

# Aggregation and Measurement Errors in Performance Evaluation

Anil Arya  
John C. Fellingham  
Douglas A. Schroeder  
*The Ohio State University*

**Abstract:** In this paper, we present a sequential production setting wherein employing aggregate measures for performance evaluation prove superior to those constructed specifically to measure individual activity. In our setting, unverifiable inputs translate into verifiable measures via two types of shocks: the first is production errors that cause outputs to deviate from inputs, and the second is measurement errors that result in outputs themselves being stated imprecisely. Agents are evaluated using either individual or aggregate measures, where the former measures the incremental output added by each link and the latter measures the cumulative output produced at the end of each stage. Aggregate measures can be preferred to individual measures because they increase the sample size available to infer upstream agents' unobservable acts and because they serve as an avenue for measurement errors to cancel.

## INTRODUCTION

The use of aggregate measures in performance evaluation is commonplace. Examples include rewarding team members based on group output, paying bonuses to workers when assembly line productivity exceeds established standards, and, more broadly, offering managers compensation contracts that are contingent on firm performance.<sup>1</sup> One reason for not always isolating each evaluatee's unpolluted output is that such measures are difficult (infeasible or too costly) to obtain. In this paper, we present another reason for why aggregate measures may be preferred to individual measures. When agents' inputs are subject to moral hazard and their outputs are subject to measurement errors, aggregate measures can be efficient because they increase the sample size available to infer upstream agents' unobservable acts and because they serve as an avenue for measurement errors to cancel.

We present our results in the context of a sequential production setting wherein the measurement question naturally arises: Is it desirable to measure the incremental output added by each link, or is it better to measure the aggregate output produced at the end of each stage? The output at the end of each stage is an aggregate number in that it depends on the cumulative effort of all upstream agents and on accumulated production shocks, i.e., it is the sum of individual (incremental) outputs.

---

We thank Rick Antle, Stan Baiman, Bala Balachandran, Joel Demski, Shane Dikolli, Ron Dye, Jon Glover, Yuji Ijiri, Pierre Liang, Brian Mittendorf, Mark Penno, Rick Young, two anonymous referees, and workshop participants at Carnegie Mellon University, Northwestern University, The Ohio State University, and Purdue University for helpful comments.

<sup>1</sup> Demski (1994) is an excellent example of an accounting text wherein performance evaluation and aggregation are a constant theme.

If outputs are available for contracting, then tracking either individual or aggregate outputs are equivalent since they are informationally identical. However, we allow for the possibility that the contracting variables differ from the outputs because the outputs are themselves subject to measurement errors. For brevity, we refer to the contracting variables in the aggregate (individual) output case as aggregate (individual) performance measures.

To see the main forces at work, consider a two-stage production chain. In this simple setup, aggregate measures, in contrast to individual measures, yield multiple observations of the upstream agent's effort. That is, under aggregation, the measure at the end of each stage is affected by the upstream agent's act, while with individual measures, the measure at the end of the second stage is not affected by the upstream agent's act. When measurement errors are uncorrelated, multiple observations reduce the risk imposed on the upstream agent. When measurement errors are significant, this benefit offsets the cost associated with using a more polluted (aggregate rather than individual) measure for evaluating the downstream agent. Alternatively, when production shocks dominate, measures of individual outputs, unpolluted by the other individual's activities, do better.

As the correlation in the measurement errors increases, the advantage of aggregate measures in providing multiple observations (and the related disadvantage of using polluted measures for the downstream agent) diminishes.<sup>2</sup> Now interest shifts to which system is more efficient in canceling measurement errors. In this case, aggregation is preferred if measurement errors are relatively small, and the benefit of aggregation is due to the reduced risk premium paid to the downstream agent. Given the upstream agent's performance, aggregate performance is more informative of the downstream agent's effort than his own individual performance. As an extreme case, note that perfectly correlated measurement errors cancel completely to precisely reveal the downstream agent's output when aggregate measures are employed. Informativeness depends on conditional controllability, not controllability (see, for example, Antle and Demski [1988] and Christensen and Demski [2003]).

Conditional controllability is the notion that variables may help in contracting with an individual even if the individual's actions do not directly affect the variable. This is because of "indirect" learning—the variable may inform about random shocks thereby permitting construction of a better proxy for the individual's actions. For example, a student's grade depends not only on his own exam performance, but also on the exam performance of his classmates. Presumably, relative grading helps lessen the impact of noise in the evaluation instrument. Similarly, a CEO's bonus is sometimes conditioned not only on his own firm's stock price, but also on the performance of the market. Again, the idea is to adjust for the common shock that affects all firms. In this paper, under aggregation, each agent's act directly impacts the realization of all measures that follow. So, not surprisingly, downstream measures can be valuable in creating a proxy for the agent's act. Moreover, the aggregation of production shocks and correlation in measurement errors imply that measures upstream to the agent may also be informative about the agent's action. It is this possibility of both direct and indirect learning that makes the analysis in the paper delicate.

Fortunately, the analysis is simplified by the equivalence between optimal linear incentive contracts and generalized least squares (GLS) estimates of the agents' unobservable acts. The close connection between statistics and control problems is well recognized. For example, sufficient statistics and informativeness are equivalent ideas (Holmstrom 1979). In our setting, the connection is again intuitive. In the control problem, the principal's goal is to come up with the best proxy for the agents' unobservable acts. In the LEN (linear contract-exponential utility-normal distribution) framework, the simple closed-form GLS estimator is the best proxy.

We note two related observations. First, when aggregate measures are optimal, *ex ante* identical agents are treated asymmetrically in that they are offered different contracts. Second, while the LEN

<sup>2</sup> The familiar statistical result that a large sample is associated with increased power is predicated on sample observations being independent.

setup allows for a crisp demonstration that neither measurement system always dominates, linearity does not drive this result as an example illustrates.

This paper augments a recurring theme in accounting on the merits of aggregation. A commonly cited reason to aggregate information is bounded rationality: limits on information transmission, reception, and processing can make aggregated information desirable. Benefits to aggregation, even with fully rational participants, include cancellation of errors in product costing (Datar and Gupta 1994), conveying information via choice of aggregation rule (Sunder 1997), protecting proprietary information (Newman and Sansing 1991), and substituting for commitment (Demski and Frimor 2000).

**MODEL**

Consider an  $n$ -stage sequential production process with  $n$  agents, one for each stage,  $n \geq 2$ . At stage  $i$ ,  $i = 1, \dots, n$ , agent  $i$  supplies an unobservable productive input (act),  $a_i$ , chosen from the set  $\{\underline{a}, \bar{a}\}$ ,  $\bar{a} > \underline{a}$ . Also, stage  $i$  is subject to a stochastic production shock  $\varepsilon_i$ . The output of each stage serves as an input to the following stage.

The incremental production (individual output) contributed by stage  $i$  is denoted  $x_i^I$ , and equals the sum of agent  $i$ 's act and the production shock  $\varepsilon_i$ .<sup>3</sup> The total production at stage  $i$  (aggregate output) is denoted  $x_i^A$ , and equals the sum of cumulative upstream acts and production shocks:

$$x_i^I = a_i + \varepsilon_i \text{ and } x_i^A = x_{i-1}^A + a_i + \varepsilon_i = \sum_{j=1}^i a_j + \sum_{j=1}^i \varepsilon_j.$$

At the end of each stage, in order to compensate the agents, one measurement is taken of the output (either individual or aggregate). We refer to measurement of  $x_i^k$  as  $m_i^k$ , where  $m_i^k$  potentially differs from  $x_i^k$  due to stage  $i$  measurement error  $e_i$ .<sup>4</sup>

$$m_i^k = x_i^k + e_i, \quad k = I, A.$$

Substituting for  $x_i^k$ , the system of equations for individual and aggregate performance measures can be represented in matrix notation as follows:

$$m^I = a + \varepsilon + e, \tag{1}$$

$$m^A = La + L\varepsilon + e, \tag{2}$$

where  $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$  and  $e = (e_1, e_2, \dots, e_n)$  are the (column) vectors of production and measurement shocks,  $a = (a_1, a_2, \dots, a_n)$  is the vector of acts chosen by the  $n$  agents, and  $L$  is the  $n \times n$  lower triangular matrix of ones.<sup>5</sup>

The  $2n \times 1$  error vector  $(e^k, e)$  is multivariate normally distributed with mean  $\theta$  and variance-covariance matrix  $V_k$ ,  $V_k = E[(e^k, e)(e^k, e)^T]$ .<sup>6</sup> Then, from Equations (1) and (2), the variance-covariance matrix in the individual and aggregate measurements cases,  $\Sigma_I$  and  $\Sigma_A$ , are:

<sup>3</sup> One may argue that individual measurements may not even be feasible; however, we show that even if feasible, they may not be optimal. Feasibility may be less of an issue when the production process has more features of a lateral setup (e.g., individual parts are tracked and produced and then assembled in one place). We thank a referee for this example.  
<sup>4</sup> In discussing different sources of randomness that heighten incentive problems, Milgrom and Roberts (1992, 207–208) write about “uncontrollable randomness in outcomes” (e.g., road construction reducing a restaurant’s sales) and that “performance evaluation measures include random or subjective elements” (e.g., due to sporadic monitoring or use of random samples to evaluate worker performance). Our modeling of production and measurement shocks is consistent with such randomness.  
<sup>5</sup> For notational convenience a column vector is sometimes written as an array whose elements are separated by commas. The  $L$ -matrix is a convenient way of representing the fact that production errors accumulate as one moves downstream in the production process.  
<sup>6</sup> With normal distributions one confronts the issue of negative variables. In our setting, an option is to view  $x_i^k$  as the true quality of output with negative numbers indicating poor quality. In this case,  $m_i^k$  is the measurement of quality, which may differ from true quality due to, say, sampling errors. More broadly, we view the  $x$ 's as a true but unobservable underlying trait and  $m$ 's as the observed measurement of that trait.

$$\Sigma_i = [I \ I] V_i [I \ I]^T \text{ and } \Sigma_A = [I \ L] V_A [I \ L]^T,$$

where  $I$  is the  $n \times n$  identity matrix. That is, in the individual measures case, the  $2n \times 2n$   $V_i$  matrix is pre-multiplied by a block matrix of size  $n \times 2n$  (since it consists of two side-by-side identity matrices) and post-multiplied by the transpose of the same block matrix. Note this yields  $\Sigma_p$ , the  $n \times n$  variance covariance matrix associated with the  $n$ -individual measures. In the aggregate measures case, the only difference is that the block matrix consists of  $I$  followed by  $L$ ; these are the coefficients on  $e$  and  $\varepsilon$  in Equation (2). Under either system, the variance-covariance matrix is free of the agents' acts. In contrast, the means of the measurements are act contingent and, hence, denoted  $E[m^k|a]$  and  $E[m^k|a]$ , where  $E[m^k|a] = a$  and  $E[m^k|a] = La$ . Denote the payment the principal makes to agent  $i$  under measurement system  $k$  by  $w_i^k$ . We assume  $w_i^k$  is linear in  $m^k$ , where  $m^k$  is the  $n$ -length vector of observables  $(m_1^k, \dots, m_n^k)$ :

$$w_i^k = \delta_i^k + \gamma_{i1}^k m_1^k + \gamma_{i2}^k m_2^k + \dots + \gamma_{in}^k m_n^k = \delta_i^k + \gamma_i^{kT} m^k.$$

Note that  $\gamma_i^k$  is the vector of weights used in agent  $i$ 's variable compensation.<sup>7</sup> The first subscript in the scalar  $\gamma_{ij}^k$  indicates that agent  $i$  is being compensated, the second subscript indicates the weight is placed on measurement  $j$ , and the superscript indicates performance measure system  $k$  is employed. Running the expectations and variance operators, we have:

$$E[w_i^k | \bar{a}] = \delta_i^k + \gamma_i^{kT} E[m^k | \bar{a}]$$

$$Var[w_i^k] = \gamma_i^{kT} \Sigma_k \gamma_i^k$$

where  $\bar{a}$  is a vector with each element equal to  $\bar{a}$ .

We assume that agent  $i$ 's preferences exhibit constant absolute risk aversion (CARA). In particular, agent  $i$ 's utility function is  $-e^{-r_i(w_i - a)}$ , where  $r_i > 0$  is the coefficient of absolute risk aversion. CARA preferences combine with normally distributed compensation to allow for a convenient certainty equivalent representation:  $E[w_i^k] - a_i - 0.5r_i Var[w_i^k]$ . Hence, the certainty equivalent of agent  $i$ , given all agents choose effort level  $\bar{a}$ , is

$$\delta_i^k + \gamma_i^{kT} E[m^k | \bar{a}] - \bar{a} - 0.5r_i \gamma_i^{kT} \Sigma_k \gamma_i^k.$$

The principal is risk-neutral. Her contracting problem under measurement system  $k$  is presented in program ( $P^k$ ). The principal minimizes the total expected payments subject to the following constraints.<sup>8</sup> First, agent  $i$  receives at least  $\theta_i$ , the certainty equivalent associated with his next best employment opportunity, i.e., the contract is individually rational ( $IR_i$ ). Second, it is in agent  $i$ 's best interest to choose  $\bar{a}$ , given all other agents choose  $\bar{a}$ , i.e., the contract is (Nash) incentive compatible ( $IC_i$ ).<sup>9,10</sup> (In ( $P^k$ ),  $\bar{a}_i$  is a  $n$ -length vector whose  $i$ th element is  $a$  and all other elements are  $\bar{a}$ .)

Program ( $P^k$ ): 
$$\text{Min}_{\delta_i^k, \gamma_i^k \forall i} \sum_{i=1}^n \delta_i^k + \gamma_i^{kT} E[m^k | \bar{a}]$$

s.t.

$$\delta_i^k + \gamma_i^{kT} E[m^k | \bar{a}] - \bar{a} - 0.5r_i \gamma_i^{kT} \Sigma_k \gamma_i^k \geq \theta_i \quad \forall i \tag{IR}_i$$

$$\gamma_i^{kT} E[m^k | \bar{a}] - \bar{a} \geq \gamma_i^{kT} E[m^k | \bar{a}_i] - a \quad \forall i \tag{IC}_i$$

<sup>7</sup> Holmstrom and Milgrom (1987) study a dynamic problem and prove it is equivalent to solving a static problem in which the agent chooses the mean of a normal distribution and in which the principal is restricted to using linear contracts. That is, in Holmstrom and Milgrom (1987) the use of linear contracts is without loss of generality.

<sup>8</sup> We assume the principal's expected benefit from production (the dollar value of the output) is sufficiently large that she chooses to hire all agents and induces each of them to supply  $\bar{a}$ .

<sup>9</sup> In writing ( $IC_i$ ), we assume measurements are revealed to the agents only at the completion of the entire production process, an assumption we later relax. Also, we write the ( $IC_i$ ) constraints after dropping the common terms which are independent of  $a_i$  from each side.

<sup>10</sup> Because the variance term is free of the agents' acts, the Nash incentive constraints can be costlessly replaced by dominant strategy incentive constraints. Furthermore, by increasing  $\gamma_{ii}^k$  by any arbitrary small positive amount, the dominant strategy incentives can be made strict. Hence, in our setting, there is no pressing tacit collusion (multiple equilibria) problem.

The solution to program ( $P^k$ ) is denoted by  $\delta_i^{k*}$  and  $\gamma_i^{k*}$ .<sup>11</sup>

**RESULTS**

The measurement systems are ranked by comparing the optimal objective function values under ( $P^I$ ) and ( $P^A$ ). Identifying the optimal contract under each program is facilitated by a connection between the control problem and the GLS problem of estimating the agents' unobservable acts.

The system of equations in (1) and (2) can be succinctly written as:

$$m^k = F^k a + F^k \epsilon + e^k, \text{ where } F^I = I \text{ and } F^A = L.$$

Plugging  $a_i$  for agent  $i$ 's act and  $\bar{a}$  for all other agents' acts implies:

$$m^k = F^k \bar{a} - F_i^k \bar{a} + F_i^k a_i + F^k \epsilon + e^k, \text{ where } F_i^k \text{ is the } i\text{th column of } F^k.$$

$$\Rightarrow m'_k \equiv m_k - F^k \bar{a} + F_i^k \bar{a} = F_i^k a_i + F^k \epsilon + e^k.$$

The left-hand side (LHS) of the last equation, consisting only of observables, is denoted  $m'_k$ .

The right-hand side (RHS) implies the variance of  $m'_k$  is  $\Sigma_k$ . The GLS estimate for  $a_i$  then follows from standard regression results (Greene 1997, 507).

**Observation:** Under measurement system  $k$ , the GLS estimate for agent  $i$ 's act is  $\omega_i^{kT} m'_k$ , where:

$$\omega_i^{kT} = (F_i^{kT} \Sigma_k^{-1} F_i^k)^{-1} F_i^{kT} \Sigma_k^{-1}$$

In a moral hazard problem, the principal seeks the most informative proxy for the unobservable act. The first part of our proposition confirms the intuition that the GLS estimate in the observation is the best proxy.

The equivalence of compensation weights with GLS weights also allows for a convenient way to compare the objective function values of ( $P^I$ ) and ( $P^A$ ). The efficiency loss (relative to first-best) is due to the risk borne by risk-averse agents. For an agent with CARA preferences, the risk premium is a scalar multiple of the variance of his compensation. From the GLS equivalence, the compensation variance is equal to the variance of the estimate of the unobservable action. This leads to part two of our proposition. (All proofs are provided in the Appendix.)

**Proposition:** (i) Under measurement system  $k$ , where  $k = I, A$ , the optimal contract is:

$$\gamma_i^{k*} = \omega_i^k$$

$$\delta_i^{k*} = \theta_i + \bar{a} + 0.5 r_i (F_i^{kT} \Sigma_k^{-1} F_i^k)^{-1} - \gamma_i^{k*T} F^k \bar{a}$$

(ii) Aggregate measures are strictly preferred if and only if:

$$\sum_{i=1}^n r_i (F_i^{AT} \Sigma_A^{-1} F_i^A)^{-1} < \sum_{i=1}^n r_i (F_i^{IT} \Sigma_I^{-1} F_i^I)^{-1}$$

The general form of  $V_k$  (and, hence,  $\Sigma_A$  and  $\Sigma_I$ ) leads to the mechanical appearance of the criteria in part (ii) of the Proposition. To provide intuition, we restrict attention in the remainder of the paper to a specific variance-covariance structure that underscores the role of measurement errors.

<sup>11</sup> The domain additivity of the utility function translates into a certainty equivalent that is additively separable. Our results are less dependent on our assumption that the disutility of  $a_i$  equals  $a_i$  rather than  $c(a_i)$ , where  $c$  is an increasing convex function. The optimal  $\gamma$ -weights in this more general case would be simply the  $\gamma$ -weights we obtain times  $\frac{c(\bar{a})-c(a)}{\bar{a}-a}$ , the marginal cost to the agent of choosing  $\bar{a}$  rather than  $a$ . On this point, also see Corollary 3.

Consider a two-stage ( $n = 2$ ) production process wherein each measurement error term is associated with variance  $\sigma_e^2$ , each production shock with variance  $\sigma_\epsilon^2$ , and all error terms, except for  $\epsilon_1^k$  and  $\epsilon_2^k$ , are uncorrelated. Let  $\rho$ ,  $-1 \leq \rho \leq 1$ , denote the correlation in the measurement errors. Also, let  $v$  denote the ratio of the measurement error to production shock variances, i.e.,  $v = \sigma_e^2 / \sigma_\epsilon^2$ . In this restricted setting, the proposition reduces to the following.

**Corollary 1:** Assume two identical agents and correlated measurement errors. Then:

(i) the optimal contract under each performance measurement system is:

	$\gamma_{11}^*$	$\gamma_{12}^*$	$\gamma_{21}^*$	$\gamma_{22}^*$
Agent 1	1	$-\frac{\rho v}{1+v}$	$\frac{1+(1-\rho)v}{1+2(1-\rho)v}$	$\frac{(1-\rho)v}{1+2(1-\rho)v}$
Agent 2	$-\frac{\rho v}{1+v}$	1	$-\frac{1+\rho v}{1+v}$	1

(ii) aggregate measurements are strictly preferred if and only if:

$$-1 < \rho < 0.5 \text{ and } v > \kappa, \text{ or}$$

$$0.5 < \rho \leq 1 \text{ and } v < \kappa, \text{ where } \kappa = \frac{(5-4\rho)^{1/2} + (1-2\rho)}{2(1-\rho^2)} \text{ if } \rho < 1.$$

$$= 1 \quad \text{if } \rho = 1.$$

The optimal contract in the two-agent case is essentially the standard “beta” regression coefficients (recall the GLS connection). With individual measures, the principal puts a weight of 1 (the marginal cost of effort) on agent  $i$ ’s own measure and a weight of negative beta ( $Cov[m_i^A, m_j^A]$ , the signal, divided by  $Var[m_j^A]$ , the noise) on the other agent’s measure. The same regression weights also apply in the aggregate measures case except that the weights are computed and applied to  $m_1^A$  and  $m_2^A - m_1^A$ .<sup>12</sup>

Under aggregate measures, it is optimal to treat even (*ex ante*) identical agents asymmetrically by offering them different compensation contracts.<sup>13</sup> Hence, unlike the case of individual measures, under aggregate measures, one agent bears more risk while the other agent bears less risk. However, the increased risk for one agent and the decreased risk for the other agent is not symmetric. The intuition for this result is particularly crisp when measurement errors are uncorrelated.

In the independent errors case ( $\rho = 0$ ), the use of aggregate measurements helps ease the control problem with the upstream agent (agent 1). The measurement at stage 2,  $m_2^A$ , provides information about the upstream agent’s action choice over and above the information provided by the measure at stage 1. There is both direct and indirect learning. The direct effect is that  $m_2^A$  depends on  $a_1$ . Indirect learning occurs because both  $m_1^A$  and  $m_2^A$  depend on the first-stage production shock. Hence,  $m_2^A$  is informative about  $a_1$ . Roughly stated, an aggregate system increases the size of the sample that is used to draw inference about the upstream agent’s act.

<sup>12</sup> In the aggregate case, the (IC) constraints for agent 1 and agent 2 are  $\gamma_{11}^* \bar{a} + \gamma_{12}^* (\bar{a} + \bar{a}) - \bar{a} \geq \gamma_{11}^* \underline{a} + \gamma_{12}^* (\underline{a} + \bar{a}) - \underline{a}$  and  $\gamma_{21}^* \bar{a} + \gamma_{22}^* (\bar{a} + \bar{a}) - \bar{a} \geq \gamma_{21}^* \underline{a} + \gamma_{22}^* (\underline{a} + \bar{a}) - \underline{a}$ , respectively. Solving these constraints as equalities yields  $\gamma_{11}^* + \gamma_{12}^* = 1$  and  $\gamma_{22}^* = 1$ , respectively. This weighting reflects the fact that while agent 1’s effort affects both measurements, agent 2’s act affects only the second measurement. Of course, this is not to say the first measurement provides no information about agent 1’s act. Informativeness and conditional controllability (not controllability) go hand in hand.

<sup>13</sup> When  $r_1 \neq r_2$ , Corollary 1 is unchanged except that  $\kappa = \frac{[4(1-\rho^2)R + (1-2(1-\rho)R) + 2(1-\rho)R - 1]}{2(1-\rho^2)}$  if  $\rho < 1$  and  $\kappa = R$  if  $\rho = 1$ , where  $R = r_2/r_1$ .

On the other hand, the use of aggregate measurements complicates the control problem with the downstream agent (agent 2). Under either individual or aggregate measurements, only the second observation depends on  $a_2$ , and hence there is no increased sample size benefit when contracting with agent 2. Instead, in contracting with the downstream agent, there is a cost associated with using the less informative aggregate measure. In particular,  $m_2^A$  reflects production shocks at both stages while  $m_2^I$  is subject to only stage 2's production shock.

More precisely, with  $\rho = 0$  under the optimal contract presented in Corollary 1, the benefit of aggregate measurements is the reduction in risk premium paid to agent 1:

$$\sigma_\varepsilon^2[(\gamma_{11}^A)^2(1 + \nu) + (\gamma_{12}^A)^2(1 + \nu)] - \sigma_\varepsilon^2[(\gamma_{11}^I)^2(1 + \nu) + (\gamma_{12}^I)^2(2 + \nu) + 2\gamma_{11}^I \gamma_{12}^I] = \frac{\nu^2}{1 + 2\nu} \sigma_\varepsilon^2 > 0.$$

Note that there is no common error term in the two measurements taken by the individual system; hence, there is no covariance term. In contrast, agent 1's production shock is common to both measurements under the aggregate system; the term  $2\gamma_{11}^I \gamma_{12}^I \sigma_\varepsilon^2$  represents this covariance.

The cost of aggregation is the increase in risk premium paid to agent 2:

$$\sigma_\varepsilon^2[(\gamma_{21}^A)^2(1 + \nu) + (\gamma_{22}^A)^2(2 + \nu) + 2\gamma_{21}^A \gamma_{22}^A] - \sigma_\varepsilon^2[(\gamma_{21}^I)^2(1 + \nu) + (\gamma_{22}^I)^2(1 + \nu)] = \frac{\nu}{1 + \nu} \sigma_\varepsilon^2 > 0.$$

The difference between the benefit and the cost of aggregate measurements can be simplified as  $\frac{(\nu + g - 1)(\nu - g)\nu}{(1 + \nu)(1 + 2\nu)} \sigma_\varepsilon^2$  where  $g$  is the golden ratio.<sup>14</sup> The benefit exceeds the cost if and only if  $\nu > g$ .

That is, when  $\rho = 0$  and  $\nu > g$ , the benefit of using multiple observations to evaluate agent 1 dominates the cost of using a more polluted measurement for agent 2, thereby making aggregate measures optimal.

Next, consider the other extreme of perfectly correlated measurement errors. With  $\rho = 1$ , the issue becomes which system is more efficient in canceling measurement errors. The system ranking reverses as an aggregate system is preferred for relatively small measurement errors, while an individual system dominates for larger measurement errors. Restating the contract weights in Corollary 1 after substituting  $\rho = 1$  reveals the intuition for this result:

	$\gamma_{11}^A$	$\gamma_{12}^A$	$\gamma_{11}^I$	$\gamma_{12}^I$
Agent 1	1	$-\frac{\nu}{1 + \nu}$	1	0
Agent 2	$-\frac{\nu}{1 + \nu}$	1	-1	1

A negative sign on the other agent's measurement is indicative of the optimal contract's attempt to cancel measurement errors. Unlike the  $\rho = 0$  case, with  $\rho = 1$  the benefit (cost) of using aggregate measurements arises in contracting with agent 2 (agent 1). With aggregate measures, the weights are chosen so as to completely adjust for the measurement error and leave only one production error term in agent 2's evaluation. In particular, the last two column weights in the last row show that agent 2 is evaluated using  $m_2^A - m_1^A = a_2 + \varepsilon_2$ . However, with individual measurements, if the weights were similarly chosen to completely eliminate the measurement error for agent 2, the result would be two production shock terms:  $m_2^I - m_1^I = a_2 - a_1 + \varepsilon_2 - \varepsilon_1$ . The production shock variance would be twice as much as with aggregate measurements. Hence, with aggregate measures, a lower risk premium is needed to compensate agent 2.

<sup>14</sup> The golden ratio is the limit of the ratios of successive Fibonacci numbers as the numbers get large. The golden ratio is  $(1 + \sqrt{5})/2 \approx 1.61803$ .

In fact, with individual measures, it is never optimal to completely eliminate measurement error from any one agent but rather to let both agents share some of the error (and the associated risk). In contrast, under aggregate measures, agent 1 alone bears the measurement error. Risk sharing is more valuable the more significant measurement errors are. Not surprisingly, then, with  $\rho = 1$ , aggregate measures are preferred only for small  $v$  values.

Figure 1 summarizes the result in Corollary 1.

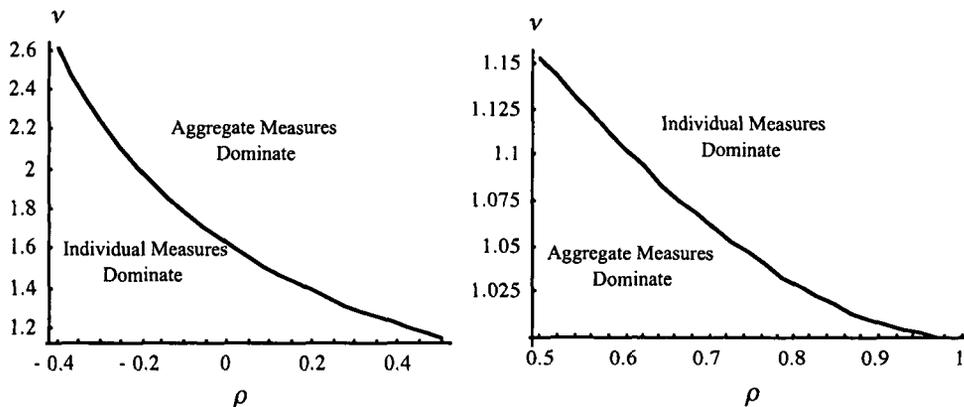
The forces at work at the extreme values of  $\rho = 0$  and  $\rho = 1$  are also at work when  $\rho$  takes on intermediate values.<sup>15</sup> From Corollary 1,  $\gamma_{11}^{A*} + \gamma_{12}^{A*} = 1$ . As  $\rho$  increases, the relative informativeness of  $m_2^A$  regarding  $a_1$  diminishes and, hence,  $\gamma_{12}^{A*}$  decreases. In the extreme, when  $\rho = 1$ ,  $m_2^A$  becomes a garbled signal of  $m_1^A$  and is useless in providing information on  $a_1$ , hence,  $\gamma_{12}^{A*} = 0$ . In short, the benefit of increased sample sizes diminishes as  $\rho$  increases.

For agent 2, when  $\rho = 0$ ,  $m_2^A$  is polluted with respect to  $a_2$  (compared to  $m_2^I$ ). The polluted component " $a_1 + \varepsilon_1$ " can be partially filtered out by assigning a negative weight to  $m_1^A$ , i.e.,  $\gamma_{21}^{A*} < 0$ . As  $\rho$  increases,  $|\gamma_{21}^{A*}|$  increases, consistent with the fact that benefits due to canceled errors increase. Thus, aggregate measures can be preferred because of the sample size effect and because they provide an avenue for measurement errors to cancel.

While the above two-agent discussion has been provided in the context of sequential production, there may be other settings to which the result is applicable. For example, in practice, some firms carefully track performance of a few, but not all, of their divisions. A reason cited for this practice is that tracking and measuring performance is a costly activity and, hence, it might be more efficient to carefully scrutinize the performance of only those divisions that are vital to the firm's operations.

The result in Corollary 1 can be interpreted as suggesting that hierarchical measurements may also be preferred if the measurement process itself introduces noise. There may be other benefits of

**FIGURE 1**  
Parameter Values Under Which Each Measurement System Dominates



<sup>15</sup> We thank a referee for this intuition.

such measurements, namely, that the use of firm profits in each manager's evaluation may help him internalize firm-wide implications of his actions (Bushman et al. 1995). More broadly stated, the creation and use of managerial hierarchies, which is somewhat akin to sequencing, may be beneficial to a firm as this organizational structure facilitates the natural development of aggregate (hierarchical) performance measures. Managers at each stage are evaluated based on overall performance as well as the performance of subordinates working for them and so on.

The arguments provided in the two-agent setting apply to the  $n$ -agent case as well. In particular, the next corollary presents necessary and sufficient conditions under which aggregate measures are preferred in the uncorrelated and the perfectly correlated measurement errors cases.

**Corollary 2:** Assume  $n$  identical agents. Then:

- (i) with uncorrelated measurement errors, aggregate measurements are strictly preferred if and only if  $\nu > \kappa$ , where  $\kappa$  solves

$$\sum_{i=1}^n [F_i^{A'} (F^A F^{A'} + \kappa F^A)^{-1} F_i^A]^{-1} = n(1 + \kappa).$$

- (ii) with perfectly correlated measurement errors, aggregate measurements are strictly preferred if and only if  $\nu < 1$ .

**EXTENSIONS**

**Nonlinear Production and Compensation**

In the previous section, we confined attention to a linear compensation rule and a linear production function. However, our basic result that neither system dominates holds even when the linearity assumptions are relaxed.

Let  $x_i, x_i' \in \{\underline{x}, \bar{x}\}$ , denote the individual (incremental) output contributed by agent  $i, i = 1, 2$ .  $x_i'$  depends on agent  $i$ 's act and the random productivity parameter  $\varepsilon_i$ . We suppress  $\varepsilon_i$  and focus directly on  $Pr(x_i' | a)$ , the probability distribution over the output, and consider either individual or aggregate measures to evaluate the agents. Under either system,  $m_i^k = x_i + e_i^k$ , where the zero mean measurement error  $e_i^k \in \{\underline{e}, \bar{e}\}, k = I, A$ .

Suppose  $\underline{x} = 300, \bar{x} = 600, \underline{a} = 0, \bar{a} = 15, Pr(\bar{x} | \bar{a}) = 0.8, Pr(\bar{x} | \underline{a}) = 0.5$ , and  $\theta_i = 0$ . Also, let  $\{\underline{e}, \bar{e}\} = \{-150, 150\}$ , with each value equally likely. Finally, assume the measurement errors are independent of each other, and independent of the production shocks. (Despite the binary structure of the variables and error independence, the optimal contract in the aggregate case requires solving for ten payments.)

Table 1 presents the principal's expected payments under both performance measurement systems corresponding to two different sets of agents' risk aversion coefficients. In each case, "\*" denotes the optimal system.

**TABLE 1**  
**Optimal Expected Payments under Each Performance Measurement System**  
**and under Alternative Risk Aversion Assumptions**

	Panel A: $r_1 = 0.01$ and $r_2 = 0.01$		Panel B: $r_1 = 0.06$ and $r_2 = 0.01$	
	Individual*	Aggregate	Individual	Aggregate*
Agent 1	18.326	17.808	65.327	64.251
Agent 2	18.326	19.027	18.326	19.027
Total	36.652	36.835	83.653	83.278

\* Optimal Measurement System.

As expected, for both sets of risk aversion parameters in Table 1, the multiple observations provided by the aggregate system reduce the expected cost of contracting with agent 1. Of course, this benefit must be compared against the increased cost of contracting with agent 2 that results because the measurement used to infer agent 2's act is more polluted under the aggregate system. The parameters are identical in the two panels of Table 1 except that agent 1's risk aversion coefficient is higher in Panel B, making the control problem with agent 1 more severe in that case. Hence, aggregate measurements are strictly preferred in Panel B.<sup>16</sup>

### Continuous Acts

While our earlier Proposition is stated for binary act choice, the paper's results extend in an analogous manner to the continuous act case. Assume  $a_i$  is selected from the interval  $[\underline{a}, \bar{a}]$  and let  $a_i^*$  denote agent  $i$ 's optimal action. To ensure the optimal acts are interior, assume agent  $i$ 's personal cost of choosing  $a_i$  is  $c(a_i)$ , where  $c$  is an increasing convex function. In effect, agent  $i$ 's utility function is  $-e^{-\gamma(m_i - c(a_i))}$ . The next corollary states that the optimal contract is essentially unchanged in this more general action environment.

**Corollary 3:** In the continuous action environment, the optimal linear contract is the same as that presented in the Proposition except that the variable compensation weights (the  $\gamma$ -weights) are multiplied by the marginal cost of effort  $c'(a_i^*)$ , where  $a_i^*$  is the value that solves  $\gamma_i^k F_i^k - c'(a_i) = 0$ .

### Timing

So far we have assumed that the agents observe measurements only at the end of the production process. A final corollary states that the result is unaffected if this assumption is relaxed, i.e., if measures at each stage become public as soon as the stage is completed.

**Corollary 4:** Assume agent  $i$  observes  $m_1^k, m_2^k, \dots, m_{i-1}^k$  prior to choosing his own act. In this case the optimal linear contract is the same as that presented in the Proposition (for binary acts) or as that presented in Corollary 3 (for continuous acts).

## CONCLUSION

Performance evaluation when agents' inputs are subject to moral hazard and their outputs are subject to measurement errors is a delicate exercise. In this paper, we study such a setting and identify conditions under which individual and aggregate performance measurement systems are each optimal. Roughly stated, aggregate measurements can be efficient because they increase the sample size available to infer the upstream agents' unobservable acts and because they provide an avenue for measurement errors to cancel.

The simultaneous modeling of both moral hazard and measurement errors was necessary for the aggregate system to be optimal. In the absence of moral hazard, the principal can directly contract on the agent's input supply. In the absence of measurement errors, the aggregate system is simply a linear transformation of the individual system.

Accounting systems maintain an archive of historical data; aggregation is a choice variable in the design of the archive. Furthermore, it is conventional for accountants to engage in data compression, i.e., to calculate and disseminate a restricted set of possible measures in the form of flow and stock variables. An extension to this paper would be to evaluate the information and incentive consequences of such time-aggregated data.

<sup>16</sup> If the two aggregate measures are interpreted as one firm-wide and one divisional performance measure, then which division's performance to track becomes an interesting question. Under this interpretation, in our example, the principal prefers to track the divisional performance of the more risk-averse agent (agent 1). In fact, if instead agent 2's divisional performance is tracked, then individual measures outperform aggregate measures.

APPENDIX

Proof of the Proposition

Consider the solution to (P<sup>k</sup>). The (IC<sub>i</sub>) constraints can be equivalently rewritten as  $y_i^{kT} F_i^k - 1 \geq 0$ . At the optimal solution, clearly the (IR<sub>i</sub>) constraint holds as an equality. Solving for expected payments from the (IR<sub>i</sub>) constraint yields:  $E[w_i^k] = \theta_i + \bar{a} + 0.5r_i Var[w_i^k]$ . Hence, the Lagrangian for (P<sup>k</sup>) is as follows, where  $\lambda_i$  denotes the multiplier on the (IC<sub>i</sub>) constraint:

$$\mathcal{L} = \sum_{i=1}^n [\theta_i + \bar{a} + 0.5r_i y_i^{kT} \Sigma_k y_i^k - \lambda_i (y_i^{kT} F_i^k - 1)]. \tag{3}$$

The first-order conditions of Equation (3) with respect to  $y_i^k$  and  $\lambda_i$  are:

$$\frac{\partial \mathcal{L}_i}{\partial y_i^k} = r_i \Sigma_k y_i^k - \lambda_i F_i^k = 0. \tag{4}$$

$$\frac{\partial \mathcal{L}_i}{\partial \lambda_i} = y_i^{kT} F_i^k - 1 = 0. \tag{5}$$

From Equation (4),  $y_i^{kT} = \frac{1}{r_i} \lambda_i F_i^{kT} \Sigma_k^{-1}$ . Substituting this in Equation (5) and solving for  $\lambda_i$  yields  $\lambda_i = r_i (F_i^{kT} \Sigma_k^{-1} F_i^k)^{-1}$ . Because  $\lambda_i > 0$ , it follows that the (IC<sub>i</sub>) constraints bind. Substituting this value of  $\lambda_i$  back into the expression for  $y_i^{kT}$  yields  $\omega_i^k$ .  $\delta_i^{k*}$  is obtained by solving the (IR<sub>i</sub>) constraint as an equality.

The risk premium paid to agent  $i$  is  $0.5r_i Var[w_i^k]$ . Substituting the optimal compensation weights derived above, implies the risk premium is  $0.5r_i (F_i^{kT} \Sigma_k^{-1} F_i^k)^{-1}$ . The Proposition then simply compares the sum of the risk premium paid to the  $n$  agents under the two measurement systems. ■

Proof of Corollary 1

In the two-agent correlated measurement errors case:

$$V_k = \sigma_e^2 \begin{bmatrix} v & \rho v & 0 & 0 \\ \rho v & v & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The above matrix defines  $\Sigma_k$  (see the formulas in the model section of the paper). The result then follows by substituting  $\Sigma_k$  and  $r_1 = r_2$  in the proposition. ■

Proof of Corollary 2

With uncorrelated measurement errors,  $\Sigma_i = \sigma_e^2 F^i + \sigma_e^2 F^i$  and  $\Sigma_A = \sigma_e^2 F^A F^{AT} + \sigma_e^2 F^A$ . Substituting these expressions in part (ii) of the Proposition, and simplifying, implies aggregate measures are preferred if and only if:

$$\sum_{i=1}^n [F_i^{AT} (F^A F^{AT} + \sigma_e^2 F^i)^{-1} F_i^A]^{-1} - n(1 + \sigma_e^2) < 0. \tag{6}$$

Part (i) of the corollary then follows from the following (detailed proofs of (a) – (c) are available from the authors):

- (a) The derivative of Equation (6) with respect to  $\sigma_e^2$  is positive when evaluated at  $\sigma_e^2 = 0$ ,
- (b) Expression (6) is concave with respect to  $\sigma_e^2$ , and
- (c) The limit of (6) as  $\sigma_e^2$  goes to infinity is negative infinity.

Part (ii) of the corollary can be proved by verifying that the optimal contract in the perfectly correlated is much as in the two-agent case, and therefore part (ii) of Corollary 1 applies. In particular, with individual measures, the optimal contract for agent  $i$  puts a weight of 1 on  $i$ 's own measure and  $-\frac{v}{1+(n-1)v}$  on all other individual measures. With aggregate measures, the optimal contract for agent  $i, i > 1$ , places a weight of 1 on stage  $i$ 's measure,  $-1$  on stage  $i-1$ 's measure, and

0 on all other measures. For agent 1, the optimal contract places a weight of 1 on the first measure and 0 on all other aggregate measures. ■

### Proof of Corollary 3

In the continuous act setup, the  $(IC_i)$  constraint is:

$$a_i \in \underset{a_i}{\operatorname{argmax}} E[w_i^k | (a_{-i}^*, a_i)] - a_i - \operatorname{Var}[w_i^k].$$

Assuming the first order approach is valid, the  $(IC_i)$  constraint can be replaced by the following condition  $\gamma_i^{k'} F_i^k - c'(a_i) = 0$ . This constraint is the same as the  $(IC_i)$  constraint in Equation (5) except that the  $\gamma$ -weights add to the marginal cost  $c'(a^*)$  rather than to 1.

### Proof of Corollary 4

Consider the contracting problem with agent  $i$  if  $i$  observes  $m_1^A, \dots, m_{i-1}^A$  prior to choosing  $a_i$ . (We provide only the proof for the aggregate case; the individual case follows directly.) The objective function and  $(IR_i)$  are the same as before since, at the time of contracting,  $m$  is not known to either party. To consider the effect on  $(IC_i)$ , partition the measures into  $(m_a, m_b)$ , where  $m_a = m_1^A, \dots, m_{i-1}^A$  and  $m_b = m_i^A, \dots, m_n^A$ . The unconditional mean,  $La$ , and variance-covariance matrix,  $\Sigma_A$ , can similarly be partitioned as follows:

$$La = \begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix} \quad \text{and} \quad \Sigma_A = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}$$

The distribution of  $m_b$  conditional on observing  $m_a$  is multivariate normal with mean and variance-covariance matrix as listed below (Greene 1997, 90):

$$E[m_b | m_a] = \mu_b + \Sigma_{aa}^{-1} \Sigma_{ab} (m_a - \mu_a) \quad \text{and} \quad \operatorname{Var}[m_b | m_a] = \Sigma_{bb} - \Sigma_{ba} \Sigma_{aa}^{-1} \Sigma_{ab}.$$

As the conditional distribution is multivariate normal we can continue to use the convenient certainty equivalent expression used before. Since  $\operatorname{Var}[m_b | m_a]$  (and  $\delta_i^A$ ) is free of  $a_i$ , it drops out of  $(IC_i)$  as before. Agent  $i$ 's own act affects  $E[m_b | m_a]$ , but only by its effect on  $\mu_b$ ; the second term in  $E[m_b | m_a]$  is based on the realization  $m_a$  and  $\mu_a$ , which are free of  $a_i$ . Hence, if agent  $i$  chooses  $\bar{a}$  rather than  $\underline{a}$ , he increases the conditional mean by  $\bar{a} - \underline{a}$ . In certainty equivalent terms, by choosing  $\bar{a}$  rather than  $\underline{a}$ , agent  $i$  increases his compensation by  $\gamma_i^{A'} F_i^A (\bar{a} - \underline{a})$  and increases his disutility by  $\bar{a} - \underline{a}$ . Hence,  $(IC_i)$  is the same as before:  $\gamma_i^{A'} F_i^A \geq 1$ . ■

### REFERENCES

- Antle, R., and J. Demski. 1988. The controllability principle in responsibility accounting. *The Accounting Review* 63 (October): 700-717.
- Bushman, R., R. Indjejikian, and A. Smith. 1995. Aggregate performance measures in business unit manager compensation: The role of intrafirm interdependencies. *Journal of Accounting Research* 33 (Supplement): 101-128.
- Christensen, J., and J. Demski. 2003. *Accounting Theory: An Information Content Perspective*. New York, NY: McGraw-Hill Irwin.
- Datar, S., and M. Gupta. 1994. Aggregation, specification and measurement errors in product costing. *The Accounting Review* 69: 567-591.
- Demski, J. 1994. *Managerial Uses of Accounting Information*. Norwell, MA: Kluwer Academic.
- , and H. Frimor. 2000. Performance measure garbling under renegotiation in multi-period agencies. *Journal of Accounting Research* 37 (Supplement): 187-214.
- Greene, W. 1997. *Econometric Analysis*. Englewood Cliffs, NJ: Prentice Hall.

- Holmstrom, B. 1979. Moral hazard and observability. *Bell Journal of Economics*: 74–91.
- , and P. Milgrom. 1987. Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*: 303–328.
- Milgrom, P., and J. Roberts. 1992. *Economics, Organization, and Management*. Englewood Cliffs, NJ: Prentice Hall.
- Newman, P., and R. Sansing. 1991. Disclosure policies with multiple users. *Journal of Accounting Research* 31: 92–112.
- Sunder, S. 1997. *Theory of Accounting and Control*. Cincinnati, OH: International Thomson.

Copyright of Journal of Management Accounting Research is the property of American Accounting Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.