COMPARING THE DISTORTIONARY EFFECTS OF ALTERNATIVE INTERGOVERNMENTAL TRANSFERS

WILLIAM JACK
Georgetown University

This article compares the distortions associated with alternative intergovernmental allocation rules when a central authority provides inputs for the provision of social services by local governments, as well as when local governments differ in their needs. Under a quantity-based mechanism, the input choices of high-need localities will tend to be distorted downward. To convince the center of their higher needs, these communities signal their status by spending too little. However, under an expenditure-based mechanism, the direction of distortion of the input choices of high-need localities depends on the price elasticity of demand for the local input. When demand is inelastic (elastic), to signal their high needs, high-need localities spend too much (little) on local inputs. If social services have positive interjurisdictional externalities, expenditure-based mechanisms are preferred, at least in the case of inelastic demand.

Keywords: *intergovernmental transfers; matching grants*

1. INTRODUCTION

Responsibility for the provision of public goods and publicly provided private goods is often shared between different levels of government. For example, in the health sector, lower levels of government might cover primary health care expenditures, while higher levels contribute to the provision of hospital care. Similarly, local governments in the United States often have responsibility for primary and secondary schooling, while state governments contribute more to tertiary education spending.

AUTHOR'S NOTE: Thanks to Matt Haag, Roger Lagunoff, and Arik Levinson, as well as to the editor and two referees, for helpful comments and suggestions. Any errors are mine.

PUBLIC FINANCE REVIEW, Vol. 33 No. 4, July 2005 488-505
DOI: 10.1177/1091142105275168

Such divisions of responsibility are often determined by balancing the desire to capture scale economies, which call for central provision, against the allocative efficiency gains that may derive from better local information (Oates 1972). More recently, researchers have examined the internal organization of government as an exercise in ensuring accountability (e.g., Bardhan and Mookherjee 1999; Seabright 1996), constraining collusion (e.g., Laffont and Martimort 1998), and providing implicit incentives through ownership (e.g., Besley and Ghatak 2001; Jack 2004).

This article takes the organization of government—that is, the degree of decentralization—as given and starts from the simple observation that the optimal level of central government financing of local public goods is likely to depend on local information. Intergovernmental transfer mechanisms will then need to be designed so as to elicit this information and to respond to it. I examine this mechanism design problem facing the higher level of government when local needs differ.

Modeling intergovernmental transfers as incentive schemes is not new (see, e.g., Bordignon, Manasse, and Tabellini 2001). What this article does is to examine the properties of two natural means by which a central government might provide incentives. Under the first scheme, local governments are held accountable for their actions, in the sense that they are rewarded according to the attainment of certain performance measures, such as primary health care coverage rates or pupil-teacher ratios. The second incentive scheme bases intergovernmental transfers instead on local government expenditures and can be thought of as a reward for revenue mobilization effort. Both of these ways of generating incentives are used in practice. Examples in the United States include the accountability provisions of the No Child Left Behind Act of 2001 and the federal funding contributions based on state Medicaid expenditures. Although both schemes considered in this article make transfers conditional on costly local government choices, I find that the nature of the distortionary effects of the two mechanisms, when optimally designed, differ qualitatively. While there is a growing literature on the analysis of intergovernmental transfers as a mechanism design problem (see, e.g., Bordignon, Manasse, and Tabellini 2001; Cornes and Silva 2003; Cremer, Marchand, and Pestieau 1996; Huber and Runkel 2003), none of these studies has compared the design of transfers based on alternative signals.

A very simple environment is assumed, in which two inputs are used by local governments to deliver an output. These inputs could be local public goods or publicly provided private goods. Local governments choose the level of the first input (e.g., primary health care), and the central government provides, or at least finances, the second input (e.g., hospital care). The local government is not required to purchase the second input from the center. Instead, it is allocated free of charge. The two inputs yield social benefits, such as an increase in population health status or productivity.[1]

The task of the central government is to allocate a fixed supply of (funding for) the second input across localities.[2] The efficient allocation of the centrally provided input could be ensured by selling it to localities at a market-clearing price. But if the input is to be distributed free of charge to localities in the form of a categorical grant, the center must design a revelation mechanism that induces localities to truthfully reveal their needs. Clearly, if the center simply offers a menu of levels of the centrally provided input, all localities will choose the highest level of financing offered, and differential allocations will not be incentive compatible.

Local needs are modeled by assuming that the productivity of the locally provided input varies across jurisdictions, taking on one of two possible values. Thus, for a given level of (centrally financed) hospital care, the impact of an increase in primary health care on health status is assumed to be smaller in localities with greater health needs. One way to interpret this is that localities with greater needs face higher local input costs measured in productivity-adjusted units. A natural source of these differential costs is the labor market, as some localities are likely to have to pay higher wages for primary school teachers and general practitioners than others.

Our focus on categorical grants or in-kind transfers is supported by Huber and Runkel (2003), who show in a similar model that neither unconditional block grants nor open-ended matching grants can implement the social optimum under asymmetric information and that categorical grants (or, equivalently, closed-end matching grants) are required. Like this article, both Cornes and Silva (2003) and Huber and Runkel (2003) consider environments with two public goods. However, in the former, transfers are based on aggregate local spending, and in the latter, they are based on local expenditures on one of the goods.

Despite the existence of self-selection constraints, in this binary model, the efficient allocation of a given aggregate level of the central input is sometimes possible. In other cases, incentive constraints preclude the attainment of an efficient allocation, but in general, it remains possible to implement a differential allocation across jurisdictions. In addition to this allocative inefficiency, the incentive compatibility constraints induce an internal production distortion at the locality level in which, conditional on the level of the centrally provided input, the input choices of localities with high costs are distorted away from their efficient levels.

However, I find that the direction of distortion depends on the signal used by the central government. When the centrally provided input is allocated on the basis of observed local-level inputs, high-cost localities signal their types by performing too little primary health care/education, given the level of the centrally provided input. Correspondingly, health status or educational attainments are inefficiently low. On the other hand, when the centrally provided input is made contingent on local expenditures on primary health care/education, high-cost localities signal their types by spending either too much or too little on the local input. They spend too much if unconstrained demand for the local input is inelastic (so that localities with higher costs, ceteris paribus, spend more), and they spend too little if demand is elastic (so higher cost localities spend less).

These comparative results show that while incentive constraints lead to local production inefficiency in general, the nature of that inefficiency—that is, the direction of the distortion to local input choices—depends on the mechanism for implementing the intergovernmental transfer. The simple intuition for the result is that localities with high costs, which are allocated higher centrally provided inputs, must convince the center that they indeed face high costs. When the *quantity* of locally provided inputs is used to condition central allocations, they can do this by using few of them, thus lowering program outcomes. On the other hand, when *expenditures* on local primary-level inputs are observable, high-cost localities can convince the center of their types by incurring either relatively high or low expenses, depending on the elasticity, thereby either "overachieving" or "underachieving" in terms of program outcomes.

There are both normative and positive implications of this analysis. First, a central government choosing between a quantity and an expenditure mechanism must assess the distortionary costs of each. A priori, it is difficult to tell whether over- or underprovision of health or education services would be more costly. However, if, as is widely believed, these services have positive externalities associated with them, then inducing high-cost localities to *over*spend on local inputs would likely be preferable. An expenditure-based intergovernmental grant system would then be preferred, as long as local demand is inelastic. Second, as an empirical matter, the model predicts that in a federal system in which states make transfers to local authorities, the average stringency of requirements placed on local governments should vary across states according to what kind of mechanism (quantity or expenditure based) each state adopts.

The next section describes the economic environment and objectives of central and local governments. Section 3 analyzes the design of intergovernmental grants based on quantity choices at the local level, and expenditure-based transfers are examined in section 4. Section 5 discusses the policy and empirical implications of the model in more detail, and section 6 concludes.

2. THE MODEL

A local government combines a locally provided input, x, and a centrally provided input, y, to produce an outcome $z = \zeta(x, y)$. The production function, ζ, is increasing and quasi-concave, with $\zeta(x, 0) > 0$ and $\zeta(0, y) > 0$ for $x, y > 0$. This outcome is noncontractible, so central allocations cannot be conditioned on it. This is a reasonable assumption when z is a social sector outcome, like health status or educational attainment. Localities have the same production function but face potentially different unit costs, θ, of providing the local input x. For simplicity, I assume that there are just two types of locality. High-cost localities have higher costs than low-cost localities, that is, $\theta_H > \theta_L$. A proportion λ of the localities have low costs.

The social value of the outcome z is $v(z)$, where $v(.)$ is increasing and concave. Denote the net social value of an arbitrary input vector (x, y) to a locality with costs θ by $w(x, y, \theta) = \phi(x, y) - \theta x$, where $\phi(x, y) \equiv v(\zeta(x, y))$. Given a centrally provided input level y, the maximal net value to the locality is

$$\omega(y, \theta) = \max_x \; w(x, y, \theta).$$

Let $x^*(y, \theta)$ be the locality's choice of local input. Then localities with higher costs choose lower local inputs, $\partial x^*/\partial\theta = 1/\phi_{xx} < 0$.

The central government has a budget Y to distribute in-kind to localities. It seeks to maximize the unweighted social value of the services that this resource produces. Thus, the center's problem is to

$$\max_{y_L, y_H} \quad \lambda\omega(y_L, \theta_L) + (1-\lambda)\omega(y_H, \theta_H)$$
$$\text{s.t.} \quad \lambda y_L + (1-\lambda) y_H = Y. \tag{1}$$

When local input costs are observable by the center (and contractible), the first best allocation is attainable using an intergovernmental transfer $y_s^* \equiv y^*(\theta_s)$ satisfying the budget constraint and the first-order condition

$$\omega_1(y_L^*, \theta_L) = \omega_1(y_H^*, \theta_H).$$

If the central government announces these allocations and solicits reports of θ, all localities will claim to have high costs, and the center's budget constraint will be violated. In the following two sections, I characterize the second-best (i.e., incentive compatible) efficient allocations under the quantity- and expenditure-based mechanisms.

3. QUANTITY-BASED MECHANISMS

In this section, it is supposed that the central government can observe an indicator of the quantity of local inputs chosen and condition its allocation on these. The central government thus chooses a pair of input vectors, (x_L, y_L), (x_H, y_H), so that if a locality chooses local input x_s, it is allocated centrally provided inputs y_s. The center's problem is then to

$$\begin{aligned}
&\max_{(x_L, y_L), (x_H, y_H)} && \lambda w(x_L, y_L, \theta_L) + (1-\lambda) w(x_H, y_H, \theta_H) \\
&\text{s.t.} && \lambda y_L + (1-\lambda) y_H = Y \qquad (2)\\
&\text{and} && w(x_s, y_s, \theta_s) \geq w(x_s, y_s, \theta_{s'}), \text{ for } s = L, H, s \neq s'.
\end{aligned}$$

To employ graphical arguments, it turns out to be simpler to write a locality's value function in terms of the outcome z, the local input x, and the cost parameter θ. Let

$$B(z, x, \theta) = v(z) - \theta x$$

be this function. It is easy to confirm that iso-value curves in (x, z)–space are increasing and convex and satisfy the Spence-Mirrlees single crossing property, $\partial^2 z / \partial\theta\partial x < 0$ (see Figure 1). For an arbitrary centrally provided input level y_s, the feasible (x, z) pairs available to the locality satisfy $z_s = \zeta(x, y_s)$.

3.1. IMPLEMENTABLE ALLOCATIONS

If a locality of type θ_s treats the level of the centrally provided input y_s as exogenous, then its best choice of local inputs is $x_s^*(y_s)$, which solves $\max_x B(z, x, \theta_s)$ s.t. $z = \zeta(x, y_s)$. On the other hand, a pair of input vectors (x_L, y_L), (x_H, y_H) satisfies the incentive constraints in (2) if and only if

$$B(z_s(x_s), x_s, \theta_s) \geq B(z_s(x_s), x_s, \theta_{s'})$$

for $s = L, H$, and $s' \neq s$. Thus, we say a central allocation (y_L, y_H) is *efficiently implementable* if and only if $(x_L^*(y_L), y_L)$, $(x_H^*(y_H), y_H)$ is in-

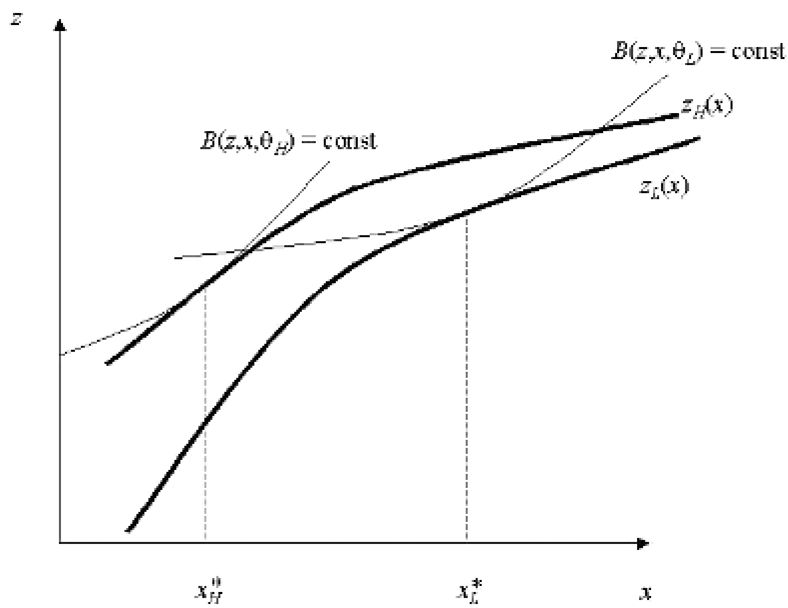

Figure 1: The Centrally Provided Inputs, y_L and y_H, Are Efficiently Implementable under the Quantity-Based Mechanism

centive compatible, and $\lambda y_L + (1 - \lambda)y_H = Y$. Figure 1 shows an efficiently implementable central allocation. Note that the uniform allocation $y_L = y_H = Y$ is trivially efficiently implementable and, by continuity, so is any feasible allocation with $y_L < Y < y_H$ and with y_L and y_H close enough to Y.

However, if the central government wishes to allocate widely differing central inputs to localities with different costs, the incentive constraint in (2) will bind. Thus, Figure 2 shows a pair of central inputs (y_L, y_H) that is not efficiently implementable since the input-outcome pair at point *A* lies above the iso-value curve of the low-cost localities passing through point *C*. To ensure implementability, the local input choice of the high-cost localities must be distorted to $\tilde{x}_H(y_L, y_H) < x^*_H$, at which the low-cost localities' iso-value curve passes through the curve $z = \zeta(x, y_H)$. Note that any value of x_H less than $\tilde{x}_H$ is also part of an implementable allocation but is constrained inefficient. In addition, there is a value $\hat{x}_H > x^*_H$ such that for $x_H > \hat{x}_H$, the pair (x^*_L, y_L), (x_H, y_H) is incentive compatible for the low-cost localities. However,

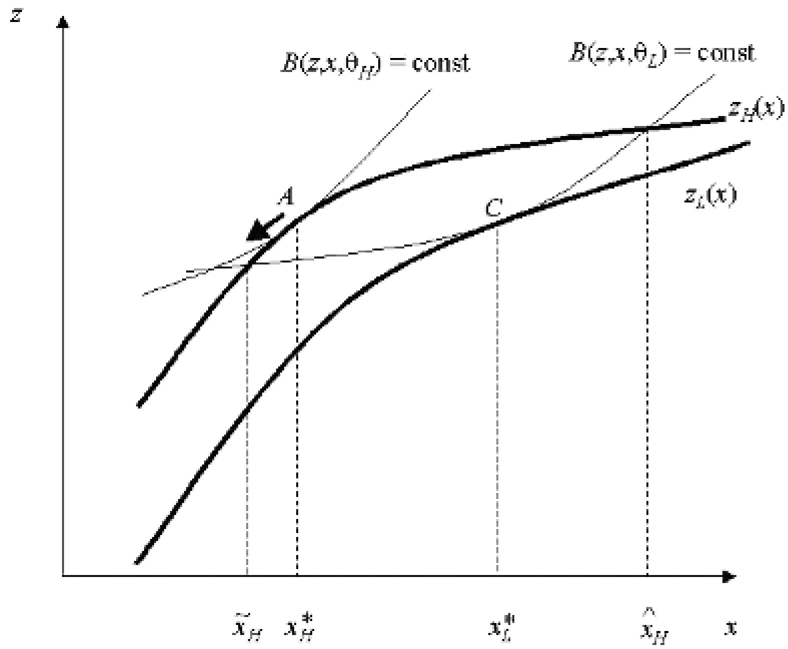

Figure 2: The Centrally Provided Inputs, y_L and y_H, Are Not Efficiently Implementable. The Local Input Choice of High-Need Localities Is Distorted Downward, to $\tilde{x}_H < x^*_H$.

such an allocation would not be incentive compatible for the high-cost localities, which prefer point *C* by the single crossing property, so it is not implementable. Finally, there is no need to distort the allocation to the low-cost localities away from $(x^*_L(y_L), y_L)$ since this only reduces their welfare level while requiring a further distortion in the allocation to high-cost localities. This discussion can be summarized in the following result.

Proposition 1: Under a quantity-based allocation mechanism, if centrally provided inputs are close enough to uniform, they can be implemented efficiently. However, if they differ sufficiently, the local input choices of high-cost localities will be distorted below their efficient levels, yielding lower health/education outcomes than would obtain in the absence of the incentive constraints. The local input choices of low-cost localities are not distorted.

3.2. OPTIMAL CENTRAL ALLOCATIONS

The center's maximization problem (1) must be reformulated to account for the incentive compatibility constraint. Thus, let $y_L = (Y - (1 - \lambda)y_H)/\lambda$ (so that (y_L, y_H) is feasible), and define

$$\tilde{\omega}(y_H, \theta_H) = \begin{cases} \omega(y_H, \theta_H) & \text{if}(y_H, y_L) \text{ is efficiently implementable} \\ \omega(\tilde{x}_H(y_L, y_H), y_H, \theta_H) & \text{otherwise} \end{cases}$$

as the implementable welfare of a high-cost locality when centrally provided inputs are (y_L, y_H). This value function exhibits a kink—and the marginal value function a jump down—at the point at which the incentive constraint binds. The center's unconstrained maximization problem is then

$$\max_{y_L, y_H} \quad \lambda\omega(y_L, \theta_L) + (1-\lambda)\tilde{\omega}(y_H, \theta_H)$$
$$\text{s.t.} \quad \lambda y_L + (1-\lambda)y_H = Y.$$

This problem is illustrated graphically in Figure 3. Three regimes are apparent. In panel (a) of the figure, the optimal division of the central input is fully efficient, and localities' own input choices are undistorted. In panel (b), the incentive constraint binds, and the allocation of central inputs is distorted toward low-cost localities, compared with the fully efficient allocation. However, this allocative inefficiency is not matched by a production inefficiency: both high- and low-cost localities continue to choose their own inputs efficiently, conditional on the central allocations. Finally, in panel (c), both central allocations and local decisions are distorted, and the mechanism is both allocatively and productively inefficient, compared with the unconstrained allocation.

4. EXPENDITURE-BASED MECHANISMS

In this section, I perform a similar exercise to section 3.1 and characterize the implementable allocations under an expenditure-based mechanism. Because the results on optimal division of the central input are qualitatively similar to those under a quantity-based mecha-

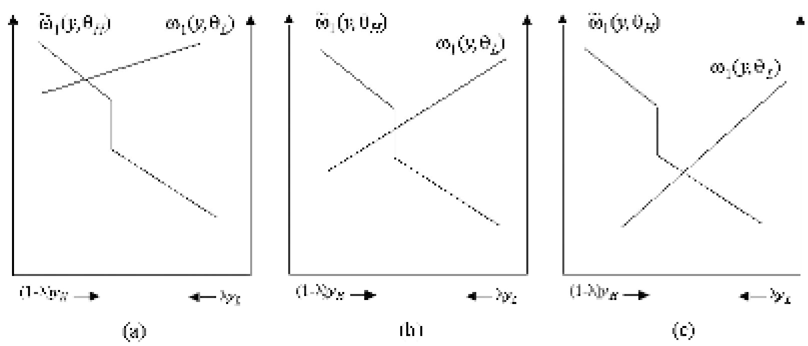

Figure 3: Optimal Centrally Provided Inputs: (a) Allocative and Production Efficiency, (b) Allocative Inefficiency but Production Inefficiency, and (c) Allocative and Production Inefficiency

nism in section 3.2, I concentrate here only on the implementation issue.

Define expenditures on the local input by $e = \theta x$, and let

$$\beta(z, e) = v(z) - e$$

be a locality's value of reaching an outcome z while spending e. Note that in (e, z)–space, all localities have the same increasing and convex iso-value curves. For a θ_s locality, the combinations of centrally provided inputs, y, and local expenditures, e, that yield outcome z satisfy

$$z = \zeta\left(\frac{e}{\theta_s}, y\right).$$

An expenditure-based allocation mechanism is a pair of vectors (e_L, y_L), (e_H, y_H) that specifies the level of centrally provided input, y_s, conditional on an incurred expenditure on local inputs e_s. This pair is incentive compatible if and only if

$$\beta\left(\zeta\left(\frac{e_s}{\theta_s}, y_s\right), e_s\right) \geq \beta\left(\zeta\left(\frac{e_{s'}}{\theta_s}, y_{s'}\right), e_{s'}\right) \tag{3}$$

for $s = L, H$, and $s \neq s'$. In this case, we say a central allocation (y_L, y_H) is *efficiently implementable* if and only if $(e^*_L(y_L), y_L)$, $(e^*_H(y_H), y_H)$ is incentive compatible, and $\lambda y_L + (1 - \lambda) y_H = Y$. Figure 4 shows an effi-

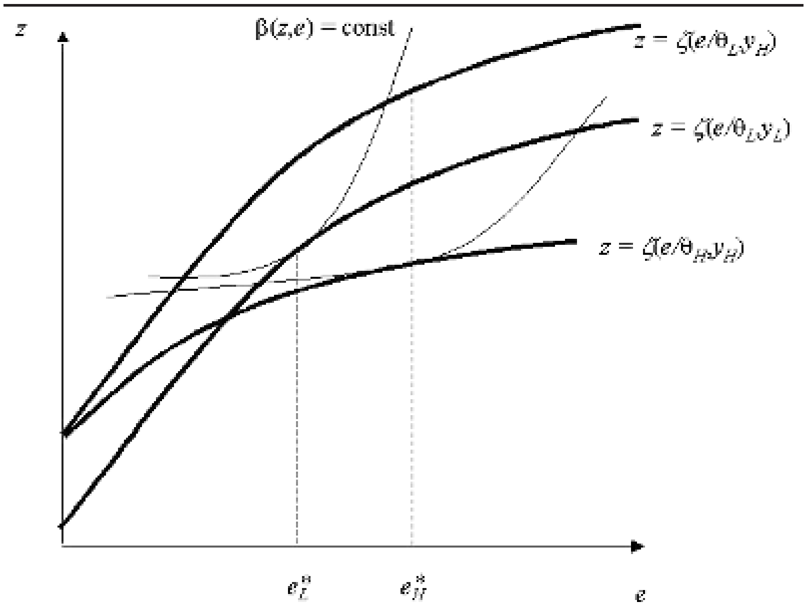

Figure 4: The Pair of Centrally Provided Inputs, y_L and y_H, Is Efficiently Implementable under the Expenditure-Based Mechanism

ciently implementable central allocation. Note once again that the uniform allocation $y_L = y_H = Y$ is trivially efficiently implementable and, by continuity, so is any feasible allocation with $y_L < Y < y_H$ and with y_L and y_H close enough to Y.[3]

As in the case of the quantity-based instrument, when the centrally provided input levels differ enough, efficient implementation is precluded. However, the direction of the induced distortion now depends on the elasticity of demand for the locally provided input. To see this, suppose demand for the locally provided input is inelastic, and expenditures increase with θ. Figure 5 shows an example of a pair of centrally provided inputs, $y_L < y_H$, that is not efficiently implementable. First, note that if both high- and low-cost localities receive y_H from the center, then their expenditures on the local input satisfy $e^*(y_H, \theta_H) > e^*(y_H, \theta_L)$. Efficient implementation of the pair of central inputs (y_L, y_H) would require low-cost localities to produce at point C and high-cost localities to locate at point A. However, given this choice, the low-cost locality can misrepresent its type and locate at point D. To ensure

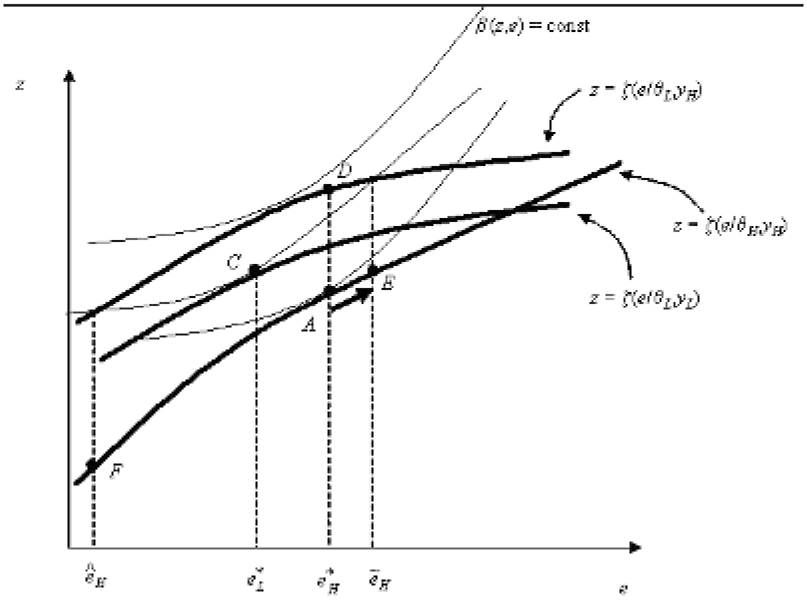

Figure 5: The Pair of Centrally Provided Inputs, y_L and y_H, Is Not Efficiently Implementable. Demand for the Local Input Is Inelastic, and the Pair Is Implemented Optimally by Distorting the Local Expenditure Choice of High-Need Localities to $\tilde{e}_H > e^*_H$.

that the low-cost locality selects the correct input bundle and produces at *C*, it is necessary to distort the production of high-cost localities to point *E*, where expenditures on the local input are $\tilde{e}_H > e^*_H$.

Alternatively, distorting the production of high-cost localities to point *F*, where expenditures on the local input are $\hat{e}_H < e^*_H$, also ensures incentive compatibility. However, it is straightforward to check that, under the elasticity condition, high-cost localities strictly prefer *E* to *F*. To see this, note that if

$$\frac{\partial^2}{\partial e \partial \theta} \phi(e/\theta, y) > 0, \tag{4}$$

then as local expenditures are increased from $\hat{e}_H$ to $\tilde{e}_H$, holding central inputs at y_H, the difference between the value of services produced by a low-cost locality, $\phi(e/\theta_L, y_H)$, and the value of those produced by a high-cost locality, $\phi(e/\theta_H, y_H)$, falls. Since, by definition, the net value to a low-cost locality is the same at $\hat{e}_H$ and $\tilde{e}_H$, it must therefore be the case that the higher expenditure level ($\tilde{e}_H$) is preferred by the high-cost locality. Finally, it is easy to confirm that condition (4) reduces to the elasticity condition,

$$\varepsilon \equiv \frac{-x\phi_{xx}}{\phi_x} < 1.$$

Conversely, if local expenditures fall as input costs increase, holding central inputs fixed, the input choices of high-cost localities are distorted downwards when central input pairs are not efficiently implementable. Figure 6 illustrates this case. Now, efficient implementation of the pair of centrally provided inputs $y_L < y_H$ requires the high- and low-cost localities producing at points *A* and *C*, respectively. With this choice, the low-cost localities can imitate those with high input costs and produce at *D*, which they prefer to *C*. To reattain incentive compatibility, the high-cost localities' production must be distorted to either point *E* or point *F*, and by a similar reasoning to that above, *E* is preferred, as long as $\varepsilon > 1$.

This discussion is summarized in the following result:

> *Proposition 2:* Under an expenditure-based allocation mechanism, if centrally provided inputs are close enough to uniform, they can be implemented efficiently. However, if they differ sufficiently, the local input choices of high-cost localities will be distorted above (below) their efficient levels as $\varepsilon <(>) 1$, yielding higher (lower) health/education outcomes than would obtain in the absence of the incentive constraints. The local input choices of low-cost localities are not distorted.

Finally, consider the knife-edge case of unit elastic demand, $\varepsilon = 1$. In this case, given a value of y, the *optimal* expenditure on the local public good is independent of a locality's input cost. However, this does not mean that the central government is unable to devise a mechanism that induces separation of high- and low-cost types with $y_L < y_H$. As in the two examples above, the expenditure of the high-cost locality must be distorted in one direction or the other. In the case of $\varepsilon > 1$, shown in Figure 6, the downward distortion (to $\tilde{e}_H$) is less costly than the upward distortion (to $\hat{e}_H$). When $\varepsilon < 1$, the upward distortion is less costly than the downward distortion. However, when $\varepsilon = 1$, both dis-

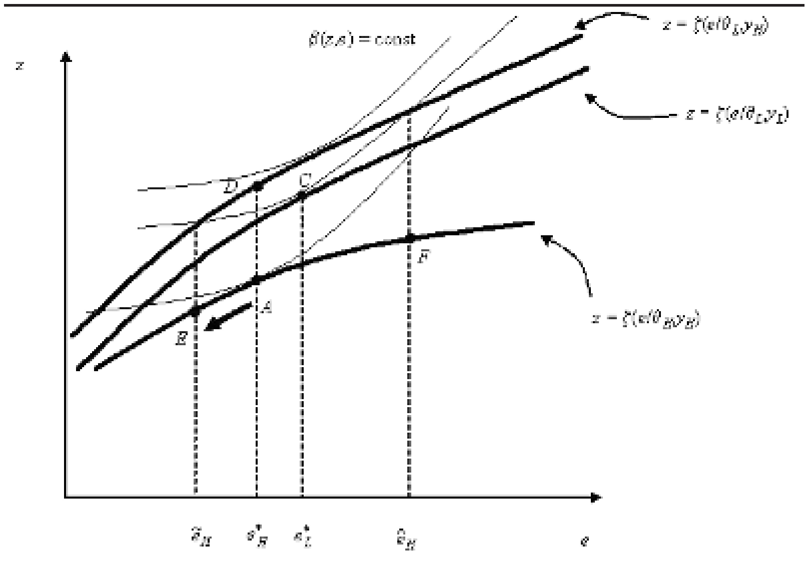

Figure 6: The Pair of Centrally Provided Inputs, y_L and y_H, Is Not Efficiently Implementable. In This Case, Demand for the Local Input Is Elastic, and the Pair Is Implemented Optimally by Distorting the Local Expenditure Choice of High-Need Localities to $\tilde{e}_H < e_H^*$.

tortions are equally costly, and the center is indifferent between inducing the high-cost locality to overspend or underspend.

5. SOME IMPLICATIONS

The normative policy implications of this exercise depend on the costs of gathering information about local government choices. Indeed, if both local quantity and expenditure decisions could be costlessly observed, then underlying local costs could be inferred, and there would be no incentive problem. On the other hand, if one kind of signal is administratively much cheaper to collect than the other, the mechanism that conditions on the cheaper signal should likely be used. However, the costs of information collection may well be nonnegligible and similar for each kind of signal. For example, complete and well-enforced accounting standards are costly to implement, and recent corporate history in the United States confirms that the scope for financial slight of hand should not be underestimated. Similarly, while counting the number of teachers or health clinics is straightforward in principle, adjusting these measures for quality (including attendance and on-the-job performance) is likely to be difficult and costly.

If collecting information on both quantities *and* expenditures is prohibitively costly, then the choice of which mechanism to use will come down to which distortion is less costly. A priori, it is difficult to tell whether over- or underprovision of health or education services would be more costly. However, if, as is widely believed, these services have positive externalities associated with them, then inducing high-cost localities to overspend on local inputs would likely be preferable. An expenditure-based intergovernmental grant system would then be preferred, as long as local demand is inelastic.

The positive implication of the comparative analysis of this article is that the precise method by which intergovernmental transfers are used to provide incentives under asymmetric information can affect the observed distribution of spending on public (or publicly provided private) goods. For example, the model predicts that, in a federal system, "holding states accountable for their actions" by mandating certain such actions (i.e., a quantity-based system) would require that high-cost states be induced to clear relatively low hurdles (compared to the first best). On the other hand, "rewarding states for mobilizing resources" (i.e., using an expenditure-based mechanism) would mean that the same states should be induced to choose relatively demanding options (assuming inelastic demand). A useful empirical test of the analysis of the model presented here could be carried out in a country with (at least) three tiers of government, such as the United States. One could focus on the incentive mechanisms that states use to allocate financing across heterogeneous counties and examine how the performance standards placed on local governments vary across states.

6. CONCLUSIONS

This article has compared the distortions associated with alternative intergovernmental allocation rules when a central authority finances one of two inputs into a local government production process. I have concentrated on the social sectors—health and education—for two reasons. First, these sectors receive a large amount of public funding in many countries, much of which is provided by both central and local governments. And second, the outputs in the health and education sectors are notoriously difficult to monitor, so it is highly likely that any incentive mechanism will use information on input use rather than observed outcomes.

I have examined the implications of alternative measures of local input use for the design and efficiency properties of central allocation mechanisms when localities differ in their costs, represented here as the unit cost of productivity-adjusted local inputs. Under the quantity-based mechanism studied, a central government can gather information on the real level of inputs employed at the local level but may not be able to accurately measure the associated expenditures, say, due to difficulties in attributing various budgetary costs. In this case, to implement differential allocations of the central input according to local costs, the input choices of high-cost localities will tend to be distorted downwards. To convince the center of their higher costs, these communities will signal their status by spending too little (compared with the efficient outcome).

Alternatively, under an expenditure-based mechanism, the central government gathers information on local expenditures but cannot measure the real level of resources employed. In this case, the direction of distortion of the input choices of high-cost localities depends on the price elasticity of demand for the local input. When demand is inelastic (elastic), to signal their high costs, high-cost localities spend too much (little) on local inputs (again, relative to the efficient outcome).

A central government choosing between the two intergovernmental transfer mechanisms must weigh the distortionary costs associated with each, although it is difficult to predict which mechanism will be less distortionary without putting much more structure on the model. However, if social sector expenditures are expected to have large positive externalities, then inducing overprovision at the local level is the preferred option, and an expenditure-based mechanism should be adopted (assuming inelastic demand).

NOTES

1. Of course, in practice, the central government would implement such grants by making financial transfers to localities and restricting the use of the funds to purchases of the second input. Such transfers are referred to as categorical grants.

2. A more general model would endogenize the aggregate quantity of the second input chosen. However, the same allocation problem would continue to confront the central government.

3. Figure 4 illustrates an efficient incentive-compatible pair of transfers in the case where demand for the local public good is inelastitc. Similarly efficient and incentive-compatible transfers are also possible when demand is elastic. To economize on space, I do not include a separate figure to illustrate this possibility.

REFERENCES

Bardhan, Pranab, and Dilip Mookherjee. 1999. Relative capture of local and central governments: An essay in the political economy of decentralization. Mimeo, Department of Economics, University of California, Berkeley.

Besley, Timothy, and Maitreesh Ghatak. 2001. Government versus private ownership of public goods. *Quarterly Journal of Economics* 116 (4): 1343-72.

Bordignon, M., P. Manasse, and G. Tabellini. 2001. Optimal regional redistribution under asymmetric information. *American Economic Review* 91:709-23.

Cornes, Richard, and Emilson Silva. 2003. Public good mix in a federation with incomplete information. *Journal of Public Economic Theory* 5 (2): 381-97.

Cremer, H., M. Marchand, and P. Pestieau. 1996. International redistribution through tax surcharge. *International Tax and Public Finance* 3:157-73.

Huber, Bernd, and Marco Runkel. 2003. Optimal design of inter-governmental grants under asymmetric information. Working paper no. 919, CESifo. www.CESifo.edu.

Jack, William. 2004. The organization of public service provision. *Journal of Public Economic Theory* 6 (3): 409-25.

Laffont, Jean-Jacques, and David Martimort. 1998. Transaction costs, institutional design and the separation of powers. *European Economic Review* 42 (3-5): 673-84.

Oates, Wallace. 1972. *Fiscal federalism*. New York: Harcourt, Brace.

Seabright, Paul. 1996. Accountability and decentralisation in government: An incomplete contracts model. *European Economic Review* 40:61-89.

William Jack is an associate professor in the Department of Economics at Georgetown University. He has worked at the Joint Committee on Taxation of the U.S. Congress, the International Monetary Fund, and the World Bank. His research areas include applied public finance, health economics, and development economics.