Optimal Decentralization in Corporations and Federations

Matthew J. Holian*
San Jose State University

October 17, 2008

Abstract

Oates' Theorem and the M-form Hypothesis are both organizational theories of decentralization, though they deal with different types of organizations. This brief note describes how the two theories complement one another, through both verbal description and mathematical models. The result is a simple but comprehensive account of the delegation problem.

1 Introduction

Industrial organization economists often think about the issue of decentralization as it pertains to multidivisional corporations; the most famous theory of the multidivisional corporation is the M-form Hypothesis, due to Williamson (1975) and Chandler (1962). A textbook treatment of the M-form Hypothesis, such as Jean Tirole (1988), describes the costs of decentralization as loss of economies of scale, and the benefit of decentralization in terms of mitigating the moral hazard problem that exists between upper and lower levels of management. Recent formal theory (Maskin et al., 2000) has also described the M-form Hypothesis in these terms. Yet a careful reading of Williamson suggests that these two forces – economies of scale and comparability – are only half the story.

Contrary to IO economists, when political economists think about decentralization, they are more likely to have in mind theories of federalism than of corporations. Like industrial organization, there is also a famous theory among these social scientists, which has become known as Oates' Decentralization Theorem (Oates, 1972). Oates describes the trade-offs of federalism as failure to internalize externalities, and

*I began this project while at the University of Michigan's EITM V workshop. I thank Massimo Morelli for his comments and encouragement. All errors are my own.

(like the M-form Hypothesis), loss of economies of scale, in exchange for better matching of service levels to preferences. The two unique aspects of the Decentralization Theorem – preference matching and internalizing externalities – when combined with the textbook description of the M-form Hypothesis, provide a fuller representation of the decentralizaton problem, and indeed a better description of the theory Williamson intended.

Comparability arises as a benefit of decentralization in the textbook description of the M-form Hypothesis due to hidden action and the need to attenuate opportunism. However the textbook description fails to incorporate other important aspects of knowledge that are core to the original theory. Meanwhile, the theory of federalism essentially deals with these left-out knowledge issues. The main body of the paper begins by describing the theories verbally and with examples, and proceeds to mathematical presentations of the models. This detailed exploration of how the textbook description of how the M-form Hypothesis and Oates' Theorem complement each other helps to differentiate among many of the different aspects of decentralization, and in this sense provides a model that is both comprehensive and applicable to a wide variety of settings.

Williamson (1975) argues that while the vertically integrated, centralized firm – the unitary or U-form organization – allows for economies of scale and division of labor, he mentions a major cost of the U-form comes in the form of "control loss" (p. 133). Williamson suggests two main channels whereby decentralizing structure mitigates control loss problems: "...the M-form organization... served both to economize on bounded rationality and attenuate opportunism." Bounded rationality comes in the form of "confounding strategic and operating decisions." (p. 137)

While the textbook model accounts for attenuating opportunism, the model of fiscal federalism can be used to illustrate both the "strategic" and "operating" implication of bounded rationality in a simple, yet powerful way. This is useful, because, as David Kreps writes, "...mathematics-based theory still lacks the language needed to capture essential ideas of bounded rationality, which are central to Williamson's concepts of transaction costs and contractual form." (Kreps, 1996, p. 562)

But beyond just an exercise in formalism, clarifying the original intention of the M-form Hypothesis also sheds light onto those properties of the decentralization problem that have practical relevance, and can be found, for example, in management books in the business literature. For example, Charles Koch describes optimal decentralization (pp.132-133) as follows:

> Proximity to a problem or process does not determine who is in the best position to make a decision...Those with local knowledge are often in a better position to solve the problem at hand...but universally decentralized decision-making has its own problems. Some decisions, if made at the local level, can be unprofitable because a broader perspective is required.

This is to say there is a tension between optimally using local knowledge and global

knowledge.[1] Having global knowledge[2] means knowing something about how the different parts of the organization fit together. It is often those at the top that will have global knowledge, but not always.

The models described below, and presented mathematically in section three, illustrate many of the core issues involved in the organizational problem, including the tension between local knowledge and global knowledge, and how organization affects both incentives and production costs (where the latter two represent the two aspects typically considered in textbook descriptions of the M-form Hypothesis.) One interesting aspect about the formal models we present is that the M-form Hypothesis is told through a cost function, whereas the Decentralization Theorem is told through an objective function. A brief summary of the paper concludes.

2 The Decentralization Theorem and the M-form Hypothesis

The key assumption of Oates' Decentralizaton Theorem is that centralized systems must provide a uniform level of service to districts. This assumption also seems to capture Williamson's notion of confounding operating decisions, as follows. A CEO, say of a two division firm, may not know enough about the conditions on the shop floor to make optimal operating decisions for both divisions; if the CEO is required to make operating decisions for both divisions, his only recourse may be to a uniform policy.

Policy uniformity often arises under centralized systems, and this is a feature of centralization that has long been recognized. Alexis de Tocqueville in the 1830's wrote about political systems:

> The federal system was created with the intention of combining the different advantages which result from the magnitude and the littleness of nations...In great centralized nations the legislator is obliged to give a

[1] Koch provides an example of this (p. 133):

> The mindless application of either approach–universally centralized or completely decentralized decision rights–is not the answer. For example, decisions about how to gain optimum throughput from a refinery at any given time probably are best made by people on site. On the other hand, people further removed, but with broader knowledge, may be better positioned to make a decision on what the most profitable product mix will be in five years.

[2] It should be noted that the term "global knowledge" is used here differently than in much of the management and strategy literature, where it is used to refers to knowledge spillovers in a national, or economy-wide setting. Here, global knowledge means knowledge of the spilloves, and the setting is the corporate.

> character of uniformity to the laws, which does not always suit the diversity of customs and of districts; as he takes no cognizance of special cases, he can only proceed upon general principles...since legislation cannot adapt itself to the exigencies and the customs of the population, which is a great cause of trouble and misery" (Vol. I, p. 163, cited in Oates, 2005).

Regarding this quotation, Oates (2005, p. 353) writes, "One might read this historic passage as placing primary emphasis on the political constraints on centralized provision, although this may reflect incomplete knowledge as well." Thus decentralization that leads to greater matching of operating decisions to conditions on the shop floor can be viewed as suiting the diversity of divisions, to paraphrase de Tocqueville. The possibility of this type of mathcing is the upside to decentralization in Oates' Theorem. The downside to decentralization in Oates' Theorem, however, can be viewed as what Williamson refered to as an inability to make strategic decisions. In Oates (1972), this comes through the notion of cross-district spillovers. Government leaders (executives) are often in the best position to internalize inter-district(departmental) externalities, for example by building a park that benefits citizens of both districts, as they are more likely to be aware of their existence; that is, because of their vantage point, executives are more likely to have the information to act on them.[3]

Just as externalities exist between governments, inter-departmental externalities exist between divisions of a firm. This is especially true because the divisions in M-form organizations are related businesses; taking advantage of these externalities has been recognized as one of the great advantage of the M-form, by Williamson and others, over holding companies. Daniel Spulber (2007, p. 237) provides a modern example; Wal-Mart takes advantage of externalities (spillovers) arising through impulse purchases:

> The combination of a department store and a supermarket offered customers the convenience of one-stop shopping, an advantage over specialized department stores and over standard supermarkets. Wal-Mart's supercenter format took advantage of impulse purchasing and in-store promotions. A customer of the department store or the supermarket might be induced to make an unanticipated purchase from the other side of the supercenter.

Thus if one division of the store advertises, the other may receive a benefit in the form of a spillover. To see how these are underprovided for in the decentralized case, imagine that for some reason, customers who buy pomegranates also make many impulse purchases. Then, if the manager of Wal-Mart.s grocery division can

[3]Of course there are issues of compensation as well, but our aim is to emphasize the knowledge problems.

attract these types of customers, say through holding a sale on this fruit, the other department (e.g. clothing or hardware) will benefit from increased sales. However, the grocery manager may not be aware of this, as he has only local, not global knowledge. Moreover he may not be motivated by externalities if his compensation is based only on his division's performance.

Internalizing inter-divisional spillovers can be thought of as making strategic decisions in the following sense. Executives that have global knowledge are often in the best position to make strategic decisions, while regional managers may have local knowledge and be in the best position to make operating decisions. In this way, the fiscal federalism model of Oates (1972) represents a basic tension in strategic decision making, discussed by Williamson (1975) with respect to bounded rationality. Textbook descriptions of the M-form Hypothesis tend to capture well Williamson's notions of "attenuating opportunism" and economies of scale, but they often leave out considerations of interdivisional spillovers; and, while they the textbook depections typically focuses on moral hazard, they do not explore deeper information issues (what some may call knowledge issues) such as matching operating decisions to shop-floor conditions.

The notion that divisions of an M-form are related is at the core of the textbook description's depiction of how the M-form hypothesis can minimize the motivational problems when worker effort is unobservable. The example of the internal structure of General Motors is illustrative. Maskin et al. (2000, p. 360) describe the structure of General Motors (GM):

> A classic example of the U-form was the Ford Motor Company before the Second World War. In those days, Ford was organized into a number of functionally specialized departments: production, sales, purchasing, and so on. In other words, the various departments carried out complementary tasks; none was independent of the others. By contrast, General Motors under Alfred Sloan became the prototypical M-form; GM comprised (and still comprises) a collection of fairly self-contained divisions, e.g. Chevrolet, Pontiac, and Oldsmobile.

For there to be a theory of firm structure, it must explain why firms are sometimes more centralized (U-form) and other times more decentralized (M-form). In the model of Maskin et al. (2000), fully formalized below, when moral hazard is particularly likely to be high, decentralization may allow for greater control, by providing benchmarks that are internal to the firm to offer indications of divisional performance. In our example, the GM central office is likely to find it easier to compare the performance of Chevrolet with Pontiac, than if GM had only one product line, and had to either make external comparisons, or comparisons across the different departments; comparing the production department with the sales department, for example, might be difficult. Therefore, when being able to make comparisons is important, the or-

ganization will decentralize. However the downside of decentralization, in Maskin et al. (2000) and elsewhere, is the inability to realize economies of scale.

In a nutshell, the M-form Hypothesis, as described by Maskin et al. (2000) and Tirole (1988), rests on the importance of the comparability of divisions versus departments, versus the importance of economies of scale that result from U-form (unitary, or centralized) organizations. In general, organizations structured along M-form lines are more comparable; if the benefit from greater comparability outweighs the benefit of economies of scale under U-form structure, the organization should decentralize its structure (choose the M-form).

Finally, while the main purpose of this article was to compare the M-form Hypothesis with the Decentralization Theorem, it is interesting to note that the trade-offs highlighted in a number of other models can represent specific subcases of a joined version of the models presented below. As one example, Alesina and Spolaore (1997) incorporate the trade-off of economies of scale versus local control. Thus the two models presented in the following section include many of the costs and benefits mentioned in the literature on structural choice.

3 Models

This section demonstrates Williamson's M-form Hypothesis and Oates' Decentralization Theorem, with the use of formal models. The following subsection contains the demonstration of the M-form Hypothesis, described through production technology. This is a simplified, but fully formalized version of Maskin et al.'s (2000) description of the M-form Hypothesis. The demonstration of the Decentralization Theorem, described through an objective function, is due to Besley and Coate (2003), and is summarized in the second subsection.

3.1 M-form Hypothesis

This subsection demonstrates the essence of Maskin et al.'s (2000) interpretation of *the M-form Hypothesis*.[4] A unit of output can be produced at cost p, but this subsection describes other cost considerations, and these costs are represented by $c(\Omega)$. These costs depend on organizational structure Ω, which can be either centralized (or U-form, $\Omega = U$), or decentralized (or M-form, $\Omega = M$). In the centralized case, two jurisdictions with heterogenous citizens exist in a unified political body. In the decentralized case, each jurisdiction operates its own government. The cost function for one jurisdiction is:

$$c(\Omega) = \begin{cases} Eb(U) + \frac{\theta}{2} & \text{for } \Omega = U \\ Eb(M) + \theta & \text{for } \Omega = M \end{cases} \tag{1}$$

[4]However the modeling framework here differs from theirs for two reasons: to formally model economies of scale, and for parsimony.

Here, $Eb(\Omega)$ is the expected bonus that needs to be paid to the bureaucrat to exert effort above a "low powered" baseline. This depends on the organizational structure. The per unit cost p does not depend on the level of the service, but there are economies of scale because of saving on fixed costs θ. That is, economies of scale result from sharing fixed costs. The total cost to the organization in both cases will be $2c(\Omega)$ because the organization as a whole has to pay for service in both entities, regardless of administrative structure.

The amount $Eb(\Omega)$ depends on structure in the following way: the executive (CEO or city official) can costlessly obtain a signal s of the manager's performance. The bonus can be based on the signal because it is observable to both parties. The probability this signal is good ($s = 1$) depends on a "common shock" and the efforts of the managers. The manager or managers make a binary decision, $\Delta = 1$ or 0, exert effort or not. The common shock is positive with probability σ and when it is positive $s = 1$ regardless of a manager's effort decision. If the common shock is negative, which occurs with probability $1 - \sigma$, then whether the signal is positive or negative depends on the manager's effort decision; in the case of a negative common shock, the signal is positive with probability equal to q_k where $q_1 > q_0$. That is, a manager (bureaucrat) is more likely to receive a good signal if he worked than if he did not.[5]

The bonus the city pays a manager in the centralized case can depend on the functional division manager's signal s_i, and in the decentralized case, a manager's bonus can depend on his own signal s_i *and* the signal of the other manager s_{-i}. Thus there are four possible states of the world with respect to observable signals in the decentralized case: both manager i's and manager $-i$'s signals are good $(1, 1)$ both bad $(0, 0)$ one good and one bad $(1, 0)$ and one bad and one good $(0, 1)$. The cost of motivating the employees to work is minimized when they are paid only when the signal of one is positive and the other is negative, or relative performance evaluation (RPE).[6] The intuition is that when one manager's signal is positive and the other's is negative, the common shock had to have been negative, so the signal is more likely to reflect that the agent with a positive shock was actually working. The minimized bonus is given by (see appendix for all details)

[5] $b(\Omega)$ depends on organizational structure because decentralization permits relative performance pay, but centralization does not. This may appear to be a rather strong assumption; in principal organizations can find benchmarks for relative performance pay from a variety of sources, including the other functional divisions which are not modeled here. This assumption, however, merely asserts that geographic (or product based) comparison is the best benchmark, and should be seen as a simplifying assumption. While in principle it is also possible that the costs of providing incentives outweighed their benefits, assuming that the benefit the city receives for effort is sufficiently high that it would like to get effort in any case, then the point is minimizing the bonus, and this also further simplifies the analysis.

[6] This part of the model borrows the general framework of Che and Yoo (2001), where optimality of RPE is also shown. RPE was demonstrated formally by Bengt R. Holmstrom in "Moral Hazard in Teams." Bell Journal of Economics and Management Science, Autumn 1982, 13(2), pp. 324-40.

$$b(M) = \frac{e}{(1-\sigma)(1-q_1)(q_1-q_0)} \tag{2}$$

The assumption of risk neutral agents is important here, because agents always work, but are paid only sometimes. They are willing to accept this arrangement because when they do get paid, the payoff is large enough. This framework can also be used to show the size of the bonus needed to motivate an employee under a centralized regime. In the centralized case, the assumption that there is no correlation between the other manager's signal means the city cannot use RPE. The optimal wage when $\Omega = U$ is to pay the bureaucrat only when his signal is good, and then to pay him a bonus equal to

$$b(U) = \frac{e}{(1-\sigma)(q_1-q_0)} \tag{3}$$

While (3) is smaller than (2), the bonus in a centralized regime is paid more often, and so the city's *expected* wage bill from paying bonuses is actually higher in a centralized regime than in a decentralized regime. The reason is that in a centralized regime when the bureaucrat is only paid when he does well, he is paid when the common shock occurs even if he did not work. Under RPE in the decentralized regime, an agent is never paid when the common shock occurs. Equations (3.13) and (3.14) in the appendix derive the expected bonuses used below in (3.4) and (3.5).

This framework can be used to express the M-form Hypothesis as interpreted by Maskin et al. (2000).

If the gains from economies of scale are less than the gains from cost savings due to a lower wage bill, then the organization decentralizes. If not, it centralizes.

This idea can be expressed formally as follows. The expected costs $Ec(\Omega)$ under decentralization and centralization are, respectively:

$$Ec(M) = 2(\frac{eq_1}{(q_1-q_0)} + \theta) \tag{4}$$

$$Ec(U) = 2(\frac{e\sigma}{(1-\sigma)(q_1-q_0)} + \frac{eq_1}{(q_1-q_0)} + \frac{\theta}{2}) \tag{5}$$

Therefore restating the claim above: if $Ec(M) < Ec(U)$ then the city chooses a decentralized structure, if not it centralizes. This occurs when

$$\theta < \frac{2e\sigma}{(1-\sigma)(q_1-q_0)} \tag{6}$$

Thus the cost function (1) can be used to express the M-form Hypothesis as interpreted by Maskin et al. (2000) and others. Notice that this demonstration of the M-form Hypothesis did not rely on differences in the objective function resulting

from the organizational form. Although this result assumes there is no correlation of performance across services, assuming there was some, but that the correlation is less than across districts, would only result in a negative term being added to the left hand side of (6).

3.2 Decentralization Theorem

This section demonstrates the *Decentralization Theorem* borrowing the modeling framework from Besley and Coate (2003). The organization is divided into two entities, which may be geographically distinct cities in the case of a county or differentiated product lines in the case of a firm, indexed by $i \in \{1, 2\}$. There are three goods (x, g_1, g_2) and g_i is the "local public good" in entity i which can represent a single service, such as police service and parks in the case of a city, or it can represent an array of types of corporate public goods, such as accounting, advertising, etc. in the case of a firm.[7] x can alternatively be thought of as citizen endowment of income in the city case or in the firm case x can be revenue that accrues to the firm regardless of investment level in the public good For the moment, ignore the cost considerations of the previous subsection. To produce one unit of either of the local public goods requires p units of the "private good", where p is a linear public good cost/price.

The objective function for entity i is given by

$$x + \lambda_i[(1-k)\ln g_i + k \log g_{-i}] \tag{7}$$

where g_i refers to the level of the public good in entity i and g_{-i} refers to the level of the public good in entity $-i$. λ^i represents either intensity of preference for the public good by the city's median voter or the effectiveness of investment in the local public good (e.g. advertising by the firm). This is to say, if λ^i is high the median voters receives more utility from the public goods or the firm finds the marginal return to advertising is higher. The objective function is increasing in the size of the public good in the other district whenever $k > 0$. That is, there are externalities, and the extent of spillovers is measured by k. For example, citizens in district i may benefit from a high level of safety in the neighboring district or from visiting a park in district $-i$ or firm i may benefit from advertising done by firm $-i$ for example, through commercials that raise overall demand for the goods as a side effect.

Utility (or revenue) does not depend on structure directly, but structure influences public good levels and tax (or cost.) When $\Omega = U$, $g_i = g_{-i}$. That is, the organization chooses the same level for both entities. This is because a centralized decision maker may have difficulty differentiating between appropriate leaves in each entity, hence

[7] Local public goods are simply Samuelsonian public goods. However, only people in a subsection of a given geographic area receive the benefits of the public good. Below, there are spillovers, making these public goods in between local and pure public goods.

choosing a "one size fits all" policy arises due to bounded rationality.[8] When $\Omega = M$ then each entity chooses its level independently, but is only concerned about the wellbeing (profit) of its own entity.

This setup can be used to illustrate Oates Decentralization Theorem:

As a benchmark case, aggregate public good surplus with public good levels (g_1, g_2) is given by:

$$S(g_1, g_2) = [\lambda_1(1-k) + \lambda_2 k] \log g_1 + [\lambda_2(1-k) + \lambda_1 k] \log g_2 - p(g_1 + g_2) \quad (8)$$

It is readily checked that the surplus maximizing public good levels are given by equation (a) in table 1 below.

However when $\Omega = M$, districts choose levels of g independently and maximize

$$\lambda_i[(1-k) \log g_i + k \log g_{-i}] - pg_i \quad (9)$$

and the outcome is given by equation (b) in table 1.

Finally, when $\Omega = U$, a uniform public good level that maximizes aggregate social welfare is chosen (by assumption) to maximize aggregate public goods surplus, which maximizes

$$[\lambda_1 + \lambda_2] \log g - 2pg \quad (10)$$

This is simply aggregate public good surplus, with the constraint that $g_1 = g_2 = g$. Solving this problem yields the public goods levels shown in equation (c) in table 1.

Compared to a benchmark case (a) that maximizes aggregate public good surplus, the public good level when $\Omega = U$ is too high in one district and too low in the other unless $\lambda_1 = \lambda_2$. When $\Omega = M$, the public good level is too low compared to the benchmark case for both districts whenever $k > 0$ is large enough. A straightforward welfare comparison can now be made; characterizing it is somewhat complicated,[9] but the intuition comes through clearly. Besley and Coate (Proposition 1), recasts Oates' Decentralization Theorem:

A centralized structure is preferred when the public good preferences of the median voters in each district are similar and/or the spillovers are large, and a decentralized structure is preferred when the public good preferences of the median voters in each district are dissimilar and/or the spillovers are small.

[8] While in the firm context it may seem strange that managers know the average level but cannot assign a specific level to each division, this assumption may better reflect limits on what they can do. Perhaps they know different divisions should have different levels, but due to time or other constraints, they must provide a single level. The following section elaborates on this point.

[9] As a simple first order approximation, one can take the absolute value of the difference between the benchmark level (a) and the public good level in cases (b) and (c), respectively; whichever difference is smaller is, by this approximation, optimal.

Note that, to demonstrate Oates' Theorem, the illustration did not depend on differences in production technology that result from organizational form.

$$(g_1, g_2) = \left(\frac{\lambda_1(1-k)+\lambda_2 k}{p}, \frac{\lambda_2(1-k)+\lambda_1 k}{p}\right) \quad \text{(a)}$$

$$(g_1, g_2) = \left(\frac{\lambda_1(1-k)}{p}, \frac{\lambda_2(1-k)}{p}\right) \quad \text{(b)}$$

$$(g_1, g_2) = \left(\frac{\lambda_1+\lambda_2}{2p}, \frac{\lambda_1+\lambda_2}{2p}\right) \quad \text{(c)}$$

Table 1 Public good levels under three regimes

4 Conclusion

This paper has argued that the text-book description of Williamson's M-form Hypothesis misses some important components; in particular, it misses the tension in the delegation problem between optimal use of local and global knowledge. The remedy to this deficiency came from perhaps an unlikely place: the political economy literature on federalism. This paper showed how the modes of Maskin et al. (2000) and Besley and Coate (2003) together provide a superior representation of the theory Williamson intended. Moreover, these models are readily integrated, as the former comes through as a cost function, and the latter as an objective function. This provides a comprehensive view of delegation problems in general settings.

5 References

1. Alesina, A. and Spolaore, E, 1997. On the number and size of nations. *Quarterly Journal of Economics*, 112, 1027-1056

2. Besley, T. and Coate, S., 2003. Centralized versus decentralized provision of local public goods: a political economy approach. *Journal of Public Economics*

3. Che and Yoo, 2001. Optimal incentives in teams. *American Economic Review*

4. Koch, C., 2007. The Science of Success. *Wiley*

5. Kreps, D.M., 1996. Markets and hierarchies and (mathematical) economic theory." *Industrial and Corporate Change*, 5 p. 561.

6. Maskin, E. Qian, Y. and Xu, Y., 2000. Incentives, information and organizational form. *Review of Economic Studies*

7. Oates, W., 1972. *Fiscal Federalism* Harcourt Brace Jovanovich New York

8. Oates, W., 2005. Toward a second-generation theory of fiscal federalism. *International Tax and Public Finance*, 12, 349–373

9. Spulber, D., 2007. *Global Competitive Strategy.* Cambridge University Press.

10. Williamson, O., 1975. *Markets and Hierarchies* Free Press

6 Appendix: Derivation of expected bonus

This section shows how to derive the size of the bonuses the city must pay manager to elicit effort, using the framework of Che and Yoo (2001).

Under a centralized regime, the city has two options: pay the one manager a bonus only when the signal is good or only when it is bad. It is easy to show the latter option is not optimal. Minimization of the bonus goes on under the constraint that

$$[\sigma + (1-\sigma)q_1]w - e \geq [\sigma + (1-\sigma)q_0]w \tag{11}$$

This is an incentive compatibility constraint. Minimization of w implies that this constraint holds with equality, thus deriving (3). Under a decentralized regime, the city can make each of the two manager's bonuses depend on the signal of both. This is optimal, because the signal of the other contains information about the other. In this case, minimization under the constraint

$$(1-\sigma)q_1(1-q_1)w - e \geq (1-\sigma)q_0(1-q_1)w \tag{12}$$

And solving this as an equality results in the equation number (2). While the wage paid under a centralized regime is lower than the bonus paid in a decentralized regime, the former is paid more often than the later. Thus the city's expected wage bill from paying bonuses is equal to

$$\frac{e}{(1-\sigma)(1-q_1)(q_1-q_0)}(1-\sigma)q_1(1-q_1) = \frac{eq_1}{(q_1-q_0)} \tag{13}$$

under a decentralized regime, and

$$\frac{e}{(1-\sigma)(q_1-q_0)}(\sigma + (1-\sigma)q_1) = \frac{e\sigma}{(1-\sigma)(q_1-q_0)} + \frac{eq_1}{(q_1-q_0)} \tag{14}$$

under a centralized regime. Thus the cost of a bonus in a centralized regime is always greater whenever $\sigma > 0$, that is, whenever there is a common component to the signal.