University of Groningen

Design of Yardstick Competition and Consumer Prices

Dijkstra, Peter T.; Haan, Marco A.; Mulder, Machiel

Published in:
Energy Economics

DOI:
[10.1016/j.eneco.2017.06.016](https://doi.org/10.1016/j.eneco.2017.06.016)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Dijkstra, P. T., Haan, M. A., & Mulder, M. (2017). Design of Yardstick Competition and Consumer Prices: Experimental Evidence. *Energy Economics*, 66, 261-271. <https://doi.org/10.1016/j.eneco.2017.06.016>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Design of yardstick competition and consumer prices: Experimental evidence[☆]

Peter T. Dijkstra^{a,*,1}, Marco A. Haan^b, Machiel Mulder^b

^a Netherlands Authority for Consumers & Markets (ACM), The Hague, The Netherlands

^b Department of Economics, Econometrics and Finance, University of Groningen, P.O. Box 800, Groningen 9700 AV, The Netherlands

ARTICLE INFO

Article history:

Received 18 December 2015

Received in revised form 12 January 2017

Accepted 24 June 2017

Available online 8 July 2017

JEL classification:

C73

C92

L13

L51

Keywords:

Collusion

Industry structure

Yardstick competition

Experiment

ABSTRACT

In this paper we analyze the effect of the design of yardstick competition on consumer prices, by means of a theoretical analysis as well as an economic experiment. We compare four different designs: the uniform yardstick, the unweighted uniform yardstick, the discriminatory yardstick, and the best-practice yardstick. The effect of a specific design on prices depends on two separate mechanisms, one which affects the incentive power to increase productive efficiency and another which affects the risk of collusion. We show theoretically that for the best-practice yardstick, which is widely applied in several industries in a number of countries, these two mechanisms point in the same direction (high prices), which is confirmed by the experiment. Both the theoretical analysis as well as the economic experiment show that the discriminatory yardstick results in lower prices than the unweighted uniform yardstick. The theory, however, does not give a clear answer on the relative performance of the discriminatory versus the weighted uniform yardstick. In the experimental analysis, we find that the advantage of the discriminatory yardstick in terms of giving incentives to improve productive efficiency exceeds the disadvantage of a relatively higher risk of collusion. This conclusion appears to be robust for different degrees of heterogeneity of the industry. Hence the discriminatory yardstick yields the lowest prices for consumers.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The revenues of firms subject to tariff regulation are often based on yardstick competition. With this type of regulation, the (maximum) price for consumers is based on the costs of a group of similar firms. The idea behind yardstick competition is that this gives each firm an incentive to produce efficiently, while the benefits of these efficiency improvements are transferred to consumers. In the US, for example, the payment a hospital received for a treatment under *Medicare*, depended on the average cost of that treatment in other hospitals (Shleifer, 1985). Yardstick competition is used

in industries as diverse as water supply and sewerage,² electricity networks,³ railways⁴ and bus transport services,⁵ to name but a few. In several cases, the introduction of yardstick regulation improved productive efficiency or resulted in lower consumer prices. In other cases, however, the experience appeared to be less successful.

There are many ways to implement yardstick competition. With a uniform yardstick, prices that firms can charge depend on the average costs of all firms. That average can be weighted or unweighted average costs. With a discriminatory yardstick, the price that an individual firm can charge depends on the average costs of all *other* firms. Each firm then faces a different price cap. With a best-practice yardstick, prices depend on the costs of the most efficient firm.

If firms behave in a noncooperative fashion, the incentive to raise productive efficiency is higher when a firm's regulated price bears less relation to its own costs. Indeed, Shleifer (1985) shows that a discriminatory yardstick then leads to the socially efficient level of cost

[☆] We thank Editor Richard Tol and three anonymous referees for their valuable comments. We are also indebted to participants of CRESSE 2015 (Crete) and seminar participants at the University of Groningen (RUG) and the Netherlands Authority for Consumers & Markets (ACM). Financial support of the Netherlands Authority for Consumers & Markets (ACM) is gratefully acknowledged. The authors are fully responsible for the contents of the paper.

* Corresponding author.

E-mail address: Peter.Dijkstra@acm.nl (P.T. Dijkstra).

¹ Previous affiliation: Department of Economics, Econometrics and Finance, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands.

² Dassler et al. (2006).

³ Haffner et al. (2010), Blázquez-Gómez and Grifell-Tatjé (2011), Jamasb and Pollitt (2007).

⁴ Mizutani et al. (2009).

⁵ Dalen and Gomez-Lobo (2002).

reduction. But a yardstick that performs well under the assumption that firms behave non-cooperatively may still backfire if it facilitates collusion. Firms may agree, perhaps tacitly, not to reduce costs or to jointly manipulate productivity reports (Tangerås, 2002). To evaluate the effectiveness of a yardstick we have to take into account the incentive to increase efficiency as well as the incentive to collude; some types of yardstick may make collusion easier than others. A priori, it is not always clear which of these two possibly counter-vailing effects is strongest. The risk of collusion in case of yardstick competition was already mentioned by Shleifer (1985), but to what extent that risk could mitigate the benefits of yardstick competition has hardly been analyzed up to now.

Hence, given the fact that regulated firms may be able to collude, it is not clear what type of yardstick would be most successful in achieving low consumer prices. To try to establish the potential impact of collusion, we conduct a laboratory experiment. First, we evaluate the relative performance of a weighted uniform, an unweighted uniform, a discriminatory, and a best-practice yardstick. We do so in an environment with two firms that have different sizes, and study the effect of yardstick type on average prices and on the incidence of collusion. In a second experiment, we zoom in on the weighted uniform, and the discriminatory yardstick. These two yardsticks perform well in our first experiment, but both theory and that first experiment do not provide a clear prediction in terms of their market performance. We compare these two yardsticks for a variety of industry structures as the distribution of firms regarding market size may have an effect on the effectiveness of yardstick competition.

This paper is not the first to conduct such experiments. Potters et al. (2004) find that the extent of collusion as well as consumer prices are higher with a discriminatory rather than a uniform yardstick. Apparently, the authors conclude, with a discriminatory yardstick the stronger incentives to collude outweigh the higher incentives to raise productive efficiency. Yet, the authors only show this for two equally-sized firms, prompting the question whether a more heterogeneous firm-size distribution would give the same result. In a companion paper (Dijkstra et al., 2017), we study the effect of industry structure on the incentive to collude with a (weighted) uniform yardstick. We find that in more heterogeneous industries, firms appear to collude less than in homogeneous industries.⁶

In our experiment subjects decide upon marginal cost levels, knowing that their output prices are set by a regulator using a specific form of yardstick competition. Our design is the same as that in Dijkstra et al. (2017). We use a simple model, loosely based on Shleifer (1985), where managers have to exert costly effort to lower their costs. Managers aim to maximize the sum of profits and managerial benefits, and do so in a repeated game. For simplicity, managerial benefits are a concave and quadratic function of marginal costs.

In our first experiment we compare the four different yardsticks for an industry with two firms with different sizes. We find that the price for consumers is highest for a best-practice yardstick, while a discriminatory yardstick yields prices that are lower (i.e. closer to the social optimum) than an unweighted uniform yardstick. The discriminatory yardstick seems to result in lower prices but more collusion than a weighted uniform yardstick, but not significantly so. In our second experiment, we compare the weighted uniform and the discriminatory yardstick for different industry structures. We find that the discriminatory yardstick yields the lowest consumer prices overall. From this we conclude that the discriminatory yardstick is the preferred yardstick in order to raise productive efficiency

of monopolistic firms and to pass on these benefits to consumers. Note that this conclusion runs counter to that in Potters et al. (2004).

As noted, in a companion paper (Dijkstra et al., 2017) we use a similar set-up to study a different question. Therefore, parts of the description of the theoretical model, the experimental set-up and the literature survey in this paper closely follows that in our companion paper.

The structure of this paper is as follows. Section 2 gives some background information, including a discussion on policy practice, as well as related experiments. Section 3 describes our theoretical model. The experimental design is presented in Section 4. Section 5 discusses the result of our experiment with 4 yardsticks, while Section 6 discusses the experiment that compares the discriminatory and (weighted) uniform yardsticks. Section 7 concludes.

2. Background

2.1. Yardstick competition⁷

The last decades saw the liberalization of several network industries that used to be state-owned vertically-integrated monopolies. Examples include telecommunications, electricity, gas, water and sewerage. Parts of these industries are natural monopolies characterized by subadditivity of costs, which calls for regulatory supervision (Viscusi et al., 2005). A key component of such supervision is the regulation of tariffs. Historically, regulated tariffs were either based on actual costs or an allowed rate of return. However, such schemes give little incentive to increase productive efficiency as lower costs directly translate to lower revenues. Hence, the incentive power is low, as a change in costs hardly affect firms' profits. Introducing price-cap regulation solves this problem as any cost decrease then fully benefits the firm. But this method also implies the risk of significant rents or losses for regulated firms. To overcome both problems, Shleifer (1985) proposed to impose tariffs based on the actual costs of a group of similar (benchmark) firms. This type of tariff regulation is called yardstick regulation. As tariffs are based on the relative performance of a firm, yardstick regulation is generally seen as a powerful tool both for giving incentives for productive efficiency as well as for rent extraction (Tangerås, 2002; Burns et al., 2005).

The incentive power of a yardstick partly depends on the exact manner in which costs of the benchmark firms determine tariffs. There are various types of yardsticks. With a *uniform yardstick* every firm faces the same cap on the tariffs it may charge, based on some average of the cost information of all firms in the benchmark group. With a *discriminatory yardstick*, every firm faces a specific cap based on the costs of all *other* firms in the benchmark group. With a *best-practice yardstick*, every firm faces the same cap, based on the cost information of the most efficient firm.

2.2. Practical implementations and experience

Yardstick regulation has been implemented in a number of industries in several countries. In the United Kingdom, yardstick competition is often based on some type of best practice. Dassler et al. (2006) show how it is applied in telecommunication, gas, electricity, water and sewerage. For electricity, the yardstick was based on the average performance in the group of similar firms, though the exact implementation differs.⁸ A reason for this different treatment is the degree

⁷ This section partly follows our companion paper, Dijkstra et al. (2017).

⁸ In their assessment of the UK regulatory scheme for electricity distribution networks, for example, Jamasb and Pollitt (2007) state that the regulatory framework included different treatments of operational costs and capital costs. For operational costs (OPEX), the price cap was based on a form of best-practice benchmarking. For capital costs (CAPEX) standardized parameters were used regarding costs of capital and depreciation methods.

⁶ The result that asymmetries might lead to less collusion is also observed in unregulated markets, see e.g. Mason et al. (1992), Fonseca and Normann (2008), Dugar and Mitra (2009), Phillips et al. (2011), and Dugar and Mitra (2016).

of similarity within the group and the transparency about the costs and outputs of firms. As [Yatchew \(2001\)](#) stated, best-practice forms of yardstick competition are more sensitive to outliers which are not comparable to the other firms in the group, while it is also more complicated to precisely determine the best practice. The application of yardstick competition in the UK electricity industry (based on best practice for the operational costs) was successful in terms of increasing efficiency, reducing end-user prices while maintaining the quality of service ([Jamasp and Pollitt, 2007](#)).

In Japan, the yardstick is in essence an unweighted average between a uniform yardstick and a firm's own costs. Needless to say, this implies little incentive power. Unsurprisingly, for the Japanese railway industry, [Mizutani and Usami \(2016\)](#) do not find evidence that the application of yardstick competition resulted in more cost reductions. In Spanish electricity distribution, the yardstick is an external benchmark that is the same for all firms, making it a combination between a uniform (in that the same yardstick applies to all firms) and a discriminatory benchmark (in that the yardstick depends on the performance of others). Yardstick competition in this industry resulted in higher levels of efficiency, although consumers hardly benefited from these improvements ([Blázquez-Gómez and Grifell-Tatjé, 2011](#)).

In the Norwegian bus industry a uniform benchmark has been used, see [Dalen and Gómez-Lobo \(2003\)](#), although the authors argue that the number of bus companies in each region is large enough to make it almost insensitive to efficiency improvements by the individual company. The authors show that the adoption of this more high-powered scheme significantly reduced operating costs. In the Netherlands, in the high-voltage network a form of best-practice is used; the regulated price is based on the average costs of the three most efficient firms ([Mulder, 2012](#)). In the US, the Prospective Payment System of Medicare uses average performance as the basis of regulation ([Shleifer, 1985](#)).

Summing up, some form of best-practice is often used as the basis for yardstick competition. Alternatively, some form of uniform benchmark is applied. Despite its attractive theoretical properties, it seems that a discriminatory yardstick is hardly applied. Often, some combination of benchmarks is used.

2.3. Experiments

Numerous experiments have studied collusion in unregulated markets. For an overview, see e.g. [Haan et al. \(2009\)](#). Some of these studies focus on the effect of asymmetries between firms. With quantity setting, [Mason et al. \(1992\)](#) find more collusion if firms have equal rather than different marginal costs. [Phillips et al. \(2011\)](#) confirm this result. With price competition, [Fonseca and Normann \(2008\)](#) find more collusion if firms have identical capacity constraints, and [Dugar and Mitra \(2009, 2016\)](#) find more collusion if the range of possible firm-specific marginal costs is smaller. [Argenton and Müller \(2012\)](#) however, find that collusion is unaffected if firms have different rather than identical cost structures. Still, overwhelmingly, experiments on unregulated markets find that asymmetries do hinder collusion.

Two experimental papers deal with yardstick competition. [Potters et al. \(2004\)](#) study the effect of yardstick design on collusion in a duopoly with symmetric firms. They compare a uniform yardstick to a discriminatory one. When firms behave non-cooperatively, the discriminatory yardstick yields lower prices and lower profits, making it more prone to collusion. In their experiment, the authors indeed find higher cost levels with a discriminatory yardstick.

There are two differences in the design of [Potters et al. \(2004\)](#) and ours. First, they use downward sloping demand, while we assume inelastic demand. Second, they allow subjects to choose cost levels as well as prices, while we only allow them to choose cost levels. We have chosen inelastic demand as it greatly simplifies the experiment.

Especially once we allow for asymmetries between firms, having downward sloping demand as well would greatly complicate the experimental set-up and instructions. We have chosen to drop the price-setting stage as [Potters et al. \(2004\)](#) find that subjects set prices equal to the price cap in about 99% of the cases, so this stage adds little insight.

In our companion paper, [Dijkstra et al. \(2017\)](#), we use the same experimental design as we do in this paper but focus on a different question; we study to what extent industry structure affects performance of a uniform yardstick. We find that more heterogeneity leads to lower prices. Surprisingly, industries with 3 symmetric firms are more collusive than those with 2 symmetric firms.

3. The model

3.1. Setup

We consider the following model. There are n firms that play an infinitely repeated game. Each firm acts as a local monopolist. For simplicity, we assume that firm i faces demand that is completely inelastic and exogenously given by mass $\alpha_i \in (0, 1)$. This is largely for simplicity: we need the experiment to be as easy and straightforward as possible. Any complication would make it harder for subjects to understand the setup.

We normalize total demand, so $\sum_{i=1}^n \alpha_i = 1$. For ease of exposition, we will refer to α_i as the market share of total demand of firm i . In each round, firm i decides on its constant marginal cost c_i for that round. Firm i 's profits then equal $\pi_i(c_i, \mathbf{c}_{-i}) = (p_i^y(c_i, \mathbf{c}_{-i}) - c_i) \alpha_i$, with p_i^y the price that firm i can set under yardstick y , and the vector $\mathbf{c}_{-i}$ consists of the cost levels chosen by the other firms. We assume that the manager of firm i maximizes $u_i = \pi_i(c_i, \mathbf{c}_{-i}) + R_i(c_i)$, with R_i a managerial benefit that is non-negative and concave: $R_i \geq 0$, $R_i'' < 0$, and $R_i(0) = 0$. Note that this specification is equivalent to [Shleifer \(1985\)](#) where managers invest in a cost-reducing technology.⁹ For simplicity, we use the following quadratic specification:

$$R(c_i) = (-ac_i^2 + bc_i) \alpha_i, \quad (1)$$

where $a, b \in \mathbb{R}^+$ are parameters and $c_i \in [0, b/a]$. Throughout, we assume that a and b are such that all expressions that we derive are well defined. The assumption that managerial benefit is proportional to market share is crucial. If firms have equal marginal costs, the size distribution of firms does not influence the total amount of managerial benefits in an industry: $\sum \alpha_i r(c) = r(c) \sum \alpha_i = r(c)$. It thus implies that the optimal mix does not depend on the size of the firm, and hence there are constant returns to scale. This allows us to investigate the effect of industry size and firm heterogeneity in isolation, without the confounding effect of already having lower average costs due to economies of scale. Our results should also be interpreted as such. Furthermore, our assumption is consistent with the available empirical evidence as economies of scale in electricity distribution are already exhausted with about 20,000 customers, which is far less

⁹ In [Shleifer \(1985\)](#), firms invest $R(c)$ in a cost-reducing technology that lowers marginal cost by c . With downward sloping demand, the monopolist cannot fully capture the social welfare generated by lower costs. Hence it will underinvest from a welfare perspective. This is no longer true, however, if demand is inelastic. In that case, the monopolist does capture the entire surplus and hence there is no underinvestment. For experimental simplicity, we do assume inelastic demand. We therefore use a slightly different set-up. We effectively assume that the manager of the firm has to exert some costly effort to lower costs. The discrepancy between the private and social optimum is then caused by the fact that the manager's costly effort is not taken into account when maximizing social welfare. This allows us to generate the same theoretical prediction (that firms underinvest), but in a simpler set-up. Note that our interpretation closely follows [Potters et al. \(2004\)](#).

than the number of customers on the Dutch market (see [Yatchew, 2000](#) and the references therein).

Note that we allow R_i to be decreasing in c_i for large enough c_i . In electricity networks for example, high marginal costs are often associated with a lack of maintenance. Such networks are prone to outages, compensation claims and political pressure, all factors that imply higher costs and do not exactly contribute to a quiet life for the manager. Also, this choice simplifies the experimental implementation of our model. The manager of firm i is thus assumed to maximize

$$u_i = (p_i^y(c_i, \mathbf{c}_{-i}) - c_i) \alpha_i + (-ac_i^2 + bc_i) \alpha_i. \quad (2)$$

Managerial benefit is maximized by

$$c^m = \frac{b}{2a}. \quad (3)$$

If consumers value the product at v , total welfare equals

$$W = (v - p) + (p - c) + (-ac^2 + bc), \quad (4)$$

where the first term is consumer surplus, and the other two terms is the total payoff to firms. Maximizing with respect to c yields the socially optimal cost level

$$c^W = \frac{b - 1}{2a}. \quad (5)$$

3.2. Yardsticks, non-cooperative behavior

In this section we derive the optimal cost level for each firm, depending on the type of yardstick competition. We consider the weighted uniform, unweighted uniform, discriminatory, and best-practice yardstick.

With a **UNIFORM** yardstick,¹⁰ the price firm i can charge equals the weighted average of all cost levels in the industry:

$$p^U(c_1, \dots, c_n) = \sum_{j=1}^n \alpha_j c_j. \quad (6)$$

Plugging this into Eq. (2) and maximizing with respect to c_i yields the non-cooperative cost level

$$c_i^{U*} = \frac{b - 1 + \alpha_i}{2a}. \quad (7)$$

Note that this is a dominant strategy, as it does not depend on the cost choices of the other firms. Also note that a firm with a higher market share α_i has a stronger influence on the regulated price, hence it will choose a higher cost level. Third, with $\alpha_i < 1$, the equilibrium cost level is strictly lower than the cost level that maximizes managerial benefits, but higher than the social optimum. The regulated price now becomes

$$p^{U*} \equiv p^U(c_1^{U*}, \dots, c_n^{U*}) = \sum_{i=1}^n \alpha_i \frac{b - 1 + \alpha_i}{2a} = \frac{b - 1}{2a} + \frac{\sum_{i=1}^n \alpha_i^2}{2a}. \quad (8)$$

¹⁰ As the regulation literature usually refers to the weighted uniform yardstick as “uniform”, we also follow that convention here.

With a uniform unweighted yardstick, the price firm i can charge equals the unweighted average of all cost levels in the industry:

$$p^E(c_1, \dots, c_n) = \frac{1}{n} \sum_{j=1}^n c_j. \quad (9)$$

In our experiment, we will refer to this treatment as **EQUALW** – for equal weights. Plugging Eq. (9) into Eq. (2) and maximizing with respect to c_i yield

$$c_i^{E*} = \frac{b - 1 + \frac{1}{n}}{2a}. \quad (10)$$

Hence, the non-cooperative cost level is now the same for all firms, as all firms have the same influence on the regulated price. Again, non-cooperative cost levels are strictly between the social optimum and the level that maximizes managerial benefits. The regulated price becomes

$$p^{E*} \equiv (c_1^{E*}, \dots, c_n^{E*}) = \frac{1}{n} \sum_{i=1}^n \frac{b - 1 + \frac{1}{n}}{2a} = \frac{b - 1}{2a} + \frac{1}{2an}. \quad (11)$$

With a discriminatory yardstick (denoted **DISCRIM** in our experiment), the price firm i can charge equals the weighted average of the cost levels of all *other* firms in the industry:

$$p_i^D(c_1, \dots, c_n) = \frac{\sum_{j \neq i}^n \alpha_j c_j}{\sum_{j \neq i}^n \alpha_j}. \quad (12)$$

Note that our notation reflects that in this case the price is firm-specific. Plugging Eq. (12) into Eq. (2) and maximizing with respect to c_i yield

$$c_i^{D*} = \frac{b - 1}{2a}. \quad (13)$$

Note therefore that this yardstick yields the social welfare optimum derived in Eq. (5). Also, all firms will choose the same cost level. The regulated price now differs per firm. Since a share of α_i of total demand is supplied by firm i , the average regulated price is

$$p^{D*} \equiv \sum_{i=1}^n \alpha_i p_i^D(c_1^{D*}, \dots, c_n^{D*}) = \sum_{i=1}^n \alpha_i \frac{b - 1}{2a} = \frac{b - 1}{2a}. \quad (14)$$

With the best-practice yardstick (**BESTPR**), the regulated price equals the lowest cost level of all firms in the industry:

$$p^B(c_1, \dots, c_n) = \min\{c_j\}. \quad (15)$$

Finding the Nash equilibrium now is somewhat more involved. Suppose all other firms set c^* . By setting some $c < c^*$, manager i earns

$$u_i(c_i, c^*, \dots, c^*) = (c - c^*) \alpha_i + (-ac^2 + bc) \alpha_i, \quad (16)$$

which is maximized by setting $c = b/2a$. This implies that if the other firms set some $c^* \leq b/2a$, this firm has no incentive to defect to a lower cost level: its maximization problem under the constraint that $c \leq c^*$ is given by the corner solution $c = c^*$.

By setting some $c > c^*$, manager i earns

$$u_i(c_i, c^*, \dots, c^*) = (c^* - c) \alpha_i + (-ac^2 + bc) \alpha_i, \quad (17)$$

which is maximized by setting $c = (b - 1)/2a$. This implies that if the other firms set some $c^* \geq (b - 1)/2a$, this firm has no incentive to defect to a higher cost level: its maximization problem under the constraint that $c \geq c^*$ is given by the corner solution $c = c^*$. This implies that we have a continuum of equilibria: any

$$c^{B*} \in [(b - 1)/2a, b/2a] \quad (18)$$

is a Nash equilibrium. Thus, possible Nash equilibria range from the social optimum to the outcome that maximizes managerial utility. By construction, Eq. (18) also gives the range of equilibrium prices. Given the multiplicity of equilibria in this case, we expect players to coordinate on the payoff-dominant equilibrium, which implies

$$p^{B*} = b/2a. \quad (19)$$

From the above, we immediately have¹¹

Theorem 1. *Regulated prices in the non-cooperative outcome are ranked*

$$p_{\text{DISCRIM}} < p_{\text{EQUALW}} \leq p_{\text{UNIFORM}} < p_{\text{BESTPR}}.$$

Only with symmetric market shares, $p_{\text{EQUALW}} = p_{\text{UNIFORM}}$.

3.3. Collusion

When managers collude they receive utility u_{ik}^y in each round, where $y \in \{U, E, D, B\}$. If manager i decides to defect instead, she receives utility u_{id}^y . Finally, if managers behave non-cooperatively they receive utility u_{in}^y . As usual, a cartel is stable if the short-run benefits of defection are outweighed by the long-term losses due to a cartel breakdown. In other words, following e.g. Friedman (1971) we assume grim trigger strategies and look for the critical discount factor δ such that all managers have an incentive to stick to the cartel agreement. With the usual arguments manager i does not defect if and only if

$$\frac{u_{ik}^y}{1 - \delta} > u_{id}^y + \frac{\delta}{1 - \delta} u_{in}^y. \quad (20)$$

This implies that we require

$$\delta > \hat{\delta}^y(\alpha_i) \equiv \frac{u_{id}^y - u_{ik}^y}{u_{id}^y - u_{in}^y}. \quad (21)$$

For collusion to be sustainable, we need that this condition is satisfied for all managers:

$$\delta > \hat{\delta}^y \equiv \max \{ \hat{\delta}^y(\alpha_1), \dots, \hat{\delta}^y(\alpha_n) \}. \quad (22)$$

We will focus on the cartel agreement in which all managers set the same cost level c_k that maximizes the sum of their utilities. Arguably, this is the most obvious and focal agreement. We will refer

to it as the perfect symmetric collusive agreement. Of course, if this would not yield a stable cartel, managers could still try to coordinate on a different agreement, either symmetric or asymmetric. In our theoretical analysis, however, we rule out this option as we feel that it would be much harder to coordinate on such an alternative.¹² Of course, in the experimental implementation of our model, firms are free to try to coordinate on whatever they can agree upon.

If firms collude on a common cost level c_k , the regulated price simply equals $p^y = c_k$ for all yardsticks, so they maximize joint utility

$$U = (p^y(c_k, \dots, c_k) - c_k) + (-ac_k^2 + bc_k) = (-ac_k^2 + bc_k), \quad (23)$$

which is maximized by setting

$$c_k^* = c^{m*} = b/2a. \quad (24)$$

As a joint cost level does not affect the price-cost margin, the cartel simply chooses to maximize its managerial benefit. This implies for all yardsticks

$$u_{ik}^y = (-a(c_k^*)^2 + bc_k^*) \alpha_i = \frac{\alpha_i}{4a} b^2. \quad (25)$$

The manager's utility in the non-cooperative outcome is

$$u_{in}^y = (p^{y*} - c_i^{y*}) \alpha_i + (-a(c_i^{y*})^2 + bc_i^{y*}) \alpha_i. \quad (26)$$

Finally, focusing on firm 1 for ease of exposition, u_{id}^y is given by

$$u_{id}^y = (p^y(c_1^{y*}, c_k, \dots, c_k) - c_1^{y*}) \alpha_1 + (-a(c_1^{y*})^2 + bc_1^{y*}) \alpha_1, \quad (27)$$

as we already showed that the one-shot optimal cost level for a firm is independent of the cost levels of the other firms. Therefore, the optimal deviation cost level is equal to the non-cooperative cost level. Defining $A_i \equiv \alpha_i/4a$, we then find the values of u_{in}^y , u_{id}^y and $\hat{\delta}^y(\alpha_i)$ for the different yardsticks that are given in Table 1.¹³

Hence industry structure only influences the critical discount factor in UNIFORM. Under BESTPR collusion is always sustainable. This is because deviation profits equal non-cooperative profits. We thus have¹⁴:

Theorem 2. *The critical discount factors are ranked*

$$\hat{\delta}^{\text{BESTPR}} < \hat{\delta}^{\text{EQUALW}} = \hat{\delta}^{\text{DISCRIM}} \leq \hat{\delta}^{\text{UNIFORM}}.$$

Only with symmetric market shares, $\hat{\delta}^{\text{DISCRIM}} = \hat{\delta}^{\text{UNIFORM}}$.

¹¹ This follows from comparing Eqs. (8), (11), (14) and (19). It is immediately clear that p^{D*} is the lowest. To see that $p^{E*} \leq p^{U*}$, define $\Delta_i \equiv \frac{1}{n} - \alpha_i$ for all i . We then have $\sum \alpha_i^2 = \sum (\frac{1}{n} - \Delta_i)^2 = \frac{1}{n} + \sum \Delta_i^2 - \frac{2}{n} \sum \Delta_i = \frac{1}{n} + \sum \Delta_i^2$ as $\sum \Delta_i = \sum (\frac{1}{n} - \alpha_i) = 1 - 1 = 0$. Hence $p^{U*} = p^{E*}$ if all $\Delta_i = 0$, hence if all firms are symmetric. Otherwise $p^{U*} > p^{E*}$. Finally, writing $p^{B*} = \frac{b-1}{2a} + \frac{1}{2a}$, it is immediate that $p^{B*} > p^{U*}$, as $\sum \alpha_i^2 < 1$ for $n > 1$.

¹² Indeed, in our experiment we also observe that groups, if anything, coordinate on the perfect symmetric collusive agreement, irrespective of the distribution of market shares.

¹³ For UNIFORM, u_{in}^U is found by plugging Eq. (7) into Eq. (26), and u_{id}^U by plugging Eqs. (7) and (24) into Eq. (27). From Eq. (21), we then have $\hat{\delta}^U(\alpha_i)$. For EQUALW, plugging Eq. (10) into Eq. (26) yields u_{in}^E . Plugging Eqs. (10) and (24) into Eq. (27) then yields u_{id}^E . For DISCRIM, plugging Eq. (13) into Eq. (26) yields u_{in}^D . Plugging Eqs. (13) and (24) into Eq. (27) then yields u_{id}^D . For BESTPR, plugging Eq. (18) into Eq. (26) yields u_{in}^B . Plugging Eqs. (18) and (24) into Eq. (27) then yields u_{id}^B .

¹⁴ First note that the critical discount factor in a treatment is the highest critical discount factor of all individual firms. Most of the results follow directly from Table 1. Note that $\hat{\delta}^U(\alpha_i) \leq 1/2$ if and only if $(1 - \alpha_i)^2 \leq 1 - \alpha_i - \sum_{j \neq i} \alpha_j^2$, hence if and only if $\alpha_i \geq \sum_j \alpha_j^2$. From Footnote 10, $\sum_j \alpha_j^2 \geq 1/n$, with equality if all firms are equally sized. As $\alpha_i \geq \sum_j \alpha_j^2$ has to be satisfied for all firms, this implies that it cannot be satisfied if $\alpha_i < 1/n$, and is satisfied with equality if all firm are of size $1/n$.

Table 1

Non-cooperative utility (u_{iN}^y), defection utility (u_{iD}^y) and critical discount factor ($\hat{\delta}^y(\alpha_i)$) for all yardstick types.

Yardstick type	u_{iN}^y	u_{iD}^y	$\hat{\delta}^y(\alpha_i)$
UNIFORM	$A_i(b^2 - 1 + 2 \sum_{j=1}^n \alpha_j^2 - \alpha_i^2)$	$A_i(b^2 + 1 + \alpha_i^2 - 2\alpha_i)$	$\frac{1}{2} \frac{(1-\alpha_i)^2}{1-\alpha_i - \sum_{j \neq i}^n \alpha_j^2}$
EQUALW	$A_i(b^2 - (\frac{n-1}{n})^2)$	$A_i(b^2 + (\frac{n-1}{n})^2)$	$\frac{1}{2}$
DISCRIM	$A_i(b^2 - 1)$	$A_i(b^2 + 1)$	$\frac{1}{2}$
BESTPR	$A_i(b^2 - 1)$	$A_i(b^2 - 1)$	0

3.4. Summing up

From the analysis above, we thus have the following. If managers indeed set their cost levels non-cooperatively, we expect to see the highest prices in BESTPR, followed by UNIFORM; EQUALW; and DISCRIM, respectively, where UNIFORM and EQUALW yield the same prices with equally sized firms. Yet, collusion is easiest to sustain in BESTPR, followed by EQUALW and DISCRIM; and UNIFORM, respectively, where the last three yield the same amount of collusion with equally sized firms.

From a theoretical perspective, we thus expect to see the highest prices in BESTPR: noncooperative cost levels are then the highest, and collusion is easiest to sustain. We also expect prices to be higher in EQUALW than in DISCRIM, as noncooperative cost levels are higher, and collusion is equally hard to sustain. But it is ambiguous how prices in UNIFORM compare to those in either EQUALW or DISCRIM. On the one hand, noncooperative cost levels are higher in UNIFORM, but on the other hand, collusion is harder to sustain, rendering the net effect ambiguous. Therefore, we use an experiment to compare the price effects of these treatments.

We thus have

Corollary 1. *Prices in BESTPR are higher than prices in any other treatment. Prices in EQUALW are higher than those in DISCRIM. All other pairwise comparisons are ambiguous from a theoretical perspective.*

4. Experimental design

In our experiment subjects play the game described above for at least 20 rounds. We set $a = 1/24$ and $b = 1$ which ensures that the non-cooperative and collusive cost levels in all our treatments are integers. To avoid possible end-game effects, the experiment ends with a probability of 20% in each round from round 20 onwards (see also Normann and Wallace, 2012). Subjects are randomly matched into groups at the beginning of the experiment and play all rounds within the same group.

Every round consists of three steps. First, subjects can communicate using a chat screen. This chat is completely anonymous. Second, subjects unilaterally choose cost levels c_i from the set $\{1, 2, \dots, 20\}$.¹⁵ Third, prices are determined using the relevant yardstick. After each round, subjects learn the profits and managerial benefits they have realized, and the cost levels that each subject has set.¹⁶ The unrestricted communication we allow for creates circumstances that

are closest to the real world. The Dutch regulatory framework for the distribution-network operators, for example, does not forbid them to mutually exchange information on operational and strategic issues. The operators together have established a joint organization (*Netbeheer Nederland*) which facilitates the cooperation among the operators besides representing the group of operators in the political debate on energy policy in general and the design of regulation in particular. Also, without communication, it is hard to sustain collusion in a cartel experiment with more than two players, see e.g. Haan et al. (2009).

Our experiment is twofold. Experiment A serves to compare the four different yardsticks, and does so in an industry structure with two firms, where one firm is twice as big as the other: with equally sized firms, there would be no difference between the UNIFORM and EQUALW treatments. Experiment B zooms in on the difference between the DISCRIM and UNIFORM. These two yardsticks perform relatively well in experiment A, whereas we do not have a clear theoretical prediction concerning the relative performance of the two, making a closer look all the more relevant. We do so by considering a number of industry structures that differ with respect to the number of firms as well as their relative size.

The treatments in experiment A are given in Table 2. For each treatment, the table also shows the non-cooperative cost level of each firm, the regulated price that results, the symmetric cost level that maximizes joint profits, and the critical discount factor.¹⁷ All these results follow from our discussion in Section 3.¹⁸

Throughout the experiment, we will use a loose interpretation of the critical discount factor, rather than the strict theoretical interpretation that there will always be collusion if the discount factor is bigger than $\hat{\delta}$ and there will never be collusion if it is smaller than $\hat{\delta}$. Instead, we interpret a higher $\hat{\delta}$ as it being less likely that there will be collusion, which is in line with other work (see e.g. Bigoni et al., 2012, but also Potters et al., 2004, or Dijkstra et al., 2017). Note

¹⁷ Strictly speaking, the critical discount factors we derive are only valid from round 20 onwards. Still, it is easy to see that this does not affect our qualitative results. If a cartel is stable from round 20 onwards, it is definitely stable in earlier rounds as well, since the incentives to defect become smaller as the first 19 rounds are played for sure. If a cartel is not stable from round 20 onwards, then subjects know it will break down in that round and hence, by backward induction, are not willing to form a cartel in an earlier round.

¹⁸ In fact, strictly speaking, in DISCRIM, the theoretical prediction would be for all firms to set a cost level equal to 0. However, we do not allow our experimental subjects to choose cost levels lower than 1; doing so would not only yield zero profits but also zero managerial benefits, which is likely to affect their behavior in ways that interfere with the purpose of our experiment. However, that also has an effect on the incentives to collude; taking into account that costs cannot be lower than 1 imply a critical discount factor of $\hat{\delta}^D = 0.54$. After careful deliberation, we have decided to work with our original theoretical predictions – effectively assuming that for the subjects, the difference between $\hat{\delta} = 0.50$ and $\hat{\delta} = 0.54$ is negligible. Using $\hat{\delta}^D = 0.54$ rather than $\hat{\delta}^D = 0.50$ would hardly affect our results. If we would use $\hat{\delta}^D = 0.54$, our statistical test when comparing the incidence of collusion should be one-sided rather than two-sided when comparing UNIFORM and DISCRIM, or when comparing EQUALW and DISCRIM. In that case, the comparison in Duo66 in Table 9 would become significant at 10% (contradicting our hypothesis), as would that in Trio444 (confirming our hypothesis). In Table 5 the comparison between EQUALW and DISCRIM would become significant at 5% rather than 10%.

¹⁵ As we saw in the theory section, both the non-cooperative and the collusive strategy are independent of the competitors' actions. In that sense, our game boils down to a relatively straightforward prisoners' dilemma game with two strategies. Still, we choose not to represent it as such in our experiment, as that would get rid of much of the regulatory context that we feel is crucial for our purpose.

¹⁶ Of course, in our set-up profit and managerial benefit are equally important in determining payoffs. This was also outlined in the experimental instructions. Subjects indeed seemed to understand this as they correctly answered control questions on this topic.

Table 2
Treatments experiment A.

Treatment	Non-cooperative cost level			Collusive cost level	$\hat{\delta}$
	c_1^*	c_2^*	Price		
UNIFORM	8	4	6.67	12	1.00
EQUALW	6	6	6.00	12	0.50
DISCRIM	1	1	1.00	12	0.50
BESTPR	12	12	12.00	12	0.00

All treatments: $\alpha_1 = 2/3$, $\alpha_2 = 1/3$.

that for our specific parametrization the critical discount factor in UNIFORM is equal to 1, which implies that the theoretical prediction is that a cartel will never be stable under that yardstick.

Experiment B consists of 10 treatments: we study 5 different industry structures and consider DISCRIM and UNIFORM in each. We now look at treatments with 2 and with 3 firms, that range from equally-sized to highly asymmetric firms. Table 3 gives an overview of the treatments. Treatments with 3 firms are denoted TrioXYZ, with X, Y and Z the relative sizes of firms 1, 2 and 3, respectively. Treatments with 2 firms are denoted DuoXY, with X and Y the relative sizes of firms 1 and 2. Treatments Duo84 are identical to those in experiment A; we include them here as well for ease of comparison. The five treatments with a uniform yardstick are also studied in a companion paper (Dijkstra et al., 2017), where we study the effect of industry structure on the performance of the (weighted) uniform yardstick. Again, the table shows the non-cooperative cost level of each firm in each treatment, the regulatory price that results, the symmetric collusive cost level, and the critical discount factor that all follow directly from the analysis in Section 3.

The experiment was conducted at the Groningen Experimental Economics Laboratory (GrEElab) at the University of Groningen in 2013. A total of 132 subjects participated in experiment A, and 423 in experiment B (of which 2 sessions are used from experiment A). All students were recruited from the University of Groningen (86%) or the Hanze University of Applied Sciences (14%), most of them in the fields of economics and business (62%). Every session consisted of one treatment and lasted between 80 and 130 min. Subjects signed in for sessions, while treatments were randomly assigned to sessions.

Every treatment with two players was played in two sessions, every treatment with three players in three. Between 14 and 18 subjects participated in a session, resulting in 15 to 18 groups per treatment. The experiment was programmed in z-Tree (Fischbacher, 2007). Printed instructions were provided and read aloud.¹⁹ On their computer, subjects first had to answer a number of questions correctly to ensure understanding of the experiment.

Participants were paid their cumulative earnings in euros. Since firm size differed between participants, exchange rates were varied such that they would receive identical amounts if they collude on cost level 12.²⁰ Also, participants received an initial endowment of €4. Average earnings were €16.22, ranging from €4.70 to €24.00.

5. Comparing the four yardsticks

In this section, we discuss the results of experiment A, where we compare 4 yardsticks in a given industry structure. We use non-parametric tests for differences between treatments as is common in experimental studies.²¹ We use the Mann-Whitney *U* test (MWU) to test for differences between two populations, and the Kruskal-Wallis test to test for an ordering of three or more populations.

Fig. 1 shows average prices over time.^{22,23} The figure suggests that BESTPR indeed yields the highest prices, while DISCRIM yields the lowest.

A Kruskal-Wallis test rejects the hypothesis that all four yardsticks lead to the same price ($p < 1\%$). For pairwise comparisons between any two treatments, we use the Mann-Whitney *U* (MWU) test. Results are given in Table 4. The entries in the right-hand panels indicate whether the row treatment yields prices that are significantly lower ($<$) or higher ($>$) than the column treatment, or whether the difference is not significant ($\approx$). We use this convention throughout this paper.

We thus have the following:

Result 1. Prices with a best-practice yardstick are higher than prices with any other yardstick. Prices with a uniform yardstick with an unweighted price are higher than those with a discriminatory yardstick.

Note that the result is in line with Corollary 1. We also find that prices in EQUALW are higher than those in UNIFORM, but not significantly so. Apparently, in this case, the collusion effect (collusion is easier to sustain in EQUALW) dominates the non-cooperative effect (non-cooperative cost levels are lower in EQUALW). The same reasoning holds for the insignificant result that prices in UNIFORM are higher than in DISCRIM.

In the above, we only tested Corollary 1, which considers the net effect on regulated prices of non-cooperative cost levels on the one hand, and the extent of collusion on the other. Yet, our experimental results also allows us to disentangle these two effects.

First consider the extent of collusion. Table 5 considers the incidence of full collusion (i.e. the percentage of cases where all firms set the collusive cost level of 12). A Kruskal-Wallis test indicates differences between all yardsticks ($p < 0.1\%$). Pairwise comparisons are given in Table 5. Theorem 2 predicts the highest incidence of collusion in BESTPR, followed by EQUALW and DISCRIM, whereas UNIFORM has the lowest incidence. This is largely what we find in Table 5: the incidence of full collusion in BESTPR is significantly higher than in any other treatment, whereas the incidence in UNIFORM is significantly lower than that in EQUALW. We do however find a difference between DISCRIM and EQUALW (albeit only significant at 10%), whereas we do not find a difference between DISCRIM and UNIFORM. A Jonckheere-Terpstra test²⁴ with the ordered alternative hypothesis that the incidence of full collusion is highest in BESTPR, followed by EQUALW, DISCRIM, and UNIFORM, respectively, is highly significant ($p < 0.1\%$).

Second, consider the effect on the non-cooperative cost level. Table 6 looks at the average price in rounds where there is no collusion in the sense defined above (i.e. it is not the case that all firms set a cost level of 12). Arguably, these are the cases in which firms behave non-cooperatively. Note that for BESTPR the results are hard to interpret, as there both the collusive and the non-cooperative cost level equals 12. The results for the other treatments are perfectly in line with Theorem 1: prices in DISCRIM are lower than those in either UNIFORM or EQUALW, whereas prices in UNIFORM are also higher than those in EQUALW. A Jonckheere-Terpstra test with the ordered alternative hypothesis that prices in UNIFORM are higher than those in EQUALW which in turn are higher than those in DISCRIM, is highly significant ($p < 0.1\%$).

¹⁹ Instructions for Trio642 Discriminatory are reproduced in the online appendix. Instructions for other treatments are similar and available upon request.

²⁰ This allows us to focus on the effect of asymmetries in industry structure without the possible interference of fairness issues.

²¹ Such tests do not impose a specific distribution on the data and are less sensitive to outliers than parametric tests.

²² Note that in DISCRIM, we take the weighted average price as defined in Eq. (14).

²³ The development of costs over time for each individual market are given in the online appendix.

²⁴ This test is equivalent to the Kruskal-Wallis test but has an unordered alternative hypothesis. See Jonckheere (1954) or Terpstra (1952).

Table 3
Treatments experiment B.

Industry structure	Market share			Non-cooperative cost level					Collusive cost level	δ^U	δ^D
				UNIFORM				DISCRIM			
	1	2	3	c_1^*	c_2^*	c_3^*	p^*	c_1^*, p^*			
Trio444	1/3	1/3	1/3	4	4	4	4.00	1.00	12	0.50	0.50
Trio633	1/2	1/4	1/4	6	3	3	4.50	1.00	12	0.64	0.50
Trio642	1/2	1/3	1/6	6	4	2	4.67	1.00	12	0.74	0.50
Duo66	1/2	1/2		6	6		6.00	1.00	12	0.50	0.50
Duo84	2/3	1/3		8	4		6.67	1.00	12	1.00	0.50

Fig. 1. Average price per round (across all groups).

6. Comparing the uniform and discriminatory benchmarks

Fig. 2 gives the average price in each round in UNIFORM and DISCRIM, for each of the 5 industry structures we consider in experiment B. In Table 7 we compare prices for each industry structure.

For most industry structures we find higher prices with UNIFORM. This difference is significant for Trio444 and Trio633. The only exception is Duo66, where DISCRIM yields higher prices – but not significantly so. Interestingly, Potters et al. (2004) also found higher prices with a discriminatory rather than a uniform yardstick, in an industry structure that is equivalent to our Duo66.

Tests for multiple industry structures indicate that DISCRIM yields significantly lower prices in triopolies and heterogeneous industries. Most importantly, taking all industry structures together, DISCRIM yields significantly lower prices at the 10%-level.²⁵

Table 8 takes a parametric approach, comparing both yardsticks by means of random-effects OLS regressions. Column (1) explains price solely by a constant and a DISCRIM-dummy. Column (2) adds dummies to correct for the effect of different industry structures.²⁶

²⁵ In this case, we have done the tests for multiple industries by simply interpreting all data generated by the UNIFORM treatments as coming from a single population, and treating all data generated by the DISCRIM treatments likewise. One alternative to make the data generated by different industry structures more comparable, is to do the following. First subtract the average price in DISCRIM/Trio444 from all observations in Trio444. Then, do the same for all other industry structures (thus, subtract the average price in DISCRIM/Trio633 from all observations in Trio633, etc.). Treating the resulting cleaned data generated by the UNIFORM treatments as coming from a single population, and treating all resulting cleaned data generated by the DISCRIM treatments likewise, allows us to filter out the effect of industry structure while still doing a non-parametric test on the difference between the two yardsticks. Performing this analysis, however, yields the same qualitative results as those given in Table 7.

²⁶ Note therefore that this is essentially the parametric version of then non-parametric test we propose in Footnote 25.

In both regressions, a DISCRIM yardstick leads to significantly lower prices at the 10%-level. We thus have

Result 2. Prices are lower with a discriminatory yardstick than with a uniform yardstick. This effect is largely driven by industries with 3 firms, and industries where market shares are heterogeneous.

Again, we can disentangle the effects of collusion and noncooperative cost levels. Table 9 looks at the incidence of collusion. In this case, we see little difference between the two yardsticks, even after aggregating them. The only significant difference is found in Trio633, where UNIFORM leads to a higher incidence of collusion, an effect that runs counter to our theoretical prediction. Average prices in non-collusive rounds are given in Table 10. Non-cooperative cost levels in UNIFORM are indeed clearly higher than those in DISCRIM, at a significance level of 0.1%. Hence, lower non-cooperative cost levels are the driving force behind the lower prices in DISCRIM.

Whereas our theoretical framework predicts a tradeoff between a non-cooperative and a collusive effect when comparing DISCRIM and UNIFORM, our experimental results indicate that the collusive effect in fact is not really there. That is remarkable, as we do find

Table 4
Price per treatment (across all rounds and groups).

Treatment	Average	EQUALW	DISCRIM	BESTPR
UNIFORM	10.76	≈	≈	<***
EQUALW	10.99		>*	<+
DISCRIM	9.80			<***
BESTPR	11.65			

Entries in the right-hand panel indicate whether the row treatment yields prices that are significantly lower (<), significantly higher (>) or that do not differ significantly (≈) from the prices in the column treatment. Differences between two treatments are tested using the MWU test for equality (two-sided for UNIFORM–EQUALW and UNIFORM–DISCRIM, one-sided for the other comparisons).

+ Significantly different at 10%.

* Significantly different at 5%.

** Significantly different at 1%.

*** Significantly different at 0.1%.

Table 5
Incidence of full collusion per treatment (across all rounds and groups).

Treatment	Average	EQUALW	DISCRIM	BESTPR
UNIFORM	43.4%	<*	≈	<***
EQUALW	71.2%		>+	<+
DISCRIM	56.1%			<***
BESTPR	92.0%			

Entries in the right-hand panel indicate whether the row treatment yields collusion rates that are significantly lower (<), significantly higher (>) or that do not differ significantly (≈) from the rates in the column treatment. Differences between two treatments are tested using the MWU test for equality (two-sided for EQUALW–DISCRIM, one-sided for the other comparisons).

+ Significantly different at 10%.

* Significantly different at 5%.

*** Significantly different at 0.1%.

Table 6

Price if no collusion, per treatment.

Treatment	Average	EQUALW	DISCRIM	BESTPR
UNIFORM	9.76	> +	> **	> *
EQUALW	8.59		> +	≈
DISCRIM	7.12			≈
BESTPR	7.11			

Only rounds are included in which not all firms in a group set the fully collusive cost level of 12. Entries in the right-hand panel indicate whether the row treatment yields prices that are significantly lower (<), significantly higher (>) or that do not differ significantly (≈) from the average prices in the column treatment. Differences between two treatments are tested using the MWU test for equality (one-sided).

+ Significantly different at 10%.

* Significantly different at 5%.

** Significantly different at 1%.

strong support for our theoretical predictions concerning the extent of collusion when we compare four yardstick in experiment A.

7. Conclusion

In this paper we have analyzed the effect of the design of yardstick competition on the prices consumers have to pay. As alternative designs, we distinguish the weighted uniform yardstick, the unweighted uniform yardstick, the discriminatory yardstick, and the best-practice yardstick. In all cases, the effect of a design on prices depends on two different mechanisms: the incentives for firms to increase productive efficiency on the one hand and the incentives for firms to collude on the other. The risk of collusion in case of yardstick competition was already mentioned by [Shleifer \(1985\)](#), but to what extent that risk could mitigate the benefits of yardstick competition has hardly been analyzed up to now. In an economic experiment, [Potters et al. \(2004\)](#) find that the benefits of a discriminatory yardstick in terms of a higher productive efficiency are more than neutralized by the costs in terms of higher prices resulting from an increased risk of collusion. Their experiment is conducted using two equally sized firms, triggering the question whether it also

Table 7

Comparison of price for UNIFORM and DISCRIM for each industry structure (averages across all rounds and groups).

Structure	UNIFORM		DISCRIM
Trio444	11.00	> +	10.48
Trio633	10.04	> *	8.45
Trio642	9.48	≈	9.06
Triopolies	10.18	> *	9.34
Duo66	10.32	≈	10.73
Duo84	10.76	≈	9.80
Duopolies	10.54	≈	10.24
Homogeneous	10.67	≈	10.61
Heterogeneous	10.08	> *	9.13
All	10.32	> +	9.71

Entries between values indicate whether prices in UNIFORM are significantly lower (<), significantly higher (>), or do not differ significantly (≈) from those in DISCRIM. Differences are tested using the MWU test for equality (one-sided for Trio444, Duo66 and Homogeneous; two-sided for other comparisons).

+ Significant at 10%.

* Significant at 5%.

holds in heterogeneous industries. In a companion paper ([Dijkstra et al., 2017](#)), we found that the degree of heterogeneity affects the probability of collusion with a uniform weighted yardstick. More specifically, we found that the probability of collusion in case of a uniform yardstick is negatively related to the level of heterogeneity of an industry. Using a similar experimental set up, we now have analyzed the ultimate price effects of different yardstick designs against the background of a heterogeneous industry.

On the basis of a theoretical analysis we find that the best-practice yardstick yields the lowest power of incentive to improve efficiency as well as the highest probability of collusion. Hence, this yardstick is expected to result in the highest prices, which was confirmed by the experimental analysis. Despite its name, the best-practice yardstick has the worst performance seen from the perspective of consumer prices. This is an important observation, as some form of this yardstick is applied in several industries in many countries.

Fig. 2. Average price per round for each industry structure (across all groups).

Table 8
Regression results for price.

	(1)	(2)
DISCRIM	−0.603 ⁺ (0.356)	−0.618 ⁺ (0.340)
Trio633		−1.487 ^{**} (0.533)
Trio642		−1.467 ^{**} (0.527)
Duo66		−0.209 (0.489)
Duo84		−0.468 (0.563)
Constant	10.32 ^{***} (0.221)	11.04 ^{***} (0.392)
Observations	3260	3260

Dependent variable: price in each group.

Clustered standard errors on group level in parentheses.

⁺ Significant at 10%.^{**} Significant at 1%.^{***} Significant at 0.1%.

Both the theoretical analysis and the economic experiment show that the discriminatory yardstick results in lower prices than the unweighted uniform yardstick. The theory, however, does not give a clear answer on the relative performance of the discriminatory versus the weighted uniform yardstick. The discriminatory yardstick gives the highest power of incentive to reduce costs, but it also has a higher probability of collusion than the uniform yardstick. In the experimental analysis, we find that the advantage of the discriminatory yardstick in terms of giving incentives to improve productive efficiency exceeds the disadvantage of a relatively higher risk of collusion. This conclusion appears to be robust for different degrees of heterogeneity of the industry. So, we conclude that the discriminatory yardstick yields the lowest prices for consumers.

From our analysis, we learn that yardstick competition may indeed result in collusion between the regulated firms, but that this effect does generally not outweigh the benefits of this type of competition for consumer prices. So, implementing yardstick competition in order to mimic competition in industries where otherwise no competition would exist, can be beneficial for consumers. Having said this, it must be acknowledged that collusion may occur and the risk of it depends on the degree of heterogeneity of the industry as well as the design of the yardstick.

Table 9
Incidence of full collusion in UNIFORM and DISCRIM for each industry structure (averages across all rounds and groups).

Structure	UNIFORM		DISCRIM
Trio444	71.7%	≈	62.8%
Trio633	58.1%	> *	26.3%
Trio642	34.4%	≈	44.0%
Triopolies	54.7%	≈	44.4%
Duo66	56.2%	≈	73.8%
Duo84	43.4%	≈	56.1%
Duopolies	49.8%	≈	64.4%
Homogeneous	64.2%	≈	68.3%
Heterogeneous	45.1%	≈	42.7%
Overall	52.8%	≈	52.8%

Entries between values indicate whether the incidence of full collusion under UNIFORM is significantly lower (<), significantly higher (>), or does not differ significantly (≈) from the price under DISCRIM. Differences between two treatments are tested using the MWU test for equality (two-sided for Trio444, Duo66, Homogeneous, Triopolies, Duopolies and Overall; one-sided for other comparisons).

* Significant at 5%.

Table 10
Prices if no full collusion in UNIFORM and DISCRIM for each industry structure (averages across all rounds and groups).

Structure	UNIFORM		DISCRIM
Trio444	8.03	≈	8.09
Trio633	7.78	≈	7.27
Trio642	8.27	> *	6.91
Triopolies	8.07	> +	7.43
Duo66	8.73	> +	6.96
Duo84	9.76	> **	7.12
Duopolies	9.26	> ***	7.06
Homogeneous	8.44	≈	7.62
Heterogeneous	8.70	> ***	7.11
Overall	8.61	> ***	7.29

Entries between values indicate whether prices under UNIFORM are significantly lower (<), significantly higher (>), or do not differ significantly (≈) from the price under DISCRIM. Differences are tested using the MWU test for equality (one-sided).

⁺ Significant at 10%.^{*} Significant at 5%.^{**} Significant at 1%.^{***} Significant at 0.1%.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.eneco.2017.06.016>.

References

- Argenton, C., Müller, W., 2012, Nov. Collusion in experimental Bertrand duopolies with convex costs: the role of cost asymmetry. *Int. J. Ind. Organ.* 30 (6), 508–517.
- Bigoni, M., Fridolfsson, S.-O., Le Coq, C., Spagnolo, G., 2012. Fines, leniency and rewards in antitrust. *RAND J. Econ.* 43 (2), 368–390.
- Blázquez-Gómez, L., Grifell-Tatjé, E., 2011. Evaluating the regulator: winners and losers in the regulation of Spanish electricity distribution. *Energy Econ.* 33 (5), 807–815.
- Burns, P., Jenkins, C., Riechmann, C., 2005. The role of benchmarking for yardstick competition. *Util. Policy* 13 (4), 302–309.
- Dalen, D.M., Gomez-Lobo, A., 2002. Regulatory contracts and cost efficiency in the Norwegian bus industry: do high-powered contracts really work? Discussion Paper 6. Norwegian School of Management.
- Dalen, D.M., Gómez-Lobo, A., 2003. Yardsticks on the road: regulatory contracts and cost efficiency in the Norwegian bus industry. *Transportation* 30, 371–386.
- Dassler, T., Parker, D., Saal, D.S., 2006, Sep. Methods and trends of performance benchmarking in UK utility regulation. *Util. Policy* 14 (3), 166–174.
- Dijkstra, P.T., Haan, M.A., Mulder, M., 2017. Industry structure and collusion with uniform yardstick competition: theory and experiments. *Int. J. Ind. Organ.* 50, 1–33.
- Dugar, S., Mitra, A., 2009. The Size of the Cost Asymmetry and Bertrand Competition: Experimental Evidence, Unpublished Manuscript.
- Dugar, S., Mitra, A., 2016. Bertrand competition with asymmetric marginal costs. *Econ. Inq.* 54 (3), 1631–1647.
- Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* 10 (21), 171–178.
- Fonseca, M.A., Normann, H.-T., 2008, Mar. Mergers, asymmetries and collusion: experimental evidence. *Econ. J.* 118 (527), 387–400.
- Friedman, J.W., 1971. A non-cooperative equilibrium for supergames. *Rev. Econ. Stud.* 38 (1), 1–12.
- Haan, M.A., Schoonbeek, L., Winkel, B.M., 2009. Experimental results on collusion. In: Hinlopen, J., Normann, H.-T. (Eds.), *Experiments and Competition Policy*. Cambridge University Press, pp. 9–33, chap. 2.
- Haffner, R., Helmer, D., van Til, H., 2010. Investment and regulation: the Dutch experience. *Electr. J.* 23 (5), 34–46.
- Jamasb, T., Pollitt, M., 2007. Incentive regulation of electricity distribution networks: lessons from the experience from Britain. *Energy Policy* 35 (12), 6163–6187.
- Jonckheere, A.R., 1954. A distribution-free K-sample test against ordered alternatives. *Biometrika* 41 (1/2), 133–145.
- Mason, C.F., Phillips, O.R., Nowell, C., 1992, Nov. Duopoly behavior in asymmetric markets: an experimental evaluation. *Rev. Econ. Stat.* 74 (4), 662–670.
- Mizutani, F., Kozumi, H., Matsushima, N., 2009, Dec. Does yardstick regulation really work? Empirical evidence from Japan's rail industry. *J. Regul. Econ.* 36 (3), 308–323.
- Mizutani, F., Usami, M., 2016. Yardstick regulation and the operators' productivity of railway industry in Japan. *Res. Transp. Econ.* 59, 86–93.
- Mulder, M., 2012. Financeability of investments and allocation of costs: an assessment of the incentive regulation of the Dutch high-voltage network. *Compet. Regul. Network Ind.* 13 (2), 160–186.
- Normann, H.-T., Wallace, B., 2012, Aug. The impact of the termination rule on cooperation in a prisoner's dilemma experiment. *Int. J. Game Theory* 41 (3), 707–718.

- Phillips, O.R., Menkhaus, D.J., Thurow, J.N., 2011. The small firm in a quantity choosing game: some experimental evidence. *Rev. Ind. Organ.* 38 (2), 191–207.
- Potters, J., Rockenbach, B., Sadriehc, A., van Damme, E., 2004, Sep. Collusion under yardstick competition: an experimental study. *Int. J. Ind. Organ.* 22 (7), 1017–1038.
- Shleifer, A., 1985. A theory of yardstick competition. *RAND J. Econ.* 16 (3), 319–327.
- Tangerås, T.P., 2002. Collusion-proof yardstick competition. *J. Public Econ.* 83 (2), 231–254.
- Terpstra, T.J., 1952. The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking. *Indag. Math.* 14 (3), 327–333.
- Viscusi, W.K., Harrington, J.E., Vernon, J.M., 2005. *Economics of Regulation and Antitrust*. 4th edn., MIT Press, Cambridge, MA., pp. 960.
- Yatchew, A., 2000. Scale economies in electricity distribution: a semiparametric analysis. *J. Appl. Economet.* 15 (2), 187–210.
- Yatchew, A., 2001. Incentive regulation of distribution utilities using yardstick competition. *Electr. J.* 14 (1), 56–60.