Screening Budgets

Alexander F. Wagner, Nolan Miller, and Richard J. Zeckhauser*

Harvard University

January 23, 2004

Abstract

How should an organization's center allocate resources units under its control which are better informed? Even with conscientious productivity reviews, important information will remain asymmetrically held. If units value their own expenditures more than those of their peers, they will seek excess budgets and expenditures. Fortunately, budget authorities can infer productivities from units' expenditure patterns across spending categories and over time. Optimal screening budgets reward more productive units with greater overall budgets. Such screening provides significant welfare gains over traditional fixed or reallocable budgets. Empirical results for a large electricity and infrastructure provider fit an important version of the model.

JEL-Code: D23, D82, M41.

Keywords: Budgeting, screening, asymmetric information

*Wagner is a Ph.D. candidate in Political Economy and Government at Harvard University. Miller is an Assistant Professor of Public Policy at the John F. Kennedy School of Government, Harvard University. Zeckhauser is the Frank Plumpton Ramsey Professor of Political Economy at the John F. Kennedy School of Government, Harvard University. We thank Philippe Aghion, Alberto Alesina, George Baker, Oliver Hart, Liz Keating, and Martin Weitzman for comments. Wagner would like to thank the Austrian Academy of Sciences for support through the APART fellowship.

Screening Budgets

1 Introduction

Information is the lifeblood of efficient governance in organizations. As Hayek (1945) noted, the central authority of any organization must delegate decisions to individual units because it lacks information about the productivity and skills of units at lower levels of the hierarchy, such as divisions in a private company or bureaus within a government agency. The central authority wants to make sure that productive units are given what they need to take advantage of the opportunities before them, but it also wants to prevent non-productive units from wasting the organization's scarce resources. Because the central authority can not fully observe the productivity of its units, these two goals inevitably conflict. Unless it has some way to sort the non-productive units from the productive ones, the center must treat all units equally; productive units receive too few resources and non-productive ones too many. This misallocation of resources is costly to the organization, to society as a whole, and even to each unit ex ante, i.e., starting at a time before they get private information about their productivities.

The center's task of identifying which units are productive is complicated by the fact that all units, regardless of their type, prefer to spend more rather than less. Thus, the center can not merely offer additional funds to any unit that wants them, since all units will jump at the opportunity. Our model posits the extreme case where the center never learns units' productivities, or learns them too late to affect any budget decisions.

This paper explores the idea that while all units prefer more spending to less, after securing private information some units wish to tilt actual expenditures relative to budget toward materials, while other units may prefer a tilt toward personnel. If this inter-category preference is related to overall productivity (e.g., highly productive units prefer materials spending), then the center can give more funds to productive units by doing it in a way that is relatively more appealing to productive units than to non-productive units. It will give a greater total budget to those who overspend on materials.

In this paper, we derive the way a central budget authority should make optimal use of differences in units' inter-category preferences. Our basic model considers a firm whose units use two inputs, materials and personnel. All units glean the same productivity from personnel spending, but the productivity of materials expenditures may be either high or low. In this case, we show that, in response to the center's optimal (i.e., profit-maximizing) scheme, high-productivity units spend less on personnel than on materials, they spend more on materials than low-productivity units, and they spend more in total. Moreover, expenditures on materials are more variable than those for personnel. We also discuss how a very similar model can be applied when some units prefer to spend early while others prefer to spend late, allowing the center to exploit information contained in the intertemporal pattern of expenditures. In both cases, the posited difference in the units' attitudes toward inter-category or intertemporal consumption patterns satisfies the "single-crossing property," which permits the central authority to separate the units' types through use of a screening contract.

We then test the predictions of the model, using a unique dataset of budgets and expenditures data for a large infrastructure and energy provider. We find support for the hypothesis that there is a systematic relationship between the pattern of spending on the one hand and the degree to which units are allowed to exceed their allocated budgets on the other. In particular, units that overspend on materials (instead of personnel) receive greater funds in the current year. These units also receive larger budgets in the following year. Consistent with the theory, we also find that materials expenditures are more variable than expenditures on personnel. Additionally, our results indicate that early spenders enjoy greater budget growth than late spenders, a pattern consistent with a plausible model where initial information is asymmetric and productivity is mean reverting.

Though use of spending data in the manner described by our model improves the center's welfare, it can only partially eliminate the consequences of information asymmetry. The deviation of the resulting second-best allocation from the first-best allocation reflects the units' wily adverse selection behavior. Adverse selection plays a major role in many economic settings. The classic reference is Akerlof (1970). Salanié (1997) provides an excellent survey of the basic methodology and applications. Recently, the adverse selection approach has been used more extensively to understand life in organizations (Milgrom and Roberts 1992). Several papers deal with the implications of incentive problems for resource allocation in organizations (Harris and Townsend 1981, Harris, Kriebel and Raviv 1982, Antle and Eppen 1985, Brown, Thornton, Buede and Miller 1992, Graziano 1995, Harris and Raviv 1996, Harris and Raviv 1998). However, these papers generally focus more on project choice instead of the optimal design of operations budgets. (An exception is Spencer (1982).)

Although the sorting procedure applied in this paper is well established, we believe that our approach – employing different revealed inter-category and intertemporal preferences of units to design optimal budgets – is new. The paper also innovates by using real company data to test and confirm the predictions of the model.

The paper is organized as follows. Section 2 describes the basic model and characterizes the optimal resource allocations under full and asymmetric information when personnel productivity is known and units have private information about materials productivity. Section 3 discusses implications for organizations' budgeting processes. Section 4 shows how the logic of screening budgets can be applied in more general settings when personnel productivities differ. It also illustrates how a very similar model can be used to exploit units' different intertemporal preferences. Section 5 reports empirical results from a real-world organization. Section 6 mentions a series of extensions – formal treatments are given in an extended version of this paper – and draws conclusions.

2 Screening via inter-category preferences

2.1 Basic Model

A principal employs an agent who produces output using two inputs, personnel and materials, using a Cobb-Douglas production function.[1] There are two types of units, where a unit's type corresponds to a specification of the parameters of its production function.[2] We denote

[1] For clarity, we refer to the principal as male and the agent as female. We will also use "principal" interchangeably with "center" and "agent" interchangeably with "unit."

[2] Results for the extension to N types are available on request.

the types as "High" and "Low." The unit's type is private information, but the center knows the prior likelihood of a unit's being High or Low. The probability that the agent is High is λ_H, where $0 < \lambda_H < 1$, while $\lambda_L = 1 - \lambda_H$ denotes the probability that the agent is Low.

Let $\theta_{iP} > 0$ and $\theta_{iM} > 0$ denote an agent i's agent's productivity with personnel and materials, respectively, where $i \in \{H, L\}$. For the production function, assumed common to both agents, we employ a Cobb-Douglas form. While the basic logic of screening budgets applies for other functions as well (shown below for intertemporal screening), we adopt the Cobb-Douglas form because it is widely used and convenient.

The production of a type i unit that spends x_P on personnel and x_M on materials is:

$$f_i(x_P, x_M) = K x_P^{\theta_{iP}} x_M^{\theta_{iM}},$$

where K is a constant.[3] We assume that there are decreasing returns to scale, i.e., $\theta_{iP} + \theta_{iM} < 1$ for both types.

The units care about their own production, but not about the costs to the center of funds expended to provide the units' inputs.[4] For example, units may have an empire-building impulse. The Cobb-Douglas production function is strictly increasing in each of its inputs. Thus, as discussed in the introduction, we have a case where all units have the same intra-category preferences: more is preferred to less. However, when $\frac{\theta_{HP}}{\theta_{HM}} \neq \frac{\theta_{LP}}{\theta L}$, the different types of units have different attitudes toward inter-category trade-offs. For our base case, we assume that the High and Low types do not differ in personnel productivity

[3] We choose K large enough to make each component of each type's allocation greater than or equal to one. Specifically, it is necessary that $K > \max_{i,j,j'} \left[1/\left(\theta_{ij}^{\theta_{ij}} \theta_{ij'}^{1-\theta ij}\right)\right]$. This is sufficient to ensure that, over the input range relevant for our problem, a higher θ value for a factor correspond to higher marginal productivity for that factor. It is a peculiarity of the Cobb-Douglas technology that, when input levels are less than one, this need not be the case.

[4] When there is sufficiently strong (and not necessarily perfect) alignment between the center's and the units' preferences about the alternative use of funds, the center may be able to approximate or even achieve the first-best allocation. More detailed results are available on request from the authors.

but do differ in their materials productivity.[5] Thus, we assume:

$$\theta_{HM} \geq \theta_{HP},\ \theta_{LP} \geq \theta_{LM},\ \theta_{HP} = \theta_{LP}. \tag{1}$$

That is, Highs expect that they will be more productive with materials than personnel, while Lows expect that they will be less productive. This case allows us to capture the basic intuitions of the framework employed in this paper. As we show below, these intuitions extend to other cases.

One way to view this assumption is to realize that it implies that the productivity parameter for materials is more variable than that of personnel. In terms of first principles, the underlying notion is that there are shocks to productivity, and that these shocks are more severe for materials. The values of the shocks are known to the units. Units with positive shocks are High types, those with negative shocks Low types.

Expression (1) implies the following inequality, which embodies the idea that the agents have different attitudes toward inter-category spending:

$$\frac{\theta_{HP}}{\theta_{HM}} \leq \frac{\theta_{LP}}{\theta_{LM}}.$$

This expression implies that, for any given allocation, High's marginal rate of technical substitution of materials expenditure for personnel expenditure is greater in absolute value than Low's; i.e., High has a stronger preference for materials expenditure than does Low.[6] Thus, with x_M plotted on the horizontal axis and x_P on the vertical, High's production isoquants are steeper than Low's.

We assume that the risk-neutral principal incurs an opportunity cost for funds $C(x_M + x_P)$ if the agent spends x_M and x_P on materials and personnel, respectively. Although the results generalize to the case where $C()$ is any increasing and convex function, for simplicity we let $C() = x_M + x_P$. The principal's objective is to maximize profit: expected production less the cost of inputs.

[5] Alternatively, the principal may be able to perfectly measure personnel productivity. Since the principal can then condition any contract on the known personnel productivity, we can assume without loss of generality that $\theta_{HP} = \theta_{LP}$.

[6] To see this, note that $MRS_i = (\partial f_i / \partial x_M) / (\partial f_i / \partial x_P) = (\theta_{iM} / \theta_{iP}) (x_P / x_M)$. Since $\theta_{HP} = \theta_{LP}$, we have that High's isoquant is steeper at each allocation.

2.2 The Full-Information Benchmark

If the principal could observe the units' types, his problem would be straightforward and a first-best outcome would be reached. The principal would simply choose x_{iM} and x_{iP} to maximize:

$$Kx_{iP}^{\theta_{iP}}x_{iM}^{\theta_{iM}} - x_{iP} - x_{iM}.$$

Thus, letting x_{iM}^{FB} and x_{iP}^{FB} denote a type-i agent's expenditure on materials and personnel, respectively, for the first-best (FB) outcome, the principal should set:

$$K\theta_{iM}\left(x_{iM}^{FB}\right)^{\theta_{iM}-1}\left(x_{iP}^{FB}\right)^{\theta_{iP}} = K\theta_{iP}\left(x_{iM}^{FB}\right)^{\theta_{iM}}\left(x_{iP}^{FB}\right)^{\theta_{iP}-1} = 1.$$

Since higher θ values correspond to higher marginal products (when K is chosen as specified above), the principal would like to give High more for both materials and personnel (because through the multiplicative form personnel expenditures indirectly count more for High even though $\theta_{HP} = \theta_{LP}$). Of course, if the principal can not observe the agent's type, he can not implement this result. Given the choice between $\left(x_{HP}^{FB}, x_{HM}^{FB}\right)$ and $\left(x_{LP}^{FB}, x_{LM}^{FB}\right)$, both High and Low would choose $\left(x_{HP}^{FB}, x_{HM}^{FB}\right)$ since it is larger on both components. In other words, if the first-best menu were offered, Low would envy High and would simply claim to be High. Note also that at the solution, the variance of materials spending across types is greater than the variance of personnel spending.

2.3 The Asymmetric Information Case: Screening Budgets

Under asymmetric information, the principal must choose expenditure bundles (x_{HP}^*, x_{HM}^*) and (x_{LP}^*, x_{LM}^*) to maximize expected profit, subject to the constraints that each type of agent selects its own intended bundle. Our analysis makes use of the Revelation Principle and follows the typical approach in the screening literature (see Salanié (1997)).[7] The principal's challenge is thus to find a way to give more resources to High without causing Low to envy High. The agents' different attitudes toward inter-category substitution provide a way to accomplish this.

[7] We concentrate exclusively on non-random schemes.

To see how, begin from a situation where all agents receive the same allocation, evenly split between the two categories. High would prefer to spend \$1 less on personnel and \$1 more on materials. Since Low has a relative preference for personnel spending, Low finds High's new bundle strictly worse than her old bundle. Thus, after shifting High's spending towards materials, the principal could also give High slightly more to spend without violating Low's incentive-compatibility constraint. The principal also benefits from shifting Low's consumption towards personnel. Holding total expenditures constant, shifting \$1 intended for Low from materials to personnel makes this bundle more attractive for Low, which allows the principal to take up the slack this creates in Low's incentive-compatibility constraint by further increasing High's total budget.

Formally, the principal solves:

$$\max_{\{x_{iM},x_{iP}\}_i} \sum_{i=H,L} \lambda_i \left[K x_{iM}^{\theta_{iM}} x_{iP}^{\theta_{iP}} - (x_{iM} + x_{iP}) \right] \tag{2}$$

subject to

$$K x_{HM}^{\theta_{HM}} x_{HP}^{\theta_{HP}} \geq K x_{LM}^{\theta_{HM}} x_{LP}^{\theta_{HP}}, \text{ and} \tag{3}$$

$$K x_{LM}^{\theta_{LM}} x_{LP}^{\theta_{LP}} \geq K x_{HM}^{\theta_{LM}} x_{HP}^{\theta_{LP}}.$$

Thus, the principal maximizes expected profit under the constraints that High must not envy Low and that Low must not envy High. We can characterize the solution, given our assumptions. Denote it as (x_{HP}^*, x_{HM}^*) and (x_{LP}^*, x_{LM}^*).

The first step in the derivation, the details of which are provided as Lemma 1 in the Appendix, is to realize that the incentive-compatibility constraint for Low must bind at the optimum. The basic idea of the proof is as follows. Since the first-best allocation is not incentive compatible and agent types have different preferences, exactly one incentive-compatibility constraint must bind. If High's incentive-compatibility constraint (IC) is binding, then the allocations given to High and Low must maximize their production conditional on their total expenditure. (Otherwise, moves along isocost curves allow increased production at identical costs without violations of ICs.) But in that case, the principal spends more on Low than on High. This can not be optimal, for the principal could always increase production by shifting some money from Low to High. Thus, Low's incentive-compatibility constraint must be the one that binds. If Low's IC is binding, High will not be on her Engel curve, i.e. she will not maximize her production conditional on her total expenditures. The question then is where the optimal allocations will be located. The following six properties characterize the solution to the principal's problem under asymmetric information.

Proposition 1 ***(Screening budget allocation)*** *Suppose (1) holds. Then, the problem (2)-(3) leads to the following allocations:*

1. $x^*_{HM} \geq x^*_{HP}$.

2. $x^*_{LP} \geq x^*_{LM}$.

3. $x^*_{LP} \geq x^*_{HP}$ *and* $x^*_{HM} \geq x^*_{LM}$.

4. $x^*_{HM} \leq x^{FB}_{HM}, x^*_{HP} \leq x^{FB}_{HP}$; $x^*_{LM} \geq x^{FB}_{LM}, x^*_{LP} \geq x^{FB}_{LP}$.

5. Given $x^*_{LP} + x^*_{LM}$, $\{x^*_{LP}, x^*_{LM}\}$ *maximizes Low's production.*

6. $x^*_{HP} + x^*_{HM} \geq x^*_{LP} + x^*_{LM}$.

Proof. See the Appendix. ■

This solution is illustrated in Figure 1.

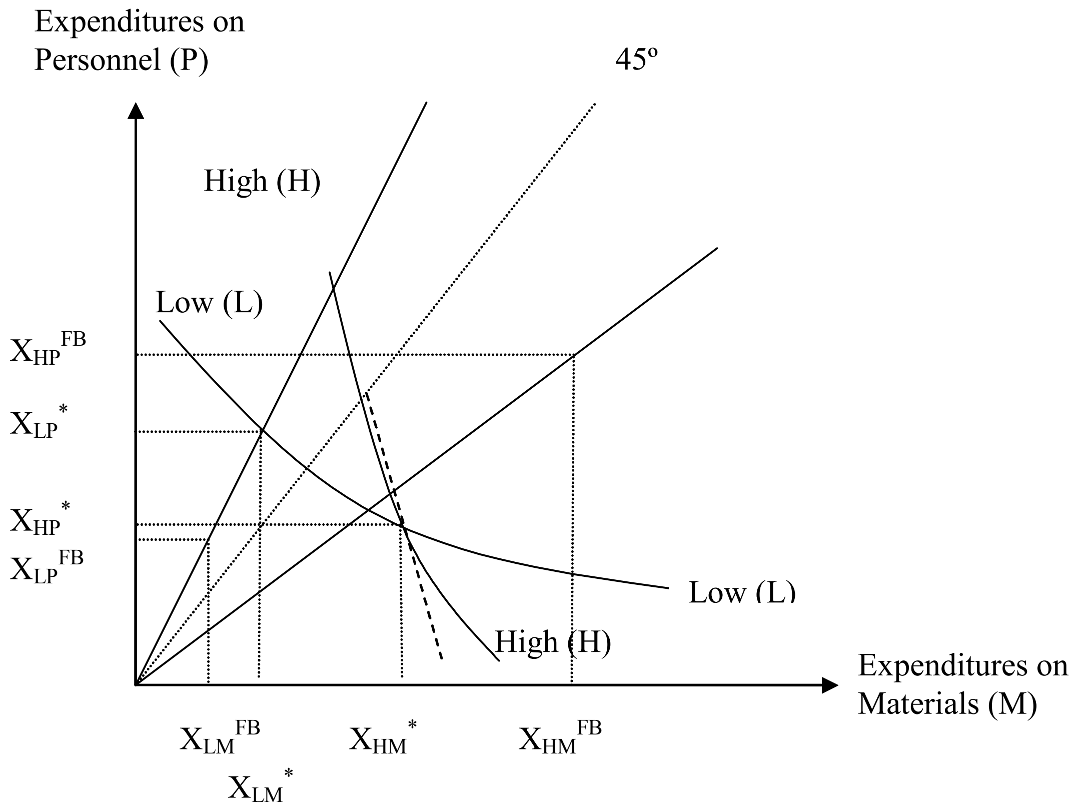

Figure 1: First-best and second-best solutions of the 2-category budget problem (Cobb-Douglas production). X_{ij} denotes type i's spending on category j, where $i \in \{L, H\}$ and $j \in \{M, P\}$. FB denotes "first-best"; * denotes "second-best." The dashed line through (x^*_{HM}, x^*_{HP}) is an isocost-line.

Starting from the first-best, the principal increases Low's production (to make her directly better off and less envious) and distorts High's allocation (to make it less attractive for Low to envy High). At the solution, High will not envy Low, and the principal will allow Low to pick her optimal bundle, located somewhere along her Engel curve. This is the principle of "no distortion at the bottom."[8] To maximize welfare subject to Low's IC, the principal induces High to spend more in the category that is relatively productive for High and relatively unproductive for Low: materials. Moreover, by moving somewhat to the right from High's

[8]Note that for the Cobb-Douglas case, property 5 can also be stated more specifically as $\{x^*_{LP}, x^*_{LM}\} = k * \{x^{FB}_{LP}, x^{FB}_{LM}\}$.

Engel curve, the principal can give High more funds. While additional funds increase High's production, they are now unfortunately spent in a distorted and inefficient way. In choosing High's optimal allocation, the principal thus accepts increased distortion to serve increased production. Note that the properties together imply that the bundle with higher total expenditure has the smaller personnel expenditure; in effect, a player who sacrifices a unit of personnel expenditure gets more than a dollar back for materials.[9]

Finally, note that materials, i.e., the category with the greater variability in the productivity parameter, has the more variable expenditures. This point is summarized by the following three related properties.

Corollary 1 *Suppose (1) holds. Then, the problem (2)-(3) also implies*

1. $Var\left(x_M^*\right) \geq Var\left(x_P^*\right).$
2. $\frac{x_{HM}^* - x_{LM}^*}{x_M} \geq \frac{x_{LP}^* - x_{HP}^*}{x_P}$, *where* $x_M = \lambda x_{HM}^* + (1-\lambda)\, x_{LM}^*$ *and* $x_P = \lambda x_{HP}^* + (1-\lambda)\, x_{LP}^*$.

Proof. See the Appendix. ■

Together, Proposition 1 and Corollary 1 allow us to make a prediction that we will test in section 5. If the priors are as assumed in this section, in evaluating divisions with similar tasks within a company, their materials spending should be more variable than their personnel spending, and the greater rewards for spending on materials should be reflected in greater total budgets for units that spend more on materials.

3 Implications for budgeting

The incentive-compatible screening budget we have proposed is promising for organizations whose units' productivities vary across categories, and that realize that handing out fixed budgets per category would give up important benefits of decentralization, assuming that units have superior information to the center.[10]

[9] That the optimal solution must have High's bundle below High's Engel curve follows from an argument similar to the proof of Lemma 1. For details see the Appendix.

[10] In theory, all problems due to asymmetric information could be avoided if budget examiners could secure information about productivity. However, the large costs of the information acquisition process may not be

Budget rules. We now evaluate the welfare benefits of screening budgets over fixed and reallocable budgets. Under *fixed budgets*, the center requires each unit to spend the same amount, independent of its type, and the unit can not reallocate funds between the two categories. Under *reallocable budgets*, the center gives the same amount of total funds to each unit, but allows units to allocate resources between the two categories, spending more where they are highly productive and less where they are less so. Thus, total budgets are inefficient, but spending within budgets is perfect.

In other words, for reallocable budgets, the principal solves

$$\max_{\{x_{iM},B^r\}_i} \sum_{i=H}^{L} \lambda_i \left[K x_{iM}^{\theta_{iM}} \left(B^r - x_{iM}\right)^{\theta_{iP}} - B^r \right], \tag{4}$$

s.t. $x_{iM} \in \arg\max_{x_{iM}} K x_{iM}^{\theta_{iM}} \left(B^r - x_{iM}\right)^{\theta_{iP}}$ for $i = L, H$. Fixed budgets add the constraint $x_{HM}^* = x_{LM}^*$. We denote the optimal fixed budget with B^f.

For screening budgets, we call $B_i = x_{iM} + x_{iP}$ the budget of agent i, and $B^s = \lambda_H B_H + (1 - \lambda_H) B_L$ the expected screening budget.

Numerical results. The budgets problem must be solved numerically (except for the first-best and for reallocable budgets in special cases). Table 1 shows the results for a scenario where the primary information asymmetry concerns materials productivities. In particular, we assume that High's and Low's personnel productivity is identical, but the sum of the exponents in High's production function is twice that of Low's.[11]

worthwhile. Finally, traditional budget inspection only deals with the excess spenders; nothing is learned about less productive units, which should have their budgets reduced.

[11]All of these results are general as long as the parameter values fulfill the conditions of our main case, asymmetry on materials productivity.

Table 1: Allocations under four different budget regimes

$\theta_{HP} = \theta_{LP} = 0.2, \theta_{HM} = 0.4, \theta_{LM} = 0.1, \lambda = 0.3, K = 8.7$

	Materials		Personnel		Total	Total	Expected	Principal's	Recovery
Budget rule	High	Low	High	Low	High	Low	budget	welfare	rate
First-best	16.000	1.000	8.000	2.000	24.000	3.000	9.299	9.700	100%
Screening	7.258	1.732	1.693	3.465	8.951	5.197	6.323	8.343	36.83%
Reallocable	3.658	1.829	1.829	3.658	5.488	5.488	5.488	7.854	14.01%
Fixed	2.682	2.682	2.559	2.559	5.241	5.241	5.241	7.552	0%

High spends less than Low on personnel, but more on materials. High spends more than Low overall. Low gets more under screening budgets than she gets in the first-best since she must be deterred from selecting High's allocation. Conversely, High gets less under screening budgets than in the first-best. Going from High's to Low's allocation results in loss of 5.526 of materials spending, while the gain in personnel spending is only 1.772. Thus, the slope between High's and Low's bundles is -3.11 and therefore far greater than one in absolute value. By contrast, reallocable budgets employ tradeoff rates of exactly one, which explains why their total budgets are so inefficient.

Note that for both the first-best and screening budgets, materials spending (both absolutely and as a fraction of the expected materials budget) is more variable than personnel spending, as property 1 of Corollary 1 predicted.[12] Property 2 can also be easily verified from the table.

The principal's welfare levels (shown in the next to last column of the table and measured as total production minus expenditures) show that the principal is significantly better off with screening budgets than with either of the two other schemes, which are both widely employed. In general, the extent to which screening budgets outperform reallocable budgets and fixed budgets depends on the degree of differences in inter-category preferences. In the

[12]For reallocable budgets, the variance is the same for the two categories. However, this is a special feature of the two-categories case; for more than two, the category with the most variable productivity parameter also has the most variable expenditures. For fixed budgets, there is of course zero variance across types for all categories.

last column, we calculate the "recovery rate," that is, the percentage of the shortfall in welfare between fixed and first-best budgets that is recovered by screening budgets. For example, in this example, because screening budgets allow the principal to obtain 0.791 (=8.343-7.552) more welfare, while the difference in welfare between first-best and fixed budgets is 1.648 (=9.7-7.552), the recovery rate of screening budgets is 37%. Similar calculations can be made to show similarly significant gains when productivities in both categories differ.[13]

4 Generalizations

4.1 Patterns of productivity

Our base case has agents' productivities differ merely for one category. The principal can exploit differences in inter-category preferences in many other situations. Highs and Lows can be profitably separated whenever their preferences exhibit "single crossing." Having High's indifference curve through any point be flatter (or steeper) than Low's, as it is in our examples, is a sufficient condition for such single crossing. This will be the case for a wide variety of reasonable specifications.[14]

For example, suppose that the principal can not observe personnel productivities. Since he can not condition on that information, θ_{HP} and θ_{LP} will now differ. The most basic extension of the results in section 2 occurs when $\theta_{HM} > \theta_{HP} > \theta_{LP} > \theta_{LM}$. With these parameters, the relative slopes of isoquants will be the same as before; the crossing will just be sharper. Thus, Proposition 1 continues to hold.

Alternatively, suppose that $\theta_{HP} > \theta_{HM} > \theta_{LM} > \theta_{LP}$. That is, High is again absolutely more productive (or expects greater positive shocks) in both categories of inputs. But, High now has a comparative advantage in personnel spending, while Low has relatively

[13] We have allowed each budgeting strategy to use its own optimal budget amount. It is also possible to ask what happens to welfare if we impose a certain (non-optimal) expected budget to be used. For example, we can see how the four budget schemes compare given the optimal fixed budget. (Results for this exercise are available on request.) In both analyses screening budgets facilitate significant welfare gains.

[14] Indeed, this is the key feature in any screening problem. See Milgrom and Shannon (1994).

better shocks in materials than in personnel. Note that this configuration fulfills the dual conditions $\frac{\theta_{HP}}{\theta_{HM}} \geq \frac{\theta_{LP}}{\theta_{LM}}$ and $\theta_{HP} > \theta_{LP}$; it thus allows the principal to separate the types, albeit with a different contract. In particular, High's isoquants would now be flatter than Low's. This means High now wants to spend more on personnel. Optimality thus requires that $x^*_{HP} > x^*_{HM}$, $x^*_{LM} > x^*_{LP}$, and $x^*_{LM} > x^*_{HM}$, $x^*_{HP} > x^*_{LP}$. In this case, the optimal tradeoff rate is less than unity. Properties 4 to 6 remain unchanged.

What if $\theta_{HP} > \theta_{LM} > \theta_{HM} > \theta_{LP}$? Such a situation arises when High's personnel productivity is much higher than Low's, but Low is absolutely more productive in materials. For example, although High hires particularly skilled workers, Low has fewer problems with machine maintenance. The results just stated still hold for this situation. The crucial feature is that High maintains a preference for personnel spending or, equivalently, that Low is comparatively more productive in materials.

By contrast, screening is not feasible when both types' materials productivities are constant multiples of their personnel productivities, i.e., when $\theta_{HM} = \alpha\theta_{HP}$ and $\theta_{LM} = \alpha\theta_{LP}$. In this case, at every allocation the High and Low types have the same marginal rate of technical substitution. Because of this, it is impossible to design two bundles such that High prefers one and Low prefers the other. One way to view these results is that, given our assumed organizational structure, the principal needs the units to have idiosyncratic shocks to permit inter-category screening.

4.2 Intertemporal screening (dynamic budgets)

Screening budgets can be employed to distill information out of spending data whenever there are two (or more) disparate categories for expenditure. An important example arises when units spend over periods of time, such as this year and next or even early and late in a budget period. For inferences to be drawn successfully, the units must have different relative productivities over time, implying different intertemporal preferences. Assume that there are two time periods, but for ease of exposition merely one category of expenditure. (Finer breakdowns would allow more information to be distilled.) For example, suppose that the center can detect current productivity, but units have private knowledge about their expected future productivity; this is similar to the first case discussed for the inter-category screening scenario. For example, if the unit is a sales department, the principal can more easily acquire information about current or near-term sales prospects than about those expected in the second half of the year. However, the units may have important private information about the latter. The High type here has high second-period expectations, while the Low type has low second-period expectations. The different attitudes toward intertemporal consumption again produce our single-crossing requirement.[15]

Formally, suppose that production is additively separable across time. Per-period production is given by an increasing, concave, and twice differentiable function $f()$, multiplied with the productivity parameter θ_{it}, where $t = 1, 2$ indexes time, e.g., first and second half of the year. Here, the principal solves

$$\max_{\{x_{i1}, x_{i2}\}_i} \sum_{i=H,L} \lambda_i \left[\theta_{i1} f(x_{i1}) + \theta_{i2} f(x_{i2}) - C(x_{i1} + x_{i2})\right] \tag{5}$$

subject to

$$\theta_{H1} f(x_{H1}) + \theta_{H2} f(x_{H2}) \geq \theta_{H1} f(x_{L1}) + \theta_{H2} f(x_{L2}), \text{ and} \tag{6}$$
$$\theta_{L1} f(x_{L1}) + \theta_{L2} f(x_{L2}) \geq \theta_{L1} f(x_{H1}) + \theta_{L2} f(x_{H2}).$$

Thus, the principal again offers an optimal contract, subject to the incentive-compatibility constraints. To mirror the analysis above, suppose that $\theta_{H2} \geq \theta_{H1} = \theta_{L1} \geq \theta_{L2}$ and thus $(\theta_{H1}/\theta_{H2}) \leq (\theta_{L1}/\theta_{L2})$. This reflects the notion that future shocks to productivity are expected to be beneficial for High, but adverse for Low. Thus, High prefers to spend late, while Low prefers to spend early. Note that the incentive-compatibility constraints take the same form as before.[16] Again, therefore, Low's incentive-compatibility constraint will be

[15] For simplicity, we assume that the agent and the principal do not discount.

[16] Indeed, due to the multiplicative separability of the Cobb-Douglas function, if we had taken the logarithm of both sides of the incentive-compatibility constraints in the inter-category screening problem, the resulting expressions would look very much like the present ones, with $f(x) = \ln(x)$.

binding. The following corollary restates Proposition 1 in this setting.[17]

Corollary 2 ***(Dynamic budget allocation)*** *Suppose that $\theta_{H2} \geq \theta_{H1} = \theta_{L1} \geq \theta_{L2}$. Then, the problem (5)-(6) leads to the following allocations:*

1. $x^*_{H2} \geq x^*_{H1}$.
2. $x^*_{L1} \geq x^*_{L2}$.
3. $x^*_{L1} \geq x^*_{H1}$ *and* $x^*_{H2} \geq x^*_{L2}$.
4. $x^*_{H2} \leq x^{FB}_{H2}, x^*_{H1} \leq x^{FB}_{H1}$; $x^*_{L2} \geq x^{FB}_{L2}, x^*_{L1} \geq x^{FB}_{L1}$.
5. *Given* $x^*_{L1} + x^*_{L2}$, $\{x^*_{L1}, x^*_{L2}\}$ *maximizes Low's production.*
6. $x^*_{H1} + x^*_{H2} \geq x^*_{L1} + x^*_{L2}$.

We call this type of intertemporal screening *dynamic budgets*. The approximate real-world correspondence to the optimal contract of equations (4)-(5) consists in the center setting a budget, and then adjusting it after first-period expenditure has been observed. The way the budget will be adjusted is announced in advance. The principal would like to give High more, but is constrained by Low's envy. To achieve the constrained optimum, he distorts High's allocation and gives Low more in both categories than in the first-best. The second-best optimum trades off these two types of distortion.

Again, alternative distributions of productivity will produce different optimal contracts. For example, if productivity shocks have both a permanent and temporary component, there will be mean reversion, with High's expected second-period productivity below her first-period productivity, but still above the norm, and the converse for Low. In that case, High should optimally consume more in the first period than in the second, more in the first period than Low in the first period, and more than Low overall. The general conditions needed for the mechanism to work discussed earlier would still hold.[18] If firms are using dynamic budgets, this should produce observable patterns relating "overspending" in first-

[17] For a more explicit proof of this corollary see the extended version of the paper. We assume commitment to this contract and do not deal with issues of renegotiation discussed in the literature following Hart and Tirole (1988), Dewatripont (1989), and Ma (1991), among others.

[18] In separate work we are exploring separation possibilities when productivities stay constant from period to period, but High and Low have different shapes to their total productivities curves within a period.

and second-period budgets to the total budget received. Thus, in the case addressed in Corollary 2, units that overspend more in the second period will get higher overall budget allocations.

5 Empirical results for a real-world organization

5.1 Introduction and data description

For many firms, operations expenditures make up the bulk of spending. Important as they are, investment expenditures often are only a relatively small portion of total spending which is decided anew each year. For example, even though the company studied in this paper is in the infrastructure business, a capital-intensive business, about 80% of all non-fixed expenditures come from spending on operations. We focus on non-fixed expenditures to draw inferences about how the company spends resources.

Most studies of budgeting in organizations are either purely theoretical, or have relied on surveys about behavior, i.e., stated preference methods.[19] In contrast, we use data of actual budgets and expenditures of units in a major company to test the predictions of the model directly. To do this, we have acquired a detailed dataset on budgets and expenditures for a large European company. We will refer to this firm, one of the largest electricity providers in its country, as "the Company." We summarize the most relevant information for this study. More detailed information and additional empirical results appear in Wagner (2003).

We wished to analyze important units with relatively homogeneous functions. For example, we did not compare distribution units to power-generation units. The Company's five service centers - which employ roughly 30% of the Company's employees and are responsible for 25% of its operating budget - provide a splendid example. They serve the five regions

[19] Budgetary slack has been the greatest concern of scholars in the literature in accounting. Studies like Young (1985) and Dunk (1993) find that when participation, budget emphasis, and information asymmetry are high, slack will be high. Osi (1973) examines the relation between behavioral variables and budgetary slack. Through surveys, he finds that budgetary slack is created as a result of the use of budgeted profit attainment as a basic criterion in evaluating performance.

where the Company conducts business, and are responsible for the distribution of electricity, planning and building of lines and local distribution centers, and providing consulting services to both businesses and private households. The units do not get direct rewards for revenues. Hence they are expense centers, not profit centers. The service centers and their managers are located sufficiently far from the center geographically that some degree of asymmetric information is inevitable. While the center's staff travel to the service centers approximately quarterly, there remains a significant degree of private information on the part of the units.

We obtained monthly, quarterly, and yearly operations budgets (planned expenditures) from January 1998 through March 2003 by expenditure category, e.g., personnel (bonus payments, extra hours payments, personnel leasing and temporary workers, travel costs, etc.) and materials.[20] We also secured actual expenditures data for this period as well as data on capital (project) budgets for January 1999 through March 2003. To supplement our understanding of how the budgeting process and general corporate culture works at the Company, we conducted numerous interviews with the company's CFO, vice president, chief controller, and several unit managers.

Planned budgets are agreed on in a process where bottom-up-planning meets top-down-goals, following a precise handbook of budget creation and adaptation. The fiscal year begins in September; budgets are approved by the board by the end of June of that year. What is important for the present paper is that units have the authority to "shuffle" personnel and materials expenditures between categories and, at least partially, across time during the year. Quarterly budget talks with each unit aim to ensure that the overall budget for the entire organization is not exceeded.

"Planned expenditures" (available to us in the Company's database) are not changed in the course of the year. In other words, the entry for planned expenditures for the fiscal year of 1999 (October 1998 to August 1999), say, is indeed what was agreed on in June 1998.

[20]Together, personnel and materials accounted for roughly 95% of the discretionary expenditures by the service centers. The remaining 5%, nondescript "other" expenses, is left aside for our empirical analysis. Fixed personnel expenditures like salaries were not included because they are not subject to decisions by the unit manager.

Our principal analysis is of actual versus planned expenditure.

5.2 Hypotheses and methodology

The principal empirical prediction of the model is that the category with greater variability in productivity will get greater rewards to expenditure and will have more variable expenditures. Absent further information, we can not state whether materials or personnel can expect greater shocks and thus more variable productivity parameters.[21] However, Proposition 1 and Corollary 1 allow us to make the following compound prediction:

a. One category will have greater variability in expenditure. Be it materials or personnel, label it C. Label the other category D.

b. The rewards to excess expenditure in one category, C, will be greater than in D.

This compound prediction of parts a and b should be compared to the null prediction, which would predict identical and probably zero rewards to excess expenditure in either category, with no connection between rewards and expenditure variability.

Part a can be checked by observing expenditure variability directly.

To test the prediction in part b, we consider the following generic regression:

$$Y_{it} = \alpha + \beta x_{it} + \gamma \eta_{it} + \varepsilon_{it},$$

where i is the unit and t is the time index. Because of the panel nature of the data, we used random and fixed-effects models. Because the outcomes are similar, we only present the results for random effects.

The dependent variable, Y_{it}, is defined as the ratio of actual expenditures to budgeted expenditures in each year. A value of this variable "Ratio overall" greater than 1 indicates spending more than was planned initially. Figure 3 in the Appendix shows a histogram of this variable; the shares of overspenders and underspenders are approximately equal.

[21]Conversations with service center managers revealed that as regards the building of electricity lines, the productivity of one dollar spent on materials often varies strongly with the particular features of landscape where a line is built. Instead, a dollar spent on personnel is expected to yield the same output no matter where it is employed. However, the information obtained is not fully conclusive in this regard.

The various explanatory variables x_{it} are defined as follows. "Ratio materials" and "Ratio personnel" are the decimal shares of actual spending on materials and personnel, respectively, in spending in the whole year.[22]

Finally, η_{it} comprises certain control variables for robustness tests (see below).

Table 2 contains descriptive statistics. This table and those that follow are compiled in the Appendix for ease of presentation.

5.3 Findings

While the sample size is limited, some interesting findings emerge from the empirical analysis. Consider Table 2. First observe that similar amounts are spent on materials and personnel. (The majority spends somewhat more on personnel than on materials (not shown).) Second, directly satisfying the assumption of our hypothesis, the standard deviation of materials spending is substantially greater than that of personnel spending. Thus, we label materials as category C.

Table 3 contains results for the tests of part b of the joint prediction. The first column shows regressions of excess spending in the year overall on excess spending on materials. Recall that if the primary shocks really concern the productivity of materials spending – as suggested by the descriptive statistics just presented – we expect positive coefficients for the variable "Ratio Materials." This is indeed what we find. The coefficient implies that when the average service center spends 10% more on materials than was budgeted, it is allowed to exceed the overall budget by 5.9% more. Consistent with the assumptions, overspending on personnel expenditures is negatively related to overall overspending (column (2)). Finally, the greater the overspending on materials, the smaller the overspending on personnel (3). Overall, the Company seems to succeed in screening units along their relative productivities in the two main categories.

Turning to our dynamic budgets predictions, we need to break expenditures into first- and second-period spending. In particular, "Ratio first half" is the decimal share of actual

[22] The results are essentially unchanged if we instead use shares of expenditure categories only for the first half of the year.

spending on discretionary costs that actual spending is of budgeted spending for the first half of the year.[23] "Ratio second half" stands for the decimal share of actual spending on discretionary costs that actual spending is of budgeted spending for the second half of the year.

For the service centers, first-half spending is more variable than second-half spending. If productivity is unmonitorable, this suggests that we should expect overspending in the second half to be negatively or not significantly related to overall overspending and to be negatively related to first period overspending. In fact, there is actually a positive correlation between early and late overspending, implying that if a unit was over in the first half it was more likely to be over in the second half.[24] In sum, Table 4 suggests that screening within the year is less effective as regards intertemporal variation of productivity than inter-category variation. One logical explanation is that it is hard for the organization to discipline late overspending within the same year (although it may be possible to do so in the following year, see the next subsection.) Unlike the results for inter-category screening, this finding for intertemporal screening is compatible with a notion of a soft budget constraint (Kornai 1980, Goldfeld and Quandt 1990, Dewatripont and Maskin 1995, Dewatripont, Maskin and Roland 1999) within firms; future research could formally combine the screening budgets and soft budget constraint frameworks. Another explanation is that the center has some, albeit limited monitoring capability, and gives bigger budgets to units that appear to have positive productivity shocks.

[23]The use of semi-annual spending, together with the fact that the fiscal year begins on October 1, implies that seasonality is not a principal concern here. Air conditioning is not widely used in the country the Company operates in.

[24]A further result (not shown in the table) may be of interest. Define a dummy variable "Excess first" to take the value of 1 when "Ratio first half" is greater than 1, and zero otherwise. "Abs dev" is defined as 1 minus "Ratio first half;" it measures the absolute deviation from how much a unit would spend to just meet its initial budget total. We find that - consistent with our model - "Abs dev" has a negative coefficient but the interaction term with the dummy variable has a positive coefficient. The numbers found typically imply that for up to an overspending of about 200%, the gains of overspending outweigh the losses for the unit.

Implications for budgets in the next period. Some stickler organizations may find it difficult to let units overspend their budgets, or force others to underspend. Such organizations can still use expenditure patterns within a period to influence the next period's budget, assuming that today's productivity helps predict that of next period.[25]

We would expect greater budget increases to go to units with greater excess spending in budget category C, given a primary information asymmetry about category C productivity. Similarly, budget increases should be bigger for units with later spending, if there is a primary information asymmetry about the future, and bigger for those with earlier spending when the center lacks information about current productivity.[26] Allowing current spending patterns to influence the next period's expenditure affords a degree of freedom not available for within period budets, and allows the center to squeeze more information from limited signals. Less distortion is necessary to infer a given amount of information. Unfortunately, such budget rewards make it more attractive for Low to follow High's pattern of expenditure. To avoid having more Lows mimic Highs, the principal will have to distort High even more.

Finally, the organization may infer information not only about a unit's productivity in operations, i.e., the day-to-day business, but also how effectively it uses funds invested in occasional projects. Thus, the same across-year hypotheses can be formulated with "project

[25] A principal who believes that everybody is the same in the next year (in expectation) will not make second-year budgets contingent on first-year spending patterns.

[26] In the mean-reversion case, the basic prediction could be overcome by differential growth rates across units. Rapid growers will tend to have higher expenditures in the second than in the first period. The principal would like to give rapid growers larger future budgets. Thus, differential growth rates among units, even if only partly discernible by the principal, will obscure the operation of the mean-reversion case. Indeed, differential growth rates work against our predictions even if the principal can observe nothing but expenditures. Say such an uninformed principal knew that units grew at different rates. He could infer that larger second- than first-period expenditures implied higher growth rates, and reward units with this pattern accordingly. Our model, which abstracts from any disparities in growth rates, produces precisely the opposite prediction. Mere short-term persistence in productivity, absent any growth, e.g., a random walk without a drift, would also produce a pattern where higher second-period expenditure led to a budget increase for periods three and four. We suspect that both growth rates and short-term productivity persistence are important. Hence, finding evidence for the predictions of the mean-reversion scenario are more noteworthy given that they must overcome these growth rate effects.

budgets" replacing "budgets." (We could not test the within-year hypothesis with capital budgets because we lacked data on actual project expenditures for the Company.)

To test these across-years hypotheses, we introduce a new dependent variable, the percentage change in budget $(B_{it+1} - B_{it})/B_{it}$, where B_{it} is the assigned discretionary budget for unit i in year t. Since we have also obtained data on capital budgets (for investments and projects), we also consider $(I_{it+1} - I_{it})/I_{it}$ where I_{it} is the assigned sum of project budgets for unit i in year t.

The results for these tests are presented in Tables 5 and 6. We find suggestive support for the notion that the Company exploits information contained in current-year spending patterns for the next year's budget allocation, although the overall explanatory power of the regressions is low. The first three regressions in Table 5 imply that greater overspending on materials this year is associated with higher budget growth, while overspending on personnel is associated with lower (although not statistically significantly lower) budget growth. The effect is more pronounced for investment budget growth, as evidenced by columns (4) to (6) in Table 5.

In Table 6, we quantify the impact of intertemporal screening on the growth of budget allocations. As expected (because first-half spending is more variable than second-half spending), greater early spending leads to greater budget growth in the following year, for both operations budgets and investment budgets. For operations budgets, the Company does not significantly (in a statistical and economic sense) reign in the budget growth of late spenders; this is consistent with the explanation based on unit growth offered above. Looking at the investment (projects) budget, first-half spending continues to be positive and significant, and second-half spending is negative and significant (at least in column (6)). This is not surprising if we think of second-half expenditures as wagging tails, where units always spend so as to use up their budgets.[27]

[27]We have explored a series of alternative specifications which we do not present here for space reasons. Because the optimal contract may be quite nonlinear, we took the logarithm of all of our explanatory variables. The fit is somewhat improved for the growth regressions, while for the within-year regressions, the fit worsens considerably. In the budget growth regressions, we find that, once we use this log-transformation, initial budget size has a significant and negative effect; bigger units grow more slowly. (This effect is not

6 Conclusions

Budgets are the most prominent tool of control within organizations. A primary rationale for budgets is that local units' information is superior to that of the center. Yet this asymmetry undermines the ability of the center to set appropriate budget levels: it can neither know in advance, nor determine as time goes along, how productively units will spend their budgets. Screening budgets can reduce the cost of information asymmetries.

We have modeled this process as one where the center sets budgets on the basis of the information available to it. Subsequent to or apart from the budget process, units get productivity shocks that are not monitorable by the center. However, by observing spending patterns, the center draws inferences about the realized productivities. The model applies to both inter-category and intertemporal screening.[28] Since Highs and Lows have comparatively high productivities for different categories or in different periods, Lows can be discouraged from mimicking Highs, and the most productive units can be given more resources. The broad implication of this analysis is that an expenditure category with large shocks will tend to have greater expenditure variability, and will also get rewarded more with greater budgets.

Data from a large company on the expenditures of its five service centers confirms our primary predictions: Expenditures on materials (the more variable expenditure category) were differentially rewarded with larger budgets both within and across periods. In the two-period case, first-half expenditures – the more variable expenditure category – predicted future budget growth for both operations and investment; second-period expenditures predicted growth for neither.

For explanatory ease, our models deal with two categories or two periods and two types of agents. The two categories or periods can be extended to n, and we could increase

significantly present for the non-log regression.) In none of our regressions did the number of employees matter.

[28] As mentioned earlier, screening is also feasible when the shapes of marginal product curves differ across subunits (whether or not productivity shifts between periods). A separate paper clarifies the required conditions.

the number of agents to a continuum. The results are expected to extend to more general concave production functions, as long as the crucial single-crossing condition is retained. For the linearly separable case or for power production functions (or utility functions showing constant proportional risk aversion), we could develop a recursive infinite-period model with budgets that depend on the last period's expenditure and current assets. We believe that the qualitative properties of such models would be similar to those discussed here.

Though our focus is on organizations, this paper also relates to the debate on cash versus in-kind transfers. Typically, economists feel that it is inefficient to impose restrictions like expenditure categories on how budgets are spent, particularly when - as in our case - there is no paternalism with regard to expenditure patterns. However, if target efficiency – getting resources to individuals who can use them most effectively – is a concern, the principal may impose a schedule where the pattern of expenditures affects total resources received (Akerlof 1978, Nichols and Zeckhauser 1982, Blackorby and Donaldson 1988). The allocation mechanism proposed in this paper is intermediate between the cash and in-kind variants to low-income individuals. It is in the spirit of a differential subsidy scheme depending on the category of expenditure. It exploits the different productivities of agents to choose a tradeoff rate (essentially tracing the indifference curve of the potentially envious agent) that allows the principal to obtain greater welfare.

The screening-budgets concept readily extends to other areas. For example, parents may have little more than inferential information about how their children at college spend their allowances. EPA is concerned with the abatement costs of firms when it sets and implements regulatory standards. Federal law currently requires foundations to spend 5% of their assets each period, a rigid constraint that prevents them from tailoring expenditures toward high productivity periods and away from low. In each of these cases, a screening budget could help direct resources toward more efficient expenditures.

7 Appendix

Lemma 1 *In the screening problem (2)-(3), Low's IC binds at the second-best solution.*

Proof. The principle underlying this proof is that at the solution to the second-best problem, it must be that the principal can not increase production with the same funds without violating the incentive compatibility constraint of one of the agents.

Step 1: *Exactly one of the incentive-compatibility constraints binds.* If no incentive-compatibility constraint binds, then the solution to the second-best problem must be the solution to the first-best problem. However, the first-best solution violates incentive compatibility. Since High and Low's indifference curves have different slopes everywhere, it can not be that both High and Low are indifferent between two particular bundles.

Step 2: *If High's incentive-compatibility constraint binds, then High's and Low's bundles must lie on their respective Engel curves.* Recall from microeconomic theory that a consumer's Engel curve traces out how the utility-maximizing bundle changes as total expenditure increases. In other words, the agents' bundles maximize their respective utilities, conditional on total expenditure. Thus, along the Engel curve, the agent's marginal rate of technical substitution is equal to one. Begin with High. Suppose High's bundle is not on her Engel curve. It must then lie above or below it. Consider Figure 2. Below High's Engel curve (like at point H1), its marginal rate of technical substitution is smaller than θ_{HM}/θ_{HP} in absolute value. Thus, holding total expenditure constant, the principal can increase expected profit by moving High's bundle toward its Engel curve. This move could take place along the dashed line, an isoprofit line. This makes the bundle more attractive to High, and, since Low's incentive-compatibility constraint does not bind, this adjustment does not violate IC. Therefore the pre-move point can not be optimal. If, instead, High's optimal bundle lies above its Engel curve (like at point H2), then High's marginal rate of technical substitution is less than θ_{HM}/θ_{HP} in absolute value. Thus, the principal can once again increase its expected profit without violating any incentive-compatibility constraints by moving High's bundle toward its Engel curve while keeping High's total expenditure constant.

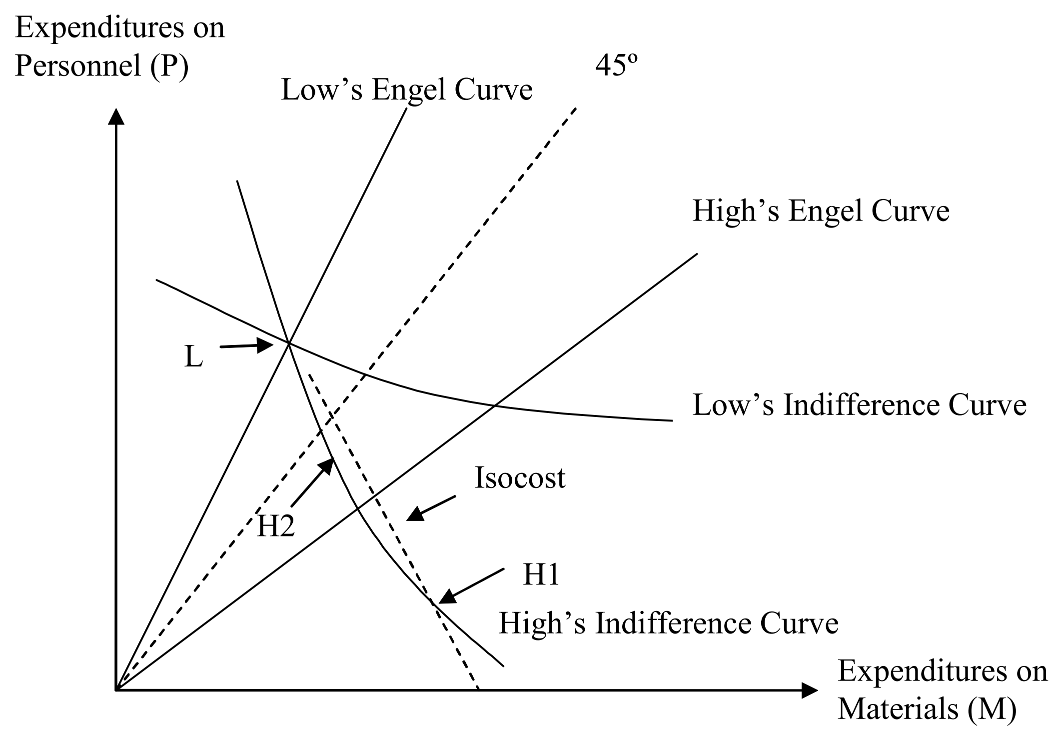

Figure 2: If High's IC binds, both Low's and High's allocations must be on their respective Engel curves.

A similar argument shows that Low's second-best bundle must also lie on her Engel curve. For any allocation below her expansion path, the marginal rate of technical substitution is less than θ_{LM}/θ_{LP} in absolute value. In this case, the principal could, without changing the funds required, increase Low's production by moving her bundle toward toward her Engel curve while keeping total expenditures constant. This violates neither incentive-compatibility constraint. Similarly, for any point above the expansion path, Low's marginal rate of technical substitution is greater than θ_{LM}/θ_{LP} in absolute value. Thus, the principal could once again improve his welfare by shifting High's expenditure towards High's Engel curve while keeping total expenditure constant. Thus, the optimal point must be on High's Engel curve.

Step 3: *If High's incentive-compatibility constraint binds, then the principal can increase profit.* By step 2, if High's incentive-compatibility constraint binds, each agent's bundle must lie on her Engel curve. Since High's indifference curves have slope greater than one in absolute value between the two Engel curves, it must be that Low's total expenditure is greater than High's. But, since both types have Cobb-Douglas production functions, it must be that if both are on their Engel curves and High is spending less than Low, then the principal can increase expected productivity while keeping the expected cost constant by increasing High's total expenditure and decreasing Low's total expenditure slightly. Thus, any allocation where High's incentive-compatibility condition binds can not be optimal.

Steps 1 - 3 suffice to prove that Low's incentive-compatibility constraint must bind. ■

Location of High's allocation. Consider the point defined by (x^*_{HM}, x^*_{HP}) in Figure 1. Because this point lies below High's Engel curve (or "expansion path"), the dashed isocost curve through it must be steeper than High's isoquant through that point. Thus, an increase in High's production would produce a move to the left. But this is not incentive-compatible for Low. By contrast, consider a candidate solution for High to the left of her Engel curve. Here, the isocost curve would be flatter. A move along it to the right, i.e., toward more materials and fewer personnel, would increase High's production and would not violate Low's IC. Therefore, any such point can not be optimal. In other words, at the second-best, High's allocation must lie to the right of her expansion path and must thus be distorted toward materials and away from personnel.

Proposition 1 (Screening budget allocation) Suppose (1) holds. Then, the problem (2)-(3) leads to the following allocations: 1. $x^*_{HM} \geq x^*_{HP}$. 2. $x^*_{LP} \geq x^*_{LM}$. 3. $x^*_{LP} \geq x^*_{HP}$ and $x^*_{HM} \geq x^*_{LM}$. 4. $x^*_{HM} \leq x^{FB}_{HM}, x^*_{HP} \leq x^{FB}_{HP}$; $x^*_{LM} \geq x^{FB}_{LM}$, $x^*_{LP} \geq x^{FB}_{LP}$. 5. Given $x^*_{LP} + x^*_{LM}$, $\{x^*_{LP}, x^*_{LM}\}$ maximizes Low's production. 6. $x^*_{HP} + x^*_{HM} \geq x^*_{LP} + x^*_{LM}$.

Proof. To some extent, the proof reflects the same logic applied in the proof of Lemma 1. Nonetheless, it is instructive to consider the properties in more detail here.

1. By way of contradiction, assume that $x^*_{HP} = a$ and $x^*_{HM} = b$, where $a > b$. Then the principal can make High (and himself) better off by switching such that $x^*_{HP} = b$ and

$x^*_{HM} = a$. This keeps expected cost constant, yet does not induce Low to mimic High since she prefers to spend on personnel. Thus, $x^*_{HP} > x^*_{HM}$ can not be optimal.

2. Suppose instead that $x^*_{LP} = a$ and $x^*_{LM} = b$, where $a < b$. Again, the principal can profitably switch the order: By this he makes Low better off and incurs the same expected cost. Since High's allocation has not changed, if Low did not envy before, she will not envy after. Thus, $x^*_{LM} > x^*_{LP}$ can not be optimal.

3. It is convenient to take logs of the IC constraints. Combining them, we have

$$(\theta_{HP} - \theta_{LP}) \left(\ln\left(x^*_{HP}\right) - \ln\left(x^*_{LP}\right)\right) + (\theta_{HM} - \theta_{LM}) \left(\ln\left(x^*_{HM}\right) - \ln\left(x^*_{LM}\right)\right) \geq 0.$$

The first term is zero. Since $\theta_{HM} \geq \theta_{LM}$, it immediately follows that $x^*_{HM} \geq x^*_{LM}$. To make Low's IC hold, it must be that $x^*_{LP} \geq x^*_{HP}$.

4. Denote with ρ the Lagrange multiplier on Low's IC constraint. Then, the first-order conditions for Low's expenditures are

$$\theta_{LM} K \left(x^*_{LM}\right)^{\theta_{LM}-1} \left(x^*_{LP}\right)^{\theta_{LP}} = \frac{1-\lambda}{(1-\lambda+\rho)}, \; \theta_{LP} K \left(x^*_{LM}\right)^{\theta_{LM}} \left(x^*_{LP}\right)^{\theta_{LP}-1} = \frac{1-\lambda}{(1-\lambda+\rho)}.$$

Since $\rho > 0$ when Low's IC is binding, $\frac{1-\lambda}{(1-\lambda+\rho)} < 1$. Thus, the net marginal profit of increasing Low's personnel expenditures is smaller than in the first-best (where it is 1). Therefore, $x^*_{LP} \geq x^{FB}_{LP}$. By the same argument, $x^*_{LM} \geq x^{FB}_{LM}$.

For High, we have

$$\begin{aligned} \theta_{HM} K \left(x^*_{HM}\right)^{\theta_{HM}-1} \left(x^*_{HP}\right)^{\theta_{HP}} &= \frac{\rho K \theta_{LM} \left(x^*_{HM}\right)^{\theta_{LM}-1} \left(x^*_{HP}\right)^{\theta_{LP}}}{\lambda} \\ \theta_{HP} K \left(x^*_{HM}\right)^{\theta_{HM}} \left(x^*_{HP}\right)^{\theta_{HP}-1} &= \frac{\rho K \theta_{LP} \left(x^*_{HM}\right)^{\theta_{LM}} \left(x^*_{HP}\right)^{\theta_{LP}-1}}{\lambda}. \end{aligned}$$

Since $\rho > 0$ and $\theta_{HM} K \left(x^{FB}_{HM}\right)^{\theta_{HM}-1} \left(x^{FB}_{HP}\right)^{\theta_{HP}} = 1$, we can infer $x^*_{HP} \leq x^{FB}_{HP}$ and $x^*_{HM} \leq x^{FB}_{HM}$.

5. Subject to Low's IC binding, the principal wants Low to do as well as possible because otherwise he could increase welfare. Thus, at Low's second best allocation, she maximizes production given her total expenditure. This can be seen formally by noting that in the equations given under point 4, the right-hand sides for the two first-order conditions for

Low's personnel and materials expenditures are both equal to $(1-\lambda)/(1-\lambda+\rho)$. By contrast, High's marginal products are not equalized.

6. Because of property 5, since Low is indifferent between her allocation and High's, it must be that $x^*_{HP}+x^*_{HM}\geq x^*_{LP}+x^*_{LM}$ since the principal would do worse if this were not the case. ■

Corollary 1. Suppose (1) holds. Then, the problem (2)-(3) also implies

1. $Var\left(x^*_M\right)\geq Var\left(x^*_P\right)$.

2. $\frac{x^*_{HM}-x^*_{LM}}{x_M}\geq\frac{x^*_{LP}-x^*_{HP}}{x_P}$,where $x_M=\lambda x^*_{HM}+(1-\lambda)\,x^*_{LM}$ and $x_P=\lambda x^*_{HP}+(1-\lambda)\,x^*_{LP}$.

Proof. 1. We need to show that

$$\left[\lambda\left(x^*_{HM}-x_M\right)^2+(1-\lambda)\left(x^*_{LM}-x_M\right)^2\right]-\left[\lambda\left(x^*_{HP}-x_P\right)^2+(1-\lambda)\left(x^*_{LP}-x_P\right)^2\right]>0.$$

Manipulation of this term yields

$$\lambda\left(1-\lambda\right)\left(\left(x^*_{HM}+x^*_{HP}\right)-\left(x^*_{LM}+x^*_{LP}\right)\right)\left(x^*_{HM}-x^*_{LM}+x^*_{LP}-x^*_{LM}\right)>0.$$

This is true because $\left(x^*_{HM}+x^*_{HP}\right)-\left(x^*_{LM}+x^*_{LP}\right)>0$ by property 6 of Proposition 1, and $x^*_{HM}-x^*_{LM}+x^*_{LP}-x^*_{LM}>0$ by properties 1 and 2 of the same Proposition.

2. Direct calculations show that

$$\frac{x^*_{HM}-x^*_{LM}}{x_M}-\frac{x^*_{LP}-x^*_{HP}}{x_P}$$

can be written as

$$\frac{x^*_{HM}x^*_{LP}-x^*_{HP}x^*_{LM}}{x_Mx_P}.$$

The numerator is positive by property 3 of Proposition 1, and the denominator is also positive (since expenditures are positive). ■

References

Akerlof, George, "The Market for 'Lemons,' Qualitative Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 1970, *84* (August), 488–500.

Akerlof, George A., "The Economics of Tagging as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning," *American Economic Review*, 1978, *68* (March), 8–19.

Antle, Rick and Gary D. Eppen, "Capital Rationing and Organizational Slack in Capital Budgeting," *Management Science*, 1985, *31*, 163–174.

Blackorby, Charles and David Donaldson, "Cash Versus Kind, Self Selection, and Efficient Transfers," *American Economic Review*, 1988, *78* (September), 691–700.

Brown, Pamela C., James R. Thornton, Dennis M. Buede, and Jeffrey B. Miller, "A revelation scheme for allocating organizational resources," *Journal of Economic Behavior and Organization*, 1992, *18*, 201–214.

Dewatripont, Mathias, "Renegotiation and Information Revelation over Time: The Case of Optimal Labor Contracts," *Quarterly Journal of Economics*, 1989, *104* (2), 589–619.

___ and Eric Maskin, "Credit and Efficiency in Centralized and Decentralized Economies," *Review of Economic Studies*, 1995, *62*, 541–555.

___ , ___ , and Gerard Roland, "Soft Budget Constraints and Transition," in Eric Maskin and A. Simonovits, eds., *Planning, Shortag, and Transformation*, Cambridge, MA: MIT Press, 1999.

Dunk, Alan S., "The Effect of Budget Emphasis and Information Asymmetry on the Relation between Budgetary Participation and Slack," *The Accounting Review*, 1993, *68* (2), 400–410.

Goldfeld, Stephen M. and Richard E. Quandt, "Output Targets, the Soft Budget Constraint and the Firm under Central Planning," *Journal of Economic Behavior and Organization*, 1990, *14*, 205–222.

Graziano, Clara, "Cost observability and renegotiation," *Journal of Economic Behavior and Organization*, 1995, *28*, 359–372.

Harris, Milton and Artur Raviv, "The Capital Budgeting Process: Incentives and Information," *Journal of Finance*, 1996, *51*, 1139–1174.

___ **and** ___ , "Capital Budgeting and Delegation," *Journal of Financial Economics*, 1998, *50*, 259–289.

___ **and Robert M. Townsend**, "Resource Allocation Under Asymmetric Information," *Econometrica*, 1981, *49* (1), 33–64.

___ **, C. H. Kriebel, and Artur Raviv**, "Asymmetric Information, Incentives and Intrafirm Resource Allocation," *Management Science*, 1982, *28* (6), 604–620.

Hart, Oliver and Jean Tirole, "Contract Renegotiation and Coasian Dynamics," *Review of Economic Studies*, 1988, *55* (2), 509–540.

Hayek, Friedrich August, "The Use of Knowledge In Society," *American Economic Review*, 1945, *35* (4), 519–530.

Kornai, Janos, *Economics of Shortage*, Amsterdam: North Hollan, 1980.

Ma, Chingto Albert, "Adverse Selection in Dynamic Moral Hazard," *Quarterly Journal of Economics*, 1991, *106* (1), 255–275.

Milgrom, Paul and Chris Shannon, "Monotone Comparative Statics," *Econometrica*, 1994, *62* (1), 157–180.

___ **and J. Roberts**, *Economics, Organization and Management*, Cambridge: MIT Press, 1992.

Nichols, Albert and Richard Zeckhauser, "Targeting Transfers through Restrictions on Recipients," *American Economic Review Papers and Proceedings*, 1982, *72* (May), 372–377.

Osi, Mohamed, "Factor Analysis of Behavioral Variables Affecting Budgetary Slack," *The Accounting Review*, 1973, *48* (3), 535–548.

Salanié, Bernard, *The Economics of Contracts: A Primer*, Cambridge, MA: MIT Press, 1997.

Spencer, Barbara, "Asymmetric Information and Excessive Budgets in Government Bureaucracies," *Journal of Economic Behavior and Organizations*, 1982, *3*, 197–224.

Wagner, Alexander F., "Screening Via Intertemporal and Intertask Preferences: An Empirical Investigation," 2003.

Young, S. Mark, "Participative Budgeting: The Effects of Risk Aversion and Asymmetric Information on Budgetary Slack," *Journal of Accounting Research*, 1985, *23* (2), 829–842.