TI 2010-124/1
Tinbergen Institute Discussion Paper

Dynamic Incentive Effects of Relative Performance Pay: A Field Experiment

Josse Delfgaauw[1,2],
Robert Dur[1,2,3]
Arjan Non[1,2]
Willem Verbeke[1,4]

[1] *Erasmus School of Economics, Erasmus University Rotterdam;*
[2] *Tinbergen Institute;*
[3] *CESifo, and IZA;*
[4] *ERIM.*

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Most TI discussion papers can be downloaded at http://www.tinbergen.nl.

Dynamic Incentive Effects of Relative Performance Pay: A Field Experiment

Josse Delfgaauw[*], Robert Dur[†], Arjan Non[‡], and Willem Verbeke[§]

December 2010

Abstract

We conduct a field experiment among 189 stores of a retail chain to study dynamic incentive effects of relative performance pay. Employees in the randomly selected treatment stores could win a bonus by outperforming three comparable stores from the control group over the course of four weeks. Treatment stores received weekly feedback on relative performance. Control stores were kept unaware of their involvement, so that their performance generates exogenous variation in the relative performance of the treatment stores. As predicted by theory, treatment stores that lag far behind do not respond to the incentives, while the responsiveness of treatment stores close to winning a bonus increases in relative performance. On average, the introduction of the relative performance pay scheme does not lead to higher performance.

JEL-Codes: C93, M52.

Keywords: Dynamic incentives; Relative performance pay; Field experiment.

[*]Erasmus University Rotterdam and Tinbergen Institute. E-mail: delfgaauw@ese.eur.nl.
[†]Erasmus University Rotterdam, Tinbergen Institute, CESifo, and IZA. E-mail: dur@ese.eur.nl.
[‡]Erasmus University Rotterdam and Tinbergen Institute. E-mail: janon@ese.eur.nl.
[§]Erasmus University Rotterdam and ERIM. E-mail: verbeke@ese.eur.nl.

1 Introduction

Non-linear pay-for-performance plans have dynamic incentive effects when employees receive intermediate performance information over the course of the incentive period. For instance, consider a salesman who can earn a bonus by attaining a monthly sales target while receiving daily or weekly sales figures. When realised sales during the month are such that it remains challenging but possible to reach the target, the bonus scheme provides strong incentives. The incentive effect is much weaker, however, when realised sales during the month are particularly high or low. High intermediate sales imply that the salesman can hardly miss the bonus, while low intermediate sales imply that the target is practically out of reach.

More generally, workers can use intermediate performance information to determine how much additional performance is necessary to obtain a bonus. This creates dynamic incentive effects, where the incentive effect of the pay-for-performance plan at each point in time depends on realised performance until then. Incentive plans based on relative performance, where prizes are awarded for outperforming sufficiently many competitors, are particularly prone to dynamic incentive effects. Sports leagues are a common example. In the workplace, examples range from employee-of-the-month contests, to beat-the-index bonuses for stock brokers, and to job promotion contests.[1]

Casas-Arce and Martinez-Jerez (2009) show formally that for contests with a large number of participants, the incentive effect of a relative performance incentive scheme is hump-shaped in lagged relative performance. Competitors who find themselves trailing far behind may perceive catching up to be impossible and consequently give up trying. Similarly, competitors who are far ahead may perceive losing as impossible and slack off as well. In contrast, incentives are highly salient for competitors who find themselves almost tied in intermediate performance. Analyzing sales contests among retailers of a commodities company, Casas-Arce and Martinez-Jerez (2009) find indeed that competitors in winning positions reduce performance when their lead increases. However, the performance of trailing competitors does not decrease when they lag further behind.[2]

[1] As a concrete example, more than half of the remuneration of the executive directors of oil company Shell is based on a ranking of Shell's performance relative to its four main competitors on four publicly available measures. The incentive plan has a three-year horizon, during which the companies regularly release the latest figures with respect to these performance measures (Royal Dutch Shell, 2009).

[2] Relatedly, Fershtman and Gneezy (2009) let kids run side-by-side and find that increasing

Testing for the presence and strength of dynamic incentive effects is hampered by two issues. First, in small contests, a competitor's optimal strategy depends on (its perception of) its competitors' strategies. A trailing competitor may be best off by accepting its loss when the other competitors keep effort high, but not when they would slack off. Second, serial correlation in performance biases estimates of the effect of intermediate relative performance on subsequent performance.

In this paper, we take a unique approach in tackling both issues by setting up a relative performance pay scheme where only one of the 'competitors' can earn a prize, while the other participants are kept unaware of their involvement. This implies that the strategies of all non-competing participants are exogenous, allowing us to use their performance as an instrument for intermediate relative performance of the competing participant. More specifically, we study the dynamic incentive effects of this relative performance pay scheme by conducting a natural field experiment in a Dutch retail chain. We provide the employees of 93 stores randomly selected from 189 of the company's stores with the opportunity to earn a bonus. The bonus is awarded when a treatment store outperforms three comparable stores *from the control condition* over the course of a four-week period (February 2010). Each week, treatment stores receive a poster with the performance of all four stores in their group. Importantly, the employees of the three comparison stores cannot earn a bonus, do not learn that another store can earn a bonus by beating their performance, and do not receive any relative performance feedback. This way the treatment stores compete against stores that are not competing. This allows us to use the performance of the three comparison stores as an instrument for trailing behind or being ahead: lagged performance of the comparison stores does affect intermediate relative performance, but does not affect current performance of the treatment store other than through lagged relative performance. Hence, using this instrument, our estimates are not biased by serial correlation in stores' own performance.

Our results are as follows. First, we find a positive effect of intermediate relative performance on current performance for stores close to the target, particularly in the last two weeks of the experiment. This effect is substantial: a one percentage point increase in intermediate relative performance increases current performance

incentives yield higher performance but also a higher fraction of kids giving up during the race. Following the early literature on tournament theory (Lazear and Rosen, 1981, Green and Stokey, 1983, Nalebuff and Stiglitz, 1983), most of the literature has abstracted from dynamic incentive effects of tournaments. A recent string of theoretical papers studies the cost and benefit to a principal of providing intermediate relative performance feedback during a contest between his agents (Aoyagi 2010, Ederer 2010, Gershkov and Perry 2009, Goltsman and Mukherjee 2010).

by 0.72 percent. Stores lagging far behind do not respond to intermediate relative performance. This suggests that the employees in these stores gave up trying to win. Hence, as predicted by theory, we find that intermediate relative performance matters more for competitors that perform close to target than for competitors that lag far behind. During the contest, hardly any treatment store managed to get far ahead of all its comparison stores. Hence, we cannot test the hypothesis that high-performers slack off as their lead increases.

Second, we find no average treatment effect of introducing the contest, neither for the four weeks taken together nor for one of the weeks separately. This contrasts with several recent findings on the incentive effects of tournaments. In another retail chain, we do find a substantial positive effect of introducing a standard tournament among shops (Delfgaauw et al., 2009), as do Erev et al. (1993) and Bandiera et al. (2009) among teams of fruit pickers and Casas-Arce and Martinez-Jerez (2009) among retailers of a commodities company. Even more striking, several recent papers suggest that the mere provision of relative performance feedback can be sufficient to trigger higher performance (Azmat and Iriberri (2010), Blanes i Vidal and Nossol (2009), Delfgaauw et al. (2009), Kosfeld and Neckermann (2010)). Bandiera et al. (2009) obtain an opposite result. A possible explanation for our divergent result is that beating unaware contestants, as in our setting, is less exciting than beating competing contestants.

Our experiment involves one incentive period of four weeks. When incentive schemes are repeated over time, as with monthly or year-on-year targets, other types of dynamic incentive effects may arise. For instance, sales may be shifted forward or backward in time around the incentive commencement date in order to meet the current target or to alleviate the difficulty of meeting the next target; see Asch (1990) and Oyer (1998) for empirical evidence and Cadsby et al. (2010) for a related lab experiment. Furthermore, when the targets in repeated incentive schemes are based on historical performance, workers have an incentive to beat the target by only a limited amount even it would be possible to greatly outperform the target. Bouwens and Knoops (2010) find evidence in line with such ratchet effects, using store-level data from a retail chain. Cooper et al. (1999) and Charness et al. (2010) find ratchet effects in the lab. Ratchet effect considerations may explain why we find no average treatment effect, as workers may have feared that a strong response to the introduction of the relative performance pay scheme would result in higher targets in their regular incentive scheme.

2 Experimental design

The experiment took place in February 2010 in a retail chain in The Netherlands that sells computer games, music, and movies. At the start of 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. The company's central management decides on the range of products sold, pricing, and advertisement. Store managers are responsible for day-to-day operations. Employees receive rather weak incentive pay on top of their base salary, based on their shop's yearly sales growth and a subjective performance evaluation. The company's management was not satisfied with the effects of this incentive scheme and wished to learn more about the effects of short-term incentives, in particular of sales contests.

We designed a relative performance incentive scheme to be implemented in a randomly selected subset of stores (the treatment condition), while the rest of the stores comprised the control condition. All employees (including the shop manager) of a store in the treatment condition could earn a bonus by sufficiently outperforming three preselected stores from the control condition. Stores in the control condition could not earn a bonus, and employees in the treatment stores were informed about this. Performance is measured as cumulative sales revenue in percentage deviation of budgeted sales in February 2010 (a period of 4 weeks).[3] Let $y_{s,w}$ be sales and $b_{s,w}$ budgeted sales of store s in week w, respectively. Weekly performance $p_{s,w}$ is given by

$$p_{s,w} = \frac{y_{s,w} - b_{s,w}}{b_{s,w}} \cdot 100\% \tag{1}$$

and cumulative performance over February 2010 is given by

$$p_s^{CU} = \frac{\sum_{w=E1}^{E4} y_{s,w} - \sum_{w=E1}^{E4} b_{s,w}}{\sum_{w=E1}^{E4} b_{s,w}} \cdot 100\% \tag{2}$$

where the summation is over the four experimental weeks $E1$ to $E4$ (i.e. week 5, 2010 to week 8, 2010).

All employees of a treatment store received a bonus of gross 150 euro when their

[3] The budgeted sales are forecasts for shops' weekly sales as determined by the company's management in October 2009 (at the start of the financial year) for a year onwards. These budgeted sales boil down to a forecast for total sales of the whole chain, with each store expected to bring in a fixed share of total sales. Hence, a combination of week and store fixed effects explains all variation in the log of budgeted sales in our data. The company gives shop managers weekly feedback on sales relative to budgeted sales, which makes it a natural measure of performance.

shop's performance in February 2010 was at least 10 percentage points higher than the performance of all three comparison stores. When a treatment store scored between 5 and 10 percentage points above all three comparison stores, its employees received 75 euro.[4] Lastly, outperforming all three comparison stores by less than 5 percentage points yielded a cake for the treatment store, but only if the treatment store also performed above budget.[5]

All communication on the experiment towards the shops went through the company's regular channels, so shop managers and employees were not aware of our involvement. Hence, our experiment classifies as a natural field experiment (Harrison and List, 2004). In January 2010, the company informed all stores that a randomly selected set of stores would get the opportunity to earn a bonus in February 2010, and that all other stores could look forward to a similar opportunity later that year. On January 22, the treatment stores were informed about the details of the relative performance incentive scheme.

During the experiment, we provided weekly feedback to the treatment stores on their relative performance in the form of a poster. The poster contained the cumulative sales relative to budget figures of the treatment shop and its three comparison shops, ranked in descending order. Furthermore, on Monday February 1, all treatment stores received a large poster, with room to glue on the four posters with weekly rankings to be received in the following weeks. Store managers were instructed to put up these posters in the store's canteen.[6] Stores in the control condition did not receive posters, nor any other type of relative performance information.

Our design has two advantages over a regular competition. First, as treatment stores only receive a bonus when they outperform comparable stores from the control condition, the payout is relatively low when the incentive has little effect on performance. This was seen as a major benefit by the company's management. Second, performance of the comparison stores is exogenous to the incentive scheme, as these stores could neither earn a bonus, nor received any relative performance feedback, and were not aware that their performance played a role in the incentive scheme.

[4] For employees who did not have a full-time contract, the size of the bonus was proportional to the contractual number of hours. Hourly wages for personnel in the shops are close to the minimum wage, which makes that receiving the high bonus would increase monthly earnings by about 10%.

[5] The latter requirement only applied for the cake, not for any of the two bonusses. This requirement was a last-minute addition by the company's management to the rules.

[6] The company's regional managers were instructed to verify that all store managers actually put up the posters in the canteen. We have not heard about a single store manager who refused to do so.

We exploit differences in comparison stores' performance during the experiment to analyse how treatment stores' intermediate relative performance affects the effect of the incentive scheme in subsequent weeks.

We used weekly sales and budget data of 194 stores[7] for the weeks 40 to 53 in 2009 to assign stores to the treatment and control conditions, and to match treatment stores with comparable stores from the control condition, as follows. First, we created four equally large strata based on store size as measured by average weekly sales revenues. Randomly, half of the stores in each stratum was assigned to the treatment condition, while the remaining half of the stores were assigned to the control condition. Subsequently, we matched each treatment store to three control stores from the same stratum. To reduce opportunities for collusion, we imposed that each treatment store was matched to control stores located in other regions, as there is frequent communication between stores within a region. Apart from this regional separation, treatment stores were matched to the control stores that were most comparable in terms of the performance measure (cumulative sales revenue relative to the budget) for the period of week 40 to week 53 in 2009. Note that a control store can be matched to multiple treatment stores. After this assignment procedure, we excluded one treatment store from the experiment as its budget figures turned out to be too unreliable. Furthermore, 3 treatment stores and 1 control store were shut down in January 2010. This leaves us with 93 stores in the treatment condition and 96 stores in the control condition. For each of these stores, we have weekly sales and budget figures for a period of in total 22 weeks, from week 40 in 2009 to week 8 in 2010. In some estimations, we separate the stores by size, by collapsing the two strata with the biggest stores together as well as the two strata with smallest stores. This yields 97 large stores and 92 small stores.

Figures 1 and 2 show weekly sales and weekly performance, respectively, averaged over all stores. Average weekly sales show two spikes in December 2009, related to Sinterklaas and Christmas festivities, respectively. Average performance hovers between plus and minus 20 percent.

The descriptive statistics in Table 1 show that average sales does not differ between treatment stores and control stores, neither for the whole period nor for the first 14 weeks in the data used to stratify the stores. The same holds for budgeted sales and for performance as measured by (1). Note that on average, sales are below

[7] The company's management excluded a specific group of 14 stores from participating in the experiment.

budget, but that variation in average performance across stores is large. Further, the number of employees per store does not differ significantly between the treatment and control stores. Lastly, in week 7 of 2010, a total of 29 stores were closed for one or two days in relation to carnival festivities, mainly in the south of The Netherlands. Treatment stores were slightly more often closed than control stores, but not significantly so. In all estimations below, we correct for the effect of carnival.

As a first hint of the overall effect of the relative performance incentive, Table 1 shows that there is no difference in average sales figures between treatment and control stores for the weeks with the bonus scheme (week 5, 2010 to week 8, 2010). Figures 3 and 4 provide further insight into the overall treatment effect, by plotting the differences between the treatment and control condition in average sales and in average performance, respectively, by week. The experiment took place in the final four weeks of the period shown. Both figures show no sign of a positive treatment effect, possibly with the exception of the final week. A second hint of the overall effect is given by the fact that only 13 stores earned a prize: 5 stores earned the high bonus, another 5 stores earned the low bonus, and three stores were entitled to cake.

3 Method

We assess the average effect of the relative performance incentive scheme using OLS with week and store-fixed effects, by estimating

$$\ln(y_{s,w}) = \alpha_s + \theta_w + \gamma B_{s,w} + \kappa F_{s,w} + \varepsilon_{s,w} \tag{3}$$

where $\ln(y_{s,w})$ is the log of sales of store s in week w. Store and week-fixed effects are given by α_s and θ_w, respectively. $B_{s,w}$ is a dummy variable that takes the value one for stores in the treatment condition from week 5 to week 8 in 2010, $F_{s,w}$ measures the number of days shop s is closed for carnival festivities in week w (this variable takes positive values only in week 7, 2010), and $\varepsilon_{s,w}$ is an error term. We estimate the effect on sales rather than on the performance measure (1) used in the incentive scheme. Budgeted sales are set in advance by the company's central management and cannot be affected by stores. This implies that shops can affect their performance only through sales. As effects on sales are more easily interpreted, we use sales to measure the effects of the relative performance incentive scheme.[8]

The main goal of this paper is to analyse the effect of intermediate relative performance on subsequent performance. First, we introduce some additional notation. Let T and C be the sets of stores in the treatment and the control condition, respectively. Further, denote by $c_t \in C$ a control store matched to treatment store $t \in T$. Lastly, let $p^{CU}_{s,w-1}$ denote the *cumulative* performance of store s during the experiment up to but not including week w, as measured by cumulative sales over budget in February 2010:

$$p^{CU}_{s,w-1} = \frac{\sum_{w=E1}^{w-1} y_{s,w} - \sum_{w=E1}^{w-1} b_{s,w}}{\sum_{w=E1}^{w-1} b_{s,w}} \cdot 100\%. \tag{4}$$

Hence, $p^{CU}_{s,w-1}$ is the performance figure for store s as depicted on the poster received at the start of week w during the experiment. The effect of intermediate performance of treatment stores relative to the best-performing comparison store on subsequent sales can be estimated by

$$\ln(y_{s,w}) = \alpha_s + \theta_w + \gamma B_{s,w} + \mu \left(p^{CU}_{t,w-1} - \max_{c_t} \left[p^{CU}_{c_t,w-1} \right] \right) B_{s,w} + \kappa F_{s,w} + \varepsilon_{s,w} \tag{5}$$

where the term $p^{CU}_{t,w-1} - \max_{c_t} \left[p^{CU}_{c_t,w-1} \right]$ gives the difference in cumulative performance during the experiment between treatment store t and its best-performing comparison store c_t up to and including the previous week.[9] Since the experiment lasted four weeks, we have three intermediate relative performance figures per treatment store, corresponding to a total of 279 treatment store-week observations. Control stores cannot earn a bonus and do not receive posters with rankings. Hence, $\max_{c_t} \left[p^{CU}_{c_t,w-1} \right]$ is fully exogenous to $y_{t,w}$. However, in case of serial correlation in the error structure of sales, $p^{CU}_{t,w-1}$ is correlated with $y_{s,w}$ (see (4)). Estimating (5) without taking account of serial correlation would yield a biased estimate of μ. Therefore, we instrument the difference in intermediate performance $p^{CU}_{t,w-1} - \max_{c_t} \left[p^{CU}_{c_t,w-1} \right]$ by the expected difference

$$D_{t,w-1} = E \left[p^{CU}_{t,w-1} \right] - \max_{c_t} \left[p^{CU}_{c_t,w-1} \right]. \tag{6}$$

The expected cumulative performance of treatment store t in the experiment $E \left[p^{CU}_{t,w-1} \right]$ is set equal to the average performance in the 18 weeks prior to the start of the ex-

[8] Using performance (1) as dependent variable instead yields similar results. This is unsurprising given the way stores' budgeted sales are determined, see footnote 3.

[9] This variable is set to zero for control stores.

periment (week 40, 2009 to week 4, 2010) while accounting for week-fixed effects in performance during the experiment:

$$E\left[p_{t,w-1}^{CU}\right] = \frac{1}{18}\sum_{w=1}^{18} p_{t,w} + \frac{\sum_{w=E1}^{w-1} b_{t,w}\theta_w^p}{\sum_{w=E1}^{w-1} b_{t,w}} \tag{7}$$

where θ_w^p is the week-fixed effect from estimating

$$p_{s,w} = \alpha_s^p + \theta_w^p + \gamma^p B_{s,w} + \kappa^p F_{s,w} + \varepsilon_{s,w}^p$$

with superscript p denoting that the estimates relate to performance as dependent variable.[10] Most importantly, this implies that for each treatment store, variation in $D_{t,w-1}$ across experimental weeks stems solely from variation in $p_{c_t,w-1}^{CU}$ and in the weighted week-fixed effects, which are both unrelated to $\varepsilon_{t,w}$ given the design of our experiment.

Equation (5) estimates a linear effect of intermediate relative performance. However, the incentive scheme is likely to have the biggest effect when treatment stores learn that they are close to the relative performance targets for winning a bonus (Casas-Arce and Martinez-Jerez, 2009). Treatment stores lagging far behind in the intermediate ranking may give up, and treatment stores far ahead may reduce their efforts when they anticipate that they can hardly miss the bonus. In the course of the experiment, we have many treatment stores that face an uphill battle, while there are only few stores that are comfortably ahead. In total, we have only 8 store-week observations where treatment stores' intermediate relative performance is more than 10 percentage points above the target for the high bonus (i.e. with $p_{t,w-1}^{CU} - \max_{c_t}\left[p_{c_t,w-1}^{CU}\right] > 0.20$). This implies that we cannot test whether stores that greatly outperform their comparison stores reduce their efforts.[11] We can test whether stores that lag far behind reduce their efforts, by allowing the effect of intermediate relative performance on current performance to differ between stores that lag far behind and stores that are still in the running.

In determining which stores still have a chance of earning a bonus, we cannot use the actual difference between the lagged cumulative performance of the treatment store and its best control, as given by (4). Serial correlation in $y_{t,w}$ would bias the estimates. Hence, we again use the estimated difference (6) to determine

[10] We weight the week-fixed effects by budgeted sales $b_{t,w}$ to account for the fact that weeks with a higher absolute budgeted sales volume have a higher weight in cumulative performance, see (4).

[11] Excluding these 8 observations from the analysis does not affect any of the results.

stores' chances of earning a bonus. Rather arbitrarily, we set the bar for being too far behind at a 5 percentage point lag relative to the best performing comparison store. Note that stores that lag 5 percentage points behind need to improve their relative performance by 5 percentage points in order to win a cake and by at least 10 percentage points to obtain a bonus. We do vary the bar to assess the robustness of the results. Let $I_{t,w-1}$ be a dummy that takes value 1 for treatment stores whenever $D_{t,w-1} > -0.05$ and zero otherwise. This yields 52 store-week observations where $I_{t,w-1} = 1$, out of a total of 279 treatment store-week observations with intermediate relative performance figures. We estimate

$$\begin{aligned} \ln(y_{s,w}) \quad = \quad & \alpha_s + \theta_w + \gamma B_{s,w} + \mu \left(p^{CU}_{t,w-1} - \max_{c_t} \left[p^{CU}_{c_t,w-1} \right] \right) B_{s,w} + \\ & + \delta I_{t,w-1} B_{s,w} + \nu \left(p^{CU}_{t,w-1} - \max_{c_t} \left[p^{CU}_{c_t,w-1} \right] \right) I_{t,w-1} B_{s,w} + \kappa F_{s,w} + \varepsilon_{s,w} \end{aligned} \tag{8}$$

again instrumenting the difference in intermediate performance $p^{CU}_{t,w-1} - \max_{c_t} \left[p^{CU}_{c_t,w-1} \right]$ by the expected difference $D_{t,w-1}$ as given by (6).

In all of our estimations, we cluster standard errors at the store level to correct for serial correlation within stores, as well as for heteroscedasticity across stores (see Bertrand et al. (2004) for a discussion of the importance of correcting for serial correlation in Difference-in-Difference estimation).

4 Results

The first column in Table 2 gives the results of estimating (3). On average, the relative performance incentive scheme did not affect sales. The second column of Table 2 shows that there is some variation in the estimated treatment effect by week, but none of the estimates differs significantly from zero.

The first three columns in Table 3 present the results of estimating (5) using OLS and IV-2SLS, respectively, with $D_{t,w-1}$ instrumenting $p^{CU}_{t,w-1} - \max_{c_t} \left[p^{CU}_{c_t,w-1} \right]$. The second column contains the first-stage regression of the IV-estimation. Actual intermediate relative performance increases one-for-one with our instrument, predicted difference (6). This instrument alone explains about 50 percent of the total variation in intermediate relative performance in the last three weeks of the experiment. Figure 5 shows the relation between the actual difference in intermediate cumulative performance between the treatment stores and their best comparison stores, $p^{CU}_{t,w-1} - \max_{c_t}\left[p^{CU}_{c_t,w-1}\right]$, and the predicted difference (6).

The OLS estimation in the first column of Table 3 shows that intermediate relative performance is significantly positively related to subsequent sales. Its point estimate suggests that a percentage point increase in lagged relative performance increases current sales by 0.26 percent. However, in the IV-2SLS estimation reported in the third column the point estimate is more than halved and is no longer significantly different from zero. Figure 6 visualises these results for the relevant subset of observations: treatment stores in the three final weeks of the experiment. It plots the residuals of the estimation of the average treatment effect (3), as presented in the first column of Table 2, against the predicted values for $p^{CU}_{t,w-1} - \max_{c_t}\left[p^{CU}_{c_t,w-1}\right]$ as estimated by the first-stage regression of the IV-2SLS estimation (second column of Table 3).

The fourth column of Table 3 shows that the effect of intermediate relative performance is concentrated in the final week of the experiment. Stores that perform about as good or even better than their best-performing comparison store in the first three weeks of the experiment manage to increase their sales in the final week. A Wald test shows that this increase is significant at the 5 percent level for stores that lag less than 4 percentage points behind their best-performing comparison store. The treatment effect in the final week increases significantly in intermediate relative performance, by 0.28 percent per percentage point.

The estimations in Table 3 assume that the effect of intermediate relative performance is linear. The first column of Table 4 reports the results of estimating (8), where the effect of intermediate relative performance is allowed to vary between stores that lag far behind and stores that are close to or above the target for winning a bonus.[12] Graphically, we allow the effect of intermediate relative performance to differ between observations to the left and right of the dashed line in Figure 6. We find that past relative performance does not affect current sales for stores that lag far behind. In contrast, current sales of treatment stores that lag less than 5 percentage points behind increases by 0.72 percent per percentage point increase in past relative performance. A Wald test shows that the overall treatment effect is significantly different from zero for stores that are at least 6 percentage points

[12] Instead of estimating (8), we could estimate a quadratic specification of intermediate relative performance. However, the estimates for the quadratic specification would be heavily affected by the many treatment store-week observations with sizable negative intermediate relative performance (see Figure 5). Hence, we would learn little about the marginal effect of intermediate relative performance for stores close to winning a bonus.

ahead of their best comparison store. The second column of Table 4 shows that the marginal effect of intermediate relative performance on current sales of relatively good-performing stores is significantly positive in the third and fourth week of the experiment, with magnitudes of 1.4 and 0.9 percent per percentage point, respectively. For stores that lag far behind, there is no such effect in any of the weeks. These results are qualitatively robust to varying the level of intermediate relative performance at which stores are deemed to stand a chance of winning between 0% and -10%. Quantitatively, the estimated effects of intermediate relative performance on current sales for stores deemed to stand a chance are larger when this level is closer to 0%.

Lastly, we examine differences in treatment effects between big and small stores. This is explorative, as store size is not exogenously determined. Hence, any differences between big and small stores may not be caused by store size per se, but by an unobserved store characteristic that causes or co-varies with store size. Unreported estimations show that we do not find a significantly positive average treatment effect in any of the strata based on store size. Columns 5 to 8 of Table 3 show the results of estimating the effect of intermediate relative performance (5) for big and small stores separately, both for the whole period as well as separated by week. For big stores, the treatment effect increases marginally significantly with intermediate relative performance. Column 6 shows that most of this effect occurs in the week where stores receive the first poster (i.e. in the second week of the experiment). For small stores, we find a sizeable treatment effect in the last week of the experiment, which increases significantly in intermediate relative performance. The point estimate of the treatment effect in the final week for small stores that perform at par with their best comparison store up to the final week is 6.4% additional sales. Columns 3 and 4 of Table 4 show that both big and small stores respond strongly to intermediate relative performance when sufficiently close to their best comparison store. For big stores, the estimated effect of a percentage point increase in intermediate relative performance on current sales is 0.78 percent. For small stores, this is 0.66 percent.[13]

Taken together, these results paint the following picture. On average, the relative performance incentive scheme had no effect on sales. Possibly, the prospect of competing against non-competitors did not excite employees in the treatment stores

[13] Across the three weeks with intermediate relative performance figures, there are 19 respectively 33 observations of small respectively big stores whose performance is at most 5 percent below their best comparison store. This implies that we cannot further disaggregate the results in columns 3 and 4 of Table 4 into estimates by week.

much. Alternatively, many stores may have perceived the relative performance targets as too ambitious. Such a perception would be reinforced after receiving the first poster with rankings, as only 23 treatment stores ranked on top of the first poster and 64 stores lagged more than 5 percent behind their best-performing comparison store on the first poster. We find that stores lagging too far behind do not respond to the incentive scheme, nor to the intermediate relative performance information. However, as stores come closer to winning a bonus through better lagged relative performance, sales increase significantly with lagged relative performance. This effect is strongest in the second half of the experiment, and is present in both big and small stores.

5 Concluding Remarks

We have reported the results of a field experiment on dynamic incentive effects of relative performance pay among stores of a retail chain. We find that intermediate relative performance feedback affects subsequent performance of stores close to the bonus target. These stores show significantly higher performance, particularly near the end of the incentive period. Stores lagging far behind do not respond to the incentive scheme, nor to intermediate relative performance. As many treatment stores happen to trail far behind bonus targets over the course of the experiment, we find no improvement in performance on average.

Our findings underline the importance of dynamic incentive effects. When in the course of a contest the target moves out of reach, people give up, which renders the incentive scheme fruitless. On the other hand, learning that intermediate performance is closer to target encourages people to increase effort. Hence, the incentive effect of competitions is path-dependent. Contests that turn out to be close are more effective than contests with large differences in early performance. From an ex ante point of view, this implies that while large common shocks make relative performance pay attractive (Lazear and Rosen, 1981, Green and Stokey, 1983, Nalebuff and Stiglitz, 1983), large idiosyncratic shocks reduce its effectiveness.

References

[1] Aoyagi, Masaki (2010), Information Feedback in a Dynamic Tournament, *Games and Economic Behavior*, forthcoming.

[2] Asch, Beth J. (1990), Do Incentives Matter? The Case of Navy Recruiters, *Industrial and Labor Relations Review,* vol. 43, pp. 89S-106S.

[3] 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.

[4] Bandiera, Oriana, Iwan Barankay, and Imran Rasul (2009), Team Incentives: Evidence from a Field Experiment, Mimeo, University College London.

[5] 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.

[6] Blanes i Vidal, Jordi, and Mareike Nossol (2009), Tournaments without Prizes: Evidence from Personnel Records, Mimeo, London School of Economics.

[7] Bouwens, Jan, and Peter Kroos (2010), Target Ratcheting and Effort Reduction, *Journal of Accounting and Economics,* forthcoming.

[8] Cadsby, C. Bram, Fei Song, and Francis Tapon (2010), Are You Paying Your Employees to Cheat? An Experimental Investigation, *B.E. Journal of Economic Analysis & Policy*, vol. 10(1), article 35.

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

[10] Charness, Gary, Peter Kuhn, and Marie-Claire Villeval (2010), Competition and the Ratchet Effect, mimeo.

[11] Cooper, David, John Kagel, Wei Lo, and Qing Liang Gu (1999), Gaming Against Managers in Incentive Systems: Experimental Results with Chinese Students and Chinese Managers, *American Economic Review,* vol. 89(4), pp. 781-804.

[12] 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.

[13] Ederer, Florian (2010), Feedback and Motivation in Dynamic Tournaments, *Journal of Economics & Management Strategy*, vol. 19(3), pp. 733-769.

[14] 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.

[15] Fershtman, Chaim, and Uri Gneezy (2009), The Trade-off between Performance and Quitting in High-Power Tournaments, *Journal of the European Economic Association,* forthcoming.

[16] Gershkov, Alex, and Motty Perry (2009), Tournaments with Midterm Reviews, *Games and Economic Behavior,* vol. 66, pp. 162-190.

[17] Goltsman, Maria and Arijit Mukherjee (2010), Interim Performance Feedback in Multistage Tournaments: The Optimality of Partial Disclosure, *Journal of Labor Economics,* forthcoming.

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

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

[20] Kosfeld, Michael, and Susanne Neckermann (2010), Getting More Work for Nothing? Symbolic Awards and Worker Performance, *American Economic Journal: Microeconomics,* forthcoming.

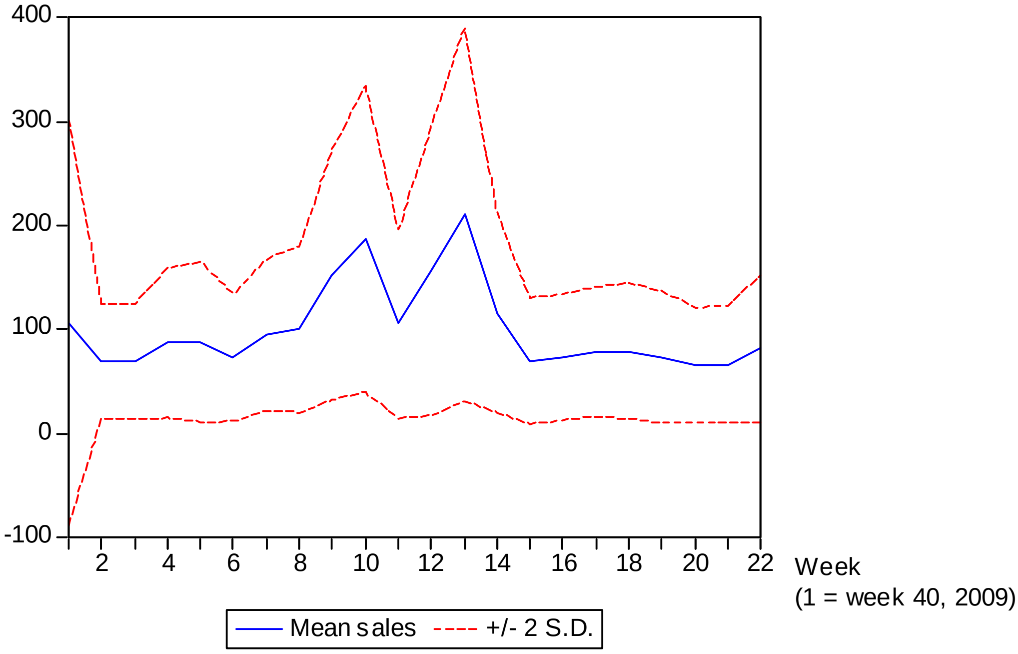
[21] Lazear, Edward P., and Sherwin Rosen (1981), Rank-Order Tournaments as Optimum Labor Contracts, *Journal of Political Economy*, vol. 89(5), pp. 841-864.

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

[23] Oyer, Paul (1998), Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality, *Quarterly Journal of Economics,* vol. 113(1), pp. 149-185.

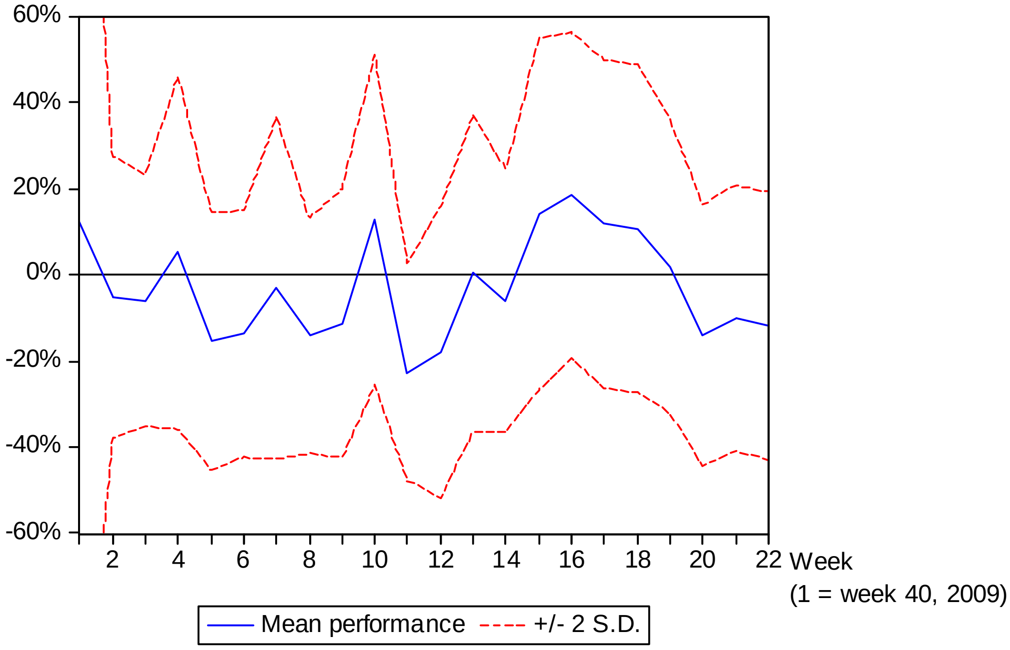
[24] Royal Dutch Shell (2009), Annual Report and Form 20-F for the Year ended December 31, 2009.

Figure 1: Average sales per store



Sales is indexed, where average sales per store-week over the whole period = 100.

Figure 2: Average weekly performance



Performance is measured as sales in % deviation from budgeted sales

Figure 3: Average sales of treatment stores divided by average sales of control stores

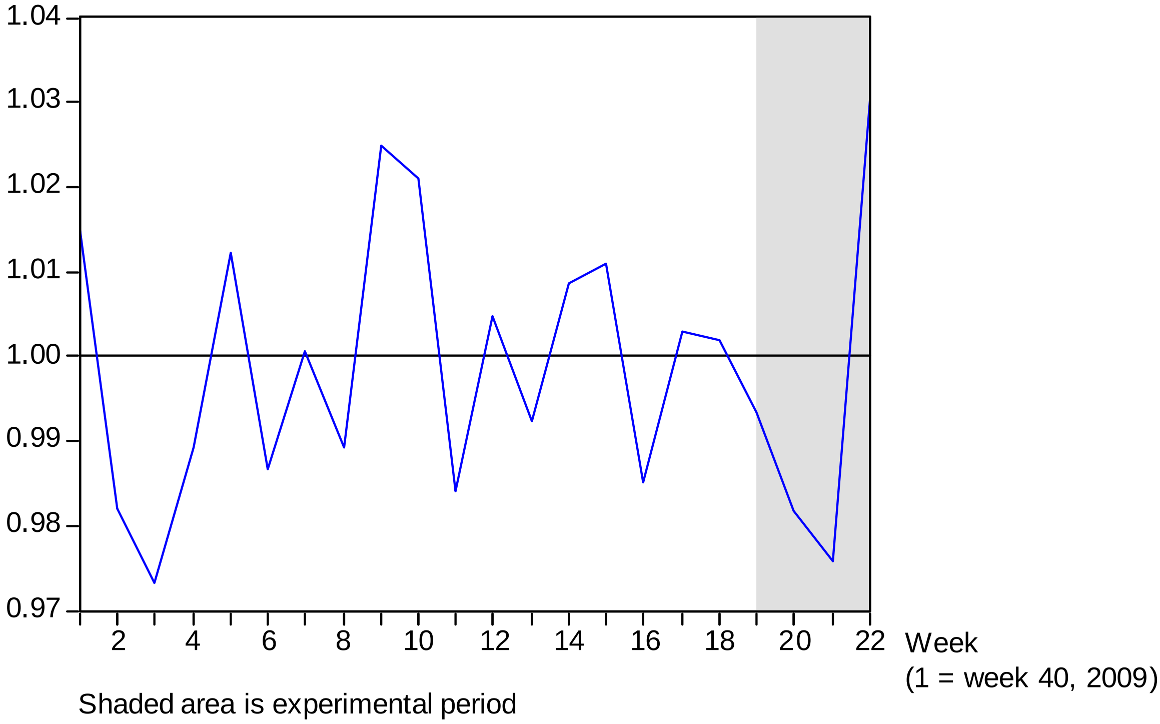


Figure 4: Difference in average performance between treatment stores and control stores

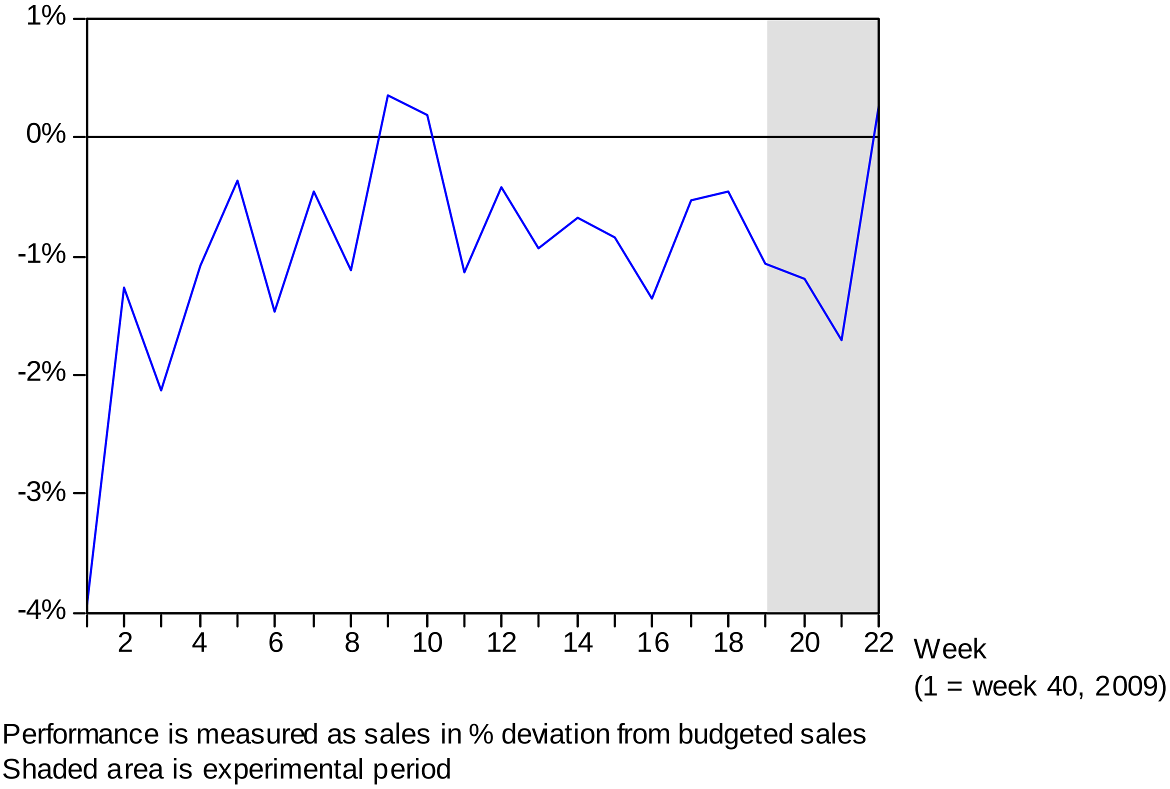
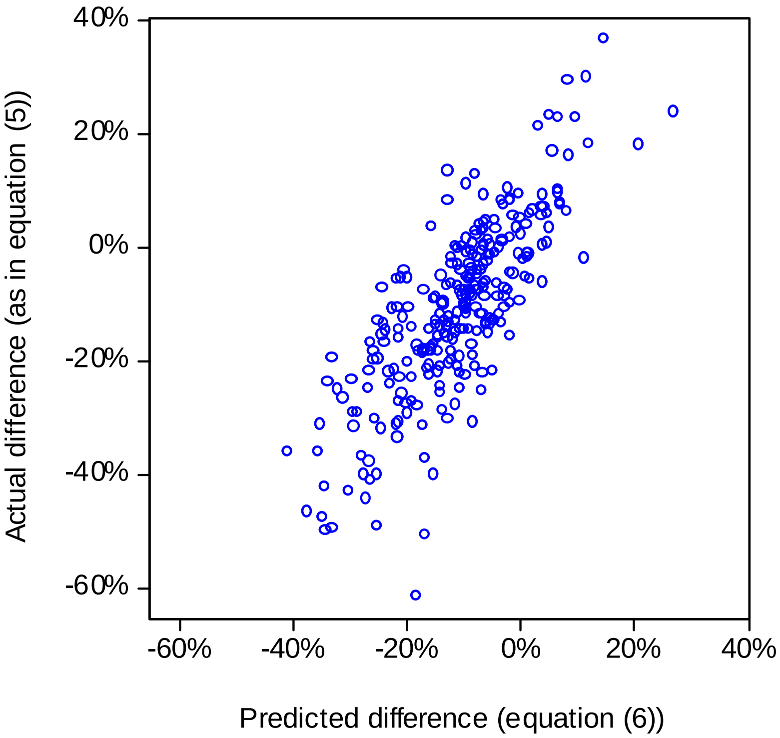


Figure 5: Actual and predicted difference in intermediate cumulative performance between the treatment store and its best comparison store $\left(p^{CU}_{t,w-1} - \max_{c_t}\left[p^{CU}_{c_t,w-1}\right]\right)$



One outlier not shown (actual difference -101%; predicted difference -36%)

Figure 6: The relation between the predicted difference $E(p^{CU}_{t,w-1}) - \max_{c_t}\left[p^{CU}_{c_t,w-1}\right]$ and the residuals from estimating (3)

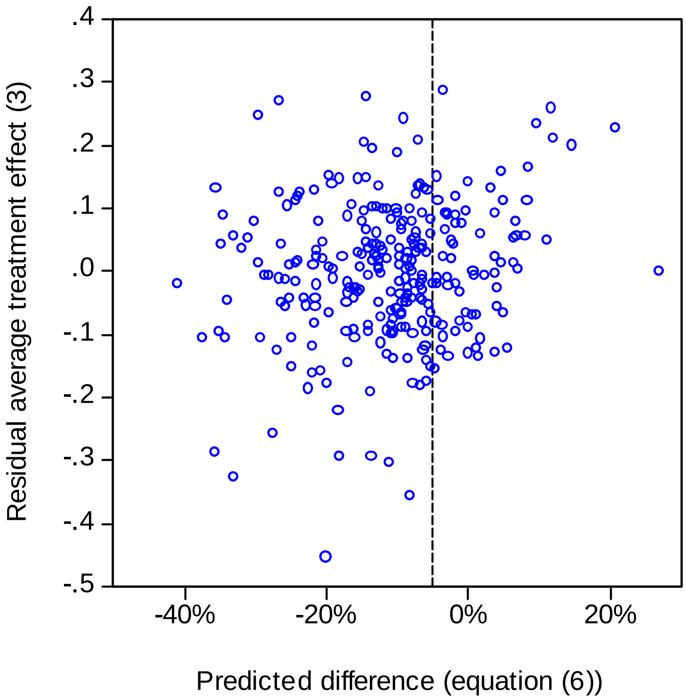


Table 1: Descriptive statistics

	All stores		Treatment stores		Control stores	
	mean	Std	Mean	Std	Mean	Std
Sales	100.00	40.31	100.04	43.77	99.96	36.89
Sales weeks 40/2009 - 53/2009	115.18	46.62	115.27	50.81	115.08	42.43
Sales weeks 5/2010 - 8/2010	71.67	29.90	71.56	31.85	71.78	28.05
Budgeted sales	104.56	41.54	105.33	45.42	103.81	37.63
Budgeted sales weeks 40/2009 - 53/2009	122.90	48.82	123.80	53.38	122.02	44.23
Budgeted sales weeks 5/2010 - 8/2010	78.52	31.19	79.10	34.10	77.96	28.86
Performance (= (sales-budgeted sales)/budgeted sales)	-0.03	0.12	-0.04	0.11	-0.02	0.12
Performance weeks 40/2009 - 53/2009	-0.06	0.11	-0.07	0.10	-0.05	0.12
Performance weeks 5/2010 - 8/2010	-0.08	0.14	-0.09	0.13	-0.08	0.14
Number of employees	5.45	1.99	5.24	1.69	5.66	2.22
Number of days closed for carnival (week 7/2010)	0.25	0.63	0.19	0.56	0.31	0.69
Number of stores	189		93		96	

For confidentiality reasons, sales and budgeted sales figures are indexed to the average sales per store per week over the whole sample.
None of the differences between treatment stores and control stores are significant at the 10%-level.

Table 2: Average treatment effect

	Dependent variable: ln(sales)	
	(1)	(2)
Treatment	-0.004	
	(0.013)	
Treatment week 1		-0.003
		(0.019)
Treatment week 2		-0.013
		(0.020)
Treatment week 3		-0.020
		(0.017)
Treatment week 4		0.021
		(0.019)
Carnival	-0.026*	-0.028*
	(0.015)	(0.015)
Store-fixed effects	yes	yes
Week-fixed effects	yes	yes
Store-week observations	4158	4158
Stores	189	189
R^2	0.9281	0.9281

Standard errors clustered at the store level in parentheses.
***, **, * denote statistically significant effects at the 1%, 5%, and 10% level, respectively.

Table 3: Dynamic incentives

	Dependent variable: ln(sales)							
	All stores				Big stores		Small stores	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV-2SLS	IV-2SLS	IV-2SLS	IV-2SLS	IV-2SLS	IV-2SLS	IV-2SLS
Stage IV-2SLS		First	Second	Second	Second	Second	Second	Second
Treatment	0.017	0.023**	0.005		0.028		-0.020	
	(0.012)	(0.010)	(0.014)		(0.019)		(0.020)	
Relative intermediate performance	0.0026***		0.0011		0.0022*		0.0003	
	(0.0005)		(0.0008)		(0.0012)		(0.0011)	
Expected intermediate relative performance, $D_{t, w-1}$ (equation (6))		1.088***						
		(0.091)						
Treatment week 1				-0.003		0.026		-0.034
				(0.019)		(0.027)		(0.026)
Treatment week 2				0.001		0.033		-0.030
				(0.021)		(0.026)		(0.027)
Treatment week 3				-0.016		0.007		-0.041
				(0.022)		(0.026)		(0.042)
Treatment week 4				0.048**		0.041		0.064**
				(0.022)		(0.031)		(0.032)
Relative performance after week 1				0.0011		0.0049***		-0.0004
				(0.0009)		(0.0016)		(0.0011)
Relative performance after week 2				0.0004		0.0005		0.0002
				(0.0013)		(0.0012)		(0.0024)
Relative performance after week 3				0.0028**		0.0009		0.0051***
				(0.0012)		(0.0018)		(0.0015)
Carnival	-0.020	-0.017**	-0.024	-0.027*	-0.028**	-0.035**	-0.025	-0.027
	(0.014)	(0.008)	(0.015)	(0.016)	(0.014)	(0.015)	(0.021)	(0.023)
Store-fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Week-fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Store-week observations	4158	4158	4158	4158	2134	2134	2024	2024
Stores	189	189	189	189	97	97	92	92
R^2	0.9284	0.6712	0.9283	0.9284	0.9050	0.9050	0.8988	0.8990

Standard errors clustered at the store level in parentheses.
The dependent variable in the first-stage regression shown in column (2) is relative intermediate performance.
***, **, * denote statistically significant effects at the 1%, 5%, and 10% level, respectively.

Table 4: Dynamic incentives separate for stores close to winning a bonus

	Dependent variable: ln(sales)			
	All stores		Big stores	Small stores
	(1)	(2)	(3)	(4)
	IV-2SLS	IV-2SLS	IV-2SLS	IV-2SLS
	Second stage	Second stage	Second stage	Second stage
Treatment	-0.006		0.023	-0.033
	(0.016)		(0.024)	(0.021)
Relative intermediate performance	0.0002		0.0017	-0.0006
	(0.0009)		(0.0017)	(0.0011)
Treatment*close	-0.008		-0.035	0.016
	(0.019)		(0.028)	(0.029)
Relative intermediate performance*close	0.0070***		0.0061***	0.0072***
	(0.0012)		(0.0023)	(0.0012)
Treatment week 1		-0.003		
		(0.019)		
Treatment week 2		-0.045		
		(0.036)		
Treatment week 3		-0.038		
		(0.028)		
Treatment week 4		0.050		
		(0.031)		
Relative performance after week 1		-0.0011		
		(0.0017)		
Relative performance after week 2		-0.0012		
		(0.0016)		
Relative performance after week 3		0.0027		
		(0.0018)		
Treatment week 2*close		0.057		
		(0.042)		
Treatment week 3*close		-0.061		
		(0.055)		
Treatment week 4*close		-0.038		
		(0.038)		
Relative performance after week 1*close		0.0048		
		(0.0030)		
Relative performance after week 2*close		0.0149**		
		(0.0059)		
Relative performance after week 3*close		0.0063**		
		(0.0028)		
Carnival	-0.024	-0.027	-0.026**	-0.027
	(0.015)	(0.017)	(0.013)	(0.021)
Store-fixed effects	yes	yes	yes	yes
Week-fixed effects	yes	yes	yes	yes
Store-week observations	4158	4158	2134	2024
Stores	189	189	97	92
R^2	0.9283	0.9282	0.9051	0.8985

Standard errors clustered at the store level in parentheses.
"Close" is a dummy variable that takes value one when the store's expected intermediate performance is at most 5 percentage points below its best comparison store, i.e. when $D_{t, w-1}$ > -.05 (see equation (6)).
***, **, * denote statistically significant effects at the 1%, 5%, and 10% level, respectively.