Meya, Johannes

Conference Paper

Dynamics of Yardstick Regulation: Historical Cost Data and the Ratchet E ffect

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Regulated Industries - Theory, No. F14-V2

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Dynamics of Yardstick Regulation: Historical Cost Data and the Ratchet Effect

Johannes Meya*

August, 2013

First Version

Abstract

Real life applications of Yardstick Regulation frequently refer to historical cost data. While Yardstick Regulation cuts the link between firms' own costs and prices firms may charge in a static setting, it does not in a dynamic setting where historical cost data is used. A firm *can* influence the price it will be allowed to charge in the future if its behavior today can affect future behavior of other firms that determines the price this firm will be able to charge later on. This paper shows that, assuming that slack, inflating of costs, is beneficial to firms, a trade-off between short term profit through abstinence from slack and the benefit of slack in (infinitely) many periods arises. A ratchet effect that Yardstick Regulation was meant to overcome can occur and firms can realize positive rents *because of* the use of historical cost data, even if firms are identical. Equilibria with positive slack can exist without any collusion between firms or threat. Moreover, this problem is more severe if the firm with lowest costs of all other firms instead of the average firm is the yardstick.

Keywords: Yardstick Regulation, Yardstick Competition, Ratchet Effect, Historical Cost Data
JEL classification: L51, L98, L97

*Chair of Public Economics, Georg-August University Göttingen, jmeya@uni-goettingen.de

1 Introduction

Natural monopolies are frequently subject to regulation. As 'natural' competition does not force prices to be a perfect competition outcome, often regulatory agencies jump in to 'regulate' profit, prices or revenue. Under traditional Rate of Return Regulation allowed profit of a firm is linked to capital employed. The well known result of Averch and Johnson (1962) is that this regulation provides incentives for the firms to employ an inefficient input mix and not to engage in cost minimizing behavior - in other words: to produce with some slack. Incentive Regulation is meant to address this issue. Price Cap Regulation, originally suggested by Littlechild (1983), decouples costs incurred and prices allowed to be charged by fixing or capping prices, no matter what costs are. The firm thus becomes the residual claimant of all costs not incurred and so has a strong incentive to produce without slack, if profit is worth more to the firm than slack is. Necessarily the question arises, how the price cap should be defined. If the regulator takes into account profits made and costs incurred the incentive structure is much less clear cut, as e.g. Train (1991) points out. The basic idea of Yardstick Regulation as described by Shleifer (1985) solves this problem by using information on costs of *other* comparable firms to define prices a firm is allowed to charge. In a static world and in every period prices and costs for each individual firm are as a consequence completely independent of each other. In absence of collusion Yardstick Regulation fosters efficient production especially if firms and circumstances of production are very similar. Tangerås (2002) summarizes: "the regulator is able to extract all surplus from firms and reach full efficiency if technologies are perfectly correlated." This paper shows that this property does not carry over into a dynamic setting if historical cost data is used. A firm can influence the price it is allowed to charge in the future via its effect on the behavior of other firms. A ratchet effect can occur under Yardstick Regulation using historical cost data without any collusion as a result of individual and independent decision making of firms.

The contribution of this paper is twofold: In a simple model with three firms and an infinite horizon we show that every firm can effectively influence the price it is allowed to charge if this price is a function of the costs of the other two firms in the period before. By this we point out a feature of real world applications of Yardstick Regulation that has not received much attention, both in academic literature and in regulatory practice: historical costs are used to define constraints. We show that steady state equilibria with positive slack are possible without any collusion if firms attach at least some positive value to slack. We furthermore compare two variants

of Yardstick Regulation: Either the firm with the lowest costs of all firms but the evaluated one or the average of the other firms can be used as the yardstick. Intuitively orientation at the best seems to be the tougher form. We show that choosing this scheme might lead to higher slack and a worse situation for society however.

In the 'real world' the price of a service is usually set before customers use the service. For Incentive Regulation this means that e.g. at the start of a regulatory period constraints are defined based on observations of costs from the regulatory period before (see e.g. the decree on Incentive Regulation for electricity and gas networks in Germany (ARegV)). In the debate prior to this decree the question which yardstick, average performance or best practice/frontier, should be chosen was intensively discussed (see e.g. Bundesnetzagentur (2006)). This indicates that it is not obvious, what the optimal yardstick is. The German legislator decided to apply a frontier approach, a well cited example for the use of average performance for regulation is Medicare (originally Shleifer (1985)).
Aspects or drawbacks of Yardstick Regulation that are subject to debate are concerns that quality might be adversely affected under Incentive Regulation in general what makes additional quality regulation necessary, see Sappington (2005) for a survey, collusion among firms (e.g. Tangerås (2002)) and investment behavior (e.g. Dalen (1998)). Another issue is that firms might lack comparability, necessary for implementation of Yardstick Regulation (e.g. Laffont and Tirole (1993)). In this paper we abstract from these issues and show that still the desired outcome, i.e. efficient production, might not be reached.
In our dynamic model with three firms, an infinite horizon and discrete time we examine the effect of the use of historical cost data under Yardstick Regulation. As we are interested in the long run effects of regulation we focus on the analysis of resulting steady state equilibria. We show that in absence of all kinds of collusion or Folk Theorem arguments steady state equilibria with positive slack, i.e. inefficient production and positive rents for firms, can exist. We furthermore show that the highest slack that can exist in such a steady state equilibrium is higher if the firm with the lowest costs of all other firms instead of the average of the other firms is used as the yardstick.
The modelling of slack, i.e. lack of costly effort, in this paper differs from a major part of contributions to the debate on Incentive Regulation, represented especially by Laffont and Tirole (1993): in these models costly effort reduces costs of production. In our model inefficiency costs, slack, are added to real, necessary costs of efficient production. The instantaneous payoff function used is very similar to the one in Blackmon (1994). This is done as it allows for straightforward interpretation of the results and explicitly models the idea that Yardstick Regulation is meant to

solve the inefficiency problem of traditional Rate of Return Regulation. In models considering only two periods, the effect driving our results does not appear: Under Yardstick Regulation using historical cost data current choices of a firm do not affect the price this firm can charge in the current and the next period. The direct effect is visible from the next but one period on, only. Like Meran and Hirschhausen (2009) we use dynamic programming techniques to account for long run effects of the decisions of firms, however, we come to differing conclusions. The key difference between their model, which is expanding the model of Shleifer (1985), and our approach that leads to these differing results is that Meran and Hirschhausen (2009) do not allow for benefit for the firms from slack, consequently firms do not gain from keeping costs high.
The remainder of this paper is structured as follows: Section 2 explains the model setup. In section 3 all possible (punishment-free) Markov-perfect steady state equilibria are characterized, existence is proven and the two regulatory schemes are compared with respect to equilibrium outcomes. Section 4 concludes.

2 Description of the model

2.1 Firms

There are three firms, labelled $j = i, o, x$, each producing a homogeneous output normalized to one. The output is bought by the consumers. One could think of demand for electricity that is very inelastic with respect to price or demand for some crucial medical treatment. These firms could be thought of catering three comparable regions with electricity grids as local monopolists. The only way they interact in 'competition' is via the regulation imposed to them. In every period the regulator assigns a price to each of the firms. Each firm may not charge more than this price for its output, so the regulator defines a price cap which is equivalent to a revenue cap under the assumption of completely inelastic demand. As demand does not react to price in this setting all firms always charge the maximum price they are allowed to. While the firms' output is directly observable the underlying cost structure is unknown to the regulator. Each firm verifiably reports its costs to the regulator who cannot distinguish between 'real' necessary cost, $C > 0$, and slack, $S_t^j \geq 0$, defined as additional costs due to inefficient use of resources, and only observes the sum of both. C does not change over time and is the same for all firms. This is equivalent to assuming that the regulator correctly and completely accounts for all heterogeneity between firms respectively (exogenous) circumstances

of production.[1] Each firm chooses its slack and may choose different slacks in different periods. Slack can be interpreted e.g. as lack of (costly) effort from managers, oversized offices or all kinds of 'unnecessary' costs that might occur under Rate of Return Regulation. As slack is inefficient production by definition, the regulator maximizing utility of society desires to avoid all slack without explicit consideration of a target function.

If firm j chooses a positive slack in period t it realizes a nonmonetary utility denoted by $B(S_t^j)$. B is twice continuously differentiable with $B(0) = 0$, $1 > B' > 0$ and $B'' < 0$. Accordingly $B(S_t^j) < S_t^j$ for all $S_t^j > 0$. If the sum of necessary costs and slack is smaller than the price the firm is allowed to charge it additionally realizes a profit. The marginal benefit from an additional unit of profit is constant and normalized to 1. Increasing profit and decreasing slack are two sides of the same medal as they add up to a constant: the price a firm charges less the necessary costs. So it is sufficient to explicitly consider just one of the two as the other one emerges as the residual. The instant payoff function of firm j is in every period given by:

$$F_t^j = P_t^j - C - S_t^j + B(S_t^j) \tag{1}$$

Firms care about profit and slack only. They discount next period's utility with δ, $0 < \delta < 1$, and maximize their intertemporal utility:

$$\sum_{t=0}^{\infty} \delta^t F_t^j \tag{2}$$

At all times firms need to break even, so that $C + S_t^j \leq P_t^j$. Not only from a perspective of the regulator or society slack is 'expensive': one marginal unit of additional profit always results in higher instantaneous utility for the firm than an additional marginal unit of slack would. The only reason why $S_t^j > 0$ could be an optimal choice of j is that it can affect the price j is allowed to charge in later periods. We consider an infinite number of periods (immortality illusion of a big firm) in order to avoid unrealistic effects of last rounds in which all slack is zero.[2] Every period there is only one choice per firm to be taken, the slack the firm chooses. The regulatory rule and break even condition are common knowledge, so are the prices of the current period. Using this knowledge firms can anticipate how their choice of slack will affect future behavior of the other firms. Accordingly all

[1] In Shleifer's (1985) one-period model accounting completely and correctly for heterogeneity leads to the efficient equilibrium.

[2] It is easy to show that a finite horizon and the corresponding backward solution will result in zero slack starting in the very first period.

three current prices are state variables for all j. Only Markov-perfect strategies are considered so firms react to the state variables they observe and do not care about the history of states. We outrule punishment between firms and arguments based on Folk Theorems, which can be seen as a form of collusive behavior, as Yardstick Regulation obviously is highly vulnerable to collusion. This paper does not strive to offer solutions for this issue but proceeds to show that even if all collusive behavior can be avoided, uncoordinated individual utility maximization by firms can result in equilibria with positive slack. Strategies are anonymous, so firm i reacts to a change in behavior of o with constant behavior of x just as it would to change vice versa. Simple renaming o into x and x into o does not change the behavior of i. Firms simultaneously choose their slack every period without observing the current choice of the others.

2.2 Regulatory rules

The price a firm is allowed to charge is derived from costs realized by the other two firms in the period before. We separately look at two regulatory schemes: Average Yardstick Regulation, under which average costs of the other firms are used as the yardstick, and Frontier Yardstick Regulation or Best Practice Regulation, under which only the costs of the best performing firm, i.e. the firm with the lowest costs, are the yardstick. So the price e.g. firm i is allowed to charge in period $t+1$ accordingly is a function of the slack o and x are choosing in t in both cases:

$$P^i_{t+1} = R^i(S^o_t, S^x_t) \tag{3}$$

Under Frontier Yardstick Regulation the price is given by

$$P^i_{t+1} = \min(C + S^o_t, C + S^x_t) = C + \min(S^o_t, S^x_t) \tag{4}$$

and under Average Yardstick Regulation by

$$P^i_{t+1} = \frac{1}{2}\sum_{j\neq i}\left(C + S^j_t\right) = C + \frac{1}{2}\sum_{j\neq i} S^j_t \tag{5}$$

Regulatory rules for the other firms and periods are defined analogously. Since necessary cost are constant, C can be factored out under both regulatory regimes and can be normalized to zero. This is equivalent to interpreting P^i_{t+1} as the amount by which the price i may charge in $t+1$ is greater than necessary costs C. In the first period of Yardstick Regulation prices are exogenously given - they could be derived from some regulatory rule that was in place before Yardstick Regulation was implemented.

Lemma 1. *Under both regulatory rules slacks and prices converge to a steady state in which all firms choose the same slack, do so in every period and realize zero profits due to regulatory mechanics. This slack may be zero.*

Proof. see appendix

As long as not all firms choose the same slack and this slack is equal to the price they are allowed to charge (C is normalized to zero) the highest slack chosen in t cannot be chosen by any firm anymore in $t+2$ the latest. Accordingly there is a downward drift of the highest slack, whenever firms choose differing slacks. As slack cannot become negative convergence is assured.

3 Equilibrium Analysis

The equilibrium concept applied in this paper is Markov-perfect in which all the firms simultaneously decide on their slack. From Lemma 1 we know that convergence to a steady state in which all firms are allowed to charge the same price, choose the same slack and make zero profits follows from regulatory mechanics. If a steady state with slack greater than zero exists, this slack must be consistent with the firms' equilibrium strategies. As we are interested in the long run effects of Yardstick Regulation using historical cost data, we focus on the analysis of the resulting steady state equilibria. From Lemma 1 we also know that asymmetric Markov-perfect steady state Nash equilibria cannot exist, so we can furthermore focus on $P_t^i = P_t^o = P_t^x \geq 0$.
It is easy to show that equilibria with very high slack could exist, given initial prices are sufficiently high, if firms punish other firms' uncooperative behavior. From the proof of Lemma 1 directly follows that under Frontier Yardstick Regulation every firm can force all firms into a steady state equilibrium with zero slack by choosing zero slack once - the worst possible steady state from the perspective of all firms. If firms know that another firm will 'burn bridges' by choosing zero slack in case of uncooperative behavior by any firm this can 'discipline' firms to cooperate. As will be shown, punishing unilaterally by a decrease of slack is not consistent with maximization of intertemporal utility by a firm. If and only if other firms punish in this sense it can be optimal for a firm to do so, too. In the spirit of the Folk Theorem[3] a threat of joint mutual punishment, i.e. choosing a very bad steady state for all firms, could be used to implement equilibria with very high slack. The analysis of

[3]See e.g. Osborne and Rubinstein (1994) for a description of Trigger Strategies and Folk Theorems.

corresponding equilibria does not offer much additional insight - as described in the introduction, Yardstick Regulation is *known* to be highly vulnerable to collusion. Joint mutual punishment, that no firm would do unilaterally, can be seen as a form of collusion. Therefore we explicitly exclude all sorts of joint mutual punishment and collusion in our analysis and show that *still* steady state equilibria with positive slack can exist that are 'punishment-free'.
In the following we implicitly define an optimal value of slack each, denoted by S^*, that maximizes intertemporal utility given current prices under both regulatory regimes that can characterize a steady state equilibrium. We furthermore derive a unique level of slack S^{A*} under Average Yardstick Regulation and S^{M*} under Frontier Yardstick Regulation that offers the highest intertemporal utility for the firms and that can exist in a steady state equilibrium, given that prices are sufficiently high. As will be shown, S^{A*} is implicitly defined by

$$B' = 1 - \frac{\frac{1}{2}\delta^2}{1 - \frac{1}{2}\delta} \tag{6}$$

and S^{M*} by

$$B' = 1 - \delta^2 \tag{7}$$

Both equations, (6) and (7), summarize the respective tradeoff between the marginal benefit of reducing slack in the current period and the corresponding marginal costs from adversely affecting future payoff each firm faces every period under both regulatory schemes. We show that that every $S^* \epsilon [0, S^{A*}]$ respectively $S^* \epsilon [0, S^{M*}]$ can occur in a Markov-perfect steady state equilibrium, provided the initial prices are high enough. Conversely, no other slack is possible in a (punishment-free) steady state equilibrium.

3.1 Optimal slack

Assume there exist a steady state equilibrium consistent with the strategies of firms i, o and x, denoted by $f^i(\mathbf{P_t})$, $f^o(\mathbf{P_t})$ and $f^x(\mathbf{P_t})$, where $\mathbf{P_t}$ is the vector of the three prices valid for firm i, o and x in t. Strategies need to be optimal in equilibrium by definition. When firms decide on their slack they consider their discounted utility in all periods to come given they decide optimally in all future periods given future states. We use the Principle of Optimality[4] to find the resulting optimal level of

[4]See e.g. Acemoglu (2009) or Stokey, Lucas with Prescott (1989) for a detailed description.

slack for firm i. So firm i solves the following maximization problem:

$$J^i(\mathbf{P_t}) = J^i(P_t^i, P_t^o, P_t^x) = \max_{S_t^i \leq P_t^i} [F(P_t^i, S_t^i) + \delta J^i(\mathbf{P_{t+1}})] \tag{8}$$

Where J^i denotes the value function of firm i and $\mathbf{P_{t+1}}$ is the vector of prices in $t+1$. While the state in t is given the state in $t+1$ is determined by the regulatory rule. Plugging the general form of the rule in leads to:

$$J^i(\mathbf{P_t}) = \max_{S_t^i \leq P_t^i} [F(P_t^i, S_t^i) + \delta J^i\Big(R^i(S_t^o, S_t^x), R^o(S_t^i, S_t^x), R^x(S_t^i, S_t^o)\Big)] \tag{9}$$

Just as i firms o and x maximize their intertemporal utility given the state variables they observe. So

$$S_t^o = f^o(\mathbf{P_t}) \tag{10}$$

and

$$S_t^x = f^x(\mathbf{P_t}) \tag{11}$$

describe the optimal slack of o and x given $\mathbf{P_t}$.
With the general forms of the strategies (10) and (11) into (9) we obtain

$$\begin{aligned} J^i(\mathbf{P_t}) = \max_{S_t^i \leq P_t^i} \Big[& F(P_t^i, S_t^i) \\ & + \delta J^i\Big(R^i(f^o(\mathbf{P_t}), f^x(\mathbf{P_t})), R^o(S_t^i, f^x(\mathbf{P_t})), R^x(S_t^i, f^o(\mathbf{P_t}))\Big)\Big] \end{aligned} \tag{12}$$

As this is a constrained maximization problem we rewrite (12) as:

$$\begin{aligned} J^i(\mathbf{P_t}) = \max_{S_t^i} \Big[& F(P_t^i, S_t^i) \\ & + \delta J^i\Big(R^i(f^o(\mathbf{P_t}), f^x(\mathbf{P_t})), R^o(S_t^i, f^x(\mathbf{P_t})), R^x(S_t^i, f^o(\mathbf{P_t}))\Big) \\ & + \lambda_t^i(P_t^i - S_t^i)\Big] \end{aligned} \tag{13}$$

With the complementary slackness conditions:

$$\lambda_t^i \geq 0 \text{ and } \lambda_t^i(P_t^i - S_t^i) = 0 \tag{14}$$

The corresponding first order condition (FOC) for the maximum problem is given by:

$$F_2(P_t^i, S_t^i) + \delta J_2^i(\mathbf{P_{t+1}}) \cdot R_1^o(S_t^i, S_t^x) + \delta J_3^i(\mathbf{P_{t+1}}) \cdot R_1^x(S_t^i, S_t^o) - \lambda_t^i = 0 \tag{15}$$

Accordingly numbers as the lower index mark derivatives and the respective number names the argument with respect to which the derivative is taken, while the upper index describes the function of which the derivative is taken. If the lower index includes a 't' it is a time index. So $R_1^o(S_t^i, S_t^x)$ describes how the price o may charge in $t+1$ reacts to a marginal change of the slack of i in t. We only need to look at derivatives to the left, i.e. reductions of slack, as starting from a steady state no firm can increase its slack without violating the break even constraint. Accordingly throughout this paper all derivatives are to be understood as left hand side derivatives, i.e. reductions of the respective variable.
Now let

$$S_t^i = f^i(\mathbf{P_t}) \tag{16}$$

describe the optimal slack of firm i given $\mathbf{P_t}$, i.e. $f^i(\mathbf{P_t})$ is the solution to (15). Inserting this into (13) leads to:

$$\begin{aligned} J^i(\mathbf{P_t}) = {} & F(P_t^i, f^i(\mathbf{P_t})) \\ & + \delta J^i\Big(R^i(f^o(\mathbf{P_t}), f^x(\mathbf{P_t})), R^o(f^i(\mathbf{P_t}), f^x(\mathbf{P_t})), R^x(f^i(\mathbf{P_t}), f^o(\mathbf{P_t}))\Big) \\ & + \lambda_t^i(P_t^i - f^i(\mathbf{P_t})) \end{aligned} \tag{17}$$

taking the derivative to the left with respect P_t^i we find with Envelope Theorem:

$$\begin{aligned} J_1^i(\mathbf{P_t}) = {} & F_1(P_t^i, S_t^i) \\ & + \delta J_1^i(\mathbf{P_{t+1}}) \cdot R_1^i(S_t^o, S_t^x) \cdot f_1^o(\mathbf{P_t}) + \delta J_1^i(\mathbf{P_{t+1}}) \cdot R_2^i(S_t^o, S_t^x) \cdot f_1^x(\mathbf{P_t}) \\ & + \delta J_2^i(\mathbf{P_{t+1}}) \cdot R_2^o(S_t^i, S_t^x) \cdot f_1^x(\mathbf{P_t}) + \delta J_3^i(\mathbf{P_{t+1}}) \cdot R_2^x(S_t^i, S_t^o) \cdot f_1^o(\mathbf{P_t}) \\ & + \lambda_t^i \end{aligned} \tag{18}$$

Analogously we find

$$\begin{aligned} J_2^i(\mathbf{P_t}) = {} & \delta J_1^i(\mathbf{P_{t+1}}) \cdot R_1^i(S_t^o, S_t^x) \cdot f_2^o(\mathbf{P_t}) + \delta J_1^i(\mathbf{P_{t+1}}) \cdot R_2^i(S_t^o, S_t^x) \cdot f_2^x(\mathbf{P_t}) \\ & + \delta J_2^i(\mathbf{P_{t+1}}) \cdot R_2^o(S_t^i, S_t^x) \cdot f_2^x(\mathbf{P_t}) + \delta J_3^i(\mathbf{P_{t+1}}) \cdot R_2^x(S_t^i, S_t^o) \cdot f_2^o(\mathbf{P_t}) \end{aligned} \tag{19}$$

and

$$\begin{aligned} J_3^i(\mathbf{P_t}) = {} & \delta J_1^i(\mathbf{P_{t+1}}) \cdot R_1^i(S_t^o, S_t^x) \cdot f_3^o(\mathbf{P_t}) + \delta J_1^i(\mathbf{P_{t+1}}) \cdot R_2^i(S_t^o, S_t^x) \cdot f_3^x(\mathbf{P_t}) \\ & + \delta J_2^i(\mathbf{P_{t+1}}) \cdot R_2^o(S_t^i, S_t^x) \cdot f_3^x(\mathbf{P_t}) + \delta J_3^i(\mathbf{P_{t+1}}) \cdot R_2^x(S_t^i, S_t^o) \cdot f_3^o(\mathbf{P_t}) \end{aligned} \tag{20}$$

Updating (19) and (20) by one period yields

$$\begin{aligned} J_2^i(\mathbf{P_{t+1}}) = \quad & \delta J_1^i(\mathbf{P_{t+2}}) \cdot R_1^i(S_{t+1}^o, S_{t+1}^x) \cdot f_2^o(\mathbf{P_{t+1}}) \\ & + \delta J_1^i(\mathbf{P_{t+2}}) \cdot R_2^i(S_{t+1}^o, S_{t+1}^x) \cdot f_2^x(\mathbf{P_{t+1}}) \\ & + \delta J_2^i(\mathbf{P_{t+2}}) \cdot R_2^o(S_{t+1}^i, S_{t+1}^x) \cdot f_2^x(\mathbf{P_{t+1}}) \\ & + \delta J_3^i(\mathbf{P_{t+2}}) \cdot R_2^x(S_{t+1}^i, S_{t+1}^o) \cdot f_2^o(\mathbf{P_{t+1}}) \end{aligned} \tag{21}$$

and

$$\begin{aligned} J_3^i(\mathbf{P_{t+1}}) = \quad & \delta J_1^i(\mathbf{P_{t+2}}) \cdot R_1^i(S_{t+1}^o, S_{t+1}^x) \cdot f_3^o(\mathbf{P_{t+1}}) \\ & + \delta J_1^i(\mathbf{P_{t+2}}) \cdot R_2^i(S_{t+1}^o, S_{t+1}^x) \cdot f_3^x(\mathbf{P_{t+1}}) \\ & + \delta J_2^i(\mathbf{P_{t+2}}) \cdot R_2^o(S_{t+1}^i, S_{t+1}^x) \cdot f_3^x(\mathbf{P_{t+1}}) \\ & + \delta J_3^i(\mathbf{P_{t+2}}) \cdot R_2^x(S_{t+1}^i, S_{t+1}^o) \cdot f_3^o(\mathbf{P_{t+1}}) \end{aligned} \tag{22}$$

Plugging (21) and (22) into the FOC (15) leads to

$$\begin{aligned} 0 = \; & F_2(P_t^i, S_t^i) \\ & + \delta R_1^o(S_t^i, S_t^x) \cdot \Big(\delta J_1^i(\mathbf{P_{t+2}}) \cdot R_1^i(S_{t+1}^o, S_{t+1}^x) \cdot f_2^o(\mathbf{P_{t+1}}) \\ & \qquad + \delta J_1^i(\mathbf{P_{t+2}}) \cdot R_2^i(S_{t+1}^o, S_{t+1}^x) \cdot f_2^x(\mathbf{P_{t+1}}) \\ & \qquad + \delta J_2^i(\mathbf{P_{t+2}}) \cdot R_2^o(S_{t+1}^i, S_{t+1}^x) \cdot f_2^x(\mathbf{P_{t+1}}) \\ & \qquad + \delta J_3^i(\mathbf{P_{t+2}}) \cdot R_2^x(S_{t+1}^i, S_{t+1}^o) \cdot f_2^o(\mathbf{P_{t+1}}) \Big) \\ & + \delta R_1^x(S_t^i, S_t^o) \cdot \Big(\delta J_1^i(\mathbf{P_{t+2}}) \cdot R_1^i(S_{t+1}^o, S_{t+1}^x) \cdot f_3^o(\mathbf{P_{t+1}}) \\ & \qquad + \delta J_1^i(\mathbf{P_{t+2}}) \cdot R_2^i(S_{t+1}^o, S_{t+1}^x) \cdot f_3^x(\mathbf{P_{t+1}}) \\ & \qquad + \delta J_2^i(\mathbf{P_{t+2}}) \cdot R_2^o(S_{t+1}^i, S_{t+1}^x) \cdot f_3^x(\mathbf{P_{t+1}}) \\ & \qquad + \delta J_3^i(\mathbf{P_{t+2}}) \cdot R_2^x(S_{t+1}^i, S_{t+1}^o) \cdot f_3^o(\mathbf{P_{t+1}}) \Big) \\ & - \lambda_t^i \end{aligned} \tag{23}$$

In equation (23) we clearly see the consequence of the use of historical cost data under Yardstick Regulation we pointed out in the introduction: The price firm i can charge in the future is influenced by its behavior today. The choice of slack of i in t does not only define its instantaneous payoff, implicitly represented by $F_2(P_t^i, S_t^i)$, but also affects the prices o and x may and will charge in $t+1$ via the regulatory rule, $R^o(S_t^i, S_t^x)$ respectively $R^x(S_t^i, S_t^o)$. Firms o and x choose their slack in $t+1$ based on the state they observe and under the restriction that they have to break even according to their strategies, $f^o(\mathbf{P_{t+1}})$ and $f^x(\mathbf{P_{t+1}})$. Via the regulatory rule

the slacks o and x choose in $t+1$ then affect P^o_{t+2} and P^x_{t+2} and *determine* the price i is allowed to charge in $t+2$, P^i_{t+2}. These three prices are the arguments of the value function of i, and in period t firm i discounts the effects in $t+2$ with δ^2.

From Lemma 1 we know that in every steady state all firms choose the same slack. Thus starting from a steady state unilateral reduction of the slack of i affects the price o and x may charge in the following period just the same way so that $R^o_1(S^i_t, S^x_t) = R^x_1(S^i_t, S^o_t)$. (We extensively deal with the derivatives of the regulatory rule in the appendix.) This reduces (23) to

$$\begin{aligned} 0 = {} & F_2(P^i_t, S^i_t) \\ & + \delta R^o_1(S^i_t, S^x_t) \cdot \Big(\delta J^i_1(\mathbf{P_{t+2}}) \cdot R^i_1(S^o_{t+1}, S^x_{t+1}) \cdot [f^o_2(\mathbf{P_{t+1}}) + f^o_3(\mathbf{P_{t+1}})] \\ & \qquad + \delta J^i_1(\mathbf{P_{t+2}}) \cdot R^i_2(S^o_{t+1}, S^x_{t+1}) \cdot [f^x_2(\mathbf{P_{t+1}}) + f^x_3(\mathbf{P_{t+1}})] \\ & \qquad + \delta J^i_2(\mathbf{P_{t+2}}) \cdot R^o_2(S^i_{t+1}, S^x_{t+1}) \cdot [f^x_2(\mathbf{P_{t+1}}) + f^x_3(\mathbf{P_{t+1}})] \\ & \qquad + \delta J^i_3(\mathbf{P_{t+2}}) \cdot R^x_2(S^i_{t+1}, S^o_{t+1}) \cdot [f^o_2(\mathbf{P_{t+1}}) + f^o_3(\mathbf{P_{t+1}})] \Big) \\ & - \lambda^i_t \end{aligned} \tag{24}$$

From Lemma 1 furthermore follows that, due to regulatory mechanics, in all steady states all firms realize zero profits, i.e. all firms choose the slack that is equal to the maximum price each firm may charge. So starting from a steady state a marginal unilateral reduction of the slack of i in t leads to $P^o_{t+1} = P^x_{t+1} < P^i_{t+1}$. Under both regulatory schemes the two other firms, o and x, will reduce their slack the next period by exactly the resulting marginal reduction of their respective price, given the price they face is not higher than the unique optimal slack S^{A*} respectively S^{M*}. We formalize this in the following Lemma considering reductions of slack only for both regulatory regimes:

Lemma 2.
Frontier Yardstick Regulation:
If $P^o_{t+1} = P^x_{t+1} \leq P^i_{t+1}$ *and* $P^o_{t+1} = P^x_{t+1} \leq S^{M*}$,
then $f^o_2(\mathbf{P_{t+1}}) + f^o_3(\mathbf{P_{t+1}}) = f^x_2(\mathbf{P_{t+1}}) + f^x_3(\mathbf{P_{t+1}}) = 1$
Average Yardstick Regulation:
If $P^o_{t+1} = P^x_{t+1} \leq P^i_{t+1}$ *and* $P^o_{t+1} = P^x_{t+1} \leq S^{A*}$,
then $f^o_2(\mathbf{P_{t+1}}) + f^o_3(\mathbf{P_{t+1}}) = f^x_2(\mathbf{P_{t+1}}) + f^x_3(\mathbf{P_{t+1}}) = 1$

Proof. see appendix

Intuitively Lemma 2 means the following: Starting from a steady state a firm has

to reduce its slack if the price this firm can charge is reduced, as it needs to break even. Given that the firm would not voluntarily unilaterally deviate from the steady state equilibrium it cannot increase its intertemporal payoff by deviating even more than necessary. That another firm also has to reduce its slack by the same amount does not change the situation here.

With Lemma 2 equation (24) reduces to

$$\begin{aligned} 0 = {} & F_2(P_t^i, S_t^i) \\ & + \delta R_1^o(S_t^i, S_t^x) \cdot \Big(\delta J_1^i(\mathbf{P_{t+2}}) \cdot [R_1^i(S_{t+1}^o, S_{t+1}^x) + R_2^i(S_{t+1}^o, S_{t+1}^x)] \\ & \qquad + \delta J_2^i(\mathbf{P_{t+2}}) \cdot R_2^o(S_{t+1}^i, S_{t+1}^x) \\ & \qquad + \delta J_3^i(\mathbf{P_{t+2}}) \cdot R_2^x(S_{t+1}^i, S_{t+1}^o) \Big) \\ & - \lambda_t^i \end{aligned} \tag{25}$$

In order to show how the solutions to this equation differ under both regulatory schemes we need to look at them separately.

Frontier Yardstick Regulation. From Lemma 1 it followed that in all steady state equilibria firms choose the same slack and the slack is equal to each firms' price due to regulatory mechanics. So in such a steady state i will choose the same slack every period, i.e. $S_{t+1}^i = S_t^i = S^*$. Every period i could deviate by reducing its slack (no firm can increase its slack in a steady state because of the break even constraint). So S^* must solve the FOC in every period. Now assume i marginally reduces its slack in t. From the FOC directly follows that it cannot be optimal for i to choose a higher slack in $t+1$ than in t. With $S_t^i < S_{t+1}^i$ the slacks of o and x would have to be smaller than the one i chooses in $t+1$ from the regulatory rule and the break even constraint. Accordingly in $t+1$ the left hand side derivatives of the regulatory rule with respect to the slack of i drop to zero if $S_t^i < S_{t+1}^i$. It follows that $S_t^i < S_{t+1}^i$ cannot describe an optimal strategy of i: the FOC could not hold in $t+1$ as $F_2(P_t^i, S_t^i) = B' - 1$ is smaller than zero and λ_{t+1}^i is nonnegative from the complementary slackness conditions. Accordingly we assume i marginally reduces its slack in periods t and $t+1$, *too*, so that $S_t^i = S_{t+1}^i < S_t^o = S_t^x$. From the regulatory rule (4) the prices o and x may charge in $t+1$ decrease to $P_{t+1}^o = P_{t+1}^x = S_t^i$ and given $S_t^i = S_{t+1}^i$ there is no additional effect on $P_{t+2}^o = P_{t+2}^x$ from this change in the behavior of o and x: The prices o and x may charge in $t+2$ are given by $P_{t+2}^o = \min(S_{t+1}^i, S_{t+1}^x)$ and $P_{t+2}^x = \min(S_{t+1}^i, S_{t+1}^o)$. So if o and x decrease their slack in $t+1$ to $S_t^i = S_{t+1}^i$ they do neither change P_{t+2}^o nor P_{t+2}^x so that in this situation the left hand side derivatives of the regulatory

rule are given by $R_2^o(S_{t+1}^i, S_{t+1}^x) = R_2^x(S_{t+1}^i, S_{t+1}^o) = 0$.[5] Intuitively i decides about its slack in t, knowing that its slack in $t+1$ will be the same. So deciding about slack in t and $t+1$ firm i knows that P_{t+2}^o respectively P_{t+2}^x are equal to S_{t+1}^i for *all* $S_{t+1}^x \geq S_{t+1}^i$ respectively $S_{t+1}^o \geq S_{t+1}^i$. Accordingly the only price in $t+2$ that is changed as a consequence of the induced reduction of the slack of o and x to $S_{t+1}^o = S_{t+1}^x = S_{t+1}^i = S_t^i$ is the price firm i itself can charge in $t+2$, P_{t+2}^i. As in addition $R_1^i(S_{t+1}^o, S_{t+1}^x) + R_2^i(S_{t+1}^o, S_{t+1}^x) = 1$ is always true under Frontier Yardstick Regulation (see appendix) equation (25) reduces to:

$$0 = F_2(P_t^i, S_t^i) + \delta R_1^o(S_t^i, S_t^x) \cdot \delta J_1^i(\mathbf{P_{t+2}}) - \lambda_t^i \tag{26}$$

We consider unilateral reductions of the slack of i starting from a steady state so $R_1^o(S_t^i, S_t^x) = 1$. Furthermore with $J_1^i(\mathbf{P_{t+2}}) = 1 + \lambda_{t+2}^i$ (Lemma 4 in the appendix) and $F_2(P_t^i, S_t^i) = B' - 1$, it follows:

$$0 = B' - 1 + \delta^2(1 + \lambda_{t+2}^i) - \lambda_t^i \tag{27}$$

as the optimization problem is the same in every period in a steady state equilibrium $\lambda_t^i = \lambda_{t+2}^i = \lambda$ solving for B' yields the implicit solution for S^*.

$$B' = 1 - \delta^2 + (1 - \delta^2)\lambda \tag{28}$$

This condition summarizes the tradeoff between marginal benefits and marginal costs of decreasing slack. The less patient firm i is, so the more weight it puts on instantaneous payoff, i.e. the smaller δ is, the greater is B' and with $B'' < 0$ the smaller is the slack i chooses. So if δ decreases the firm cares less about slack in the future but grasps profit today. A more detailed intuition based on an infinite geometric series if given in the appendix. If $\lambda > 0$ the constraint must be binding from the complementary slackness condition. So B' must then be greater and with $B'' < 0$ slack must be smaller. So if the constraint is binding firm i has to choose a smaller slack than it would otherwise do. If λ is zero, the solution to the constrained maximization problem is equal to the solution to the unconstrained maximization problem, i.e. the slack S^{M*} firm i chooses in equilibrium if all prices are sufficiently high. So the implicit definition for S^{M*} is given by:

$$B' = 1 - \delta^2 \tag{29}$$

Average Yardstick Regulation. Under this regulatory rule *all* derivatives of the regulatory rule are *always* $\frac{1}{2}$ as each price is the average of two slacks (see appendix).

[5] Derivatives would be greater than zero for *further* decreases of their slack though.

Using this and that in all steady state equilibria the FOC must hold in every period, we can update the FOC (15) by one period and plug it into (25) to find

$$0 = F_2(P_t^i, S_t^i) + \delta\frac{1}{2} \cdot \left(\delta J_1^i(\mathbf{P_{t+2}}) - F_2(P_{t+1}^i, S_{t+1}^i) + \lambda_{t+1}^i\right) - \lambda_t^i \tag{30}$$

Applying the same reasoning as above with $S_{t+1}^i = S_t^i = S^*$, $J_1^i(\mathbf{P_{t+2}}) = 1 + \lambda_{t+2}^i$, $F_2(P_t^i, S_t^i) = B' - 1$ and $\lambda = \lambda_t = \lambda_{t+1}^i = \lambda_{t+2}^i$ we find:

$$0 = B'(1 - \frac{1}{2} \cdot \delta) + 1(\frac{1}{2}\delta^2 + \frac{1}{2}\delta - 1) + \lambda(\frac{1}{2}\delta + \frac{1}{2}\delta^2 - 1) \tag{31}$$

and solving for B' it follows the implicit solution for S^*:

$$B' = \frac{(1 - \frac{1}{2}\delta^2 - \frac{1}{2}\delta) + \lambda(1 - \frac{1}{2}\delta^2 - \frac{1}{2}\delta)}{1 - \frac{1}{2}\delta} \tag{32}$$

Under Average Yardstick Regulation B' also decreases in δ, so the slack i chooses increases in the weight the firm puts on future payoff. Again B' increases in λ so the slack chosen if the constraint is binding is smaller than the slack chosen if all prices are sufficiently high. The solution to the corresponding unconstrained maximization problem, i.e. the slack S^{A*} firm i chooses in equilibrium if all prices are sufficiently high, does not include λ. So S^{A*} is implicitly defined by:

$$B' = 1 - \frac{\frac{1}{2}\delta^2}{1 - \frac{1}{2}\delta} \tag{33}$$

3.2 Steady state equilibria

From Lemma 1 it followed that there cannot exist any asymmetric steady state equilibrium. It is straightforward that the above analysis can analogously be done for firms o and x. Taking the optimal strategies of firms o and x as given we show that it is optimal for i to choose the same strategy. Doing this, we proof existence of the equilibria characterized above.
Assume optimal strategies of firms o and x under Frontier Yardstick Regulation are given by

$$f^o(\mathbf{P_t}) = f^x(\mathbf{P_t}) = \min(S^{M*}, P_t^i, P_t^o, P_t^x) \tag{34}$$

So o and x choose a slack equal to S^{M*} if all prices of the current period are at least S^{M*}. If this is not the case they choose a slack equal the lowest price. Obviously,

it cannot be optimal then to choose any slack greater than S^{M*} for firm i, as it could reduce its slack to S^{M*} without affecting any price in $t+1$. As $F_2(P_t^i, S_t^i) < 0$ this would always result in higher instantaneous and intertemporal payoff. The same is true for any $S_t^i > S_t^o = S_t^x$, as i's slack does not affect future prices if $S_t^i > S_t^o = S_t^x$ from the regulatory rule. Accordingly the FOC cannot hold with $S_t^i > S_t^o = S_t^x$ as $F_2(P_t^i, S_t^i) < 0$ and λ_t^i is nonnegative from the complementary slackness conditions. Thus it is never optimal for i to choose a slack higher than o and x under Frontier Yardstick Regulation and the optimal strategy of i given $\mathbf{P_t}$ must be $f^i(\mathbf{P_t}) \leq \min(S^{M*}, P_t^i, P_t^o, P_t^x)$.
By Theorem 6.4 of Acemoglu (2009) together with the relaxed assumption 6.3 that the instant payoff function is (weakly) concave (Acemoglu (2009), p. 189) it follows that the value function is concave in the state variables, as Assumptions 6.1, 6.2 and and the relaxed version of Assumption 6.3 from Acemoglu (2009) hold in our setup. As furthermore F is strictly concave in slack and the left hand side derivative of the regulatory rule with respect to the slack of i must be equal to one in all equilibria with $S^* > 0$, $\lambda > 0$ in all equilibria with $S^* < S^{M*}$. Accordingly the steady state described by S^{M*} is strictly preferred by firm i over all other steady states with lower slack. (Obviously all steady state equilibria with positive slack are preferred by i over the zero slack equilibrium.) From the concavity of the value function and the strict concavity of F concerning slack it also follows that λ decreases in the steady state value of slack for all $S^* < S^{M*}$. As a consequence firm i never unilaterally deviates by reducing slack from a situation where all firms choose the same slack, given $S_t^i \leq S^{M*}$: If firm i unilaterally reduces its slack starting from such a situation in t, the constraint is not binding that period, so λ_t^i needs to be zero from the complementary slackness conditions. With the concavity of the value function and strict concavity of F with respect to slack this cannot be optimal, as the FOC could not hold. Then $f^i(\mathbf{P_t}) = \min(S^{M*}, P_t^i, P_t^o, P_t^x)$ is the optimal strategy given the strategies of o and x and every slack $S^* \epsilon [0, S^{M*}]$ can describe a steady state equilibrium under Frontier Yardstick Regulation. Assuming firms o and x would not choose, 'jump to', S^{M*} in the first period given prices are sufficiently high, but would simply choose the slack equal to the lowest price of all in t, we can furthermore show that there cannot exist any other steady state equilibrium. As λ is nonnegative from the complementary slackness conditions, the derivatives of the regulatory rule cannot be greater than 1, and the value function is concave in the prices, while F is strictly concave in slack, the FOC could not hold in any steady state with slack greater than S^{M*}. So firm i would unilaterally deviate by reducing its slack, which contradicts the existence of (punishment-free) Markov-perfect steady state equilibria with slack higher S^{M*}. In every steady state with slack S^{M*} marginal benefits of unilaterally reducing slack would be greater than marginal costs of doing so.

Assume further that optimal strategies of firms o and x under Average Yardstick Regulation are given by

$$f^o(\mathbf{P_t}) = f^x(\mathbf{P_t}) = \min(S^{A*}, P_t^i, P_t^o, P_t^x) \tag{35}$$

Strategies given by equations (34) respectively (35) differ only by the unique optimal value of slack, given prices are sufficiently high. Accordingly the corresponding proof for Average Yardstick Regulation is are very similar to the one above: It is not optimal for firm i to choose a slack higher than the one o and x choose given their above strategies. Under Average Yardstick Regulation all derivatives of the regulatory rule are equal to $\frac{1}{2}$, as every price is the average of the slacks of the other two firms of the period before. Still it is not optimal for i to choose $S_t^i > S_t^o = S_t^x$. Neither o nor x would choose a higher slack in $t+1$ as then $\min(P_{t+1}^i, P_{t+1}^o, P_{t+1}^x) = P_{t+1}^i = S_t^o = S_t^x$. It follows that the highest possible slack from $t+2$ on would not be greater than $S_t^o = S_t^x$ for all slacks $S_t^i > S_t^o = S_t^x$. As $F_2(P_t^i, S_t^i) < 0$, i could increase its instantaneous and intertemporal payoff by decreasing its slack and choosing $S_t^i = S_t^o = S_t^x$. The rest of the proof is a straightforward repetition of the arguments above using S^{A*} and the corresponding derivatives of the regulatory rule.
We summarize these findings in the following proposition:

Proposition 1. *There exists an optimal value of slack, denoted by S^*, that maximizes intertemporal utility given current prices under both regulatory regimes, that can describe a steady state equilibrium. There furthermore exists a unique level of slack S^{M*} under Frontier Yardstick Regulation and S^{A*} under Average Yardstick Regulation that offers the highest intertemporal payoff of all possible (punishment-free) Markov-perfect steady state equilibria for all firms given prices are sufficiently high. Every slack $S^* \epsilon [0, S^{M*}]$ under Frontier Yardstick Regulation and $S^* \epsilon [0, S^{A*}]$ under Average Yardstick Regulation can characterize a steady state equilibrium. There are no other (punishment-free) Markov-perfect steady state equilibria.*

It is important to note, that the regulator cannot induce the zero slack steady state by simply setting all prices to zero. In our analysis necessary costs have been normalized to zero - and the reason why regulatory schemes like Yardstick Regulation exist is that the regulator does not know necessary costs of production. So by setting too low prices in the first regulatory period, the regulator risk firms going bankrupt, as they cannot break even anymore.

3.3 Comparative dynamics

From Proposition 1 we know that every slack between 0 and S^{M*} under Frontier Yardstick Regulation respectively 0 and S^{A*} under Average Yardstick Regulation

can describe a steady state equilibrium and that there cannot exist (punishment-free) Markov-perfect steady state equilibria with higher slack under the respective regulatory regime. By comparing the implicit solutions for S^{M*} and S^{A*} we find that all (punishment-free) Markov-perfect steady state equilibria under Average Yardstick Regulation can be equilibria under Frontier Yardstick Regulation, while the reverse is not true. This leads to the following Proposition:

Proposition 2. *The highest slack that can be realized in a (punishment-free) Markov-perfect steady state equilibrium is greater under Frontier Yardstick Regulation than under Average Yardstick Regulation.*

Proof. S^{A*} is implicitly defined by (33) and the corresponding value under Frontier Yardstick Regulation, S^{M*}, is implicitly defined by (29). As $B' > 0$ and $B'' < 0$, $S^{M*} > S^{A*}$ if the following inequality holds:

$$1 - \frac{\frac{1}{2}\delta^2}{1 - \frac{1}{2}\delta} > 1 - \delta^2 \tag{36}$$

$$\delta^2 > \frac{\frac{1}{2}\delta^2}{1 - \frac{1}{2}\delta} \tag{37}$$

$$1 - \frac{1}{2}\delta > \frac{1}{2} \tag{38}$$

$$1 > \delta \tag{39}$$

(39) is always true, what completes the proof. □

Intuitively, orientation at the performance of 'the best' of all other firms rather than the average of all other firms to define constraints for a firm under Yardstick Regulation seems to be the tougher regulation. Incentives to produce efficiently, i.e. without slack, should be high. Proposition 2 questions this intuition. Using historical cost data of other firms allows each firm to influence the yardstick, and as this influence is greater under Frontier Yardstick Regulation all firms could be less willing to 'push' the other firms, because they will have to 'push back'.

4 Conclusion

While Shleifer's (1985) version of Yardstick Regulation uses *current* performance of other firms to find *current* constraints for an evaluated firm, real life applications of Yardstick Regulation frequently *ex ante* define constraints, e.g. prices allowed to be

charged, based on data from the regulatory period(s) before. This use of historical cost data in Yardstick Regulation enables a firm to affect the price it can charge in the future. By affecting other firms' constraints and thus behavior the current performance of a firm is directly linked to its own future constraints. This analysis showed in a simple model framework that inefficient steady state equilibria in which all firms choose positive slack can exist under Yardstick Regulation without any form of collusion if historical cost data is used. Furthermore the highest slack that can exist in a (punishment-free) Markov-perfect steady state equilibrium is higher under Frontier Yardstick Regulation, where the firm with the lowest costs of all but the evaluated firm defines the yardstick, than if the average of all other firms is used. This challenges the perception that the best of all other firms should be the yardstick in Yardstick Regulation using historical cost data.

Appendix

Proofs

Proof of Lemma 1

As regulatory rules are anonymous only 5 relevant different cases can be distinguished, potentially with indices changed and updated though. (C is normalized to zero):

(I) $P_t^i = P_t^o = P_t^x = S_t^i = S_t^o = S_t^x$

(II) $S_t^i = S_t^o > S_t^x$

(III) $S_t^i > S_t^o = S_t^x$

(IV) $S_t^i > S_t^o > S_t^x$

(V) $S_t^i = S_t^o = S_t^x < P_t^i = P_t^o = P_t^x$

The reasoning is explained in detail for case (II) under Frontier Yardstick Regulation and Average Yardstick Regulation, the remaining is then a straightforward application along these lines.

Frontier Yardstick Regulation
To (I):
If all three prices and all three slacks are the same in t the regulatory rule does not force any change. Prices in $t+1$ are the same and the same slack as in t is possible for all firms.

To (II):

$$P_{t+1}^i = P_{t+1}^o = S_t^x < P_{t+1}^x = S_t^i = S_t^o$$

$$\Rightarrow$$

$$S_{t+1}^i \leq S_t^x$$
$$S_{t+1}^o \leq S_t^x$$
$$S_{t+1}^x \leq S_t^i = S_t^o$$

$$\Rightarrow$$

$$S^i_{t+2} \leq P^i_{t+2} \leq S^x_t$$
$$S^o_{t+2} \leq P^o_{t+2} \leq S^x_t$$
$$S^x_{t+2} \leq P^x_{t+2} \leq S^x_t$$

then either (I) or one of (II)-(V) applies.

Under Frontier Yardstick Regulation the price a firm is allowed to charge is the minimum of the slack the other two firms chose in the period before. So if firms i and o choose the same slack in t and x chooses a smaller one, the price i and o are allowed to charge in $t+1$ is equal to S^x_t, while P^x_{t+1} is equal to the slack i and o choose in t. In $t+1$ x may as the consequence choose any slack that is not greater than $P^x_{t+1} = S^i_t = S^o_t$, while i and o have to choose a slack not greater than $P^i_{t+1} = P^o_{t+1} = S^x_t$ what is smaller than $P^x_{t+1} = S^i_t = S^o_t$. In $t+2$ the price i, o and x may charge is not greater than the smallest slack in t, i.e. S^x_t. Because even if x, the only firm that can choose a higher slack than this in $t+1$, does so, the smaller one of any two slacks in $t+1$ cannot be greater than S^x_t. In $t+2$ either all three firms choose the same slack and this slack is equal to the price they may charge or one of (II) to (V) applies.

To (III):

$$S^i_{t+1} \leq P^i_{t+1} = S^x_t = S^o_t$$
$$S^o_{t+1} \leq P^o_{t+1} = S^x_t = S^o_t$$
$$S^x_{t+1} \leq P^x_{t+1} = S^o_t = S^x_t$$

then either (I) or one of (II)-(V) applies.

To (IV):

$$P^i_{t+1} = S^x_t$$
$$P^o_{t+1} = S^x_t$$
$$P^x_{t+1} = S^o_t$$

$$\Rightarrow$$

$$S^i_{t+1} \leq S^x_t$$
$$S^o_{t+1} \leq S^x_t$$
$$S^x_{t+1} \leq S^o_t$$

$$\Rightarrow$$

$$S^i_{t+2} \leq P^i_{t+2} \leq S^x_t$$
$$S^o_{t+2} \leq P^o_{t+2} \leq S^x_t$$
$$S^x_{t+2} \leq P^x_{t+2} \leq S^x_t$$

then either (I) or one of (II)-(V) applies.

To (V):

$$P^i_{t+1} = P^o_{t+1} = P^x_{t+1} = S^i_t = S^o_t = S^x_t$$

then either (I) or one of (II)-(V) applies.

Average Yardstick Regulation
To (II):

$$P^i_{t+1} = \frac{S^o_t + S^x_t}{2}$$
$$P^o_{t+1} = \frac{S^i_t + S^x_t}{2}$$
$$P^x_{t+1} = \frac{S^i_t + S^o_t}{2} = S^i_t = S^o_t$$

$$\Rightarrow$$

$$S^i_{t+1} \leq \frac{S^o_t + S^x_t}{2} < S^i_t$$
$$S^o_{t+1} \leq \frac{S^i_t + S^x_t}{2} < S^o_t$$
$$S^x_{t+1} \leq \frac{S^i_t + S^o_t}{2} = S^i_t = S^o_t \text{ [and } S^x_{t+1} \gtreqless S^x_t]$$

$$\Rightarrow$$

$$S^i_{t+2} \leq P^i_{t+2} \leq \frac{S^i_t + S^x_t + S^i_t + S^o_t}{4} < S^i_t = S^o_t$$
$$S^o_{t+2} \leq P^o_{t+2} \leq \frac{S^o_t + S^x_t + S^i_t + S^o_t}{4} < S^o_t = S^i_t$$
$$S^x_{t+2} \leq P^x_{t+2} \leq \frac{S^o_t + S^x_t + S^i_t + S^x_t}{4} < S^i_t = S^o_t$$

So the highest slack chosen in t cannot be chosen by anyone in $t+2$. Then either (I) or one of (II)-(V) applies.
Under Average Yardstick Regulation the price a firm may charge is equal to the average of the slacks the other two firms chose in the period before. So if firms i and o choose the same slack in t and x chooses a smaller one, the price i and o are allowed to charge in $t+1$ is smaller than the one x may charge and smaller than the slack i and o choose in t. Accordingly they have to choose a smaller slack in $t+1$. x may in $t+1$ choose a slack that is greater than S_t^x but not greater than the slack i and o choose in t. In $t+2$ all prices are smaller than the greatest slack in t so that this slack cannot be chosen anymore. Then either all three firms choose the same slack and this slack is equal to the price they may charge or one of (II) to (V) applies.

To (III):

$$P_{t+1}^i = \frac{S_t^o + S_t^x}{2} = S_t^o = S_t^x$$
$$P_{t+1}^o = \frac{S_t^i + S_t^x}{2}$$
$$P_{t+1}^x = \frac{S_t^i + S_t^o}{2}$$

$$\Rightarrow$$

$$S_{t+1}^i \leq \frac{S_t^o + S_t^x}{2} = S_t^o = S_t^x < S_t^i$$
$$S_{t+1}^o \leq \frac{S_t^i + S_t^x}{2} < S_t^i \text{ [and } S_{t+1}^o \gtreqless S_t^o]$$
$$S_{t+1}^x \leq \frac{S_t^i + S_t^o}{2} < S_t^i \text{ [and } S_{t+1}^x \gtreqless S_t^x]$$

So the highest slack chosen in t cannot be chosen by anyone in $t+1$. Then either (I) or one of (II)-(V) applies.

To (IV):

$$
\begin{aligned}
P^i_{t+1} &= \frac{S^o_t + S^x_t}{2} \\
P^o_{t+1} &= \frac{S^i_t + S^x_t}{2} \\
P^x_{t+1} &= \frac{S^i_t + S^o_t}{2}
\end{aligned}
$$

$$
\Rightarrow
$$

$$
\begin{aligned}
S^i_{t+1} &\leq \frac{S^o_t + S^x_t}{2} < S^i_t \\
S^o_{t+1} &\leq \frac{S^i_t + S^x_t}{2} < S^i_t \\
S^x_{t+1} &\leq \frac{S^i_t + S^o_t}{2} < S^i_t
\end{aligned}
$$

So the highest slack chosen in t cannot be chosen by anyone in $t+1$. Then either (I) or one of (II)-(V) applies.

To (V):

$$
P^i_{t+1} = \frac{S^o_t + S^x_t}{2} = P^o_{t+1} = \frac{S^i_t + S^x_t}{2} = P^x_{t+1} = \frac{S^i_t + S^o_t}{2} = S^i_t = S^o_t = S^x_t
$$

$$
\Rightarrow
$$

$$
\begin{aligned}
S^i_{t+1} &\leq P^i_{t+1} = S^i_t = S^o_t = S^x_t \\
S^o_{t+1} &\leq P^o_{t+1} = S^o_t = S^i_t = S^x_t \\
S^x_{t+1} &\leq P^x_{t+1} = S^x_t = S^i_t = S^o_t
\end{aligned}
$$

Then either (I) or one of (II)-(V) applies.

As long as slacks differ in period t in $t+2$ the latest the highest slack of t cannot be chosen by any firm anymore under both regulatory schemes. (Under Frontier Yardstick Regulation the latest in $t+2$ no slack higher than the smallest of t can be chosen.) So the maximum of the three slacks monotonically decreases, potentially

with a delay that is not greater than two periods. Furthermore all slacks are bounded below at zero. It follows that slacks necessarily have to converge. As the price for each firm is in every period the minimum respectively the average of the slack of the other two firms in the period before, prices converge, too. Prices and slacks cannot converge to different values, as this would decrease prices the next period, so in every steady state profits of all firms must be zero.

□

Derivatives

Regulatory Rules

Frontier Yardstick Regulation. We again focus on the example of firm i. The corresponding derivatives for the other firms are found analogously. The regulatory rule is given by (4):

$$P^i_{t+1} = C + \min(S^o_t, S^x_t) = R^i(S^o_t, S^x_t)$$

The derivatives of this function are:

$$R^i_1(S^o_t, S^x_t) = \frac{\partial R^i(S^o_t, S^x_t)}{\partial S^o_t} = \begin{cases} 1 & \text{for } S^o_t < S^x_t \\ 0 & \text{for } S^o_t > S^x_t \end{cases} \tag{40}$$

$$R^i_2(S^o_t, S^x_t) = \frac{\partial R^i(S^o_t, S^x_t)}{\partial S^x_t} = \begin{cases} 0 & \text{for } S^o_t < S^x_t \\ 1 & \text{for } S^o_t > S^x_t \end{cases} \tag{41}$$

and

$$R^i_1(S^o_t, S^x_t) + R^i_2(S^o_t, S^x_t) = \frac{\partial R^i(S^o_t, S^x_t)}{\partial S^o_t} + \frac{\partial R^i(S^o_t, S^x_t)}{\partial S^x_t} = 1 \text{ for all } S^o_t, S^x_t \tag{42}$$

To derive (42) for $S^o_t = S^x_t$ and simultaneous changes of slack of o and x let

$$\widetilde{S^o_t} = S^o_t + \epsilon$$

and

$$\widetilde{S^x_t} = S^x_t + \epsilon$$

where $\epsilon \neq 0$.
For $S_t^o = S_t^x$ and $\widetilde{S_t^o} = \widetilde{S_t^x}$ we see that

$$\min(S_t^o, S_t^x) = S_t^o = S_t^x \text{ and } \min(\widetilde{S_t^o}, \widetilde{S_t^x}) = \widetilde{S_t^o} = \widetilde{S_t^x}$$

then

$$\min(\widetilde{S_t^o}, \widetilde{S_t^x}) - \min(S_t^o, S_t^x) = \epsilon$$

In analogy to the definition of the derivative we find

$$\lim_{\epsilon \to 0} \frac{\min(S_t^o + \epsilon, S_t^x + \epsilon) - \min(S_t^o, S_t^x)}{\epsilon} = 1 \tag{43}$$

(Starting from $S_t^o = S_t^x$ even a marginal reduction of S_t^o holding S_t^x constant results in $\min(S_t^o, S_t^x) = S_t^o < S_t^x$ and vice versa.)

Average Yardstick Regulation. The regulatory rule is given by (5):

$$P_{t+1}^i = \frac{1}{2} \sum_{j \neq i} \left(C + S_t^j \right) = C + \frac{1}{2} \sum_{j \neq i} S_t^j = R^i(S_t^o, S_t^x)$$

So we see that all changes in slack of any firm will result in changes in the prices the other two firms may charge in the following period of half the magnitude of the aforementioned change. Spelt out for firm i this is

$$R_1^i(S_t^o, S_t^x) = R_2^i(S_t^o, S_t^x) = \frac{1}{2} \tag{44}$$

All other derivatives of the regulatory rule under Average Yardstick Regulation are equal to $\frac{1}{2}$, too.

Proof of Lemma 2

Recall the FOC, equation (15),

$$F_2(P_t^i, S_t^i) + \delta J_2^i(\mathbf{P_{t+1}}) \cdot R_1^o(S_t^i, S_t^x) + \delta J_3^i(\mathbf{P_{t+1}}) \cdot R_1^x(S_t^i, S_t^o) - \lambda_t^i = 0$$

and complementary slackness conditions (14):

$$\lambda_t^i \geq 0 \text{ and } \lambda_t^i(P_t^i - S_t^i) = 0$$

Assume firms are in a steady state so that $P_t^i = P_t^o = P_t^x = S_t^i = S_t^o = S_t^x$ and $P_t^i \leq S^{A*}$ under Average Yardstick Regulation respectively $P_t^i \leq S^{M*}$ under Frontier

Yardstick Regulation. If i's choice of slack is optimal, the FOC and complementary slackness conditions must hold. Now assume one of the other firms, e.g. firm o, instead chooses a marginally smaller slack in t so that $P^i_{t+1} = P^x_{t+1} < P^o_{t+1}$ and $P^i_{t+1} < S^{A*}$ under Average Yardstick Regulation respectively $P^i_{t+1} < S^{M*}$ under Frontier Yardstick Regulation. From the break even condition we know that i has to reduce its slack by at least that marginal change of the price it may charge in $t+1$ so that the left hand site derivative of $f^i(\mathbf{P_{t+1}})$ with respect to i's own price cannot be smaller than one. Clearly then the sum of the left hand site derivatives $f^i_1(\mathbf{P_{t+1}}) + f^i_3(\mathbf{P_{t+1}}) \geq 1$, too. (As throughout the paper we are only considering *reductions* of slack here.) If it is optimal for i to decrease its slack by even more than that marginal change of his price to any $\underline{S}^i_{t+1} < P^i_{t+1}$, the constraint in $t+1$ is not binding and it follows $\lambda^i_{t+1} = 0$ from the complementary slackness conditions. With $\underline{S}^i_{t+1} < S^i_t$ it furthermore follows $F_2(P^i_{t+1}, \underline{S}^i_{t+1}) > F_2(P^i_t, S^i_t)$ as F is strictly concave in slack. Also the value function is concave in the state variables (see section 3.2) accordingly $J^i_2(\mathbf{P_{t+1}}) \geq J^i_2(\mathbf{P_t})$ and $J^i_3(\mathbf{P_{t+1}}) \geq J^i_3(\mathbf{P_t})$. Looking at both regulatory schemes separately we can show the following:

Under Average Yardstick Regulation all derivatives of the regulatory rule are equal to $\frac{1}{2}$. Then in $t+1$ we have $\lambda^i_{t+1} = 0$ and all other terms of the FOC are not smaller than the respective values in t. It follows that the FOC cannot hold in $t+1$ what contradicts the assumption that $\underline{S}^i_{t+1}$ is the optimal choice of i.

Under Frontier Yardstick Regulation the left hand side derivatives of the regulatory rule relevant for i are equal to 1 if i chooses a slack that is smaller than the slack of the other two firms. So if $\underline{S}^i_{t+1} < S^o_{t+1}$ and $\underline{S}^i_{t+1} < S^x_{t+1}$ the FOC in $t+1$ does not hold, what again shows that $\underline{S}^i_{t+1}$ cannot be the optimal slack. So i does not unilaterally reduce its slack more than necessary.

Here we can clearly observe the vulnerability of Yardstick Regulation against the threat of joint mutual punishment and collusion in general, as discussed at the beginning section 3: If at least one of the other firms chooses a very small slack under Frontier Yardstick Regulation, it is optimal for i to do so, too: From the proof of Lemma 1 we know that the smallest slack of all in t is an upper bound for all prices and slacks from $t+2$ onwards. So if $S^i_{t+1} > \min(S^o_{t+1}, S^x_{t+1})$, it directly follows that i could reduce its slack and by this increase instantaneous payoff without adversely affecting future prices. Carrying this argument to the extreme: if any of the firms chooses zero slack, it is optimal for all other firms to do so, too. At the same time clearly all firms prefer all other possible steady state equilibria over the zero slack equilibrium. Furthermore we just have proven that in this setting it is not optimal for i to unilaterally choose *any* slack smaller than the price it can charge in $t+1$. Accordingly $f^i_1(\mathbf{P_{t+1}}) + f^i_3(\mathbf{P_{t+1}}) > 1$ could result from joint mutual punishment

behavior only and thus cannot be the optimal punishment-free strategy of firm i.[6] We can apply the same reasoning with indices changed, i.e. $P^i_{t+1} = P^o_{t+1} < P^x_{t+1}$, and for firm o and x what completes the proof.

□

Lemma 3.
Frontier Yardstick Regulation:
If $P^i_{t+1} = P^x_{t+1} < P^o_{t+1}$ *and* $P^i_{t+1} \leq S^{M*}$, *then* $f_2'^i(\mathbf{P_{t+1}}) = 0$
Average Yardstick Regulation:
If $P^i_{t+1} = P^x_{t+1} < P^o_{t+1}$ *and* $P^i_{t+1} \leq S^{A*}$, *then* $f_2'^i(\mathbf{P_{t+1}}) = 0$

Proof. Assume $P^i_t = P^x_t < P^o_t$ and $P^i_t \leq S^{M*}$ under Frontier Yardstick Regulation respectively $P^i_t \leq S^{A*}$ under Average Yardstick Regulation. As i optimally decides on its slack the FOC holds and $S^i_t = P^i_t$. If P^o_t decreases to P^o_{t+1}, with $P^i_{t+1} = P^i_t = P^x_{t+1} = P^x_t \leq P^o_{t+1} < P^o_t$ and i reacts to this by choosing any smaller slack $\underline{S}^i_{t+1} < S^i_t$, i.e. $f_2'^i(\mathbf{P_{t+1}}) > 0$, the constraint is not binding, so that $\lambda^i_{t+1} = 0$. Applying the same reasoning as in the proof of Lemma 2 it follows that $\underline{S}^i_{t+1}$ is not an optimal choice, so that $f_2'^i(\mathbf{P_{t+1}}) = 0$. □

An intuition for Lemma 3 under Frontier Yardstick Regulation is the following: Firm i knows that the lowest slack in $t+1$ describes an upper bound for all slacks and prices from $t+3$ onwards. So as long as the slacks of the other two firms are not smaller than the one i chooses, this upper bound is the same for every slack o and x choose. Then the marginal benefits and costs of a reduction of slack do not depend on these slacks, thus the decision of i is not affected. Again the same reasoning applies for $P^i_t = P^o_t < P^x_t$ as well as for firms o and x, too, with indices changed.

[6] It is easy to show that with a threat of i 'to burn bridges' in case of uncooperative behavior of the other firms, what corresponds to $f^i_1(\mathbf{P_{t+1}}) + f^i_3(\mathbf{P_{t+1}}) > 1$ in our setting, equilibria with even *higher* slack could exist.

Lemma 4.
Frontier Yardstick Regulation:
If $P^o_{t+2} = P^x_{t+2} < P^i_{t+2}$ *and* $P^o_{t+2} = P^x_{t+2} \leq S^{M*}$, *then* $J^i_1(\mathbf{P^{t+2}}) = 1 + \lambda^i_{t+2}$
Average Yardstick Regulation:
If $P^o_{t+2} = P^x_{t+2} < P^i_{t+2}$ *and* $P^o_{t+2} = P^x_{t+2} \leq S^{A*}$, *then* $J^i_1(\mathbf{P^{t+2}}) = 1 + \lambda^i_{t+2}$

Proof. Recall equation (18):

$$\begin{aligned} J^i_1(\mathbf{P_t}) = {} & F_1(P^i_t, S^i_t) \\ & + \delta J^i_1(\mathbf{P_{t+1}}) \cdot R^i_1(S^o_t, S^x_t) \cdot f^o_1(\mathbf{P_t}) + \delta J^i_1(\mathbf{P_{t+1}}) \cdot R^i_2(S^o_t, S^x_t) \cdot f^x_1(\mathbf{P_t}) \\ & + \delta J^i_2(\mathbf{P_{t+1}}) \cdot R^o_2(S^i_t, S^x_t) \cdot f^x_1(\mathbf{P_t}) + \delta J^i_3(\mathbf{P_{t+1}}) \cdot R^x_2(S^i_t, S^o_t) \cdot f^o_1(\mathbf{P_t}) \\ & + \lambda^i_t \end{aligned}$$

The way the value function of i is affected by a change of the price i may charge depends on how the other two firms react to this change. Using Lemma 3 for the reactions of o and x, inserting $F_1(P^i_t, S^i_t) = 1$ and updating (18) by two periods completes the proof. □

Intuitively Lemma 4 says that if firms are in the steady state equilibrium described by S^{M*} respectively S^{A*} and $\lambda = 0$, they would not change their slack if their price was higher but would realize a positive profit that period. So the discounted sum of the utility of i increases by 1 if the price i is allowed to charge in t increases by one unit. In any steady state equilibrium with a slack smaller than S^{M*} respectively S^{A*}, we have $\lambda > 0$. So firms would like to move to a steady state equilibrium with higher slack, but cannot because of the (binding) break even constraint.

Intuition for S^{M*} based on geometric series

When firm i decides on the slack in t it considers that its slack defines an upper bound for all prices from $t+2$ onwards under Frontier Yardstick Regulation given $S^i_t \leq \min(S^o_t, S^x_t)$. From the proof of Lemma 2 we know that o and x choose the highest slack they are allowed to, given $P^o_{t+1} = P^x_{t+1} \leq P^i_{t+1}$ and $P^o_{t+1} = P^x_{t+1} \leq S^{M*}$, in their optimal decision. Firm i has to trade off profit in t and $t+1$ against slack in $t, t+1, t+2, ..., \infty$ when it decides about $S^i_t = S^{M*}$. (As the price i may charge in $t+1$ is unaffected by S^i_t it can 'cash in' the profit from reducing slack twice.) In the steady state equilibrium described by S^{M*}, implicitly defined by (29), marginal costs of reducing slack and marginal benefits from doing so must be equal to each

other. So that

$$1 + \delta \cdot 1 = \sum_{z=o}^{\infty} \delta^z \cdot B' \tag{45}$$

With $\delta < 1$ it follows that

$$\begin{aligned} 1 + \delta \cdot 1 &= B' \frac{1}{1-\delta} \\ B' &= (1+\delta)(1-\delta) \\ B' &= 1 - \delta^2 \end{aligned} \tag{46}$$

Which is the same equation as (29).

References

Acemoglu, D., 2009. Introduction to Modern Economic Growth. Princeton University Press, Princeton, NJ.

Averch, H., Johnson, L. R., 1962. Behavior of the Firm under Regulatory Constraint. *American Economic Review* 52(5), 1059-1069.

Blackmon, G., 1994. Incentive Regulation and the Regulation of Incentives. Kluwer, Boston, MA.

Bundesnetzagentur für Elektrizität, Gas, Telekommunkation, Post und Eisenbahnen, 2006. Bericht der Bundesnetzagentur nach §112a EnWG zur Einführung der Anreizregulierung nach §21a EnWG, Bonn.

Dalen, D. M., 1998. Yardstick Competition and Investment Incentives. *Journal of Economics & Mangement Strategy* 7(1), 105-126.

Laffont, J.-J., Tirole, J., 1993. A Theory of Incentives in Procurement and Regulation. The MIT Press, Cambridge, MA.

Littlechild, S. C., 1983. Regulation of British Telecommunications' Profitability, Report to the Secretary of State, Department of Industry. Her Majesty's Stationery Office, London.

Meran, G., Hirschhausen, C. v., 2009. A Modified Yardstick Competition Mechanism. *Journal of Regulatory Economics* 35(3), 223-245.

Osborne, M. J., Rubinstein, A., 1994. A Course in Game Theory. The MIT Press, Cambridge, MA.

Sappington, D., 2005. Regulating Service Quality: A Survey. *Journal of Regulatory Economics* 27(2), 123-154.

Shleifer, A., 1985. A Theory of Yardstick Competition. *RAND Journal of Economics* 16(3), 319-327.

Stokey, N. L., Lucas, R. E., with Prescott, E. C., 1989. Recursive Methods in Economic Dynamics. Harvard University Press, Cambridge, MA.

Tangerås, T. P., 2002. Collusion-proof Yardstick Competition. *Journal of Public Economics* 83(2), 231-254.

Train, K. E., 1991. Optimal Regulation. The Economic Theory of Natural Monopoly. The MIT Press, Cambridge, MA.