spread on the response to tournament incentives, for neither of the two measures of heterogeneity.

over a period of 8 weeks. The average treatment effect on the performance measure APC (Average number of Products Sold per Client) was about 1.5 percent. In the end, the company's management cares about profits and – as an important intermediate – about sales revenues rather than APC. While we do not have data on profits, we can assess the effect of the tournament on sales. Perhaps surprisingly, we find no effect of the experiment on sales. Apparently, workers have means to increase APC without increasing revenue, suggesting that APC is prone to gaming.

References

[1] Altmann, Steffen, Armin Falk, and Matthias Wibral (2012), Promotions and Incentives: The Case of Multi-Stage Elimination Tournaments, *Journal of Labor Economics,* vol. 30(1), pp. 149-174.

[2] Audas, Rick, Tim Barmby, and John Treble (2004), Luck, Effort, and Reward in an Organizational Hierarchy, *Journal of Labor Economics*, vol. 22(2), pp. 379-396.

[3] Auriol, Emmanuelle, and Régis Renault (2008), Status and Incentives, *RAND Journal of Economics*, vol. 39(1), pp. 305-326.

[4] Azmat, Ghazala, and Nagore Iriberri (2010), The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment using High School Students, *Journal of Public Economics,* vol. 94(7-8), pp. 435-452.

[5] Bandiera, Oriana, Iwan Barankay, and Imran Rasul (2009), Team Incentives: Evidence from a Field Experiment, *Journal of the European Economic Association,* forthcoming.

[6] Becker, Brian E., and Mark A. Huselid (1992), The Incentive Effects of Tournament Compensation Systems, *Administrative Science Quarterly*, vol. 37(2), pp. 336-350.

[7] Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004), How Much Should We Trust Differences-in-Differences Estimates?, *Quarterly Journal of Economics*, vol. 119(1), pp. 249-275.

[8] Besley, Timothy, and Maitreesh Ghatak (2008), Status Incentives, *American Economic Review*, vol. 98(2), pp. 206–211.

[9] Blanes i Vidal, Jordi, and Mareike Nossol (2011), Tournaments without Prizes: Evidence from Personnel Records, *Management Science,* vol. 57(10), pp. 1721-1736.

[10] Bognanno, Michael L. (2001), Corporate Tournaments, *Journal of Labor Economics*, vol. 19(2), pp. 290-315.

[11] Bull, Clive, Andrew Schotter, and Keith Weigelt (1987), Tournaments and Piece Rates: An Experimental Study, *Journal of Political Economy*, vol. 95(1), pp. 1-33.

[12] Casas-Arce, Pablo, and F. Asis Martinez-Jerez (2009), Relative Performance Compensation, Contests, and Dynamic Incentives, *Management Science,* vol. 55(8), pp. 1306-1320.

[13] Charness, Gary, and Peter J. Kuhn (2010), Lab Labor: What Can Labor Economist Learn from the Lab? Forthcoming in: Orley Ashenfelter and David Card (Eds.), *Handbook of Labor Economics,* Vol. 4, Amsterdam: Elsevier.

[14] Delfgaauw, Josse, Robert Dur, Arjan Non, and Willem Verbeke (2010), Dynamic Incentive Effects of Relative Performance Pay: A Field Experiment, Tinbergen Institute Discussion Paper 10-124/1.

[15] Delfgaauw, Josse, Robert Dur, Joeri Sol, and Willem Verbeke (2009), Tournament Incentives in the Field: Gender Differences in the Workplace, Tinbergen Institute Discussion Paper 09-061/1.

[16] DeVaro, Jed (2006), Internal Promotion Competition in Firms, *RAND Journal of Economics*, vol. 37(3), pp. 521-542.

[17] Ehrenberg, Ronald G., and Michael L. Bognanno (1990), Do Tournaments Have Incentive Effects?, *Journal of Political Economy*, vol. 98(6), pp. 1307-1324.

[18] Erev, I., G. Bornstein, and G. Rachely (1993), Constructive Intergroup Competition as a Solution to the Free Rider Problem: A Field Experiment, *Journal of Experimental Social Psychology,* vol. 29(6), pp. 463-478.

[19] Eriksson, Tor (1999), Executive Compensation and Tournament Theory, *Journal of Labor Economics*, vol. 17(2), pp. 262-280.

[20] Frank, Douglas, and Tomasz Obloj (2011), Reference Points and Organizational Performance: Evidence from Retail Banking, mimeo, INSEAD.

[21] Freeman, Richard B., and Alexander M. Gelber (2010), Prize Structure and Information in Tournaments: Experimental Evidence, *American Economic Journal: Applied Economics,* vol. 2(1), pp. 149-164.

[22] Frey, Bruno S., and Susanne Neckermann (2008), Awards - A View from Psychological Economics, *Journal of Psychology*, vol. 216(4), pp. 198-208.

[23] Fu, Qiang, and Jingfeng Lu (2009), The Optimal Multi-stage Contest, *Economic Theory,* forthcoming.

[24] Gibbs, Michael (1995), Incentive Compensation in a Corporate Hierarchy, *Journal of Accounting and Economics,* vol. 19(2-3), pp. 247-277.

[25] Gibbs, Michael (1996), Promotions and Incentives, mimeo, University of Chicago.

[26] Green, Jerry R., and Nancy L. Stokey (1983), A Comparison of Tournaments and Contracts, *Journal of Political Economy*, vol. 91(3), pp. 349-364.

[27] Groh, Christian, Benny Moldovanu, Aner Sela, and Uwe Sunde (2010), Optimal Seedings in Elimination Tournaments, *Economic Theory,* forthcoming.

[28] Harbring, Christine, and Bernd Irlenbusch (2003), An Experimental Study on Tournament Design, *Labour Economics*, vol. 10(4), pp. 443-464.

[29] Harrison, Glenn, and John A. List (2004), Field Experiments, *Journal of Economic Literature*, vol. 42(4), pp. 1009-1055.

[30] Höchtl, Wolfgang, Rudolf Kerschbamer, Rudi Stracke, and Uwe Sunde (2011), Effort versus Sorting: That's the Question. Two-Stage Elimi-

nation Tournaments in Theory and Experiment, mimeo, University of Innsbruck.

[31] Knoeber, Charles R., and Walter N. Thurman (1994), Testing the Theory of Tournaments: An Empirical Analysis of Broiler Production, *Journal of Labor Economics,* vol. 12(2), pp. 155-179.

[32] Kosfeld, Michael, and Susanne Neckermann (2011), Getting More Work for Nothing? Symbolic Awards and Worker Performance, *American Economic Journal: Microeconomics,* vol. 3(3), pp. 86-99.

[33] Lazear, Edward P. (1989), Pay Equality and Industrial Politics, *Journal of Political Economy*, vol. 97(3), pp. 561-580.

[34] Lazear, Edward P., and Paul Oyer (2009), Personnel Economics. Forthcoming in: Robert Gibbons and John Roberts (Eds.), *Handbook of Organizational Economics.* Princeton: Princeton University Press.

[35] Lazear, Edward P., and Sherwin Rosen (1981), Rank-Order Tournaments as Optimum Labor Contracts, *Journal of Political Economy*, vol. 89(5), pp. 841-864.

[36] Lim, Noah, Michael J. Ahearne, and Sung H. Ham (2009), Designing Sales Contests: Does the Prize Structure Matter?, *Journal of Marketing Research*, vol. 46(3), pp. 356-371.

[37] McCue, Kristin (1996), Promotions and Wage Growth, *Journal of Labor Economics*, vol. 14(2), pp. 175-209.

[38] Moldovanu, Benny, Aner Sela, and Xianwen Shi (2007), Contests for Status, *Journal of Political Economy*, vol. 115(2), pp. 338–363.

[39] Murphy, Kevin J. (1985), Corporate Performance and Managerial Remuneration: An Empirical Analysis, *Journal of Accounting and Economics*, vol. 7(1-3), pp. 11-42.

[40] Nalebuff, Barry J., and Joseph E. Stiglitz (1983), Prizes and Incentives: Towards a General Theory of Compensation and Competition, *Bell Journal of Economics*, vol. 14(1), pp. 21-43.

[41] O'Keeffe, Mary, W. Kip Viscusi, Richard J. Zeckhauser (1984), Economic Contests: Comparative Reward Schemes, *Journal of Labor Economics,* vol. 2(1), pp. 27-56.

[42] Rosen, Sherwin (1986), Prizes and Incentives in Elimination Tournaments, *American Economic Review,* vol. 76(4), pp. 701-715.

[43] Schotter, Andrew, and Keith Weigelt (1992), Asymmetric Tournaments, Equal Opportunity Laws, and Affirmative Action: Some Experimental Results, *Quarterly Journal of Economics,* vol. 107(2), pp. 511-539.

[44] Sheremeta, Roman M. (2010), Experimental Comparison of Multi-stage and One-stage Contests, *Games and Economic Behavior,* vol. 68(2), pp. 721-747.

[45] Van Dijk, Frans, Joep Sonnemans, and Frans van Winden (2001), Incentive systems in a real effort experiment, *European Economic Review,* vol. 45(2), pp. 187-214.

[46] Vogt, H. (1983), Unimodality of Differences, *Metrika*, vol. 30(1), pp. 165-170.

[47] Waldman, Michael (2008), Theory and Evidence in Internal Labor Markets. Forthcoming in: Robert Gibbons and John Roberts (Eds.), *Handbook of Organizational Economics.* Princeton: Princeton University Press.

Figures and Tables

Figure 1: Mean of Average number of Products per Customer (APC) across all stores, by week

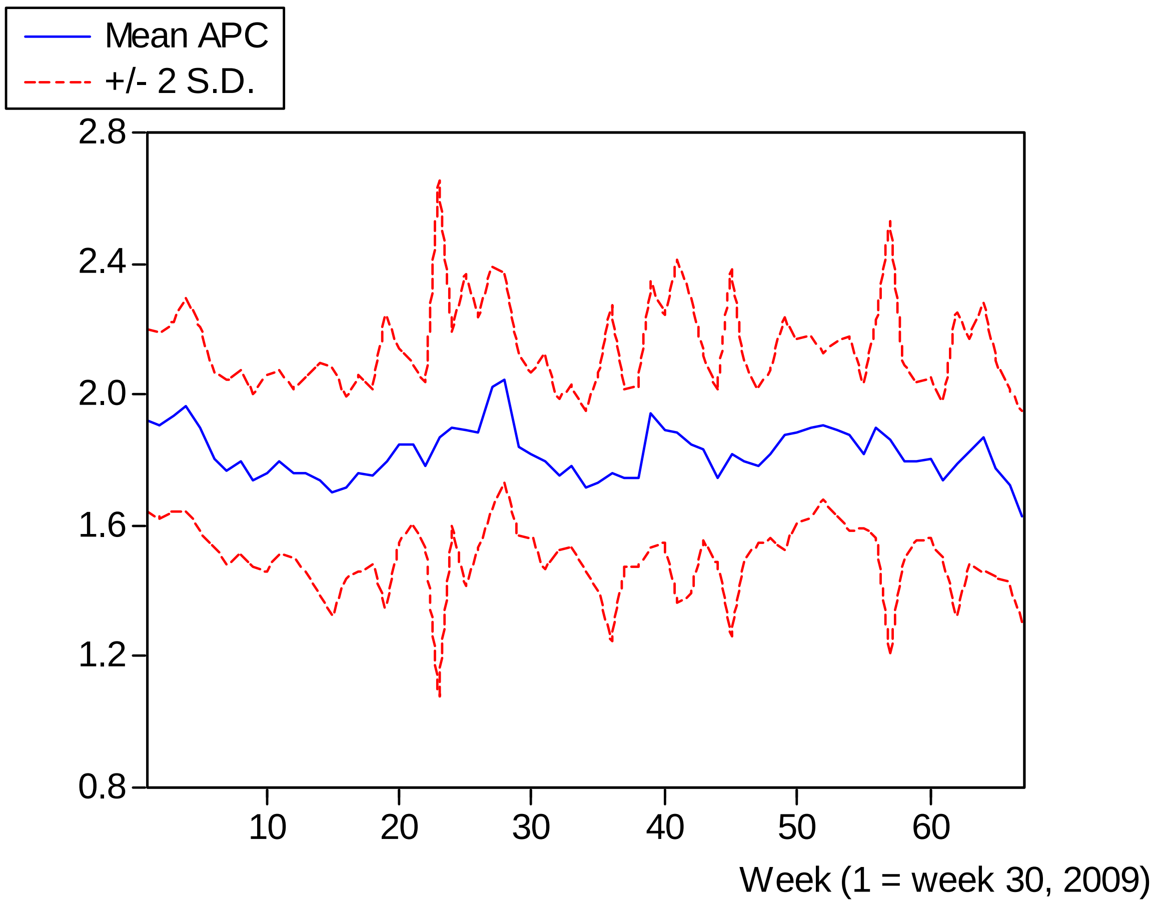
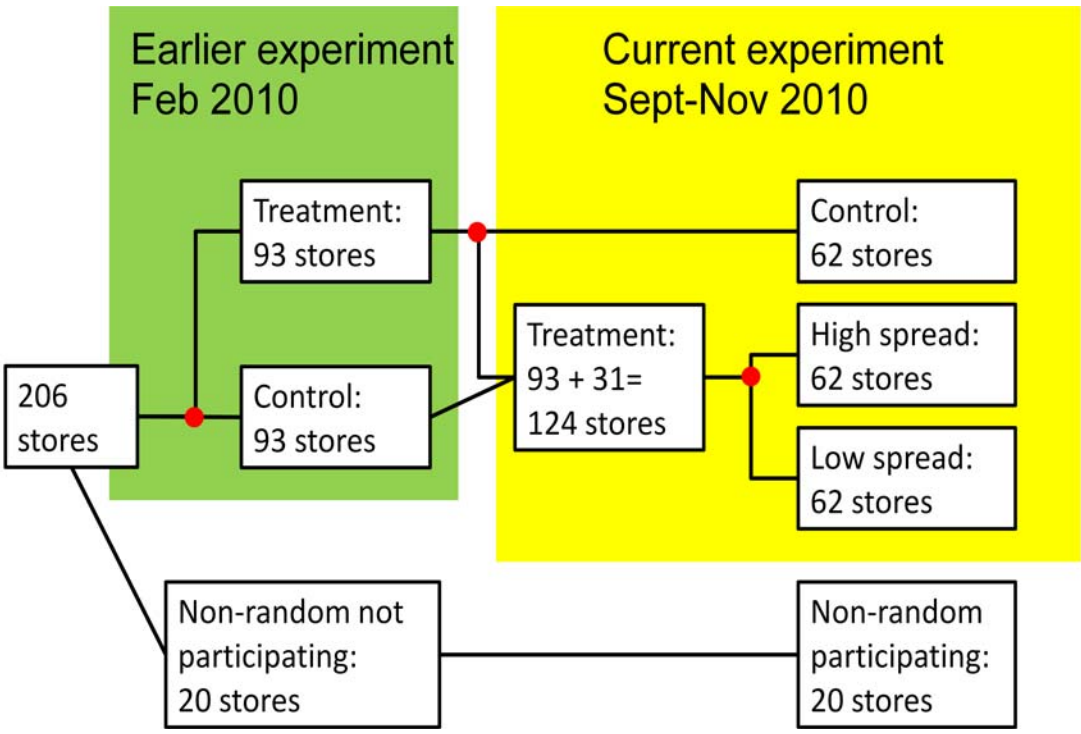


Figure 2: Assignment to treatment groups and control group



Bullets indicate points of randomization.
Two stores not shown (control group first experiment, non-randomly not participating in the current experiment).

Figure 3: Kernel density of within store standard deviation of APC in the assignment period (August 2009 - August 2010), by treatment group

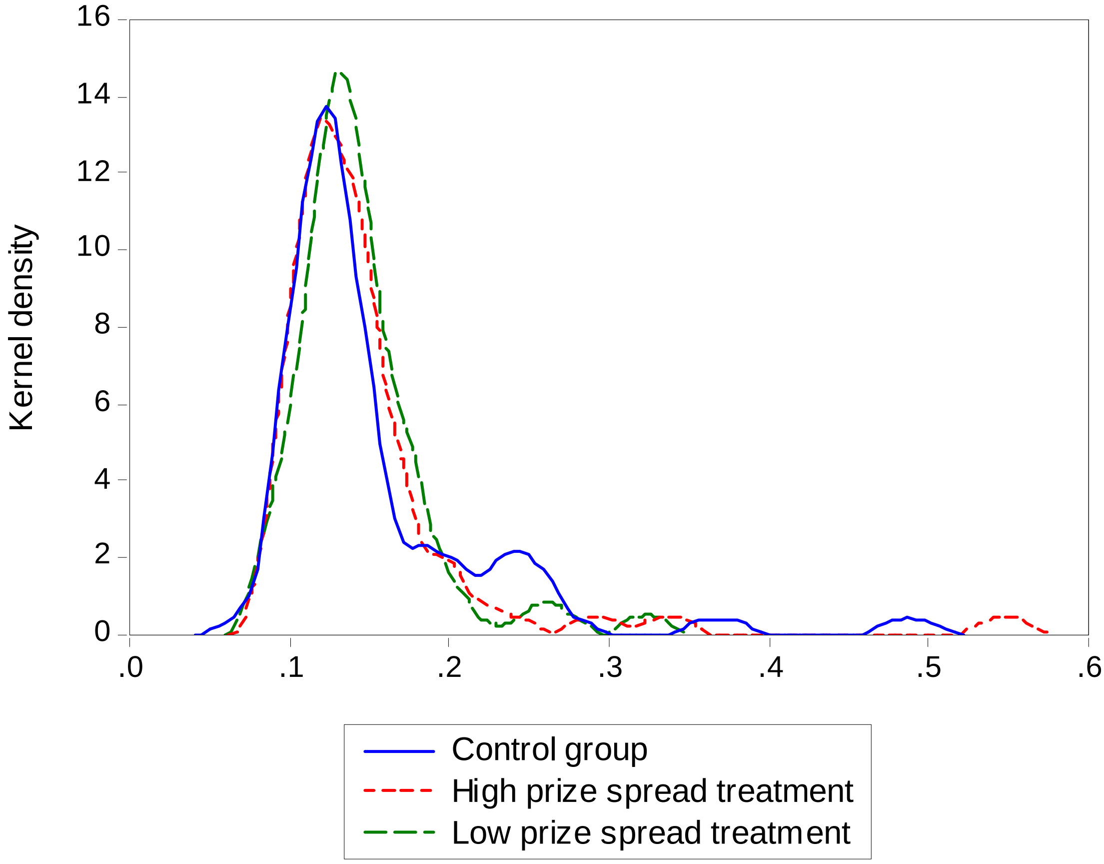
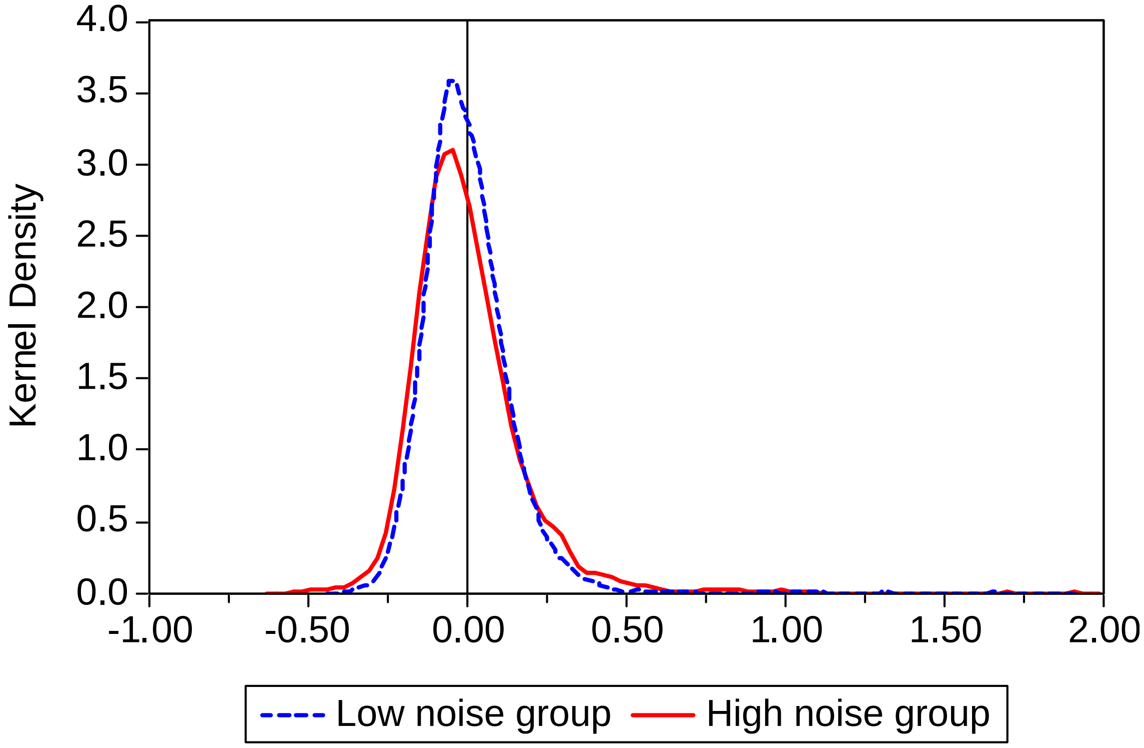


Figure 4: Kernel densities of the residuals of regressing APC on store-fixed effects, by noise-group



Based on APC-data from 57 weeks prior to the tournament.
The graph is cut off at 2 for reasons of visibility.
The kernel density of the high-variance group runs to 3.6,
based on 6 observations between 2 and 3.6.

Figure 5: Kernel density of performance during the second round

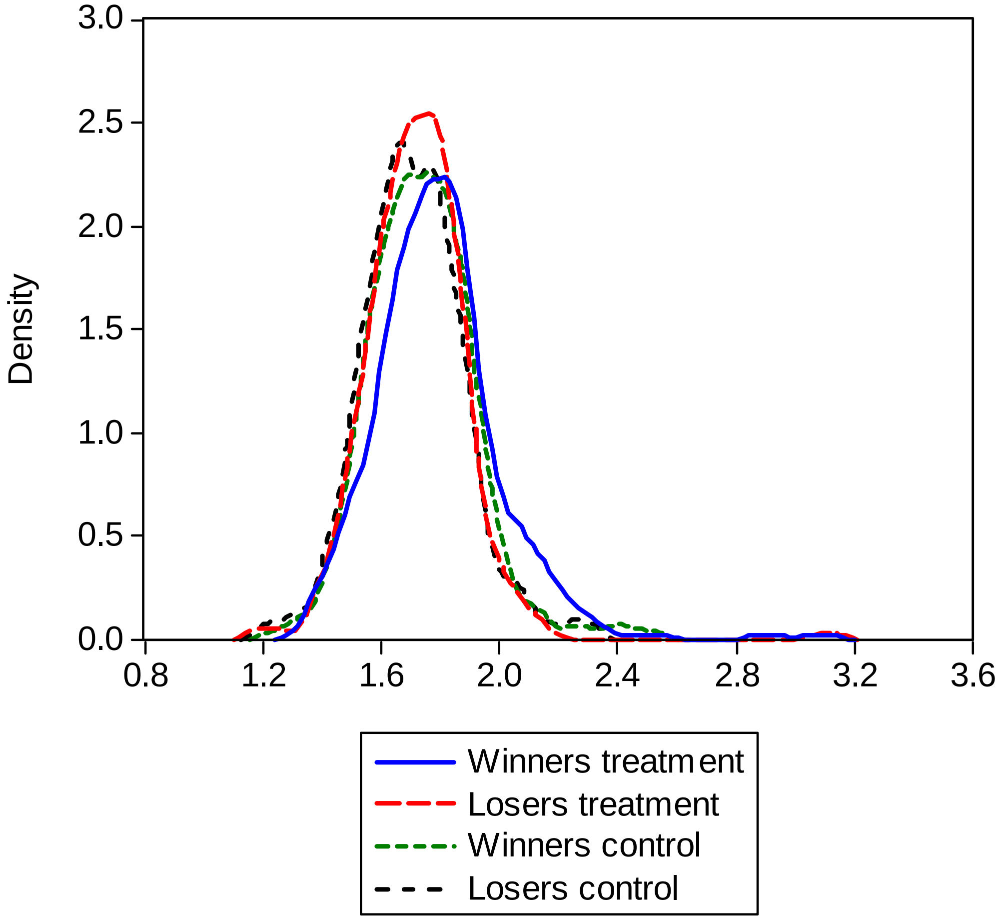
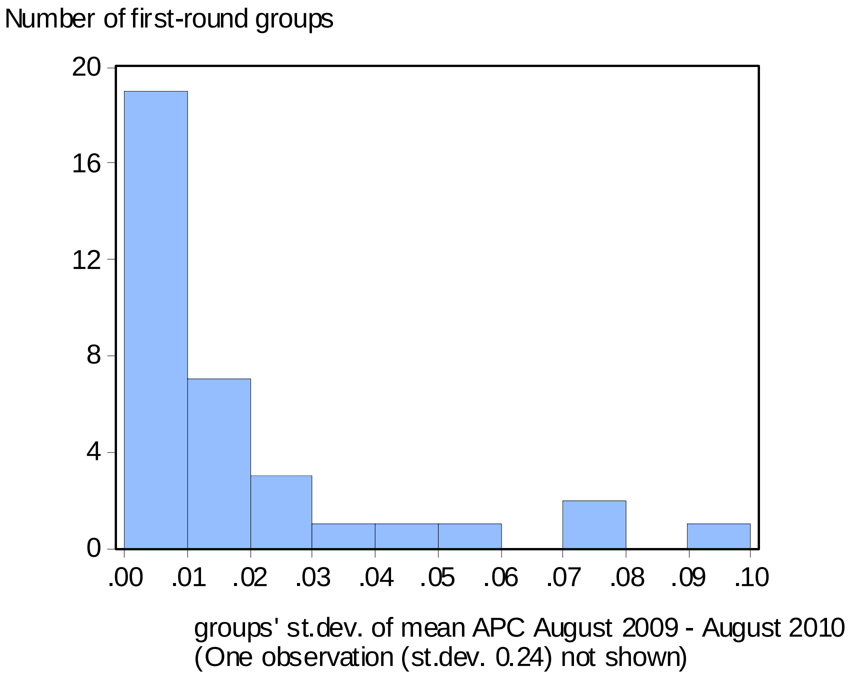


Figure 6: Distribution of groups' standard deviation of stores' mean historical performance



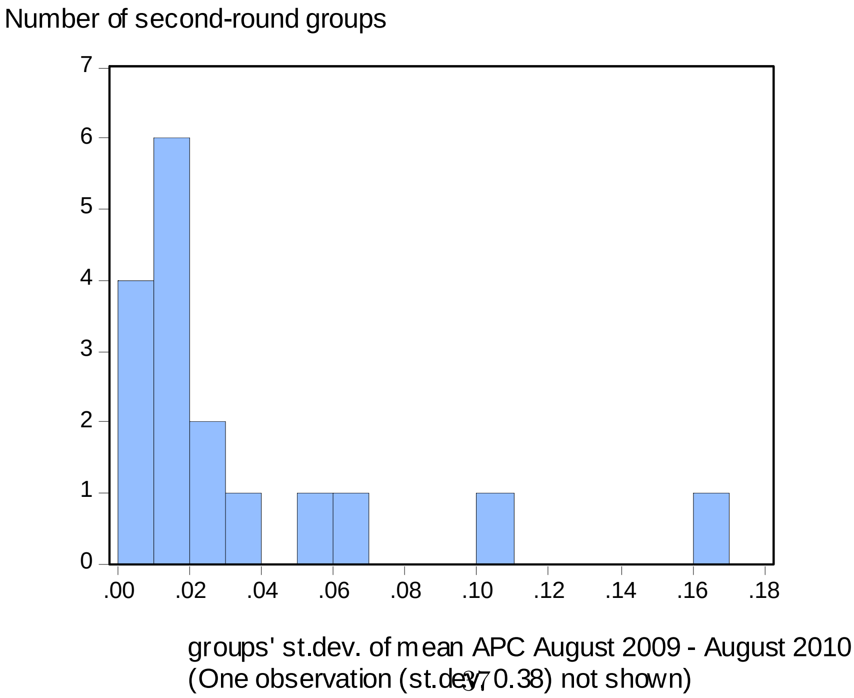
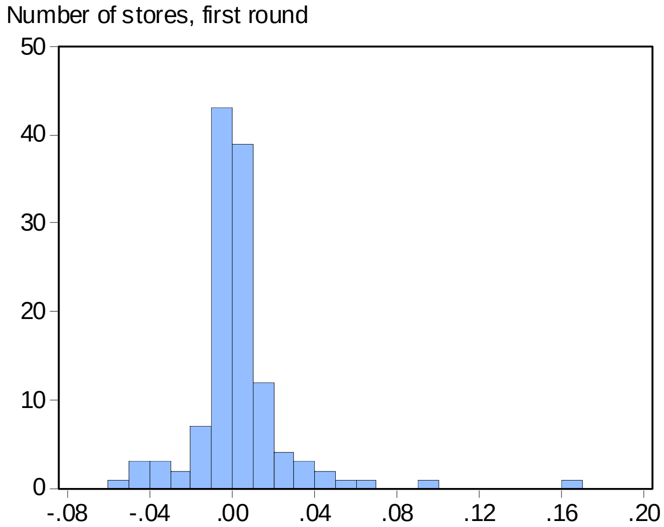
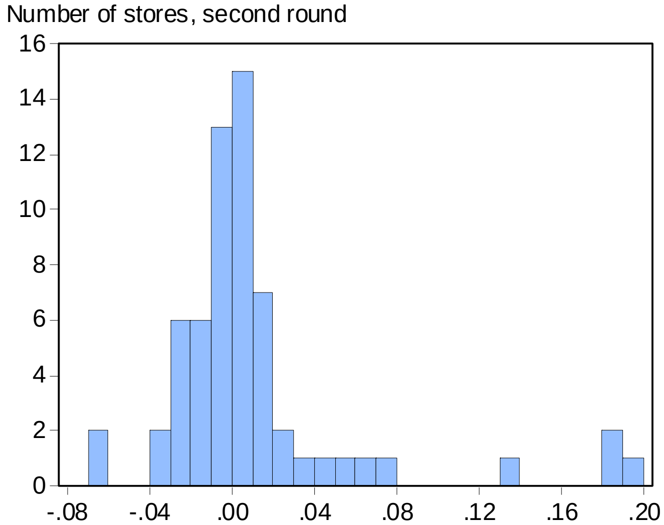


Figure 7: Distribution of stores' leads and lags in mean historical performance relative to the average of the historical performance of the stores ranked second and third in their group



Difference between a store's mean historical APC and the average mean historical APC of the stores ranked second and third in the group
(One observation not shown (value 0.56))



Difference between a store's mean historical APC and the average mean historical APC of the stores ranked second and third in the group
(One observation not shown (value 0.74))

Table 1: Descriptive statistics

	All stores		Control group		Treatment group: Low-spread		Treatment group: High-spread		Noise group: Low noise		Noise group: High noise	
	mean	Std	mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Performance (store averages)												
APC (Average number of Products per Customer), all weeks++	1.82	0.10	1.81	0.09	1.82	0.12	1.83	0.08	1.81	0.08	1.85	0.12
APC assignment period (weeks 32/2009 - 30/2010)++	1.83	0.10	1.82	0.09	1.83	0.12	1.83	0.08	1.81	0.08	1.85	0.12
APC Round 1 (weeks 36/2010 - 39/2010)	1.78	0.10	1.77	0.09	1.80	0.12	1.79	0.08	1.79	0.10	1.80	0.10
APC Round 2 (weeks 42/2010 - 45/2010)	1.75	0.13	1.73	0.11	1.76	0.15	1.76	0.12	1.76	0.13	1.75	0.14
APC Round 2, first-round winners					1.79	0.18	1.80	0.11	1.79	0.15	1.80	0.15
APC Round 2, first-round losers					1.72	0.11	1.72	0.11	1.73	0.11	1.71	0.11
Noise												
Within-store standard deviation of APC (noise)+++ in the assignment period (weeks 32/2009 - 30/2010)	0.15	0.06	0.15	0.07	0.14	0.04	0.15	0.07	0.12	0.01	0.18	0.06
Store characteristics												
Number of employees	5.03	2.17	4.93	1.82	5.00	2.11	5.15	2.55	5.11	2.55	5.03	2.11
Percentage female employees**	0.36	0.25	0.31	0.25	0.42	0.25	0.36	0.25	0.39	0.25	0.39	0.25
Mean tenure of employees (months)	38.09	27.96	35.11	31.06	38.55	29.90	40.56	22.37	38.68	28.83	40.42	23.74
Mean age of employees	24.49	3.68	24.41	3.91	24.32	3.76	24.74	3.40	24.86	4.05	24.20	3.02
Number of stores	186		62		62		62		62		62	

For one store in the control group, store characteristics were not available.
***, **, * denote statistically significant differences at the 1%, 5%, and 10% level, respectively, between control, low-spread and high-spread stores (F-test).
+++, ++, + denote statistically significant differences at the 1%, 5%, and 10% level, respectively, between control, low-noise and high-noise stores (F-test).

Table 2: Average treatment effect

	Dependent variable: APC	
	(1)	(2)
Treatment * (Round 1 + Winners * Round 2)	0.028	
	(0.011)**	
Treatment * Round 1		0.014
		(0.011)
Winners * Round 2		0.047
		(0.018)***
Losers * Round 2	-0.014	-0.005
	(0.015)	(0.015)
Store-fixed effects	yes	yes
Week-fixed effects	yes	yes
Store-week observations	12079	12079
Stores	186	186
R^2	0.4471	0.4473

Standard errors clustered at the store level in parentheses.
Winners and Losers refer to the outcome of the first round of the tournament for the treatment stores.
***, **, * denote statistically significant effects at the 1%, 5%, and 10% level, respectively.

Table 3: Carry-over effects

	Dependent variable: APC	
	(1)	(2)
Treatment * (Round 1 + Winners * Round 2)	0.029	
	(0.013)**	
Treatment * (Round 1 + Winners * Round 2) *	-0.006	
Participant earlier experiment	(0.016)	
Treatment * Round 1		0.017
		(0.011)
Treatment * Round 1 * Participant earlier experiment		-0.016
		(0.014)
Winners * Round 2		0.043
		(0.021)**
Winners * Round 2 * Participant earlier experiment		0.020
		(0.028)
Losers * Round 2	-0.014	-0.005
	(0.015)	(0.015)
Store-fixed effects	yes	yes
Week-fixed effects	yes	yes
Store-week observations	12079	12079
Stores	186	186
R^2	0.4471	0.4474

Standard errors clustered at the store level in parentheses.
Winners and Losers refer to the outcome of the first round of the tournament for the treatment stores.
Participant earlier experiment is a dummy variable that takes value 1 for treatment stores assigned to the treatment group in an earlier experiment ran in February 2010.
***, **, * denote statistically significant effects at the 1%, 5%, and 10% level, respectively.

Table 4: Noise in performance and the treatment effect

	Dependent variable: APC	
	(1)	(2)
Treatment * (Round 1 + Winners * Round 2)	0.024	
	(0.010)**	
Treatment * (Round 1 + Winners * Round 2) * StDev	-0.321	
	(0.198)	
(Round 1 + (Winners + Control) * Round 2) * StDev	-0.318	
	(0.104)***	
Treatment * Round 1		0.011
		(0.010)
Treatment * Round 1 * StDev		-0.188
		(0.199)
Round 1 * StDev		-0.365
		(0.111)***
Winners * Round 2		0.041
		(0.017)**
Winners * Round 2 * StDev		-0.632
		(0.282)**
(Winners + Control) * Round 2 * StDev		-0.269
		(0.140)*
Losers * Round 2	-0.015	-0.006
	(0.014)	(0.014)
Store-fixed effects	yes	yes
Week-fixed effects	yes	yes
Store-week observations	12079	12079
Stores	186	186
R^2	0.4494	0.4497

Standard errors clustered at the store level in parentheses.
Winners and Losers refer to the outcome of the first round of the tournament for the treatment stores.
StDev is a store's standard deviation of APC over the period August 2009 to August 2010.
This variable is mean-centered.
***, **, * denote statistically significant effects at the 1%, 5%, and 10% level, respectively.

Table 5: Noise uncorrelated with observables

	Dependent variable: APC	
	(1)	(2)
Treatment * (Round 1 + Winners * Round 2)	0.022	
	(0.011)**	
Treatment * (Round 1 + Winners * Round 2) * Residual noise	-0.342	
	(0.212)	
(Round 1 + (Winners + Control) * Round 2) * Residual noise	-0.225	
	(0.168)	
Treatment * Round 1		0.009
		(0.010)
Treatment * Round 1 * Residual noise		-0.203
		(0.184)
Round 1 * Residual noise		-0.230
		(0.150)
Winners * Round 2		0.036
		(0.017)**
Winners * Round 2 * Residual noise		-0.800
		(0.409)*
(Winners + Control) * Round 2 * Residual noise		-0.219
		(0.211)
Losers * Round 2	-0.015	-0.006
	(0.014)	(0.014)
Store-fixed effects	yes	yes
Week-fixed effects	yes	yes
Store-week observations	12017	12017
Stores	185	185
R^2	0.4485	0.4489

Standard errors clustered at the store level in parentheses.
Winners and Losers refer to the outcome of the first round of the tournament for the treatment stores.
Residual noise refers to the residuals of the OLS regression of stores' standard deviation of APC on all observable store-characteristics, as presented in Table A1. This variable is mean-centered.
***, **, * denote statistically significant effects at the 1%, 5%, and 10% level, respectively.

Table 6: Estimated treatment effects: prize spread

	Dependent variable: APC		
	(1)	(2)	(3)
Low spread * (Round 1 + Winners * Round 2)	0.030		
	(0.016)*		
High spread * (Round 1 + Winners * Round 2)	0.026		
	(0.013)**		
Low spread * Round 1		0.021	
		(0.015)	
High spread * Round 1		0.006	
		(0.010)	
Control * Pseudo-winners * Round 1			0.072
			(0.012)***
Low spread * Losers * Round 1			0.015
			(0.012)
Low spread * Winners * Round 1			0.094
			(0.022)***
High spread * Losers * Round 1			0.011
			(0.013)
High spread * Winners * Round 1			0.071
			(0.012)***
Low spread * Winners * Round 2		0.038	0.041
		(0.025)	(0.025)
High spread * Winners * Round 2		0.057	0.058
		(0.023)**	(0.023)**
Losers * Round 2	-0.014	-0.005	-0.007
	(0.015)	(0.015)	(0.015)
Store-fixed effects	yes	yes	yes
Week-fixed effects	yes	yes	yes
Store-week observations	12079	12079	12079
Stores	186	186	186
R^2	0.4471	0.4474	0.4495

Standard errors clustered at the store level in parentheses.
Winners and Losers refer to the outcome of the first round of the tournament for the treatment stores.
Control * Pseudo-winners refers to the stores in the control group that 'won' the pseudo-competition.
Reference category for first-round effects in Column 3 are the 'losers' of the pseudo-competition.
***, **, * denote statistically significant effects at the 1%, 5%, and 10% level, respectively.

Table 7: Robustness checks: heterogeneity

	Dependent variable: APC		
	(1)	(2)	(3)
Low spread * Round 1	0.021	0.018	0.027
	(0.015)	(0.014)	(0.014)**
High spread * Round 1	0.006	-0.003	0.008
	(0.010)	(0.011)	(0.011)
Low spread * Winners * Round 2	0.038	0.043	0.053
	(0.025)	(0.025)*	(0.026)**
High spread * Winners * Round 2	0.057	0.056	0.064
	(0.023)**	(0.022)**	(0.023)***
Treatment * (Round 1 * First-round group's st.dev. +		-0.468	
+ Winners * Round 2 * Second-round group's st.dev.)		(0.227)**	
Treatment * (Round 1 * First-round distance to threshold (ahead) +			-0.420
+ Winners * Round 2 * Second-round distance to threshold (ahead))			(0.049)***
Treatment * (Round 1 * First-round distance to threshold (behind) +			-0.214
+ Winners * Round 2 * Second-round distance to threshold (behind))			(0.585)
Losers * Round 2	-0.005	-0.005	-0.004
	(0.015)	(0.015)	(0.015)
Store-fixed effects	yes	yes	yes
Week-fixed effects	yes	yes	yes
Store-week observations	12079	12079	12079
Stores	186	186	186
R^2	0.4471	0.4485	0.4489

Standard errors clustered at the store level in parentheses.
Winners and Losers refer to the outcome of the first round of the tournament for the treatment stores.
First-Round and Second-round group's st.dev. refers to measure η_i, the standard deviation of stores' mean performance APC before the experiment (August 2009 - August 2010) by group, in the first and second round, respectively. This variable is mean-centered.
First-round and second distance to threshold (ahead) refer to measure Δ^+, the difference between a store's mean performance before the experiment (APC over the period August 2009 to August 2010) and the average of the mean historical performance of the stores ranked second and third in the group, for stores that are above this threshold (ranked first or second in the group on the basis of historical performance), in the first and second round, respectively.
First-round and second distance to threshold (behind) is similarly defined for stores that lie behind (ranked third and fourth in the group).
***, **, * denote statistically significant effects at the 1%, 5%, and 10% level, respectively.

Table A1: OLS of store characteristics on noise

	Dependent variable: standard deviation of APC
	(1)
Mean APC	0.183
	(0.044)***
Number of employees	0.001
	(0.003)
Number of employees in full-time equivalents	0.000
	(0.000)
Average age employees	-0.001
	(0.002)
Percentage of female employees	0.010
	(0.017)
Average tenure of employees	0.005
	(0.003)*
Brand 2	-0.033
	(0.018)*
Constant	-0.165
	(0.083)**
Regional dummies	yes
Stores	185
R^2	0.2453

Standard errors in parentheses.
Mean standard deviation of APC are based on the period August 2009 to August 2010.
The personnel variables are extracted from the companies' database as of September 1, 2010.
The personnel information is missing for 1 store in the analysis.
Brand 2 is a dummy variable for stores operating under the companies' smaller brand name.
***, **, * denote statistically significant effects at the 1%, 5%, and 10% level, respectively.