Optimal Forecasting Incentives

Kent Osband

Rand Corporation

An agent of unknown expertise is requested to forecast the mean of an uncertain outcome. The agent can refine forecasts at a constant marginal cost per unit precision, but neither cost nor precision can be verified by the planner. The problem is to induce both truthful revelation and an appropriate degree of learning so as to minimize the expected sum of direct planning losses and agent payments. Optimal contracts are derived with and without self-screening of expertise and with and without competition between agents. Self-screening tends to be much less valuable than competition.

I. Introduction

When outcomes are uncertain, planning must be based on forecasts—quite often, on forecasts submitted by others. Naturally, the planner wishes to ensure that these forecasts are prepared honestly and with an appropriate degree of care. But how is this to be done? Even if the outcome departs significantly from what was predicted, one can rarely conclusively infer that the forecaster was intentionally negligent or deceptive. To take extreme examples, a space shuttle explosion or major nuclear reactor accident was deemed very unlikely. Now that these events have occurred, forecasts of the likelihood of similar accidents will presumably be revised upward. This need not mean that previous forecasters were lazy or even that their probability forecasts were necessarily "wrong." On the other hand, it is possible that forecasting incentives were skewed toward careless or overly optimistic reports.

I wish to thank Richard Gilbert, Stefan Reichelstein, Eric Maskin, and especially Jean Tirole for criticisms of earlier drafts of this paper.

[*Journal of Political Economy*, 1989, vol. 97, no. 5]

Considering the topic's practical importance, remarkably little has been written about forecasting incentives. What has been written focuses on incentives for *honest* reporting, without regard for other incentive issues. The agent's utility compensation, often called the "score" in statistical decision theory, is tied to the forecast and to the future observed outcome. The agent is presumed to choose the forecast so as to maximize the expected score. Scoring rules are called "proper" if they always encourage sincere reporting, regardless of the agent's underlying beliefs. Savage (1971), Thomson (1979), Haim (1982), and Osband and Reichelstein (1985) characterize the full set of proper scoring rules ("incentive-compatible compensation schemes," in the terminology of principal-agent literature) for specific forecasting measures such as the mean or median. In Osband (1985), this characterization is extended to arbitrary types of forecasts.

Almost without exception, these articles skirt issues of optimality: the choice of rules within a class. Presumably the planner would be interested in minimizing expected payoff for truthful reports subject to some sort of income or expected income floor for the agent. But this consideration does not by itself lead to a very interesting choice. Given a forecast Y (which may be a vector) and outcome x, let the net utility payoff for a bounded proper scoring rule be denoted by $H(Y, x)$, with B as a lower bound. Then for any positive ϵ and any Z, $\epsilon[H(Y, x) - B] + Z$ is strictly proper with utility payoffs of at least Z. By setting Z at the agent's minimum acceptable expected utility and choosing ϵ close to zero, strictly proper scoring rules can be squeezed arbitrarily close to the minimum—and only weakly proper—flat fee.

Scoring rules appear superfluous in this instance for two reasons. First, the forecaster is assumed to know the outcome distribution without incurring any learning costs, or at least without choosing what those costs will be. Second, the forecaster is assumed unable to influence that distribution by exerting or withholding effort. Relax either of these assumptions and a real trade-off arises between information transfer costs and incentives for effort.

In this article agent learning costs are explicitly incorporated into a forecast elicitation model, while the assumption of agent-independent outcomes is maintained. The forecaster, who is asked to report the mean of an uncertain or random outcome, will begin with a rough estimate. By exerting additional effort, the forecaster can refine this estimate to any desired degree of precision, at a cost varying with his expertise. The planner may not know how expert the forecaster is. Contracts for inducing an expected degree of learning are derived, and their administrative costs analyzed. At the optimum, the marginal expected benefit to the planner from more precise estimates must be balanced by the marginal expected administrative cost.

Section II presents the basic model. The model incorporates adverse selection (the unknown expertise of the forecaster), moral hazard (the unobserved learning effort), and "white noise" (the inherent unpredictability of the outcome). Given the multiple layers of uncertainty, one might suspect that the optimal payoff contract would be very complicated. As it turns out, the best contract, provided forecasters are risk-neutral, is quite simple: quadratic in the reported mean and linear in the outcome, with easily calculated parameters.

When expertise is known, the optimal contract induces a first-best solution: expected total planning cost is no more than it would be with complete and costless monitoring. But when, as is likely, the forecaster's learning expertise is not precisely known, more expert forecasters stand to earn an "information rent." To reduce these rents, the planner tends to sacrifice forecasting precision. The net efficiency loss can be substantial, Section III shows: up to 40 percent of total planning cost, given a uniform prior on expertise.

Section IV investigates the cost-reducing potential of "self-screening" contracts, which induce forecasters to reveal their expertise beforehand. Efficiency gains turn out to be remarkably meager, less than one part in 500 for any uniform expertise distribution. Indeed, for many prior distributions of expertise, self-screening offers no advantages at all.

Section V puts the preceding results into a more general principal-agent perspective. Forecaster self-screening is reinterpreted, somewhat surprisingly, as a special case of the Laffont-Tirole (1986) model of a regulated firm with ex post observable costs. The reinterpretation helps to account for the stiff second-order conditions.

Section VI explores the potential for using competition among forecasters to reduce planning costs. Not only does competition offer significant savings over monopolistic self-screening, but it also considerably simplifies contract implementation. In some cases, contracting is as simple as selecting a single parameter in a quadratic reward schedule and then auctioning off the right to be paid according to this schedule.

II. The Model

Suppose that a planner must rely on some estimate Y of an uncertain outcome (an exchange rate, say) in order to formulate a plan. Should the actual value x deviate from the estimate, an opportunity loss L is incurred proportional to the square of the discrepancy. Thus $L(Y, x) = c(Y - x)^2$ for some positive c. If the true distribution of x has mean μ and variance σ^2, expected planning loss is $c(Y - \mu)^2 + c\sigma^2$. It is minimized at $Y = \mu$, leaving only $c\sigma^2$, the planning loss anticipated from inherent outcome riskiness.

Suppose that the planner does not know μ exactly. Instead she possesses some initial estimate Y_0 of μ, believed to be an unbiased predictor of μ with precision (the inverse of the variance) r. Overall expected planning loss using Y_0 is $c\sigma^2 + (c/r)$. To reduce this, the planner hires a forecaster. The forecaster, who is presumed to be risk-neutral, begins with the same prior as the manager but can avail himself of a forecasting technology. This technology increases forecast precision at cost b per unit precision. (For example, suppose that the forecaster takes independent unbiased "measurements" of μ at cost b apiece, with each measurement having variance one, and forms the best linear unbiased estimator [BLUE] on the basis of the results. After $n - r$ measurements, the BLUE will have precision n.) The forecaster knows b, which will be called his "expertise" (so that more expert forecasters have lower b). He is risk-neutral and willing to enter any contract offering an expected profit of at least zero.

Let n denote final estimate precision. If the planner could costlessly verify n and b, she would choose n to minimize $b(n - r) + (c/n) + c\sigma^2$, where the first term represents direct forecasting costs. This is achieved at $n^* = (c/b)^{1/2}$, for total expected planning cost $C(n^*)$ of $-br + c\sigma^2 + 2(bc)^{1/2}$. This is the "first-best" solution.

Of the three components of $C(n^*)$, the first reflects the costlessness of the first r units of precision, while the second stems from the inherent randomness of x. The remaining component, which will be called the expected administrative cost (EAC), includes both "full" measurement costs (inclusive of r) and the expected loss from having an imperfect estimate of μ, with each contributing half. The EAC rises with measurement cost and the cost of forecast error, but proportionately only half as fast as either.

If only b could be directly verified, the first-best solution could still be obtained as follows. For report $\tilde{Y}$, consider the payment schedule ("contract")

$$H(\tilde{Y}, x) = -c(\tilde{Y} - x)^2 - br + c\sigma^2 + 2(bc)^{1/2}. \tag{1}$$

Here the forecaster bears the planner's entire loss and receives enough side compensation to achieve the reservation expected utility zero. Writing this contract requires knowledge of σ^2, however. Should σ^2 not be known, the same effect can be achieved with the contract

$$H(\tilde{Y}, x) = c(2x\tilde{Y} - \tilde{Y}^2) - br - c\left(Y_0^2 + \frac{1}{r}\right) + 2(bc)^{1/2}. \tag{1'}$$

Despite the differences of form, contracts (1) and (1′) are essentially the same. Both consist of a multiple c of $(2x\tilde{Y} - \tilde{Y}^2)$ for incentive purposes and an entrance fee (possibly random but always independent of report) to tax away expected rent. The two entrance fees differ by $c[Y_0^2 + \sigma^2 + (1/r) - x^2]$, for which the expectation conditional on Y_0 is zero.

More realistically, suppose that the planner cannot directly verify b, n, or Y. Let the planner offer the forecaster a contract

$$H(\tilde{Y}, x) = Q(2x\tilde{Y} - \tilde{Y}^2) - R + S(Y_0 - x), \tag{2}$$

for some Q, R, and S. In Appendix A, this is shown to be the only contract type possessing all three of the following features: (*a*) It always induces truthful reporting of Y; that is, for any distribution of x, expected reward is maximized by reporting the perceived mean. (*b*) Expected payoff depends only on the mean, not on the variance or any other aspect of the distribution. (*c*) Induced final estimate precision depends only on b, not on any other aspect of the forecasting technology. The term in S amounts to a fair lottery in x and can for our purposes be ignored. The term R is the entrance fee, while the degree of forecasting care is determined completely by Q. If $Y = \mu + h$ is forecast instead of $Y = \mu$, expected payoff will be reduced by Qh^2, so higher Q induces more diligent forecasting.

The simplicity of the contracts above coupled with their satisfaction of properties *a*–*c* makes them particularly easy to administer. The planner has only to select an appropriate Q and R in (2). In fact, as we shall see later, a contract of type (2) is optimal in the class of all contracts.

Let the planner's beliefs about b be described by a cumulative distribution function $G(\cdot)$, so that $G(b_0)$ denotes the subjective probability that b does not exceed b_0, and let β be the perceived upper bound to b. If the planner is committed to hiring the forecaster regardless of type, she should adjust (2) until a type β forecaster would be indifferent between working and not working. This implies an entrance fee of

$$R = \beta r + Q\left(Y_0^2 + \frac{1}{r}\right) - 2(\beta Q)^{1/2} \tag{3}$$

for any chosen Q, by analogy with (1′).

Since (1) induces final estimate precision $(c/b)^{1/2}$ for an agent of expertise b, (2) must induce final estimate precision $n(b) = (Q/b)^{1/2}$. Forecaster b's gross expected payoff is

$$-\frac{Q}{n(b)} - \beta r + 2(\beta Q)^{1/2} = Q^{1/2}(2\beta^{1/2} - b^{1/2}) - \beta r.$$

Subtracting the investigation cost $b[n(b) - r]$ yields the net expected payoff

$$2Q^{1/2}(\beta^{1/2} - b^{1/2}) - r(\beta - b). \tag{4}$$

Thus for a given b, expected forecasting rent rises linearly with final estimate precision. Rent is positive for b less than β and rises at an increasing rate as b falls.

For a given b and Q, the planner's costs equal the forecaster's gross expected payoff plus the direct planning loss $c\sigma^2 + [c/n(b)]$, or

$$(2\beta^{1/2} - b^{1/2})Q^{1/2} - \beta r + c\sigma^2 + cb^{1/2}Q^{-1/2}. \tag{5}$$

The planner chooses Q^* to minimize the expectation of (5). For δ the expectation of $b^{1/2}$, we have

$$Q^* = \frac{c}{\Delta^2}, \tag{6}$$

where $\Delta \equiv [(2\beta^{1/2}/\delta) - 1]^{1/2}$. Again, it must be emphasized, the derivation assumes that c is high enough to warrant hiring even the worst available forecaster. Otherwise, the planner might be willing to completely forgo high-cost investigations in order to reduce rents for more expert forecasters. This turns β into a choice variable, and the resulting first-order condition (since δ need no longer be constant) has no general closed-form solution.

When (6) applies, expected cost $\mathscr{C}^* = \mathscr{C}(Q^*)$ equals

$$2c^{1/2}\delta^{1/2}\Delta - \beta r + c\sigma^2. \tag{7}$$

At the optimum, forecasting error $c\delta Q^{*-1/2}$ accounts for half of EAC (the first term in [7]), the same proportion as under certainty. For r small, the remainder is split between expected "full" measurement costs $\int bn(b)dG(b)$ and forecasting rents in ratio $\delta{:}2(\beta^{1/2} - \delta)$.

III. Costs of Agency

The costs of agency under uncertainty—or, perhaps I should say instead, the costs of uncertainty under agency—are measured by the value Δ. Consider as a benchmark the optimal allocation when b is randomly distributed according to $G(\cdot)$ but costlessly verified prior to contracting. The EAC in (7) is Δ times as high as in the benchmark, while for every value of b, final estimate precision is a fraction $1/\Delta$ of its value in the benchmark.

The lower δ is relative to its upper bound, the greater the relative advantages of full information. By Jensen's inequality, the expectation of a square root is less than or equal to the square root of the expectation. Hence δ is at most $\phi^{1/2}$, where ϕ is the mean of b, and the greater the spread of $G(\cdot)$ about its mean, the greater the gap between δ and $\phi^{1/2}$ will be. Indeed, if we take the binomial expansion of $[\phi + (b - \phi)]^{1/2}$, δ is seen to equal the expectation of

$$\phi^{1/2}\left[1 - \frac{(b-\phi)^2}{8\phi^2} - \frac{5(b-\phi)^4}{32\phi^4} - \cdots\right],$$

which clearly rises as $G(\cdot)$ is compressed toward its mean. One not surprising implication is that the share of forecasting rent in expected administrative cost ($\delta/[4\beta^{1/2} - 2\delta] = 1/2\Delta^2$) tends to rise both with the variance of expertise and with the difference between the mean expertise and its upper bound.

For a numerical example, suppose that b is uniformly distributed on $[\alpha, \beta]$. The efficiency advantage $\Delta - 1$ to full information about b is 15 percent for $\alpha = \beta/2$, 25 percent for $\alpha = \beta/4$, and 33 percent for $\alpha = \beta/9$. Another probability measure we shall have occasion to refer to again is defined by $G(b) \equiv (b/\beta)^l$ for $b \in [0, \beta]$, $l > 0$ ("log-linear"). Its density is increasing for l less than one, decreasing for l greater than one, and uniform for l equal to one. Here Δ works out to $[1 + (1/l)]^{1/2}$, so that the advantages to full information increase without limit as density is weighted toward the origin. For r small, expected forecasting rent amounts to $1/l$ of expected direct measurement cost.

IV. Self-Screening of Expertise

In the preceding derivations, no communication was allowed between principal and agent about the agent's expertise. Could such communication be used to reduce expected administrative costs? If so, how substantial are the likely savings? This section addresses these questions.

For $\tilde{b}$ the agent's reported expertise, expand contracts (2) to take the form

$$H(\tilde{b}, \tilde{Y}, x) = Q(\tilde{b})(2x\tilde{Y} - \tilde{Y}^2) - R(\tilde{b}) \tag{8}$$

expertise. Final estimate precision will be $[Q(\tilde{b})/b]^{1/2}$, yielding an expected net payoff to the agent of

$$br + Q(\tilde{b})\left(Y_0^2 + \frac{1}{r}\right) - R(\tilde{b}) - 2[bQ(\tilde{b})]^{1/2}.$$

For this to be maximized at $\tilde{b} = b$, we must have

$$R'(b) = \left\{Y_0^2 + \frac{1}{r} - \left[\frac{b}{Q(b)}\right]^{1/2}\right\}Q'(b) \quad \forall\, b. \tag{9}$$

To solve this for $R(\cdot)$, make the change of variable $n(b) \equiv [Q(b)/b]^{1/2}$, so that $n(\cdot)$ equals final estimate precision given truthful revelation. Integrating and checking the efficiency boundary condition of zero net payoff at $\tilde{b} = b = \beta$ establishes that

$$R(b) = \left(Y_0^2 + \frac{1}{r}\right)bn^2(b) - \int_b^\beta n(z)dz - 2bn(b) + \beta r. \tag{10}$$

Expected net payoff to the forecaster works out to

$$\int_b^\beta n(z)dz + (b - \beta)r. \tag{11}$$

The planner strives to minimize the sum of expected forecasting loss and expected gross payoff to the forecaster, or

$$\begin{aligned}&\int_\alpha^\beta \left[\frac{c}{n(b)} + \int_b^\beta n(z)dz + bn(b)\right]g(b)db \\ &= \int_\alpha^\beta \left[\frac{c}{n(b)} + \frac{G(b)}{g(b)}\, n(b) + bn(b)\right]g(b)db,\end{aligned} \tag{12}$$

where $g(\cdot)$ is the density of $G(\cdot)$ on support $[\alpha, \beta]$. The solution to (12) is

$$n^*(b) = \left(\frac{c}{b\{1 + [G(b)/bg(b)]\}}\right)^{1/2} \equiv \left\{\frac{c}{b[1 + (1/\epsilon_G)]}\right\}^{1/2}, \tag{13}$$

where ϵ_G denotes the elasticity of $G(\cdot)$ with respect to b. This is less than the first-best solution by a ratio $[1 + (1/\epsilon_G)]^{-1/2}$. Precision is sacrificed to reduce the "information rents" accruing to all but the worst forecasters. Prior beliefs about expertise exert their influence through the ratio $G(b)/g(b)$, which closely resembles the inverse hazard rate appearing in the solution to most adverse selection problems (Maskin and Riley 1984). Indeed, had we defined the cumulative distribution as the probability—currently $1 - G(b)$—of an expert worse than b, as is customary for adverse selection problems, then the inverse hazard rate would have appeared.

In the standard terminology of the literature, $b[1 + (1/\epsilon_G)]$ is the forecaster's "virtual" type; that is, as part of the trade-off for truthful revelation, the planner agrees to act as if the forecaster were of the virtual rather than the true type. To say this another way, the game of hiring forecasters under uncertainty is equivalent to the game of hiring their virtual type counterparts under certainty.

The preceding derivation is misleading in one important respect. We have looked only at the first-order condition for truthful revelation of b when a second-order condition must be met too. The condition is $2bn'(b) + n(b) \leq 0$, which together with (13) requires

$$\frac{d\epsilon_G}{db} \leq 0. \tag{14}$$

In other words, the proportional gap between the virtual type and the true type must rise with b or stay constant. If not, (13) will not define the optimum policy. There will be partial or complete pooling of types; that is, the optimal contract will be identical along some interval of expertise. A more detailed discussion, including instructions for calculating the constrained optimum, is found in Guesnerie and Laffont (1984). It can also be shown, either by checking directly for stationary points or through applying Guesnerie and Laffont's more general methods, that satisfaction of (14) guarantees the optimality of the contract defined by (8), (10), and (13).

To get some intuition for (14), consider that the $n(\cdot)$ of (13) implies

$$Q(b) = \frac{c}{1 + (1/\epsilon_G)}. \tag{15}$$

So the second-order condition can be reinterpreted as requiring that the marginal incentives for precise forecasting rise with the claimed prowess (fall with b). Otherwise a good forecaster could increase marginal payoffs by feigning inability. This kind of "monotonic sorting" requirement is typical for adverse selection problems; see Mirrlees (1971) for the original illustration with income taxation.

If ϵ_G is constant or globally increasing, communication of expertise will thus have no value whatsoever. Moreover, even when communication is worthwhile, it may not be worth very much. Suppose, for example, that expertise is normally distributed between α and one. Spreadsheet calculations show that the percentage EAC savings from relying on communication is maximized at an α of 0.31, where it equals 0.174 percent, or less than one part in 500. Thus the savings from communication about expertise, while important theoretically, are minuscule in practice given a uniform prior distribution.[1]

[1] Communication has a slightly larger though still small impact on forecast precision and the distribution of expected rents across agents. With a uniform distribution of expertise, separation of contracts according to expertise induces the most skilled forecasters to take more measurements than they would otherwise and less skilled forecasters to take fewer measurements than they would otherwise, with overall estimate precision slightly higher on average. With $\alpha = 0.31$, e.g., final estimate precision is 23 percent higher for the most skilled agent. It falls, rapidly at first but at a decreasing rate, to 11 percent for an expertise of 0.4, 4 percent for an expertise of 0.5, and eventually to -6 percent for the least skilled agent, for an average precision gain of 3 percent. Notwithstanding the more thoroughgoing investigation, expert forecasters expect less rent—up to 2.7 percent less with $\alpha = 0.31$ and r small—with self-screening contracts; on the other hand, rents are higher for the worst agents than they would be otherwise—up to 8.8 percent higher for $\alpha = 0.31$ and r small (the limiting value, $[2/(2 - \alpha)]$, is found using L'Hôpital's rule). Because of the extra measurement cost for more precise estimates, the EAC ratio with and without self-screening is not monotonic in expertise: for $\alpha = 0.31$ the savings start at 0.2 percent for the most expert agent, increase to 1.1 percent at an expertise of 0.45, and fall to -1.4 percent for an expertise of 1.

V. A Broader Perspective on Forecaster Self-Screening

The preceding model of forecaster self-selection seems fairly complicated as risk-neutral principal-agent models go. Consider, by way of comparison, Laffont and Tirole's (1986) treatment of a regulated firm. There both cost outcome and product "quality" (which may be quantity) are observed by the regulator. Here, however, only the outcome is verified, not the quality (precision) of the forecast. Even knowledge of the mean-squared discrepancy between the outcome and the forecast would not indicate quality since that measure combines the variance $1/n$ of the estimator with the variance σ^2 of the outcome. From this perspective, it is perhaps less surprising that coaxing a revealing message out of the agent is so costly than that there is ever any merit to self-selection.

Despite these differences, the results of Section IV can nevertheless be derived as a special case of the Laffont-Tirole model, as Jean Tirole has perceptively pointed out. Using their terminology, let us relabel the forecasting precision n as the effort e, the agent's disutility $b(n - r)$ as $\Psi(e)$, and the direct expected planning cost c/n (excluding payments to the forecaster and the unavoidable component $c\sigma^2$) as $\mathscr{C}$. The term $\mathscr{C}$ is nonlinear in e, unlike Laffont and Tirole's expected cost function, but a footnote in their paper suggests how to generalize their results.[2]

Suppose it were possible to peg the agent's money transfer $t(\cdot)$ directly to $\mathscr{C}$. The agent's utility is $U(b) = \max[t(\mathscr{C}) - \Psi(e(\mathscr{C}, b))]$, where $\Psi(e(\mathscr{C}, b)) = b[(c/\mathscr{C}) - r]$. First- and second-order conditions for agent maximization require

$$t'(\mathscr{C}) = -\frac{bc}{\mathscr{C}^2}, \tag{16}$$

$$t''(\mathscr{C}) - \frac{2bc}{\mathscr{C}^3} \leq 0. \tag{17}$$

The principal maximizes $E[-\mathscr{C} - t(\mathscr{C})]$ subject to $U'(b) = r - (c/\mathscr{C})$, $U(\beta) = 0$, and (16) and (17). Ignoring for a moment (17), we have

$$\begin{aligned} E[-\mathscr{C} - t(\mathscr{C})] &= E[-C - \Psi(e(\mathscr{C}, b)) - U(b)] \\ &= \int\left[-\mathscr{C} - b\left(\frac{c}{\mathscr{C}} - r\right) - \int_b^{\beta}\left(\frac{c}{\mathscr{C}} - r\right)dx\right]g(b)db \qquad (18) \\ &= \int\left[-\mathscr{C} - b\left(\frac{c}{\mathscr{C}} - r\right) - \left(\frac{c}{\mathscr{C}} - r\right)\frac{G(b)}{g(b)}\right]g(b)db, \end{aligned}$$

[2] Rogerson (1987) follows up on Laffont and Tirole's footnote and provides a general characterization of what is implementable through linear schemes.

where the last equality follows by reversing the order of integration. Equation (18) is maximized by setting $-1 + (bc/\mathscr{C}^2) + (bc/\mathscr{C}^2\epsilon_G) = 0$ for every b, so that

$$\mathscr{C} = \left[bc\left(1 + \frac{1}{\epsilon_G}\right)\right]^{1/2} \text{ or } b = \frac{\mathscr{C}^2}{c[1 + (1/\epsilon_G)]}. \tag{19}$$

Differentiate (16) and substitute into (17) to see that $b'(\mathscr{C})$ must be positive. If it is, (19) indicates the optimum. It remains to be seen whether the optimum can be "implemented," to use Laffont and Tirole's expression, by a menu of contracts linear in $\mathscr{C}$. A necessary and sufficient condition for implementation is convexity of $t(\cdot)$ since $t(\cdot)$ can then be expressed as the envelope of linear contracts. From (16) and (19), $t(\cdot)$ is convex if and only if $b/\mathscr{C}^2 = \{c[1 + (1/\epsilon_G)]\}^{-1}$ is decreasing in $\mathscr{C}$. Hence ϵ_G must be decreasing in $\mathscr{C}$ and in b (since $b'(\mathscr{C}) > 0$), which is condition (14) again. Moreover, (14) and (19) together imply $b'(\mathscr{C}) > 0$, so the latter condition may be dropped as redundant. In economic terms, the inability to peg rewards to forecasting precision forces a stronger second-order condition on the model than would otherwise apply.

From here it is easy to verify that the contract described in (8), (9), and (15) is optimal. Note that this is optimality in the class of all contracts, not just optimality in the class of equation (2) type contracts as shown earlier. Thus application of the Laffont-Tirole methodology strengthens the previous results and helps illuminate the difficulties with self-selection.

VI. Competition between Forecasters

While theoretically intriguing, the results of the previous two sections are disappointing from a practical standpoint. Self-screening is much less valuable than one might have hoped. This section pursues a different tack to reducing planning costs: using competition among forecasters. Competition improves efficiency both directly, by enabling the manager to hire the best available agent, and indirectly, by reducing the expected rents paid a given forecaster. As we shall see, both the direct and indirect savings can be substantial.

Let $G(b)$ continue to denote the probability that a randomly chosen agent will have expertise b or better. Instead of facing a single applicant, however, the planner will now be allowed to choose one applicant from a pool. Two polar cases will be considered: one in which the forecaster pool is assembled prior to contracting and cannot be expanded later, and another in which the forecaster pool is infinitely expandable at cost A per forecaster. Appendix B addresses the possibility of splitting up investigations among forecasters.

As Laffont and Tirole (1987) have pointed out, the optimal competition procedure for the regulation-with-cost-observation model, with agent risk neutrality, is a variant on a second-price auction. In the present context, have every forecaster i announce his or her expertise b_i. For b the lowest bid and B the second-lowest, award the contract $H^*(b, Y, x; K(\cdot, B))$ to the forecaster with the lowest bid b, where $H^*(\tilde{b}, \tilde{Y}, x; G(\cdot))$ denotes the optimal contract in the single-agent case (given by [8], [9], and [15]) and $K(\cdot; B) \equiv G(\cdot)/G(B)$ denotes the conditional distribution of the minimum bid given a second-lowest value of B. For a brief justification of this choice, consider that no rejected bidders would want such a contract since it would be unprofitable, and that for the sole remaining bidder of distribution $K(\cdot, B)$, the planner cannot expect to do better.

Since ϵ_K equals ϵ_G, we see from (15) that the $Q^*(\cdot)$ term in the chosen contract is independent of B. The auction can be reformulated as (i) ask forecasters to submit their $Q = c/[1 + (1/\epsilon_G)]$, (ii) choose the highest Q for Q^*, and (iii) auction off the right to be rewarded Q^* times $2xY - Y^2$ (alternatively, the right to be fined $Q^*[Y - x]^2$). When $G(b) = (b/\beta)^l$, Q^* is independent of b as well, and the optimal procedure reduces to one step: auction off the reward schedule $[cl/(l + 1)](2xY - Y^2)$, or $-[cl/(l + 1)](Y - x)^2$.

Obviously, competition between potential forecasters offers potential savings to the manager. But it is not obvious how much of the savings are due to the more favorable expertise distribution of the chosen forecaster and how much are due to savings on forecaster rents. To clarify the distinction, let us compare planning costs in four different frameworks: Auction: an auction among T available forecasters, each with expertise drawn from a log-linear distribution: $G(b) \equiv (b/\beta)^l$ on $[0, \beta]$; Unscreened Expert: one forecaster, who is probabilistically just as expert as the best forecaster in Auction, with no self-screening of expertise; Self-screened Expert: same as Expert, but with self-screening of expertise; First-Best: same expertise distributions as above but with costless independent screening of expertise.

The EAC for First-Best equals $2c^{1/2}\int b^{1/2}dL(b)$, where $L(\cdot) \equiv 1 - [1 - G(\cdot)]^T$ is the distribution of the minimum value $b_{\min}$. Integration by parts establishes that

$$\int f(b)dL(b) = T!\left[\int^{\beta} \cdots \int^{b_3}\int^{b_2} f(b_1)dG(b_2) \ldots dG(b_T)\right]$$

for any function $f(\cdot)$. (In words, the unconditional expectation of $f(b_{\min})$ equals the expectation conditional on the one in $T!$ probability event $b_1 < b_2 < \ldots < b_T$.) It follows that the expectation of $(b_{\min})^k$ ($k \neq -1$) equals

$$\frac{1 \cdot 2 \cdot \ldots \cdot T \cdot l^T}{(k + l)(k + 2l) \ldots (k + Tl)} \beta^k. \tag{20}$$

Substituting $k = \frac{1}{2}$ into (20), we see that access to the Tth agent stands to reduce EAC by $100/(2Tl + 1)$ percent. For example, when expertise is uniform on $[0, \beta]$, EAC is 20 percent less with two agents than with one and 14 percent less with three agents than with two.

In framework Auction, the contract pivots on the expertise B of the second-best forecaster. Conditional on B, EAC is

$$2c^{1/2}\Delta(B, K(\cdot; B)) \int b^{1/2} dK(b; B) = 2c^{1/2}\left(1 + \frac{1}{l}\right)^{1/2} \int b^{1/2} dK(b; B), \tag{21}$$

where

$$\Delta(\beta; K(\cdot; B)) \equiv \left[\frac{2B}{\int b^{1/2} dK(b; B)} - 1\right]^{1/2} \tag{22}$$

is defined analogously to the Δ in (6). The EAC is $[1 + (1/l)]^{1/2}$ times the corresponding conditional value for First-Best, and since this ratio is constant, it must hold unconditionally as well. For a uniform distribution on $[0, \beta]$, the percentage discrepancy is 41 percent.

For Unscreened Expert, EAC is $\Delta(\beta, L(\cdot))$ times the EAC for First-Best, where $\Delta(\beta, L(\cdot))$ equals

$$\left[\frac{2(\frac{1}{2} + l) \ldots (\frac{1}{2} + Tl)}{T! l^T} - 1\right]^{1/2}. \tag{23}$$

For a uniform distribution on $[0, \beta]$, the percentage EAC increment for Unscreened Expert over First-Best is 41 percent for $T = 1$, 66 percent for $T = 2$, 84 percent for $T = 3$, 98 percent for $T = 4$, and 110 percent for $T = 5$.

For Self-screened Expert, the EAC integral does not appear to be solvable analytically. Clearly, for any given T, EAC(Unscreened Expert) $\geq$ EAC(Self-screened Expert) $\geq$ EAC(Auction) $\geq$ EAC(First-Best). The advantages of Self-screened Expert over Unscreened Expert increase with the number of competitors, as $d\epsilon_L/db$ is more negative for higher T. Obviously, for $T = 1$ there is no advantage at all to screening as long as ϵ_G is constant.

Figure 1 graphs the results for $l = 1$. Competition in Auction squeezes out some but not all forecasting rents. And all these values decline toward zero as T rises (more generally, toward the EAC of the best possible forecaster). For $l = 1$, the ability to extract second-best

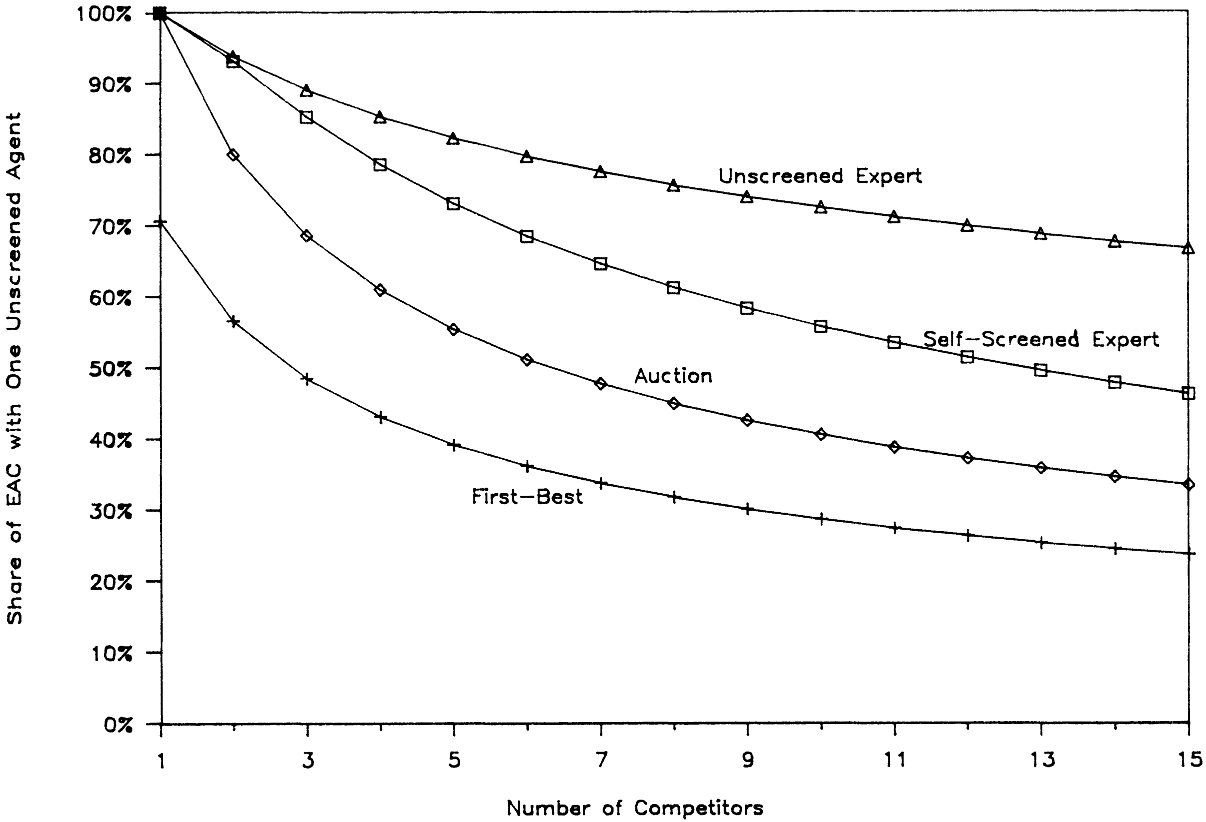

FIG. 1.—Expected administrative costs; b uniform on $[0, \beta]$

levels of rent offers significant savings, comparable to or exceeding the gains from better forecaster selection. An auction with three participants captures about half the difference in EAC between no screening and full screening, whereas self-screening captures only one-tenth.

As the number of competitors increases, expected second-best expertise is driven closer to first-best expertise, raising the relative efficiency of an auction. Thus with six participants an auction captures about two-thirds of the difference between no screening and full screening, whereas self-screening captures about a quarter.

Additional insights into the value of competition can be gleaned from examination of auctions with flexible pool size. Suppose that the planner can decide sequentially whether to elicit another bid, at constant search cost A per forecaster. In this case the manager should hire the first agent revealing a b below some predetermined threshold B. The optimal screening rule takes this form because of the "memoryless" nature of the process: previously incurred costs and revealed expertise levels have no effect on future agents' screening costs and likely expertise levels.

The first agent appearing with $b \leq B$ will accept $H^*(\cdot)$ and the search will end. Again, since ϵ_K equals ϵ_G, final estimate precision is $n^*(b)$ in (13) independent of B. The expected number of agents screened, $T(B)$, equals one times the probability $G(B)$ that the first agent is taken plus $1 + T(B)$ times the probability that the first agent is not taken. Thus $T(B) = 1/G(B)$. For $\epsilon_G = l$, the optimum B can be shown to satisfy

$$\left[\frac{cl(l + 1)}{B(l + \frac{1}{2})^2}\right]^{1/2} - lA\beta^l B^{-l-1} - r = 0, \tag{24}$$

so that for r small,

$$B \cong \left[\frac{l^{1/2}(l + \frac{1}{2})A\beta^l}{(l + 1)^{1/2}c^{1/2}}\right]^{2/(2l+1)}.$$

The lower the marginal search cost and the higher the cost of estimate imprecision, the lower the threshold should be set, which means that the manager will tend to screen more agents. The elasticities of the threshold with respect to A or $1/c$ are equal and less than one-half. The threshold rises with β, but B/β falls, so that as the support of $G(\cdot)$ widens, more agents tend to be screened.

The absolute savings from access to competition vary with A and other parameters in a predictable manner and will not be examined further here. We shall focus instead on the expected relative shares of search costs, direct measurement costs, estimate imprecision losses, and information rents. Surprisingly, these shares are independent of

A, c, or β. Straightforward calculations show that optimal expected search costs amount to $1/(2l + 1)$ of the total administrative costs. The rest of EAC is divided among imprecision, measurement, and rent in ratio $l + 1 : l : 1$, so that the losses from imprecise specification of the mean equal total expected payment to forecasters. (This last property can be shown to hold for all optimal contracts, regardless of $G(\cdot)$.) For $l = 1$, search costs, imprecision losses, and forecaster payments each account for a third of EAC; expected forecaster payments in turn are divided evenly between direct measurement costs and information rents.

Figure 2 illustrates how the composition of EAC varies with l. When l is very small so that the distribution is weighted toward very expert forecasters, it is worthwhile to set the threshold very low and spend the bulk of planning funds on screening. As l rises, search declines in importance. Rent shares rise with l up to $l = \sqrt{1/2}$ ($\cong 0.7$), where they reach 17 percent, and decline thereafter as the effect of the bunching of expertise toward the upper bound begins to dominate the effect of looser selection criteria. Losses from estimate imprecision always exceed direct measurement costs, but the gap decreases as l rises, until in the limit (one forecaster of known expertise β) each factor accounts for half of EAC.

VII. Conclusion

Previous literature on forecast elicitation has tended to assume that forecasters' beliefs are immutable. In reality, a forecaster generally begins with some rough ideas about the event in question and undertakes further investigation in order to refine them. How much refinement will occur depends on both the costs of investigation and the potential compensation.

When learning is costly, a planner doing her own forecasting must weigh the marginal cost of investigation against the expected marginal benefit of more precise information. A planner who hires someone else to do the forecasting must in addition provide an appropriate compensation scheme. This scheme must induce the forecaster to simultaneously take the desired number of investigations and report the final estimate truthfully. If the planner is uncertain about the forecaster's expertise, the problem is even more complicated. A contract for inducing a given amount of learning is likely to offer the forecaster expected rents, and the magnitude of these rents must be considered in choosing an optimal contract.

Despite the multiple layers of uncertainty, the optimal contract need not be complex, provided both parties are risk-neutral. However, the rents involved are likely to be substantial. Sometimes it is

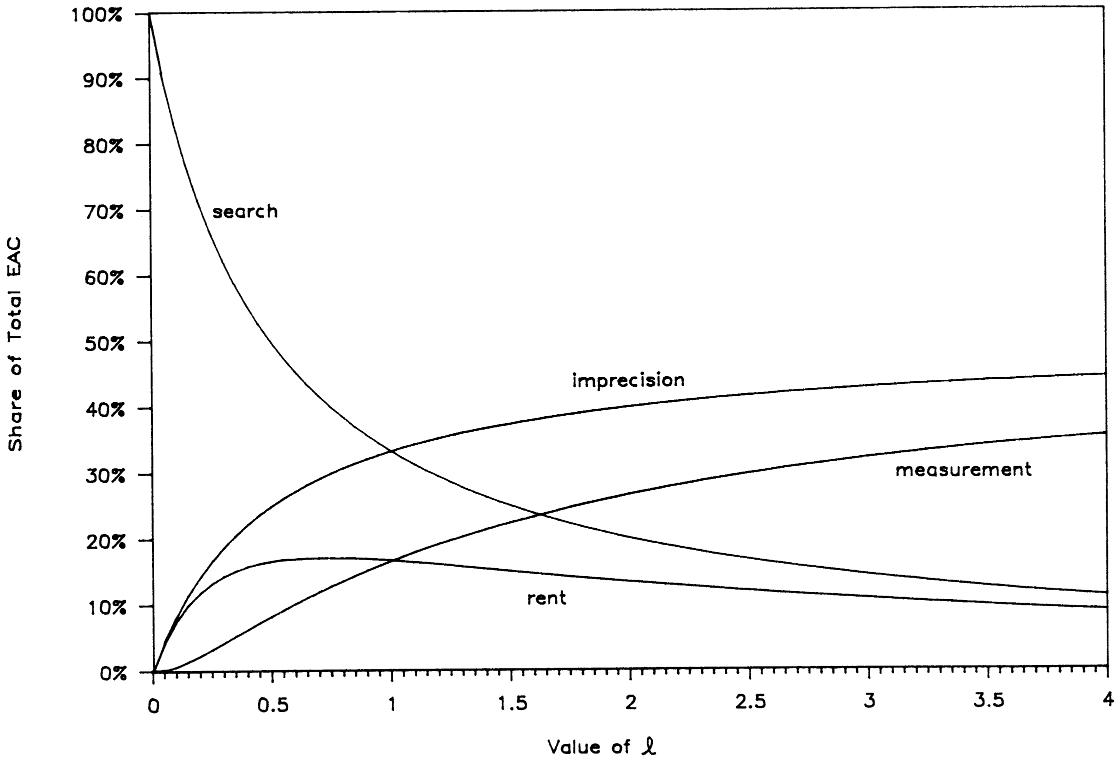

FIG. 2.—Shares in total EAC as functions of l

possible to reduce agency costs through forecaster self-screening, but not always. Even when self-screening does have value, the value may not be large enough to warrant the added contractual complexity.

A much more promising alternative is to compete two or more forecasters against each other. The planner may simply offer a multiple of her own forecast evaluation weights (the variant with self-screening lets forecasters submit their own evaluation weights) and auction off the right to be selected. Competition between forecasters reduces planning costs both by improving the odds of selecting an especially good forecaster and by reducing that forecaster's expected rents. In many cases, the second factor is at least as important as the first. With possibilities of sequential search, an optimizing manager may find it worthwhile to invest a significant share of total planning budgets on screening or training potential forecasters.

Why does self-selection achieve so little in this problem, when competition offers so much? The core of the problem seems to be the planner's inability to verify forecast precision. Because of that inability, the planner is forced to try to simulate the "ideal" contract by an envelope of contracts linear in ex post cost. Success requires that the "ideal" contract be convex in expected cost. This imposes a more stringent second-order condition on what is otherwise a straightforward application of Laffont and Tirole's (1986) work on firm regulation under ex post cost and quality observation.

I shall close with a few remarks on the broader implications of the analysis. The analysis suggests that a planning hierarchy faces a costly trade-off between improving forecast quality and reducing administrative expenses. This source of planning inefficiency does not appear to have been addressed before in the literature. In many planning hierarchies, performance would probably be improved by giving forecasters a chance to profit from accurate estimates. The drawback is that rents are likely to be substantial. In government, especially, where incentives for economizing are weak, suspicions of waste or corruption would be raised. Distinguishing "honest" from "dishonest" rents could indeed be difficult.

The analysis also suggests that organizations that operate on a "need-to-know" principle, so that forecasting expertise is restricted to one or a handful of individuals, do so at a cost of reduced planning efficiency. Perhaps this is a factor in the disappointing efficiency record of Soviet central planning. Carrying "redundant" forecasting expertise would seem to be wasteful. Yet redundancy can make for more vigorous competition, which improves selection and squeezes forecasting rents. At the very least, the notion that planning obviates inefficient competitive information gathering must be supplemented with its opposite: that competition in information gathering helps to remedy inefficient planning.

Appendix A

Optimality of Quadratic/Linear Contracts

We are looking for contracts $H(Y, x)$ that possess the following three features: (*a*) Truthful reporting of Y is encouraged; that is, expected reward is always maximized by reporting the perceived mean of x. (*b*) Expected payoff is the same for all distributions of x having mean Y. (*c*) Induced final estimate precision is the same for all forecasters having expertise b. Savage (1971) has shown that all contracts satisfying condition a must take the form

$$H(Y, x) = V(Y) - V'(Y)(Y - x) + W(x) \quad \text{with } V(\cdot) \text{ convex.} \tag{A1}$$

For an economic interpretation of (A1), see Osband (1985). In general, incentives for investigation will vary with the convexity of $V(\cdot)$. Suppose that the agent thinks that the mean is either $Y + \omega$ or $Y - \omega$ with equal probability and that an investigation costing b could determine the mean exactly. Without investigation, expected payoff is maximized by reporting Y; its value there exclusive of the $W(\cdot)$ term is $V(Y)$. With investigation, expected payoff exclusive of $S(\cdot)$ and investigation cost would be either $V(Y + \omega)$ or $R(Y - \omega)$, each with probability ½. So a risk-neutral agent should undertake an investigation if and only if $\{[V(Y + \omega) + R(Y - \omega)]/2\} - V(Y) \geq b$. The left-hand expression is just the distance from $\langle Y, V(Y)\rangle$ to the midpoint of the chord connecting $\langle Y - \omega, V(Y - \omega)\rangle$ with $\langle Y + \omega, V(Y + \omega)\rangle$. The more convex $V(\cdot)$ is, the larger ω is, and the smaller b is, the more attractive the investigation will be (see fig. A1).

More generally, let the postinvestigation estimate correction ω have distribution $J_Y(\cdot)$. For example, if the current unbiased estimate has precision n and the forecasting technology involves taking an additional conditionally independent measurement of precision ϵ at cost $b\epsilon$, $J_Y(\cdot)$ will have precision $(n/\epsilon)(n + \epsilon)$, or n^2/ϵ in the limit (Zellner [1971] or any standard textbook on Bayesian inference). Measurement will be worthwhile if and only if

$$\int V(Y + \omega)dJ_Y(\omega) \geq V(Y) + b\epsilon. \tag{A2}$$

Condition c says that satisfaction of (A2) should depend only on $b\epsilon$ and n^2/ϵ. We claim that this implies that $V(\cdot)$ is quadratic. To do so we make use of the following lemma, proved in Osband and Reichelstein (1985).

LEMMA. If the expectation of a function $g(\cdot)$ is zero for all distributions such that the expectations of $h_1(\cdot)$, $h_2(\cdot)$, . . . are zero, then $g(\cdot)$ is a linear combination of the $h_i(\cdot)$'s.

To apply the lemma, let $B\epsilon$ be the threshold value at which the forecaster is indifferent between measurement and nonmeasurement. By condition c we have

$$\int [V(Y + \omega) - V(Y) - B\epsilon]dJ_Y(\omega) = 0$$

for all $J_Y(\cdot)$ with mean zero and variance ϵ/n^2. Substituting $g(\omega) \equiv V(Y + \omega) - V(Y) - B\epsilon$, $h_1(\omega) = \omega$, and $h_2(\omega) = \omega^2 - (\epsilon/n^2)$ into the lemma establishes that $V(Y - \omega) = V(Y) + B\epsilon + j\omega + k\omega^2 - (\epsilon k/n^2)$ for some constants j and k, so

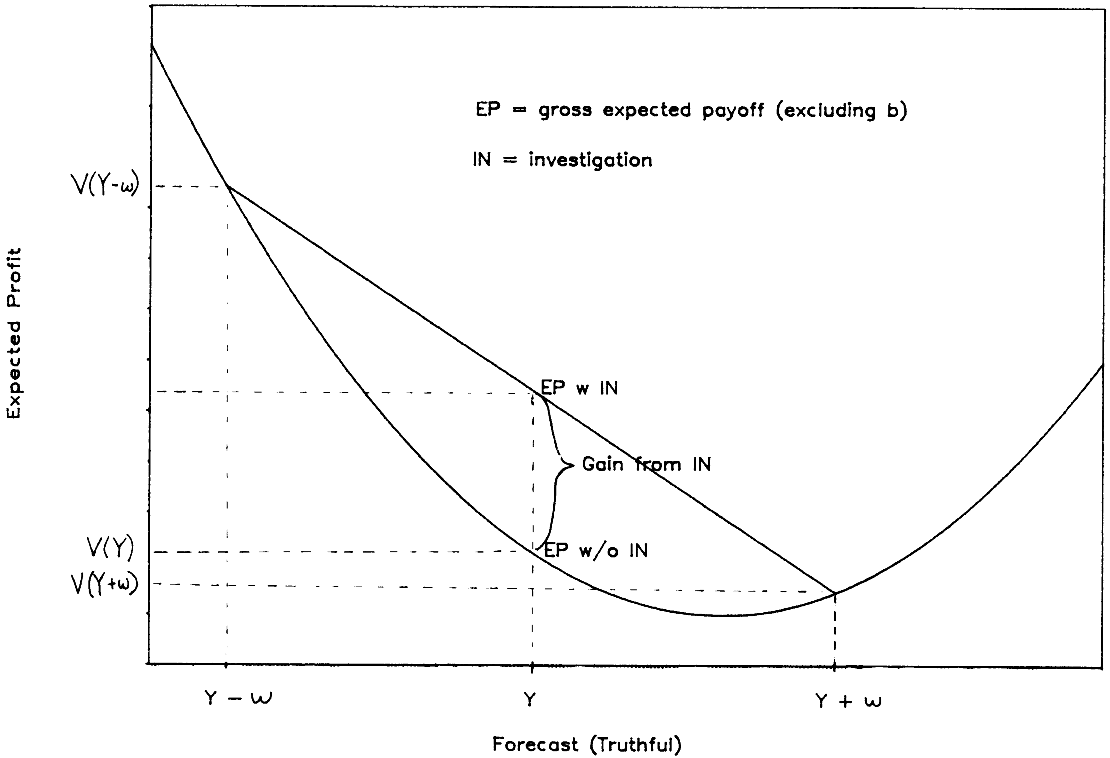

FIG. A1.—Benefits of investigation

that $V(\cdot)$ is quadratic with coefficient k on the square. Evaluating at $\omega = 0$ shows that $k = Bn^2$. The terms in $V(Y)$ and $V'(Y)$ in (A1) thus reduce to $Bn^2(2xY - Y^2)$ plus a linear term in x.

It remains to show that $W(\cdot)$ is linear. By condition b, the expectation of $W(\cdot)$ must equal $W(\cdot)$ of the expectation. Another application of the lemma with $g(x) \equiv W(x) - W(Y)$ and $h(x) \equiv x - Y$ yields the result.

Appendix B

On Hiring More than One Forecaster

The treatment of competition in the text assumes that the planner ultimately hires only one forecaster. Hiring one forecaster is always best provided "start-up" costs for forecasters are sufficiently high. Even without start-up costs, however, hiring one forecaster may be optimal.

Consider the following forecasting technology: starting from the unbiased estimate Y_0 of μ having precision r, take successive independent "measurements" of μ, each having precision one and costing b, and calculate the best linear unbiased estimator (the precision-weighted average of the measurements and prior). Then if measurements are independent across forecasters, one person's forecast cannot be used to flush information out of another. In the constant Q, R contracts described in equations (2)–(6), it costs the planner an expected $[2(b\beta)^{1/2} - b]n - Br$ to secure an n precision forecast from a type b agent, given threshold bid B. This payment is linear in n and increasing in b. Therefore, the planner will find it cheaper, no matter what the total final precision is, to have the best forecaster perform all the measurements. In an unrestricted model based on (8), gross expected payment for an n_j precision forecast from b_j type forecaster j is

$$\int_{b_j}^{B} n_j(z, B_j)dz + (b_j - \beta)r + b_j n(b_j, B_j),$$

where (b_i, B_i) is the vector $(b_1, \ldots, b_T)$. If priors are shared, total final estimate precision N equals $\Sigma n_j(b_j, B_j) - (T - 1)r$. The planner must then choose the $n_j(\cdot)$'s and N to minimize

$$\int \ldots \int \left[\frac{c}{\Sigma n_j - (T - 1)r} + \Sigma \int_{b_j}^{B} n_j(z, B_j)dz + \Sigma b_j n_j \right] dG(b_1) \ldots dG(b_T) \quad \text{(B1)}$$

subject to $n_j \geq r$ and $\Sigma n_j = N$. Set up a Hamiltonian for (B1) with multipliers λ_j on the $n_j \geq r$ constraints. First-order conditions are readily seen to require λ_j to be positive for all but the lowest b_j; that is, only the most expert forecaster should be induced to do any additional investigation.

It remains an open question whether splitting investigations might be superior when measurements are correlated across forecasters: say, that measurement 1, if taken, would be the same for all forecasters, that measurement 2 would be the same for all forecasters, and so on.

References

Guesnerie, Roger, and Laffont, Jean-Jacques. "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-managed Firm." *J. Public Econ.* 25 (December 1984): 329–69.

Haim, E. "Characterization and Construction of Proper Scoring Rules." Ph.D. dissertation, Univ. California, Berkeley, 1982.

Laffont, Jean-Jacques, and Tirole, Jean. "Using Cost Observation to Regulate Firms." *J.P.E.* 94, no. 3, pt. 1 (June 1986): 614–41.

———. "Auctioning Incentive Contracts." *J.P.E.* 95 (October 1987): 921–37.

Maskin, Eric S., and Riley, John G. "Monopoly with Incomplete Information." *Rand J. Econ.* 15 (Summer 1984): 171–96.

Mirrlees, James A. "An Exploration in the Theory of Optimal Income Taxation." *Rev. Econ. Studies* 38 (April 1971): 175–208.

Osband, Kent. "Eliciting Forecasts: A General Characterization of Scoring Rules." Discussion Paper no. 1194. Cambridge, Mass.: Harvard Univ., Inst. Econ. Res., November 1985.

Osband, Kent, and Reichelstein, Stefan. "Information-eliciting Compensation Schemes." *J. Public Econ.* 27 (June 1985): 107–15.

Rogerson, William P. "On the Optimality of Menus of Linear Contracts." Santa Monica, Calif.: Rand Corp., January 1987.

Savage, Leonard J. "Elicitation of Personal Probabilities and Expectations." *J. American Statis. Assoc.* 66 (December 1971): 783–801.

Thomson, William. "Eliciting Production Possibilities from a Well-Informed Manager." *J. Econ. Theory* 20 (June 1979): 360–80.

Zellner, Arnold. *An Introduction to Bayesian Inference in Econometrics.* New York: Wiley, 1971.