



No. 20-2013

Ivo Bischoff and Frédéric Blaeschke

**Incentives and Influence Activities in the Public Sector: the
Trade-off in Performance Budgeting and Conditional
Grants**

This paper can be downloaded from
http://www.uni-marburg.de/fb02/makro/forschung/magkspapers/index_html%28magks%29

Coordination: Bernd Hayo • Philipps-University Marburg
Faculty of Business Administration and Economics • Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

Incentives and Influence Activities in the Public Sector: the Trade-off in Performance Budgeting and Conditional Grants

Ivo Bischoff* and Frédéric Blaeschke[†]

This version: March 15, 2013

Abstract

Performance budgeting schemes in the public sector have to operate with imperfect performance measures. We argue that these imperfections generate incentives for the potential recipients of performance-based funds to use up resources in socially wasteful influence activities. We develop a game-theoretical model to analyse the trade-off between the efficiency-enhancing effect of performance budgeting and the social waste it induces. Comparing a performance signal based on recipients' effort to a signal based on their output shows that a) the former evokes more social waste while the latter amplifies inequality in the amount of public services across districts. Performance budgeting schemes using the output-based signal yield welfare gains in a wide range of parameter constellations while the applicability of PB-schemes using the effort-based signal is limited. We also show that welfare losses arise when the government is opportunistic. Our model can be applied to the very similar trade-off emerging with the use of conditional grants in federalist countries.

Key words: Performance budgeting, rent-seeking, bureaucracy, public-sector production, conditional grants, opportunistic government

JEL: D 7, H 77, H 5, H 11

* Department of Economics, University of Kassel, Nora-Platiel-Strasse 4, 34109 Kassel, E-mail: bischoff@wirtschaft.uni-kassel.de, Tel. ++49 561 8043033, Fax. ++49 561 8042818.

[†] (corresponding author) Department of Economics, University of Kassel, Nora-Platiel-Strasse 4, 34109 Kassel, E-mail: blaeschke@uni-kassel.de, Tel. ++49 561 8043034, Fax. ++49 561 8042818.

1. Introduction

The notion that public sector production is inefficient is widely accepted in the economics and public administration literature. The underlying cause is the asymmetric distribution of information between policy-makers who choose the level of public services and allocate funds and the administrative units that produce these services. Administrative units make use of this asymmetry and pursue their own goals rather than public interest (e.g., de Groot, 1988; Moe, 1997, 2006). Since the 1990s, many countries introduced performance budgeting (hereafter PB) to reduce this asymmetry in information and improve public sector efficiency (e.g., OECD, 2007; Robinson, 2007a). Under PB, the subordinate units have to provide policy-makers with information about their performance (e.g., Jordan and Hackbart, 1999; Kelly and Rivenbark, 2003: ch. 5-7; Robinson, 2007b; Lockwood and Porcelli, 2011). Policy-makers can use this information to allocate funds based on performance, thereby inducing competition among administrative units and making them improve performance (e.g., Robinson, 2007c). It is widely acknowledged that the available numerical performance indicators are imperfect in that they capture only a part of the benefits generated by public services (e.g., Joyce, 1993; Crain and O’Roark, 2004; Robinson, 2007b; Jones and McCaffery, 2010). The largest part of the existing PB-schemes uses additional information from performance reports and other sources (e.g., Kelly and Rivenbark, 2003: ch. 5-7; Gilmour and Lewis, 2006; OECD, 2007: ch. 3).

While this additional information contributes to a more comprehensive view of the unit’s performance, it causes a serious new problem: The rich informational content of performance indicators and especially of performance reports cannot be appreciated in simple allocation rules. Its appreciation requires a complex process in which policy-makers allocate funds in discretionary and often difficult or even arbitrary decisions (e.g., Moynihan, 2005). This in turn provides incentives for the administrative units to engage in socially wasteful fund-

seeking activities: They can try to attract funds by sugarcoating the performance report or directing efforts to highly visible measures. These activities will hereafter be called window-dressing.¹ If policy-makers are opportunistic and use PB-funds to extract rents, the administrative units can improve their chance of receiving funds by supporting the policy-makers election campaigns. Following Milgrom (1988), we use the generic term “influence activities” for window-dressing and campaign support. Influence activities aim at redirecting rents but do not yield any net benefit for the public. Thus, like the lobbying activities of firms and interest groups in the theory of rent-seeking (e.g. Tullock, 1980; Congleton, 2008), resources spent on influence activities represent social waste. Unlike in the rent-seeking models, however, not all activities that potentially attract funds are wasteful: Any effort that increases true performance may attract funds and increase overall welfare at the same time. In other words, there is a trade-off in PB: Distributing funds based on performance indicators and reports increases performance but also increases social waste from influence activities.

In this paper, we provide a theoretical model that analyses this trade-off in public sector PB.² The units’ performance and influence activities as well as the policy makers’ choice of PB-scheme are explained endogenously in the model. Our main results can be summarized as follows: First, we compare PB-schemes in which every sub-unit receives a performance bonus depending on its relative performance to PB-schemes based on a contest for a limited number of “prizes” awarded according to relative performance. We show that they are equivalent with respect to their impact on the units’ behavior and overall output. Second, we show that the optimal PB-scheme differs fundamentally for different performance signals. We compare two different performance signals: The effort-based signal judges units’ performance by looking at

¹ In general the term window dressing is used in the sense of “making something appear better than it is”. The accounting literature uses it to refer to activities such as accounting tricks to make balance sheets and business performance reports appear better (e.g. Stickney, 1975). But the term is also used in a variety of fields where economic appearance or labels may play a prominent role such as investment banking (e.g. Morey and O’Neal 2006), structural policy (e.g. Li-Wen, 2010) or corporate social responsibility (e.g. Amazeen 2011).

² Michaels (1988) models a rent-seeking contest with multiple inputs. However, all his inputs cause social waste. The aim of his paper was not to model the trade-off that we focus on here but to show that relative prices determine the mix of different forms or rent-seeking effort.

their organizational procedures, the general features of the process of public service production and the specific projects they launch. This information serves as an indicator for the effort of the unit leader and thus for expected performance. The output-based signal contains information about the intermediate output that units produce using the basic lump-sum funding. Both signals can be manipulated by influence activities. Under the effort-based signal, PB evokes substantial social waste from influence activities. Net gains in welfare are only feasible if the performance signal is sufficiently insensitive to influence activities and the units' effort to increase efficiency is sensitive to budgetary incentives. Under the output-based signal, the level of wasteful window-dressing is much lower. It is even possible to resolve the trade-off between efficiency gains and wasteful influence-activities if the government distributes a sufficiently high share of all available funds based on performance (in the model this always requires a share of more than 50 percent). The factors that limit the applicability of PB under the effort-based performance signal do not apply. A government that wants to maximize overall output and has the opportunity to choose between performance-signals should use the output-based signal and distribute the highest-possible share of funds based on performance – leaving just sufficient block grants to receive a performance signal. However, this increase in overall output comes at a price: PB-schemes based on the output-based signal produce systematic inequality in public services across the districts to which the PB-scheme applies. This inequality increases in the share of performance-based funds. Thus, the dominance of the output-based signal is weakened if welfare is sensitive to inter-district inequality. Under a Rawlsian welfare-function, the output-based signal is inferior in situations when there is a high inter-temporal correlation and/or substantial variance in the exogenous factors that influence unit output. Finally, we account for opportunistic policy makers. If their aim is to extract rents policy makers misuse PB-schemes and cause welfare losses.

The remainder of the paper is organized as follows: Section 2 outlines the concept of PB and reviews the related literature. In section 3, we present our baseline model and a number of

extensions. In section 4, we discuss the results and their implications for PB. Furthermore, we argue that the logic of the model applies to conditional grants in federalist countries. Section 5 concludes.

2. The concept of PB and related literature

The concept of PB originates from an exchange between scholars from management science, administrative sciences and practitioners in public administration. Since the 1990s, a large number of countries have installed elements of PB (OECD, 2007). Their numerical performance indicators include output measures such as number of students per year for schools and universities, average unemployment duration for employment agencies and performance ratings from surveys among customers of the administrative unit (e.g., Game, 2006). PB-schemes are widespread in the field of higher education (e.g., Guthrie and Neumann, 2007; Orr et al., 2007; Shin, 2010, Wilkesman and Schmid, 2011). Some countries apply similar schemes to local hospitals or cadastre services (OECD, 2007; ch. 3). Despite the importance of numerical indicators in some fields, their overall importance is limited. PB-schemes that rely on a fully-specified allocation formula based on numerical indicators only are the rare exception (OECD, 2007; Robinson, 2007a).

Two main arguments stand against allocating funds by a formula that is based on numerical performance-indicators only. The first argument is the heterogeneity in public services: Many administrative units within one constituency are monopolists for a particular public service. Given the wide range of public services, a common numerical indicator to compare the performance across services does not exist. Even if it existed, the allocation of funds across different services requires additional information (e.g., Greiling, 2005; OECD, 2007). On the other hand, there are many other public services that are produced by more than one administrative unit within the same jurisdiction. Here, the second argument applies: Tying funds to imperfect indicators sets incentives for the administrative units to maximize indicator

scores rather than performance. At the same time, they are incentivized to neglect their duties in all fields of public service production that do not influence indicator scores (e.g., Joyce, 1993; Cragg, 1997; Gilbert and Rocaboy, 2004; Gilmour and Lewis, 2006). To avoid these wrong incentives, efficiency-enhancing PB-schemes must employ other pieces of information next to the numerical indicators to assess the performance and allocate funds.

There are a number of models in the economics and public choice literature that are related to PB. The first strand of literature is theory of bureaucracy (see.g., Niskanen, 1971; Moe, 1997 and Wintrobe, 1997). Bureaucrats are assumed to know the public benefits generated by these services as well as the costs of producing them while policy-makers only know the benefits. This asymmetry gives bureaucrats a strong bargaining position in the budget process and results in inefficiently high output and/or x-inefficiencies and managerial slack. Another strand of literature applies the standard principal-agent models to the relationship between policy-makers (principals) and bureaucrats (agents) (for a review, see Moe, 2006).³ In these models, bureaucratic output depends on the bureaucrat's effort and an exogenous factor that is beyond his control. As both are unobservable for the policy-maker, neither output nor effort is contractible. The fact that policy-makers and bureaucrats follow different objective functions causes moral hazard. The literature on contest theory suggests that contests for prizes or extra funds can induce efficiency gains (e.g., Clark and Riis, 1998). McCubbins et al. (1987) argue that administrative procedures within the bureaucratic apparatus serve as instruments to control the bureaucrats and thereby reduce the bureaucrats' possibility to exploit their informational advantage. PB may be seen as a mixture of administrative procedures and incentive schemes. While the numerical performance indicators and reports are an element of control, performance-based funding sets incentives to improve efficiency.

³ A number of papers discuss similar matters in the context of federal structures in which the supra-ordinate government delegates task to local jurisdictions.

The literature sketched above ignores the fact that PB-schemes based on imperfect performance signals set incentives for the bureaucrats to waste resources in influence activities. The rent-seeking literature provides powerful instruments by which we can model the administrative units' competition for performance-based funds (e.g. Tullock, 1980; Congleton, 2008). Yet it stresses only the negative aspects of this competition – i.e. the resources wasted in this competition. A number of theoretical papers address both positive and negative aspects of (monetary) incentives in a unified framework. In particular, the works of Paul Milgrom and co-authors on the incentives in labor contracts and organizations is noteworthy. Milgrom and Roberts (1988) argue that employees can engage in influence-activities – activities that do not produce any surplus for the employer but have the sole purpose to increase the employee's probability of being rewarded, e.g. by professional advancement or higher wages. One way to limit influence activities is to limit managerial discretion (e.g., Milgrom, 1988). In the context of PB, rigid and highly formalized fund-allocation procedures are suitable means to limit managerial discretion. However, such procedures eliminate the discretion necessary to develop a comprehensive picture of the units' performance and allocate funds accordingly. If it is costly to limit influence activities the employer may prefer to give up incentive schemes altogether.⁴ Though Milgrom (1988) points out that the essential logic applies to incentives in the public sector as well, his studies have found little echo in the public economics literature so far.⁵

Here our paper comes in. We provide a theoretical model of the trade-off between positive and negative aspects of PB-induced competition for funds. We follow Milgrom (1988) in

⁴ The seminal papers by Paul Milgrom and co-authors have inspired numerous articles in the field of incentives in organizations (e.g., Prendergast, 1999, Inderst et al., 2005). Another strand of literature interprets lobbying and rent-seeking as an information-revealing process (e.g., Austen-Smith, 1997; Lagerlöf, 2007) and shows that lobbying and rent-seeking can be beneficial from a societal perspective if the information revealed in the contest for rents is sufficiently valuable. However, the contests they model are structurally different to the contest for PB-funds.

⁵ A cited reference search in the ISI database reveals only very few citations of Milgrom's contribution in leading public economics journals. For instance, it has been cited by 5 articles published in the *Journal of public economics*, by two articles published in *Public Choice* and by no articles published in the *Journal of public economic theory*.

assuming that the agent can engage in socially wasteful influence activities. Specific features of public sector production have to be accounted for. In PB, unit leaders do not compete for personal pay or professional advancement but for a larger budget for their units. More importantly, we have to account for the possibility that social welfare is sensitive to the distribution of output across districts.⁶

3. A model of influence activities in performance budgeting

3.1 The baseline model

Consider a jurisdiction that is divided into N districts. The government of this jurisdiction (the principal) is in charge of providing citizens in all districts with the public service X . It has delegated the task to N administrative units (hereafter units) – one unit per district (the agents). Citizens from district i can only consume the services X_i produced by unit i ($i = 1, \dots, N$). Let $h[X_i]$ be the welfare that X_i generates among the citizen in district i . We assume $h[X_i] = X_i$. Furthermore, let X_i be described by the following production function:

$$X_i = \alpha_i (v_i - \lambda_i) \cdot \varepsilon_i \quad (1)$$

Here, v_i depicts the gross funds of unit i and λ_i represents the amount of resources spent on preparing numerical performance indicators and writing the performance report (hereafter PB-reporting). To keep the model tractable, we assume an output-elasticity of 1 for funds used in production ($v_i - \lambda_i$), and for the effort (α_i) the unit-leader in district i puts into ensuring that these inputs are used efficiently ($\alpha_i \geq 0$). The more effort he exerts, the higher the unit's output – all other things equal. The parameters α_i and λ_i are controlled by leader of the administrative unit and are not observable to the jurisdictional government. The random term ε_i ($E(\varepsilon_i) = 1$) is not observable either. It is assumed to follow a symmetric distribution within

⁶ Milgrom (1988) accounts for “distributional effects” of the reward allocation. Yet this refers to the difference in valuation of the reward rather than the emergent inequality among competitors.

the limits $[1-d, 1+d]$, $d < 1$, and to be uncorrelated with all other determinants of output. It captures all factors that influence the unit's output yet are beyond the control of the unit leader. These factors include environmental factors from outside the administrative unit but also factors from within. Following the principal-agent literature, we assume that these factors cause a substantial variance in output and thus the unit's output X_i is not contractible.

The amount of resources available in district i (v_i) consists of basic funds identical for all units plus additional performance-based funds that vary across units. Unit i in district i can only spend these resources for PB-reporting and public service production. The level of basic funds and the distribution-scheme for performance-based funds is determined on constituency level. The individual unit leader can influence the expected amount of performance-based funds in his district by choosing the effort level α_i and the amount of resources spent on PB-reporting λ_i . Without PB-scheme, the resources used for PB-reporting $\lambda_i = 0$ and all available resources are spent on public service production. If a PB-scheme is installed, units have to spend resources on PB-reporting. There is a lower limit λ^{min} for the resources spent on PB-reporting. It comprises of the unavoidable costs of collecting information, preparing numerical performance indicators and writing an "uninspiring and any old how" performance report that meets the minimum requirements for such reports. However, the unit can also go beyond λ^{min} and spend additional resources on window-dressing. All resources spent on PB-reporting are no longer available for public service production and thus reduce output. This holds for the unavoidable λ^{min} as well as for any resources spent beyond λ^{min} .

Performance indicators and the performance report provides the government with the imperfect performance signal π_i for unit i . In the baseline model, we assume that the jurisdictional government receives the following signal:

$$\pi_i \equiv \pi(\lambda_i, \alpha_i) = \lambda_i^r \alpha_i^{1-r}, \quad r \in [0, 1]. \quad (2)$$

This signal is hereafter called effort-based signal. The size of the signal increases in the effort of the unit leaders to produce efficiently (α_i) and in the resources that unit i spends on PB-reporting (λ_i). The performance reports contain information about α_i because it describes the unit's organisational procedures and the main features of the production process. In addition, it lays out the various projects that the unit is running respectively plans to run and explains how these contribute to the overall output. Based on this information, the jurisdictional government gets a picture of the level of activity and the modernity of activities within the unit. Other things equal, the picture is more positive the more effort the unit leader exerts, i.e. the higher α_i . At the same time, the unit can positively influence the picture by polishing up the report and other forms of window-dressing. The more resources they spend on window-dressing, the more impressive the activities and the expected performance appears to be. In the specification chosen here, PB-reporting/window-dressing and the unit leader's effort cross-fertilize: resources spent on PB-reporting have a larger impact on the overall performance signal if the unit leader's effort is high and vice versa. Parameter r represents the signal elasticity of PB-reporting (i.e. $r = \partial\pi_i/\pi_i/\partial\lambda_i/\lambda_i$). The higher r , the more the performance signal responds to resources used for window-dressing. For the time being, we assume that policy-makers are not susceptible to lobbying, nor do they misuse their discretionary power to extract rents. Thus, signal elasticity r is exogenous and measures the unavoidable degree of imperfection in the performance signal. It can be interpreted as a parameter describing the degree of asymmetry in information between policy-makers and unit leaders (next to the dispersion of the random term ε described by d). It is important to note that this performance signal is imperfect because it does not capture the opportunity costs of window-dressing. Furthermore, it is input-based in that it is insensitive to all exogenous factors driving output but solely accounts for the unit leader's effort (next to their second input – i.e. resources spent on window dressing).

Let the unit leaders be risk-neutral and maximize the following objective function.

$$\Theta_i = X_i - a(\alpha_i)^z, \quad a \geq 0, z > 1 \quad (3)$$

The leader of unit i is interested in the output of his unit X_i for altruistic reasons (e.g., Francois, 2000; Besley and Ghatak, 2005), for reasons of prestige or due to career concerns (e.g., Dewatripont et al., 1999).⁷ The more effort he puts into increasing his units' efficiency, the higher X_i – other things equal. On the other hand, producing efficiently has opportunity costs because it reduces the possibility to provide benefits for himself and for his staff and requires him to put pressure on his staff to reduce slack. This impairs a good relationship between unit leader and his staff and thereby reduces his utility. By setting $z > 1$, we assume increasing marginal opportunity costs of effort.

In the baseline model, we assume a benevolent jurisdictional government that maximizes the simple utilitarian welfare function:

$$WF = \sum_{i=1}^N h[X_i] = \sum_{i=1}^N X_i \quad (4)$$

When introducing PB, the government can choose between two types of PB-schemes. First, it can initiate a contest for a performance prize given to $K < N$ units. In the extreme case where $K = 1$, the full amount of performance-based funds is given to one winning unit. These highly competitive contests may reward the “best-performing school” or the “most beautiful or lively city district”. The German “Excellence Initiative” for universities is an example for a prize-scheme in which the government spreads performance-based funds among a limited number of “well-performing units” (i.e. $K > 1$). Second, the government can decide to provide every unit with a performance bonus that depends on relative performance. This procedure rewards performance but avoids the substantial differences in funds between units that automatically

⁷ The jurisdictional government is not the primary and most immediate provider of career prospects as implied in Milgrom (1988). By running a well-performing unit, the unit leader may build up a reputation visible for many other governments beyond the own jurisdiction.

arises in performance-prize schemes – potentially even between units whose effort and window-dressing expenditures is similar.

Next to choosing between these PB-schemes, the government has to set the share of funds (f) that are distributed based on performance. For reasons of simplicity, we assume that all funds for the production of X are provided by the jurisdictional government. Let the total amount of funds be given by F . Then the total amount of performance based funds amounts to fF . The remaining fraction $(1-f)$ is distributed as block fund of equal size to all N units ($b = (1-f)F / N$).⁸

In the case of a performance-bonus scheme, unit i can expect a bonus that amounts to:

$$pb_i = \frac{fF}{N} \cdot \frac{\pi_i}{\pi_{av}} \quad \text{with} \quad \pi_{av} = \frac{1}{N} \sum_{j=1}^N \pi_j. \quad (5)$$

The higher the performance signal of a certain unit i relative to the average signal, the more funds this unit can expect. When modelling the procedure of allocating performance prizes, we follow the theory of rent-seeking and prize-contests (e.g., Tullock, 1980; Clark and Riis, 1996). We assume that all K winners receive the same performance prize ($pp = fF/K$) and we model the process of prize-allocation as if it was a lottery in which K performance-prizes are awarded consecutively. The rationale behind this as-if assumption is the following: The individual unit leader is uncertain about the performance signal of the other units. Even if he had this information, he could not anticipate the outcome of the complex allocation process on the jurisdictional level. From his perspective, it is similar to a lottery. Building on the works by Clark and Riis (1996), the probability p_i that unit i wins in one of the K rounds is given by:

$$p_i = p_i^1 + \sum_{j=1}^{K-1} \left[\prod_{s=1}^j (1 - p_i^s) \right] p_i^{j+1} \quad (6)$$

⁸ For reasons of simplicity, we assume that PB-schemes entail administrative costs on the units' level while their costs on jurisdictional level are negligible.

Here, p_i^1 stands for the probability that unit i wins the first prize, $(1 - p_j^s)$ represents unit i 's probability of not having won one of the previous prizes and p_i^j represents unit i 's probability to win in round $j > 1$. We assume that all probabilities strictly depend on relative performance such that $p_i^1 = \pi_i / \sum_{v=1}^N \pi_v$ and p_i^j is given by the ratio of π_i and the sum over the performance signals of all units still in the contest (Clark and Riis, 1996). The expected amount of performance based funds of unit i is given by:

$$E[pp_i] = \frac{fF}{NK} \cdot \frac{\pi_i}{\pi_{av}} + \frac{fF}{NK} \cdot \sum_{j=1}^{k-1} \left[\prod_{s=1}^j (1 - p_i^s) p_i^{j+1} \right] \quad (7)$$

Again, unit i 's expected funds are higher the higher its performance signal relative to the average signal. If one winner takes the entire pot (i.e. $K = 1$), the expected performance price is equal to the performance bonus under a performance-bonus scheme with the same pot size of fF/N (see expression (5)). Hereafter, we denote performance boni by the PB-scheme with $K = N$. Thus, all possible PB-schemes can be described as a combination of f and K .

The interaction between the government and its administrative units can be modelled as a sequential game consisting of four stages (see table 1): In stage 1, the government designs the PB-scheme by setting the share of performance-based funds f and the number of recipients K and provides all N units with the block-fund. In stage 2, the unit leaders decide about the effort level α_i and the amount of resources to spend on PB-reporting (λ_i). Given this decision, they use block funds net of PB-reporting expenditures ($v_i - \lambda_i$) to start providing citizens with public services. At this stage, the government receives the performance-signal π_i . In stage 3, the government distributes the remaining funds according to the PB-scheme set in stage 1. Finally, in stage 4, the units that receive performance prizes respectively boni use these additional funds to expand the production of public services. We assume that the unit leader

keeps up the initial effort level α_i . Consequently, the final output of every unit i is given by expression(1).

[insert table 1 about here]

We assume that the government knows the production function, the function of the performance signal and the unit leaders' objective function. Based on this information, the government can form rational expectations concerning the unit leaders' choice of effort, the resources used for PB-reporting and their expected output for all possible PB-schemes (i.e., combination of f and K). The government receives a unit-specific performance signal π_i in stage 2 of the game and observes the unit-specific output X_i in stage 4. Given the unobservable stochastic component ε_i these two pieces of information (π_i, X_i) are insufficient to calculate the true unit-specific effort α_i and the resources used for PB-reporting λ_i even after all information is revealed. Thus, the government cannot differentiate between those units that only spent the unavoidable amount λ^{min} of resources on PB-reporting and those units that wasted additional resources on window-dressing. Therefore, it cannot install an incentive scheme that punishes excessive PB-reporting ex post. Instead, it has to live with it and choose the PB-scheme that maximizes welfare.

The government solves the game by backward induction. For this purpose, it is necessary to develop the unit's reaction functions, $\alpha_i = \alpha_i(f, K)$ respectively $\lambda_i = \lambda_i(f, K)$. At stage 2, unit leaders choose their effort level and the resources for PB-reporting to maximize expected utility given the value of f and K they face:

$$\max_{\lambda_i, \alpha_i} \{E[\Theta_i] : \lambda_i = 0 \vee \lambda^{min} \leq \lambda_i \leq b; 0 \leq \alpha_i\} \quad (8)$$

In the absence of PB (i.e. $f = 0$), they do not spend any resources on PB-reporting and choose the effort level α_i that maximizes their objective function:

$$\alpha_i^{aut} = \left[\frac{F}{N} \frac{1}{az} \right]^{1/(z-1)} \quad (9)$$

Expression (9) shows that the effort level increases in the amount of resources available.

For reasons of simplicity, we assume that all units are identical with respect to their objective function Θ_i and the form of the production function $X_i(\cdot)$. Consequently, we assume that $\pi_i = \pi_j \forall i, j$ holds in the Nash-equilibrium (see Berry, 1993; Clark and Riis, 1996). We hereafter drop the subindex i to save notation. For ease of the argument, we furthermore assume that the lower limit for the resources used for PB-reporting λ^{min} is sufficiently small so there exists an equilibrium in pure strategies. Assuming an interior solution for the unit leaders choice of α_i and λ_i , we arrive at the following Nash-equilibrium solutions for the units' choice when facing a performance-prize scheme (expression (10)) respectively performance-bonus scheme (expression (11)):

$$\lambda_{pp}^* = r \frac{fF}{KN} M \quad \alpha_{pp}^* = \alpha^{aut} \left(1 + (1-2r)M f/K \right)^{1/z-1} \quad (10)$$

$$\text{with } M = \left[\frac{K(N-1)}{N} - \sum_j^{K-1} \frac{K-j}{N-j} \right]$$

$$\lambda_{pb}^* = r \frac{fF}{N} \frac{(N-1)}{N} \quad \alpha_{pb}^* = \alpha^{aut} \left(1 + (1-2r) f(N-1)/N \right)^{1/z-1} \quad (11)$$

If the signal elasticity r takes on the critical value $r = 0.5$, the unit leader's effort level is not changed by the PB-scheme but remains at the autonomy level α^{aut} . In this case PB wastes resources for PB-reporting without yielding any benefit. PB-schemes only increase the unit-leaders effort (i.e. yield $\alpha^* > \alpha^{aut}$) if $r < 0.5$. For all constellations satisfying this condition, expressions (10) and (11) reveal the essential trade-off: Increasing the share of performance-based funds improves overall welfare by causing all units to be more efficient ($\partial \alpha^* / \partial f > 0$). At the same time, a concomitant increase in resources spent on PB-reporting ($\partial \lambda^* / \partial f > 0$)

reduces welfare. Under performance-prizes, a similar trade-off exists when the government changes the number of recipients K as both effort and PB-reporting expenditures decrease in K . For a given share of performance-based funds f , a performance-prize for $K = 1$ units leads to the same effort level and PB-reporting expenditures as performance boni. Regardless of whether prizes or boni are used to incentivize performance, both increase in per-unit funds F/N . The effort level α^* decreases in the disutility the unit leaders witnesses as their effort increases (depicted in z and a), while these parameters are irrelevant for the resources used for PB-reporting (λ^*). The poorer the quality of the performance signal (i.e. the higher r), the more resources are spent on PB-reporting and the lower the increase of the effort induced by PB.

This pattern of unit behavior in stage 2 serves as input to the optimization problem of the jurisdictional government in stage 1. It chooses the PB-scheme, i.e. the combination of f and K , that maximizes overall welfare:

$$\text{Max}_{f,K} \left\{ E \left[WF(F, \lambda^*, \alpha^*) \right] \quad \text{s. t. } 0 \leq f \leq 1, \lambda^* = \lambda(f, K), \alpha^* = \alpha(f, K), \lambda^* \leq b \right\} \quad (12)$$

We arrive at the following interior solution for share of funds f to be used for performance prizes (expression (13)) respectively for performance boni (expression (14)).

$$f_{pp}^* = \frac{K(1-2r) - (z-1)r}{M(1-2r)zr} \quad (13)$$

$$f_{pb}^* = \frac{N(1-2r) - (z-1)r}{(N-1)(1-2r)zr} \quad (14)$$

Expression (13) shows that the government has a degree of freedom when choosing f and K for performance prizes. The optimal share f_{pp}^* increases with the number of recipients K . Comparing expression (13) and (14) also reveals that the optimal performance-bonus uses the same share of funds as the optimal performance-prize with $K = 1$. The optimal share f^* for

both types of PB-schemes increases in the quality of the performance signal (i.e. decrease with r) but decreases in the unit leader's reluctance to exert effort (depicted in a and z). However, it is insensitive to the amount of funds F/N available in total.

Assuming an interior solution for λ^* , α^* and f^* for all values of K and substituting (13) in (10) and expression (14) in (11) yields expressions for the level of PB-reporting and effort in the social optimum. These are the same for performance bonus and performance prizes:

$$\lambda^* \Big|_{f_{pp}^*, f_{pb}^*} = \frac{F(1-2r) - r(z-1)}{Nz(1-2r)} \quad \alpha^* \Big|_{f_{pp}^*, f_{pb}^*} = \alpha^{aut} \left(1 + \frac{(1-2r) - r(z-1)}{zr} \right)^{1/(z-1)} \quad (15)$$

In other words, all efficient performance-prize-schemes have the same effect on unit behavior and expected output. Furthermore, the efficient boni-scheme yields the same results as any efficient prize-scheme. All optimal PB-schemes yield the same expected overall output and welfare. Thus, performance prizes and performance boni are perfect substitutes. When plotting the welfare-surface, the optimal (f^*, K^*) -combinations for performance prizes ($K = 1, 2, \dots, N-1$) and performance bonus ($K = N$) form a fin-shaped roof-ridge. The optimal amount of PB-funds increases in K as long as $K < N$. Furthermore, it is identical for “the winner takes it all” prize scheme ($K = 1$) and the performance bonus scheme ($K = N$). Figure 1 shows this roof-ridge for an arbitrary parameter constellation.

[insert figure 1 about here]

In the theory on rent-seeking, the welfare-effects are expressed by the degree of rent-dissipation – the ratio of aggregate social waste from rent-seeking to total rent (e.g., Tullock, 1980). In the context of our paper, the degree of rent-dissipation is given by:

$$\frac{N\lambda^*}{f^*F/N} = r \frac{N-1}{N} \quad (16)$$

The degree of rent-dissipation rises in the signal elasticity of PB-reporting and the number of units. As this number increases, the degree of rent-dissipation approaches the signal elasticity.

The results above are based on the assumption that the maximization problems of unit leaders and of the jurisdictional government yield interior solutions. An interior solution applies if three conditions are met. First, the PB-scheme must cause an increase in the unit leaders' effort to produce efficiently (i.e. $\alpha^* > \alpha^{aut}$). By rearranging the numerator in expressions (13) and (14), it is easy to show that this requires the following condition to be satisfied:

$$z \leq 1/r - 1 \quad (17)$$

The more imperfections blur the performance-signal (i.e. the larger signal elasticity r), the more willing the unit leaders have to be to exert effort (i.e. the smaller z) in order to make PB socially beneficial (and vice versa). For all cases where $z \geq 1/r - 1$, the optimal solution for the government is not to make use of PB (i.e. $f^* = 0$). This holds for all situations where $r \geq 0.5$.

If condition (17) is met, there are upper limits for f^* . First, the resources spent on PB-reporting cannot exceed the block grant (i.e. $\lambda^* \leq b$).⁹ This leads to the following restrictions:

$$f_{pp}^* \leq \frac{1}{1+rM/K}, \quad f_{pb}^* \leq \frac{N}{N+r(N-1)} \quad (18)$$

The smaller signal-elasticity r is, the higher this limit. As r approaches 0, the limit converges to the absolute limit of $f_{pp}^* \leq 1$, $f_{pb}^* \leq 1$. As the ratio M/K decreases in K , an interior solution may not exist for performance-prize schemes with $2 < K < N-1$ even if it exists for $K = 1$ and $K = N$. In many parameter constellations, the set of optimal PB-schemes consists of performance bonus and performance prizes with a small number of winners K only. The roof-ridge in figure 1 loses the top of the fin. In extreme cases, the set is restricted to a performance

⁹ Theoretically, the government could make use of this restriction and set very high values for the share of performance-based funds f to raise efficiency and at the same time limit PB-reporting to the size of the block grant. Taking this argument to the limit, it would set a value of f just below 1 and thereby dissolve the trade-off described in expressions (10) and (11). We do not allow for this solution for two reasons. First, if $\lambda = b$, the units have no resources to start production of public services in stage 2 of the game (see table 1). Production in stage 2 is, however, a likely precondition for the existence of the performance signal. Thus, PB-schemes that deliberately exploit the λ -restriction destroy their own informational basis. Second, the solution lacks empirical relevance.

bonus and a performance-prize with $K = 1$, both operating with the same share of performance-based funds f^* .

For a given number of units, the applicability of PB depends on the size of signal elasticity r and on the unit leaders' reluctance to improve efficiency (expressed by parameter z). Figure 2 below assumes that a performance bonus or a performance prize with $K = 1$ is applied in a jurisdiction with $N = 20$ districts. Figure 2a visualizes the range of (r, z) –constellations for which it is optimal not to apply PB, for which an interior solution exists and for which the upper limits in expression (18) applies. Figure 2b introduces a third dimension and plots the optimal share of performance-based funds on the vertical axis. For constellations where the upper limit in expression (18) applies, it displays the critical value $f = N / (N + r(N - 1))$.

[insert figure 2 about here]

If z and/or r is small, the upper limit applies. In this case, the jurisdictional government maximizes overall output by choosing the highest possible share of performance-based funds f that still leaves block funding b sufficiently high to receive a performance signal. The corresponding (r, z) -constellations are located in the area south-west of the solid line in figure 2 and always go along with a high share of performance-based funds (see figure 2b). On the other hand, there is a broad range of (z, r) -combinations for which PB-schemes do not yield net benefits because r and/or z is too large (see expression (17)). This applies to all (r, z) -constellations north-east of the dashed line in the left graph of figure 2a. The strip between the dashed and the dotted line marks the (r, z) -constellations with an interior solution.

3.2 An alternative performance-signal

The performance signal used so far contains information concerning the unit leader's effort to produce efficiently (at the margin) – blurred by window-dressing. This gives PB-schemes based on this signal high legitimacy, especially when the quality of the performance signal is

high (i.e. r is low). The major shortcoming of this signal is that it does not account for the negative impact that PB-reporting (including window-dressing) has on final output. In this section, we analyse a performance signal that avoids this shortcoming. We assume that policy-makers receive an output-based performance signal based on the output generated by unit i while using the block b in stage 2 of the game:

$$\pi_i^X \equiv \pi(\lambda_i, X_i) = (\lambda_i)^r (\alpha_i (b - \lambda_i) \cdot \varepsilon_{i,S2})^{1-r}, \quad r \in [0, 1]. \quad (19)$$

This signal does not focus on the unit leader's effort as an input to public service production but on the resulting output (in stage 2). As a result, however, it is sensitive to exogenous factors: The random term $\varepsilon_{i,S2}$ captures the impact that factors beyond the control of the unit leaders have on the production in stage 2. We assume that $\varepsilon_{i,S2}$ follows the same distribution as the random term ε_i that disturbs final output in stage 4. The correlation between the two random terms is given by ρ_ε . In the following analysis, the assumptions from the baseline model continue to apply – unless stated otherwise. In particular, we assume that the Nash-equilibrium in stage 2 is symmetric in that all unit leaders choose the same effort level and the same amount of resources for PB-reporting. This implies that the heterogeneity due to the random term $\varepsilon_{i,S2}$ materializes only after unit leaders have made their decisions.

Hereafter, we restrict the analysis to performance-bonus scheme to save space.¹⁰ Assuming an interior solution, the units' optimization leads to the following choice in equilibrium:

$$\lambda^{\pi X} = \frac{fF(N-1)}{2N^2} + \frac{b}{2} - \sqrt{\left(\frac{fF(N-1)}{2N^2} + \frac{b}{2}\right)^2 - rb \frac{fF(N-1)}{N^2}} \quad (20)$$

$$\alpha^{\pi X} = \alpha^{aut} \left(1 - \lambda_{pb}^{\pi X} + (1-r) f(N-1)/N\right)^{\frac{1}{z-1}} \quad (21)$$

¹⁰ The equivalence of performance prize and performance bonus schemes holds under the modified performance signal.

Compared to the effort-based signal, window-dressing has a smaller impact on the expected funds ($\partial \pi_i^x / \partial \lambda_i < \partial \pi_i / \partial \lambda_i$) and thus unit leaders waste fewer resources on PB-reporting. More importantly, the resources used for PB-reporting do not rise monotonically in the share of performance-based funds f but fall beyond a threshold value. It can be shown that this threshold value is larger than 0.5 for all parameter constellations (proof, see Appendix A). Beyond this threshold, a further increase in performance-based funds no longer evokes additional window-dressing yet continues to have a positive impact on unit leaders' effort. Consequently, the strategy that maximizes overall output is to choose the highest possible share of performance-based funds f that still leaves block funds b sufficiently high to receive a performance signal. This strategy is optimal regardless of the values of model parameters $r, z, a, F, N, \rho_\varepsilon$. Based on the output-based performance signal, PB is applicable even when the unit leaders are reluctant to put effort into increasing efficiency (i.e., z, a are high) and/or the performance signal is of low quality (i.e., r is high) – constellations in which PB-schemes based on effort-based signals cannot yield welfare gains.

The output-based signal is different to the effort-based one in a second, very important respect: Due to the impact of random factors on output, the signal differs across units even if unit leaders choose the same effort level (α_i) and resources used for PB-reporting (λ_i). Substituting the equilibrium values for effort-level and PB-reporting (see expression (20) and (21)) into the expression for the output-based performance signal (see expression (19)) yields the performance bonus that unit i can expect under the modified performance signal:

$$pb_i = fF \left(\left(\varepsilon_{i,S2} \right)^{1-r} / \sum_{j=1}^N \left(\varepsilon_{j,S2} \right)^{1-r} \right). \quad (22)$$

Unit i 's bonus depends only on factors beyond the control of its leader. The distribution of performance boni is unequal. Inequality becomes larger the better the performance signal is (i.e. the lower r). The impact of this unequal distribution on overall output depends on the

correlation ρ_ε . If $\rho_\varepsilon = 0$, expected overall output does not depend on the initial inter-district distribution of performance-based funds. If $\rho_\varepsilon > 0$, the output-based performance signal provides policy makers with an important additional piece of information: It indicates which district can expect an amplification of output by favorable exogenous circumstances (and which districts can expect mitigating effects). Allocating PB-funds according to this information increases overall output. The higher the correlation ρ_ε between random terms, the stronger the impact of PB-funds on overall output. This provides an additional rationale for output-maximizing policy makers to distribute a large share of funds based on performance under the output-based performance signal and to use this signal rather than the effort-based one.

3.3 Performance-signals and inter-district inequality

In the previous sections, we operate with a simple welfare function that defines jurisdiction-wide welfare as sum of output over all N districts. It permits an analytical solution of the government's optimization problem and allows for comparative statics. However, it is insensitive to the distribution of output across districts. This assumption is likely to drive the results derived so far. In this section, we assume that the jurisdictional government aims at guaranteeing a high level of public services in all districts. Consequently, we use the Rawlsian welfare function:

$$WF^R = \min \{X_i\} \quad (23)$$

Based on this welfare function, we re-evaluate the optimality of different PB-schemes for different performance signals and parameter constellations. The change in welfare function does not change the unit's reaction functions $\alpha_i = \alpha_i(f, K)$ respectively $\lambda_i = \lambda_i(f, K)$ but changes the results of the government's optimization problem. Two results are noteworthy: First, the PB-scheme that maximizes overall output also maximizes welfare as measured by

the Rawlsian welfare function if the effort-based signal is used. The limits for the applicability and the comparative statics for performance boni under the effort-based signal are not sensitive to the welfare function used. This holds because a) the random terms in stage 2 do not influence the performance bonus a certain unit receives and b) inter-district inequality in outcome solely depends on the distribution of random terms in stage 4. Neither the correlation with possible random terms in stage 2 nor the share of performance-based funds influences the degree of inter-district inequality that materializes in the end.

Second, the output-based performance signal does not always dominate the effort-based signal under the Rawlsian welfare function. The question which signal yields higher welfare depends on the parameter values. If the output-based signal is used, the parameters describing the distribution in random terms (d, ρ_ε) have a negative impact a) on the performance bonus of the unit i with the most unfavourable random term ($\varepsilon_{i,s2} = (1-d)$) and b) on the final output of this unit. The higher d and ρ , the more severely this district suffers from unfavorable exogenous factors: These do not only mitigate this unit's PB-funds but further reduce the output achievable with the small amount of available funds. As these negative effects increase in the share of performance-based funds, they have the potential to outweigh the positive impact of an output-based performance signal on overall output. The larger the dispersion in random terms (i.e., the larger d) and the larger the correlation ρ_ε between random terms across stages, the more likely the negative effects limit the welfare-maximizing share of performance-based funds under an output-based signal. Therefore, the output-based signal is less likely to be superior the larger d and ρ_ε are – other things equal. Figure 3 shows critical combinations of these two parameters for which the superior performance signal changes for an arbitrary constellation of other parameters.

[insert figure 3 about here]

Different lines show critical combinations for different values of the other parameters.¹¹ If the quality of the performance signal is low (i.e. r is high) and unit leaders are reluctant to increase efficiency (i.e., a and z are high), the effort-based performance signal evokes massive social waste from PB-reporting per unit efficiency gain and PB is limited in its ability to increase output. Under the output-based performance signal, these factors do not limit the ability to increase output. Therefore, the output-based signal is more likely to be superior the larger the parameters r , a and z are – other things equal.¹² Consequently, the area where the output-based signal is superior increase in the parameters r , a and z .

3.4 Opportunistic policy-makers

So far, the jurisdictional government is assumed to be benevolent. Let us drop this assumption and assume that government is interested in extracting rents (McChesney, 1997; Page, 2005). Rent-extraction does not necessarily mean that the government demands unit leaders to “buy” funds by illegal favors or cash payments. Instead, the government may concentrate PB-funds in those units that support its election campaign. In any case, its utility increases in the amount of resources that the administrative units transfer back to it. In the context of the model, the resources transferred are captured by λ_i . A government that wants to extract large amounts of rents will reserve a large share of performance-based funds f and prefer the effort-based performance signal over the output-based one. More importantly signal elasticity r no longer represents the unavoidable degree of imperfection resulting from window-dressing that the government cannot identify as such and thus correct for. Instead it becomes a strategic parameter controlled by the government. The latter will send signals to the units that the PB-scheme operates at high values of r . As a result, the unit leader’s effort is of minor importance

¹¹ As the optimal share of performance-based funds under the output-based signal cannot be derived analytically, it is impossible to define the conditions under which one or the other performance signal is superior in general terms. Figure 2 has thus been generated by simulations. We assume a symmetric triangular distribution of all random terms. The critical values were generated using 50 randomly created samples for the random term vectors for every single combination of $[\varepsilon_H - \varepsilon_L]$ and ρ_ε . (Simulation routines and results are available with the author).

¹² The primary impact of the ration F/N is on α^{out} and applies to both signals in a similar way.

for the distribution of PB-funds. The performance signal no longer informs about performance but represents the unit's input in the fund-seeking contest equivalent to the rent-seeking effort in the standard rent-seeking model (Tullock, 1980).

4. Discussion

4.1 Model results and their implication for PB

The previous sections presented a game-theoretical model that captures the interaction between a jurisdictional government and its subordinate administrative units under PB. We analysed the welfare-effects of PB-schemes that operate with an imperfect performance signal which can be manipulated through window-dressing (and campaign support). First, we find that performance boni and performance prizes (for any number of recipients) are perfect substitutes when welfare in a jurisdiction is measured by overall output across districts. When the welfare function is sensitive to inter-district inequality in output, performance boni dominate performance prizes while the properties of the optimal performance bonus scheme does not depend on the form of the welfare function. Second, we find that the results fundamentally depend on the performance-signal available. Under the effort-based signal, the optimal share of performance-based funds increases in the quality of the performance signal and decreases the more reluctant unit leaders are to exert effort. The degree of rent-dissipation is proportional to the signal-elasticity of PB-reporting and converges to the value of this elasticity as the number of units increases. As the performance-signals available in reality are characterized by a substantial degree of imprecision, PB-schemes using an effort-based signal go along with substantial social waste from window-dressing. Beyond a certain level of imprecision, it is optimal not to apply PB. Somewhat paradoxically, the critical level of imperfection becomes lower the more reluctant the unit leaders are to improve efficiency and thus the more inefficient public service production is without PB.

Compared to the effort-based signal, PB-schemes that rely on the output-based signal can incentivize the unit leaders' effort at substantially lower waste from window-dressing. PB-schemes with large shares of performance-based funds ($f > 0.5$) can even dissolve the trade-off between productivity-enhancing effects and wasteful window-dressing. This makes PB-schemes using the output-based signal applicable to a much wider range of parameter constellations. Policy makers maximize output by choosing the highest possible value share of performance-based funds that still leaves basic funding b sufficiently high to get a performance signal. If both signals are available, the PB-scheme that maximizes overall output employs the output-based signal. On the other hand, PB-schemes using the output-based signal produce substantial inequality in output across districts. If this inequality matters for jurisdictional welfare, the output-based signal no longer dominates the effort-based signal in all constellations. Under a Rawlsian welfare-function, the effort-based signal is superior when the dispersion in random terms and the correlation in random terms across time is high. Finally, the analysis shows how rent-extracting governments will deliberately design PB-schemes that evoke excessive influence activities and reduce welfare.

The analysis above is based on a number of simplifying assumptions. First, we assume an output-elasticity of 1 for the effort of the unit leader respectively for the ability to increase efficiency (as expressed in parameter α). If we assume a lower elasticity instead, this will impose even stricter limits to the applicability of PB-schemes using an effort-based performance-signal. On the other hand, this logic has an interesting implication for PB in times of fiscal adjustment. In times when governments have to cut expenditures, the unit leaders will reduce their effort (as $\partial \alpha^{out} / \partial (F / N) > 0$ in expression (9)) but at the same time elevate the marginal gains in output the government can induce through PB. This may justify the use of PB-schemes in times of austerity even when it was not welfare-enhancing before.

Second, the baseline model assumes that unit leaders' utility increases in their output. Instead, following Niskanen (1971), we may assume that unit leaders are motivated by their disposable budget ($v_i - \lambda_i$) (e.g., Tullock, 1980; Wintrobe, 1997; Chang and Turnbull, 2002). The objective function thus reads $\Theta_i^N = v_i - \lambda_i - a(\alpha_i)^z$, $a \geq 0, z > 1$. It can be shown that the PB-reporting expenditures λ_i are always higher than in baseline model (respectively the effort (α_i) is lower). Thus, the degree of rent-dissipation is higher. At the same time, $\alpha^{aut} = 0$ and thus PB is always beneficial – regardless of the performance signal and its quality.¹³

Third, we assumed that the unavoidable amount of resources used for PB-reporting λ^{min} is sufficiently small to allow for solutions in pure strategies. This implies that all interior solutions involve some degree of window-dressing. If we account for the fact that λ^{min} may in many cases be substantial, there are a number of additional parameter constellations in which the welfare losses from the unavoidable administrative costs of PB-reporting cannot be outweighed by the gains in efficiency from PB. Under the effort-based performances signal, this applies to parameter constellations for which the optimal share of performance-based funds is small. As λ^{min} increases, the range of parameter constellations in which PB-schemes are welfare-enhancing shrinks. Under the output-based performance signal, significant fixed costs of PB-reporting are more likely to become binding because the Nash-equilibrium level of PB-reporting is much lower here. They limit the possibility to reduce social waste from window-dressing by increasing the share of performance-based funds. To loosen the restrictions from λ^{min} and broaden the applicability of PB, the government can reduce the frequency at which the units' performance is assessed. Existing PB-schemes generally initiate comprehensive performance assessments on a multi-annual basis (e.g., OECD, 2007: ch. 2).

¹³ Apart from that, the essential results from the baseline model hold. Results are available with the authors upon request. Note that using gross funds v_i rather than disposable funds ($v_i - \lambda_i$) in the objective function would automatically imply that units use the all block funds for window-dressing and thus overall outcome and welfare would be zero for all N-K units that do not receive a prize in the end. Thus, either $f = 0$ or $f = 1$ are optimal from the policy-makers point of view.

In addition, we model PB as a one-shot game. In reality, PB-schemes are applied over longer periods of time and involve consecutive performance reports and decisions about the allocation of funds. Thus, PB-schemes can employ information about unit-specific output X_i from previous periods as an additional numerical performance indicator. This makes it costly for the units to spend resources on window-dressing because it reduces expected performance-based funds in the next round. More importantly, the unit leader's effort has a positive influence on expected funds and output in the current *and* in future budgetary periods. Thus the ratio of effort to social waste is more favourable if PB is applied repeatedly. Other things equal, this widens the scope of parameter constellations for which PB-schemes using effort-based signals yield welfare-gains. Drawing from the logic on output-based performance signals (see section 3.2), the use of past performance as performance indicator can further increase overall output if the unit-specific random terms $\varepsilon_i (= \varepsilon_{i,t})$ are (positively) correlated inter-temporally. On the other hand, the main limitation of the output-based signal applies: Given the substantial variance of the random factors in both stages, the resulting interdistrict inequality in output is substantial. Moreover, a positive correlation in random terms implies that there are districts in which output is systematically lower even if unit leaders choose the same level of influence activities and exert the same effort as the leaders in other units. Allocating less PB-funds to these districts means lowering output in already disadvantaged districts. This destroys the legitimacy of PB-schemes and runs against the principles of fiscal equalization as well as performance budgeting (e.g., Gilmour and Lewis, 2006).

4.2 Influence-activities and conditional grant schemes in federal systems

Essential insights gained in the analysis above reach beyond the context of PB. In this section, we argue that our model opens a new perspective on conditional grants in federalist countries. Conditional grants represent an incentive-scheme for lower-level governments that is similar

to the performance prizes or boni (e.g., Boadway et al., 1999; Bessho and Terai, 2011). Local governments in most countries receive block grants depending on population size (and possibly a number of other indicators). In addition, supra-ordinate governments support them with conditional grants for specific projects that improve the quality of local public services. In their application for competitive grants, municipalities have to provide a broad account of their performance in public service production and the value-added by the prospective project. Whenever the demand for project funds exceeds the means reserved for them, some applications have to be turned down and some projects have to be downsized., This generates a competition for scarce funds. And just like PB-schemes, this competition in turn incentivizes welfare gains from improved public service production yet evoke wasteful influence activities (Borck and Owings, 2003; Bessho and Terai, 2011).

The literature on fiscal federalism names two major arguments to justify conditional grants: regional spillovers and local government failure. In the case of regional spillovers, the elements of the trade-off differ somewhat. Most importantly, conditional grants yield gains in efficiency even when the local governments are benevolent (Oates, 1999). While the main logic of the paper also applies to conditional grants to internalize spillovers, a thorough analysis requires some changes to the model. This lies beyond the scope of the current paper.

Such changes are not needed when conditional grants aim at fighting local government failure: Since the “*Federalist Papers*”, it is acknowledged that the local level is more prone to government failure than higher levels of government (e.g., Schultze, 1974; Chernick, 1979; Byrnes and Dollery, 2002; Ahmad and Brosi, 2006). Frequently, conditional grants schemes are explicitly justified by government failure. For example, the EU cohesion programs assume that economically weak regions may suffer from poor institutional quality and tie financial support to the improvement in institutional quality (e.g., Bähr, 2008). The European strategy 2020 for upcoming decade strengthens this nexus. In the US, conditional grants are arguably

introduced to ensure that certain technological and organizational standards are met in the educational sector (e.g., Fisher and Papke, 2000; Cascio et al., 2008). The same holds for the funding of regional and local youth welfare policy in Germany (e.g., Greese, 1998). Government failure may be caused by opportunistic local governments that are interested in extracting benefits (e.g., Belleflamme and Hindriks, 2005) or by regional pressure groups pressing local authorities for favourable yet inefficient solutions in public service production (e.g., Dougan and Kenyon, 1988; Austen-Smith, 1997; Lorz, 2001). Local governments have an incentive to give in to this pressure in order to ensure the support of these interest groups in the next election. On the other hand, giving in to these groups reduces the level or quality of public services for the citizens. If citizens vote retrospectively (e.g., Paldam, 2004), the local government loses their support in the next election. Thus, the objective-function used in the baseline model (see expression (3)) captures the situation of a local incumbent striving for re-election. Parameter α_i describes the local incumbent's effort to reduce slack in public service production and limit the influence of local interest groups. Parameter λ_i refers to the resources that the incumbent spends on the application for conditional grants. Developing and polishing up a prospective project proposal and application requires substantial resources. The supraordinate government's task to choose the best municipal project proposal(s) is a complex one – similar to the fund-allocation decision under PB. If interest groups are stronger on the local level than on the supra-ordinate level, conditional grants can increase net welfare. They can be a valuable instrument to reduce local government failure especially in situations where the preconditions for effective yardstick competition are not met, e.g., because mobility costs are high or citizens lack the necessary information to evaluate the local government's performance (e.g., Besley and Case, 1995; Besley, 2006).

A number of authors argue that the granting governments pursue their own political goals when distributing grants. They may prioritize certain regions to maximize political support (e.g., Grossman, 1994; Worthington and Dollery 1998). If the prioritized recipients know this,

they will be less motivated to reduce government failure but also spend less resources on window-dressing. The same holds for those regions that do not expect to be prioritized. Consequently, conditional grants will be less effective but also less harmful – the net effect is undetermined. If, however, the supra-ordinate government is interested in extracting rents, conditional grants are just as harmful as PB-schemes (see section 3.4).

5. Conclusion

Since its introduction in the 1990s, PB has spread continuously and is now applied for a substantial number of different public services in many countries. The existing PB-schemes assess the performance of administrative units in a complex procedure that uses numerical performance indicators but also draws on information from performance reports and others sources. While this is an adequate way to cope with the limitations of numerical indicators, the complex assessment procedure evokes socially wasteful influence activities. We develop a theoretical model to analyze this trade-off. We show that the impact of PB and the optimal PB-scheme differs fundamentally for different performance signals. PB-schemes using the effort-based signal can be expected to generate substantial social costs from influence activities. This imposes strict limits to applicability of PB is limited if policy-makers have to work with an effort-based performance signal. The level of wasteful window-dressing is much lower if policy makers use an output-based performance signal. Therefore PB-schemes based on this signal are more widely applicable and yield higher gains in overall output. However, PB-schemes based on the output-based signal come at the price of substantially higher inter-district inequality in public services. Once policy-makers account for this inequality, the output-based performance signal no longer dominates the effort-based signal. Depending on the constellations, the optimal PB-scheme employs the one or the other signal. We also show that policy makers whose aim it is to extract rents, misuse PB-schemes and cause welfare losses. Beyond PB, the main insights from our model can be generalized to governments in

federalist systems applying conditional grants to increase efficiency on the subordinate level, especially if inefficiencies result from government failure.

There are interesting questions for future research. From a theoretical perspective, the model can be extended to a world with heterogeneous recipient units of PB-funds (or conditional grants) and the role of targeting. Targeting rules can help governments to prioritize certain (disadvantaged) recipients (e.g. Stein 2002). Once influence activities are taken into account, they can also play an important role in reducing overall costs for influence activities. From an empirical perspective, one challenge lies in the development of an indicator for the intensity of influence activities and the resources spent on these. Such an indicator can help to differentiate wasteful from less wasteful PB-schemes. Especially in those cases where only effort-based signals are available, the question of whether PB-schemes should be implemented at all crucially depends on the degree of rent-dissipation. So far, the empirical basis to answer this question is missing entirely. Developing indicators that help quantify the costs of influence activities is an important endeavor of high and immediate political relevance. This analysis may have valuable implications for the future of the new strategy “Europe 2020”. Europe 2020 sets incentives to improve the quality of strategies chosen on national and regional level. At the same time, it builds on a system of performance reports that relies argely on ex ante reports concerning the effort of national and regional governments.

References

- Ahmad, E. and Brosi, G. (2006). Introduction. Fiscal federalism – review of developments in the literature and policy. In: Ahmad, E. and Brosi, G. (eds): Handbook of fiscal federalism, Cheltenham, Edward Elgar Publishers: 1-32.
- Amazeen, M. (2011). Gap (RED): Social responsibility campaign or window dressing? *Journal of Business Ethics* 99: 167 – 182.
- Austen-Smith, D. (1997). Interest groups: Money, information and influence. in: Mueller (ed.): 296-321.
- Bähr, C. (2008). How does sub-national autonomy affect the effectiveness of structural funds? *Kyklos* 61, 1: 3-18.
- Belleflamme, P. and Hindriks, J. (2005). Yardstick competition and political agency problems. *Social Choice and Welfare* 24: 155-169.
- Berry, S. K. (1993). Rent-seeking with multiple winners. *Public Choice* 77: 437-443.
- Besley, T. J. (2006). *Principled agents? The political economy of good government*. Oxford: Oxford University Press.
- Besley, T. J., Case, A. C., (1995). Incumbent behavior: vote seeking, tax setting and yardstick competition. *American Economic Review* 85 (1): 25–45.
- Besley, T. and Ghatak, M. (2005). Competition and Incentives with Motivated Agents. *American Economic Review* 95 (3): 616-636.
- Bessho, S. and Terai K. (2011). Competition for private capital and central grants: the case of Japanese industrial parks. *Economics of Governance* 2: 135-154.
- Boadway, R. Horiba, I. and Raghbendra, J. (1999). The provision of public services by government funded decentralized agencies. *Public Choice* 100: 157-184.
- Borck, R. and Owings, S. (2003). The political economy of intergovernmental grants. *Regional Science and Urban Economics* 33: 139-156.
- Byrnes, J. and Dollery, B. (2002). Local government failure in Australia? An empirical Analysis of New South Wales. *Australian Journal of Public Administration* 61(3): 54-64.
- Cascio, E., Gordon, N., Lewis, E., Reber, S. (2008). *Paying for Progress: Conditional Grants and the Desegregation of Southern Schools*. NBER Working Paper No. 14869.

- Chang, C. and Turnbull, G. K. (2002). Bureaucratic behaviour in the local public sector: A revealed preference approach. *Public Choice* 113: 191-209.
- Chernick, H. A. (1979). An economic model of the distribution of project grants. in: Mieszkowski, P. and Oakland, W. H. (eds). *Fiscal federalism and grants-in-aid*. Washington, D.C: The Urban Institute: 81-103.
- Clark, D. J. and Riis, C. (1996). A multi-winner nested rent-seeking contest. *Public Choice* 87: 177-184.
- Clark, D. J. and Riis, C. (1998). Competition over more than one prize. *American Economic Review* 88(1): 276-289.
- Congleton, Roger D. (ed.) (2008). *40 years of research on rent seeking*. Berlin etc.: Springer.
- Cragg, M. (1997). Performance incentives in the public sector: Evidence from the Job Training Partnership Act. *Journal of Law Economics, and Organization* 13: 147-168.
- Crain, W. M. and O'Roark, J. B. (2004). The impact of performance-based budgeting on state fiscal performance. *Economics of Governance* 5: 167-186.
- de Groot, H. (1988). Decentralization decisions in bureaucracies as a principal-agent problem. *Journal of Public Economics* 36: 323-337.
- Dewatripont, M., Jewitt, I. D. and Tirole, J. (1999). The economics of career concerns, part I: comparing information structures. *Review of Economic Studies* 66(1): 183-198.
- Dougan, W. R. and Kenyon, D. A. (1988). Pressure groups and public expenditures: The flypaper effect reconsidered. *Economic Inquiry* 26: 159-170.
- Fisher, R. C. and Papke, L. E. (2000). Local Government Responses to Education Grants. *National Tax Journal* 53, 1: 153-168.
- Francois, P. (2000). 'Public service motivation' as an argument for government provision. *Journal of Public Economics* 78: 275-299.
- Game, C. (2006). Comprehensive performance assessment in English local government. *International Journal of Productivity and Performance Management* 55: 466-479.
- Gilbert, G. and Rocaboy, Y. (2004). The central government grant allocation problem in the presence of misrepresentation and cheating. *Economics of Governance* 5: 137-147.

- Gilmour, J. B. and Lewis, D. E. (2006). Does performance budgeting work? An examination of the Office of Management and Budget's PART scores. *Public Administration Review* 66: 742-752.
- Greese, D. (1998). Kommunale Kinder- und Jugendpolitik. in: Wollmann, H. and Roth, R. (eds): *Kommunalpolitik. Politisches Handeln in den Gemeinden*. Opladen: Leske + Budrich: 717-731.
- Greiling, D. (2005). Performance measurement in the public sector: the German experience. *International Journal of Productivity and Performance Management* 54: 551-567.
- Grossman, P. J. (1994). A political theory of intergovernmental grants. *Public Choice* 78: 295-303.
- Guthrie, J. and Neumann, R. (2007). Economic and non-financial performance indicators in universities – The establishment of a performance-driven system for Australian higher education. *Public Management Review* 9: 231-252.
- Inderst, R., Müller, H. M. and Wärneryd, K. (2005). Influence costs and hierarchy. *Economics of Governance* 6: 177-197.
- Jones, L. R. and McCaffery, J. L. (2010). Performance budgeting in the U.S. Federal government: history, status and future implications. *Public Finance and Management* 10: 482-523.
- Jordan, M. M. and Hackbart, M. M. (1999). Performance budgeting and performance funding in the States: a status assessment. *Public Budgeting and Finance*: 68-88.
- Joyce, P. G. (1993). Using performance measures for federal budgeting: proposals and prospects. *Public Budgeting and Finance*, Winter: 3-17.
- Kelly, J. M. and Riverbark, W. C. (2003). *Performance budgeting for state and local government*. New York, London: M. E. Sharpe.
- Lagerlöf, Johan N. M. (2007). A theory of rent-seeking with informational foundations. *Economics of Governance* 8: 197-218.
- Lockwood, B. and Porcelli, F. (2011). Incentives schemes for local government: theory and evidence from comprehensive performance assessment in England. CESifo Working Paper No. 3483.
- Lorz, O. (2001). On the effects of capital mobility on local infrastructure policy and rent-seeking. *Regional Science and Urban Economics* 31: 319–337.

- Li-Wen, L. (2010). Corporate social responsibility in China: window dressing or structural change? *Berkeley Journal of International Law* 28: 64–100.
- McChesney, F. S. (1997). *Money for nothing: politicians, rent extraction and political extortion*, Cambridge and London: Harvard University Press.
- McCubbins, M. D., Noll, R. G., and Weingast, B. R. (1987). Administrative procedures as instruments of political control. *Journal of Law, Economics, and Organization* 3: 243–77.
- Michaels, R. (1988). The design of rent-seeking competitions. *Public Choice* 56: 17-29.
- Milgrom, P. (1988). Employment contracts, influence activities, and efficient organization design. *Journal of Political Economy* 96: 42-60.
- Milgrom, P. and Roberts, J. (1988). An economic approach to influence activities in organizations. *American Journal of Sociology* 94 (suppl.): S154-S179.
- Moe, T. M. (1997). The positive theory of public bureaucracy. in Mueller (ed): 455-480.
- Moe, T. M. (2006). Political control and the power of the agent. *Journal of Law, Economics, and Organization* 22: 1–29.
- Morey, M. R. and O'Neal, E. S. (2006). Window dressing in bond mutual funds. *Journal of Financial Research* 29: 325–347.
- Moynihan, D. P. (2005). What do we talk about when we talk about performance? Dialogue theory and performance budgeting. *Journal of Public Administration Research and Theory* 16(2): 151-168.
- Mueller, D. C., (ed) (1997). *Perspectives on public choice. A handbook*. Cambridge: Cambridge Univ. Press.
- Niskanen, W. A. (1971). *Bureaucracy and representative government*. Chicago: Aldine-Atherton.
- Oates, Wallace E. (1999). An essay on fiscal federalism. *Journal of Economic Literature* 37: 1120-1149.
- OECD (2007). *Performance budgeting in OECD countries*. OECD Publishing.
- Orr, D., Jaeger, M. and Schwarzenberger, A. (2007). Performance-based funding as an instrument of competition in German higher education. *Journal of Higher Education Policy and Management* 29: 3-23.

- Paldam, M. (2004). Are vote and popularity functions economically correct? In Rowley, C. K. and F. Schneider (eds), *The Encyclopedia of Public Choice*, Vol. 1: 49-59. Dordrecht: Kluwer Academic Publishers.
- Page, S. (2005). What's new about the New Public Management? Administrative change in human services. *Public Administration Review* 65: 713-27.
- Prendergast, C. (1999). The provision of Incentives in Firms. *Journal of Economic Literature* 37(1): 7-63.
- Robinson, M. (ed.) (2007a). *Performance budgeting*. New York. Palgrave Macmillan.
- Robinson, M. (2007b). Result information. In: Robinson (2007a): 26-45.
- Robinson, M. (2007c). Making performance budgeting work. In: Robinson (2007a): 113-127.
- Schultze, C. (1974). Sorting out the social grant programs: An economist's criteria. *American Economic Review* 64, 2: 181-189.
- Shin, J. C. (2010). Impacts of performance-based accountability on institutional performance in the US. *Higher Education* 60: 47-68.
- Stickney, C. P. (1975). Window dressing the interim-earnings report: An empirical assessment for firms initially going public. *Journal of Business* 48: 87 – 97.
- Stein, William E. (2002). Asymmetric rent-seeking with more than two contestants. *Public Choice*, 113(3): 325–36.
- Tullock, G. (1980). The costs of transfers. in: Buchanan, J. M. et al. (eds). *Towards a theory of the rent-seeking society*. College station: Texas A&M University Press: 269-282.
- Wilkesmann, U. and Schmid, C. (2011). The impacts of new governance on teaching at German universities. Findings from a national survey. *Higher Education* 61 (forthc.).
- Wintrobe, R. (1997). Modern bureaucratic theory. In: Mueller (ed.): 373-390.
- Worthington, A. C. and Dollery, B. E. (1998). The political determination of intergovernmental grants in Australia. *Public Choice* 94: 299-315.

Appendix A: Proof that the critical threshold ($f^{X,t}$) exceeds 0.5

Solving the contest for the output based performance signal under the assumption that the random terms are perfectly uncorrelated, we can derive two possible solutions for λ :

$$\lambda_{+/-}^{\pi X} = \frac{fF(N-1)}{2N^2} + \frac{(1-f)F}{2N} \pm \sqrt{\left(\frac{fF(N-1)}{2N^2} + \frac{(1-f)F}{2N}\right)^2 - r \frac{f(1-f)F^2(N-1)}{N^3}}$$

Only the negative square root solution (hereafter: $\lambda^{\pi X}$) is valid because the positive square root solution satisfies the upper limit for λ only for extremely small values of f (if at all).

From the FOC of the units' optimization problem it follows that:

$$\alpha^{\pi X} = \left[\frac{F}{Naz} + \frac{fF}{az} \frac{N-1}{N^2} - \frac{1}{az} \lambda \right]^{\frac{1}{z-1}}$$

$$\frac{\partial \alpha^{\pi X}}{\partial f} = \frac{1}{z-1} \frac{\left[\frac{F}{Naz} + \frac{fF}{az} \frac{N-1}{N^2} - \frac{1}{az} \lambda \right]^{\frac{1}{z-1}}}{\left[\frac{F}{Naz} + \frac{fF}{az} \frac{N-1}{N^2} - \frac{1}{az} \lambda \right]} \left(\frac{F}{az} \frac{N-1}{N^2} - \frac{1}{az} \frac{\partial \lambda}{\partial f} \right)$$

$$\frac{\partial \alpha^{\pi X}}{\partial \lambda} = \frac{1}{z-1} \frac{\left[\frac{F}{Naz} + \frac{fF}{az} \frac{N-1}{N^2} - \frac{1}{az} \lambda \right]^{\frac{1}{z-1}}}{\left[\frac{F}{Naz} + \frac{fF}{az} \frac{N-1}{N^2} - \frac{1}{az} \lambda \right]} \left(-\frac{1}{az} \right)$$

Taking f to its limits we can get $\alpha^{\pi X}|_{f=0} = \left[\frac{F}{Naz} \right]^{\frac{1}{z-1}}$ and $\alpha^{\pi X}|_{f=1} = \left[2 \frac{F}{Naz} - \frac{F}{N^2 az} \right]^{\frac{1}{z-1}}$.

In the permitted interval $f \in [0;1]$ the graph of $\lambda^{\pi X}$ is inversely U-shaped and has exactly two roots at $f_1^0 = 0$ and $f_2^0 = 1$. The first derivative of $\lambda^{\pi X}$ with respect to f reads:

$$\frac{\partial \lambda^{\pi X}}{\partial f} = \frac{F}{2N} \left[-\frac{1}{N} - 0.5 \frac{1}{\sqrt{A}} \left(2f \left(\frac{N-1}{N} \right)^2 - 2(1-f) + (2-4f-4r+8rf) \frac{N-1}{N} \right) \right]$$

$$\text{where } A \equiv f^2 \frac{(N-1)^2}{N^2} + (1-f)^2 + 2f(1-f) \frac{N-1}{N} - 4rf(1-f) \frac{N-1}{N}$$

Evaluating expression $\lambda^{\pi X}$ at $f=0$ and $f=1$ yields:

$$\left. \frac{\partial \lambda^{\pi X}}{\partial f} \right|_{f=0} = r \frac{F}{N} \frac{N-1}{N} > 0 \quad \left. \frac{\partial \lambda^{\pi X}}{\partial f} \right|_{f=1} = -r \frac{F}{N} < 0$$

Evaluating expression $\lambda^{\pi X}$ at $f=0.5$ yields:

$$\left. \frac{\partial \lambda^{\pi X}}{\partial f} \right|_{f=0.5} = \frac{F}{2N} \left[-\frac{1}{N} + 0.5 \frac{1}{\sqrt{A|_{f=0.5}}} \left(1 - \left(\frac{N-1}{N} \right)^2 \right) \right]$$

$$\text{where } A|_{f=0.5} = \frac{1}{4} \frac{(N-1)^2}{N^2} + \frac{1}{4} + \left(\frac{1}{2} - r \right) \frac{N-1}{N} = \frac{4N^2 - 4N + 1}{4N^2} - r \frac{N-1}{N}$$

We show that $\left. \frac{\partial \lambda^{\pi X}}{\partial f} \right|_{f=0.5} > 0$ by proving that $\left. \frac{\partial \lambda^{\pi X}}{\partial f} \right|_{f=0.5} \leq 0$ is impossible for all relevant

parameter constellations:

$$\begin{aligned} & \frac{F}{2N} \left[-\frac{1}{N} + 0.5 \frac{1}{\sqrt{\frac{4N^2 - 4N + 1}{4N^2} - r \frac{N-1}{N}}} \left(1 - \left(\frac{N-1}{N} \right)^2 \right) \right] \leq 0 \\ & \Leftrightarrow \left(\frac{4N^2 - 4N + 1}{4N^2} \right) \leq \left(\frac{4N^2 - 4N + 1}{4N^2} - r \frac{N-1}{N} \right) \Leftrightarrow r \leq 0 \end{aligned}$$

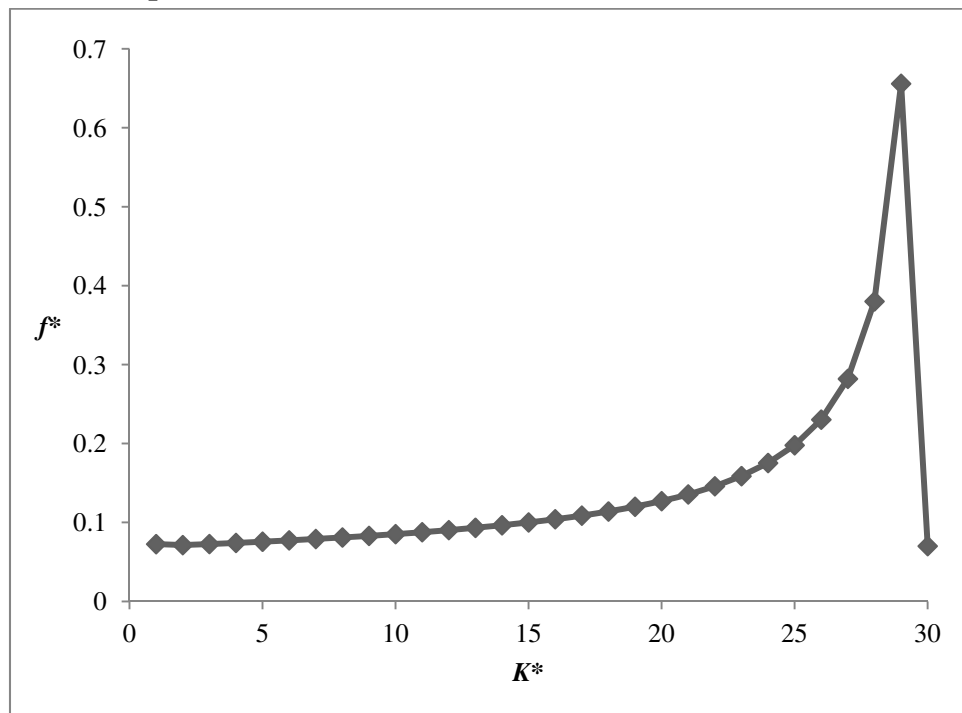
This result contradicts the assumption of the positive range of $0 < r \leq 1$. The only valid case would be a perfect performance signal ($r = 0$), which is irrelevant for our analysis. We may

conclude that $\left. \frac{\partial \lambda^{\pi X}}{\partial f} \right|_{f=0.5} > 0$. Thus, the value of f that maximizes λ lies between 0.5 and 1.

Table 1: Game structure

Stage	Activity	Information
1	Policy-makers sets f and K and allocate the block fund to the units.	f, K
2	The units choose λ_i (expenditures for window-dressing) and α_i (effort level) given f and K and start the production of X using $b_i - \lambda_i$ and α_i . Policy-makers observe the performance signal π_i	π_i
3	Policy-makers observe distribute the performance-based funds according to the PB-scheme.	
4	The units receiving additional funds use these to expand production.	WF, X_i

Figure 1: Fin-shaped roof ridge of output-maximizing performance prize and performance-bonus schemes in the baseline model



The exogenous parameter values used in this figure are: $N = 30$, $F = 15$, $z = 1.9$, $a = 1.7$, $r = 0.34$.

Figure 2: Critical combinations of r and z under the effort-based performance signal

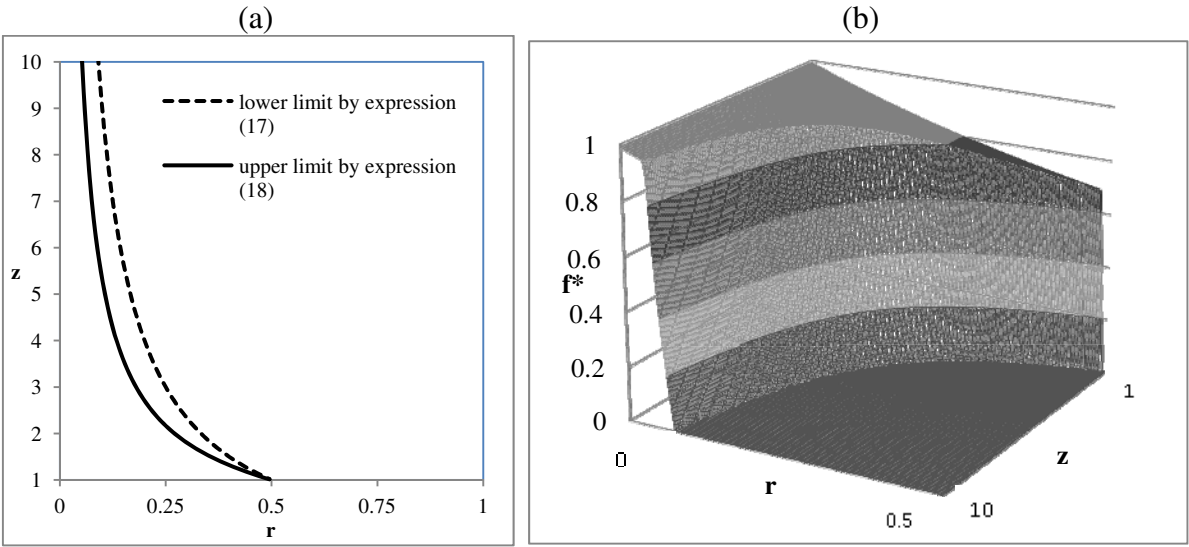
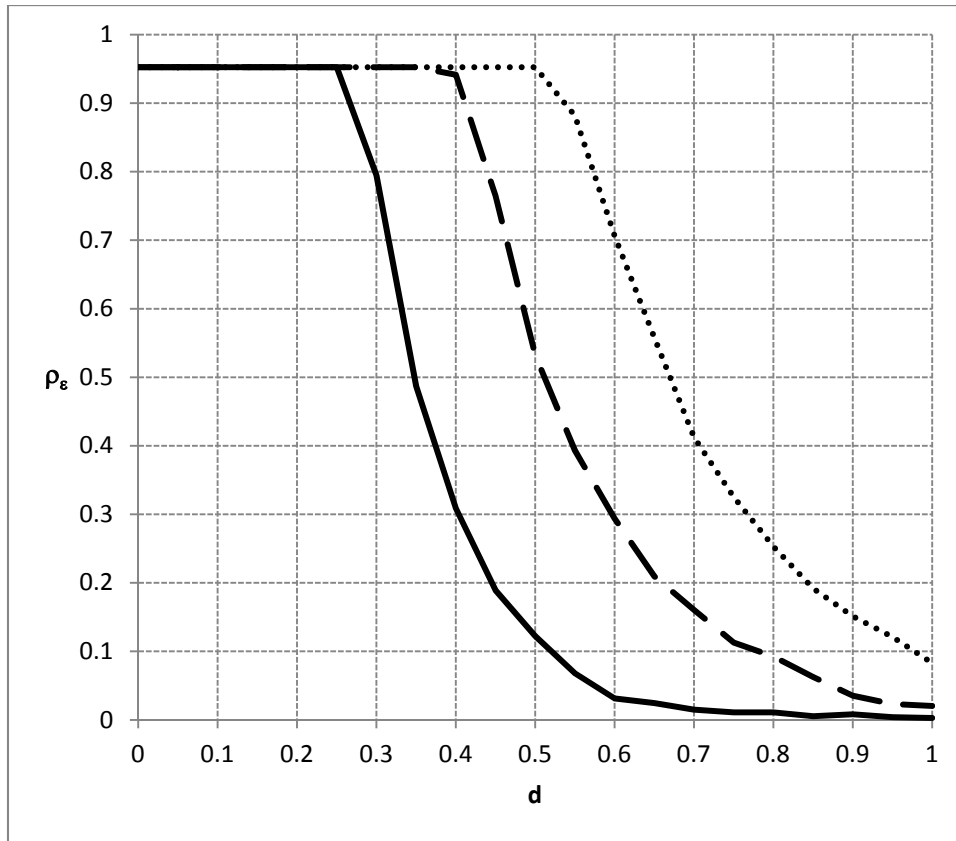


Figure 3: Effort-based versus output-based performance signal under a Rawlsian welfare function: critical values for d and ρ_ε



For all combinations (d, ρ_ε) on the left of resp. below the lines, the output-based signal is superior to effort-based model. The opposite is true for all combinations to the right resp. above the lines. The exogenous parameter values used in this figure are: $N = 32$, $F = 15$, $z = 1.7$, $a = 0.55$, $r = 0.1$ (solid line), $r = 0.15$ (dashed line) resp. $r = 0.2$ (dotted line).