INSEAD
R & D

Working Papers

HIERARCHY SIZE AND ENVIRONMENTAL UNCERTAINTY

by

K. MEAGHER *
H. ORBAY **
and
T. VAN ZANDT †

2001/18/EPS

* School of Economics, University of New South Wales, Sydney, NSW 2052, Australia.

** Sabanci University, Bankalar Cad. No 2 Karakoy, Istanbul 80020, Turkey.

† Professor of Economics at INSEAD, Boulevard de Constance, 77305 Fontainebleau Cedex, France.

Printed at INSEAD, Fontainebleau, France.

Hierarchy Size and Environmental Uncertainty

Kieron Meagher, Hakan Orbay, Timothy Van Zandt*

6 April 2001

Abstract

We examine how a firm's changing environment and the information constraints of its managers interact as determinants of the size of the firm's administration. Following the recent decentralized information processing literature, we assume that it takes individual managers time to process information. A consequence is that it takes time for a firm to aggregate information, even when this task is shared. This delay increases with the amount of information that is aggregated, leading to the following trade-off: the more data the firm samples each period (and hence the larger its managerial staff), the more precisely it can estimate the state that its environment was in when the sample was taken *but* the more the environment has changed by the time these data are used to estimate the current state. We explore this trade-off for two computation models and for both a benchmark case of costless managers and the case of costly managers. When managers are costless, the size of the administrative staff increases monotonically as the environment becomes more stable. In contrast, when managers are costly, optimal managerial size first increases and then decreases as a function of environmental stability.

Authors' addresses:

Kieron Meagher School of Economics University of New South Wales Sydney, NSW 2052 Australia	Voice: Fax: Email:	+61 (2) 9385 1145 +61 (2) 9313 6337 k.meagher@unsw.edu.au
Hakan Orbay Sabancı University Bankalar Cad. No 2 Karakoy Istanbul 80020 Turkey	Voice: Fax: Email:	+90 (216) 483 9664 +90 (216) 455 3563 hakan@sabanciuniv.edu
Timothy Van Zandt INSEAD Boulevard de Constance 77305 Fontainebleau CEDEX France	Voice: Fax: Email: Web:	+33 1 6072 4981 +33 1 6074 6192 tvz@econ.insead.edu zandtwerk.insead.edu

*This research supported in part by grant IRI97-11303 from the National Science Foundation.

1 Introduction

A firm's administration consists of many agents who collectively process information and make decisions. These tasks involve aggregating information about a changing environment, with delay even when this task is shared (decentralized). Optimal organizational structure must be adapted to the changing nature of the environment and to human information processing constraints.

Radner (1993), in a seminal paper, employs an explicit model of information processing in order to study organization structure. A main theme is that decentralization of information processing reduces delay but raises managerial costs. He observes that there remain inexorable constraints on the use of recent information even when there is decentralization.

Van Zandt and Radner (2001) consider how these trade-offs affect the optimal scale of firms. They show that firm size, as measured by the scope of activities or the number of markets in which it participates (which determines the scale of the firm's centralized decision problem), may be limited by the inability to quickly aggregate information in order to keep up with a changing environment.

This paper takes up a different exercise. We hold fixed a firm's decision problem (and hence its scale), but look at different information gathering and aggregation procedures. The firm may choose to have a large administrative staff that aggregates large amounts of information, thereby obtaining precise estimates of the state of the environment when the information was collected, or to have a lean administration that aggregates fewer data but therefore uses more recent data in its decisions.

The decision problem we consider is to forecast a single stochastic process. The firm can gather and aggregate noisy observations of this process. We restrict attention to two classes of policies. In policies without recall, the firm periodically gathers a sample and computes a new forecast, at which point previously accumulated information is disregarded. In policies with recall, the firm periodically gathers a sample and computes an update of the forecast, thereby combining the new information with previous information. The parameters of these policies are the size of each sample, the delay before a new forecast is computed from each sample, and the interval between samples.

These parameters cannot be chosen freely but instead are constrained by the information processing capabilities of the potential managers. Furthermore, different policies have different managerial costs. Given a class of policies and a computation model, the organization design problem is to choose a decision procedure that optimally trades off these policy parameters and managerial costs. We characterize optimal organizations and how they vary with the volatility of the environment.

We do this exercise for two computation models. The first is a PRAM, which is the simplest model of decentralized information processing. It allows us to see the main trade-offs but does not paint a clear picture of the structure of the organization. The second is a variant of the periodic processing model in Radner (1993), with a constraint that the processing be stationary in the sense that each sample of data is processed in the same way by the same hierarchy. The properties of this information processing model are characterized in Orbay (2001).

For each model, we first consider a benchmark case in which the cost of managers is zero. We show that sample size and managerial size increase with the stability of the environment. As the environment becomes more stable, delay is less costly and hence sample size increases; this results also in an increase in managerial size. This is consistent with the notion that firms adopt leaner managements when faced with a rapidly changing environment. When the managerial wage is positive, small managerial staffs are also optimal when the environment changes very slowly. It is then possible to keep up with a slowly changing environment by processing only a small amount of information and thereby economizing on managerial costs. Larger hierarchies are optimal in the intermediate region, when the environment changes quickly enough that there is value to processing more information each period but not so quickly that aggregate information has little value because of delay.

2 Decision problem

We consider a firm (or other type of organization) whose randomly changing environment is parameterized by a stochastic process $\{x_t\}_{t=0}^{\infty}$. The firm's profit each period depends on its decisions and on the state. As a reduced form, we assume that the key decision task is to form an estimate $\hat{x}_t$ of x_t and that the firm's expected profit in period t is a fixed level minus the mean-squared error $L_t \equiv E\big[(\hat{x}_t - x_t)^2\big]$ of the estimate (called the "loss"). The estimate is calculated from data about the environment by a managerial staff whose cost in period t is denoted W_t. The total cost C_t in period t is the decision-theoretic cost L_t plus the managerial cost W_t. The managerial staff and information processing procedure are chosen to minimize the long-run average value of $\{C_t\}$.

The state of the environment is assumed to follow a random walk,

$$x_{t+1} = x_t + v_t\,, \tag{1}$$

where v_t, the innovation term, has mean 0 and variance σ_v^2 and is uncorrelated with v_{t+s} for $s \neq 0$. The parameter σ_v^2 is called the *volatility* of the environment.

In order to estimate the state, the firm collects samples of imperfect observations. Observation i in a sample gathered at time t is denoted by

$$y_{it} \;=\; x_t + \epsilon_{it}\,. \tag{2}$$

The measurement error ϵ_{it} has mean 0 and variance σ_ϵ^2, and it is uncorrelated with other measurement errors and with the stochastic process $\{x_t\}$ across all time periods.

Remark 1 The random walk given in equation (1) is not a stationary stochastic process. Following a common practice, we consider a limiting case in which the unconditional variance of x_t goes to infinity. Formally, we could have the process run from $t = -\infty$ and either (a) specify a known starting value x_{t_0} and let $t_0 \to -\infty$; (b) let the unconditional variance of x_0 increase to infinity; or (c) replace the random walk by a stationary AR(1) process $x_t = \beta x_{t-1} + v_t$ and let $\beta \uparrow 1$.

The samples are processed by the firm's managers to form predictions according to a *decision procedure*, which specifies both the policy that is computed and the way in which it is computed by the managers. We restrict attention to two classes of cyclic policies with linear estimators. The policies in each class are parameterized by the size n of each sample, the delay d between when a sample is taken and when it is first incorporated into the estimate, and the interval k between samples. In Section 3, we define these parameterized classes of policies and then calculate, for each (n,d,k), the long-run average loss $\mathcal{L}(n,d,k)$ of the corresponding policy. In Sections 5 and 6, we describe two models of managerial capabilities and of the computation of policies. Each computation model determines which of these policies are feasible (computable) and what the minimum managerial cost $\mathcal{W}(n,d,k)$ of each feasible policy is. The reduced form of the organization design problem, given a class of policies and given a computation model, is to choose (n,d,k) from the set of feasible values so as to minimize $\mathcal{C}(n,d,k) \equiv \mathcal{L}(n,d,k) + \mathcal{W}(n,d,k)$.

3 Policies

3.1 Overview

Our motive for restricting attention to limited classes of policies is the following. The computation models in Sections 5 and 6 are quite flexible, so the sets of possible decision procedures are vast and unstructured. By restricting the class of policies, we reduce the design problem to a several-variable discrete optimization problem and obtain a tractable formula for the loss of any allowable policy. Furthermore, for the computation model in Section 6, we can assure that the computation procedures have a recognizable structure.

The policies we consider are linear estimators. Specifically, let Φ_t be the data used to calculate $\hat{x}_t$. Then $\hat{x}_t$ is the linear projection of x_t on $(1, \Phi_t)$ and is the affine function of Φ_t with the lowest mean-squared error; we denote this by $\hat{E}[x_t \,|\, \Phi_t]$. We use the well-known formulae for linear projections and for the resulting mean-squared errors. (If the random variables are Gaussian, then $\hat{E}[x_t \,|\, \Phi_t] = E[x_t \,|\, \Phi_t]$ and the linear estimator minimize the expected loss conditional on Φ_t.)

3.2 Policies without recall

In a policy *without recall*, the management gathers a sample of size n every k periods, computes an estimate from each sample in d periods, and then uses this estimate for k periods (until a newer estimate is available).

Suppose that a sample $\varphi_t = \{y_{1t}, y_{2t}, \ldots, y_{nt}\}$ of size n is gathered in period t. Consider first the linear estimator of x_t based on this sample. In the limiting case described in Remark 1, the formula for $\hat{E}[x_t \,|\, \varphi_t]$ converges to the sample average $(1/n)\sum_{i=1}^{n} y_{it}$, which we denote by y_t, and the loss $E\big[(x_t - y_t)^2\big]$ converges to σ_ϵ^2/n.

Suppose that φ_t is the data used to estimate x_{t+s}, where $s > 0$. Given that the process follows a random walk and that the innovations from $t+1$ to $t+s$ are not correlated with φ_t, it follows that $\hat{E}[x_{t+s} \,|\, \varphi_t] = \hat{E}[x_t \,|\, \varphi_t] = y_t$. The loss increases by the total variance $s\sigma_v^2$ of the innovations that occur between when the sample is gathered and when its mean is used as an estimate. Hence, if it takes d periods to compute the first estimate from φ_t and if the estimate is used over a k-period planning cycle, then the average loss over the planning cycle is

$$\text{(3)} \qquad \mathcal{L}^{\text{nr}}(n, d, k) \;\equiv\; \frac{1}{k}\sum_{j=0}^{k-1}\left(\frac{\sigma_\epsilon^2}{n} + (d+j)\sigma_v^2\right) \;=\; \frac{\sigma_\epsilon^2}{n} + \left(d + \frac{k-1}{2}\right)\sigma_v^2 \,.$$

(The superscript "nr" stands for "no recall".) This describes the policy for all but the first d periods, so this average is also the long-run average loss.

3.3 Policies with recall

A policy *with recall* is similar to a policy without recall except that $\hat{x}_t$ equals the projection of x_t on the last sample from which it is computed *and on all preceding samples.* Owing to our statistical assumptions, this requires only a simple updating rule (an example of a Kalman filter).

Suppose first that a policy with recall has no computational delay. Let t be a period in which a sample is taken, let φ_t be the sample gathered that period, and let φ_t^p denote all previous samples. Let $\hat{x}_t^n = \hat{E}[x_t \,|\, \varphi_t]$ and $\hat{x}_t^p = \hat{E}[x_t \,|\, \varphi_t^p]$; let Σ^n and Σ^p be the respective mean-squared errors of these estimates. The errors of these estimates are uncorrelated (the error for $\hat{x}_t^n$ is the sum of the sample errors and the error for $\hat{x}_t^p$ is the sum of previous sample errors and previous innovations). Therefore, from the projection formulae,

$$\hat{E}[x_t \,|\, \varphi_t^n, \varphi_t^p] \;=\; \frac{\Sigma^n}{\Sigma^p + \Sigma^n}\hat{x}_t^p + \frac{\Sigma^p}{\Sigma^p + \Sigma^n}\hat{x}_t^n \,, \tag{4}$$

and the mean-squared error is $(\Sigma^p \Sigma^n)/(\Sigma^p + \Sigma^n)$.

For period s as far back as $t-k$ (when the previous sample was gathered), $\hat{x}_t^p$ is also the estimate of x_s. Furthermore, as explained in Section 3.2, $\hat{x}_t^n$ is the sample average of φ_t. Hence, $\hat{x}_t$ is calculated by summing $y_t = (1/n)\sum_{i=1}^n y_{it}$ and then averaging

$$\hat{x}_t \;=\; (1-\alpha)\hat{x}_{t-k} + \alpha y_t \,, \tag{5}$$

where $\alpha = \Sigma^p/(\Sigma^p + \Sigma^n)$.

Suppose that the loss is the same in each period in which a sample is gathered; let Σ^* be its steady-state value. Then $\Sigma^* = (\Sigma^p\Sigma^n)/(\Sigma^p + \Sigma^n)$. Since the last sample was gathered in period $t-k$, we have $\hat{x}_t^p = \hat{E}[x_{t-k} \,|\, \varphi_t^p]$. Thus, the mean-squared error of $\hat{x}_t^p$ as an estimate of x_{t-k} is also Σ^*; as an estimate of x_t, the error of $\hat{x}_t^p$ is augmented by the variance of the intervening innovations, so $\Sigma^p = \Sigma^* + k\sigma_v^2$. We thus have the identity $\Sigma^* = (\Sigma^* + k\sigma_v^2)\Sigma^n/(\Sigma^* + k\sigma_v^2 + \Sigma^n)$, which simplifies to $(\Sigma^*)^2 + k\sigma_v^2\Sigma^* - k\sigma_v^2\Sigma^n = 0$. The positive solution to this quadratic equation is $\Sigma^* = (-k\sigma_v^2 + \sqrt{(k\sigma_v^2)^2 + 4k\sigma_v^2\Sigma^n})/2$. This expresses the steady-state loss Σ^* of the zero-delay policy *with* recall as a function of the loss Σ^n of the zero-delay policy *without* recall.

With computational delay d before new information is incorporated into the estimate, the rule for updating the estimate is the same whereas the loss in each period in which new information is incorporated is increased by $d\sigma_v^2$. There is an additional loss of σ_v^2 for each subsequent period of the planning cycle in which no new information is used. Recall from Section 3.2 that $\Sigma^n = \sigma_\epsilon^2/n$. Hence, the average loss over the planning cycle is

$$\begin{aligned} \mathcal{L}^{\mathrm{r}}(n,d,k) \;&\equiv\; \frac{1}{k}\sum_{j=0}^{k-1}\Big(\big(-k\sigma_v^2 + \sqrt{(k\sigma_v^2)^2 + 4k\sigma_v^2\sigma_\epsilon^2/n}\,\big)/2 + (d+j)\sigma_v^2\Big) \\ &=\; \Big(\sqrt{(k^2/4 + k\sigma_\epsilon^2/(n\sigma_v^2)} + d - 1/2\Big)\,\sigma_v^2 \,. \end{aligned} \tag{6}$$

(The superscript "r" stands for "recall".)

Thus, $\mathcal{L}^{\mathrm{r}}(n,d,k)$ is the limiting steady-state loss as the unconditional variance of x_t goes to infinity, as described in Remark 1. Because the estimation problem starts in period 0, the average loss over the planning cycle is not at its steady-state value but rather converges to it as $t \to \infty$; nevertheless, $\mathcal{L}^{\mathrm{r}}(n,d,k)$ is the long-run average loss.

4 Remarks on the methods for comparative statics

To derive our comparative statics results, we use a combination of numerical methods and the analytic tools of "monotone comparative statics" developed in Topkis (1979), Vives (1990), Milgrom and Roberts (1990), and Milgrom and Shannon (1994).

Here is a summary of these analytic tools. We have a cost-minimization problem $\min_z \mathcal{C}(z;p)$, where $z = (z_1,\ldots,z_m)$ is a vector of (perhaps discrete) choice variables and $p = (p_1,\ldots,p_n)$ is a vector of parameters. Let $\psi(p)$ denote the solution correspondence. We say that $\psi(p)$ is increasing (resp., decreasing) when, for $p' \geq p$ (resp., for $p' \leq p$), if

$z \in \psi(p)$ and $z' \in \psi(p')$ then $\inf\{z, z'\} \in \psi(p)$ and $\sup\{z, z'\} \in \psi(p')$. (When ψ is singleton-valued, this is the usual definition of a weakly increasing function.) The cost function $\mathcal{C}$ is submodular in z if $\mathcal{C}(\inf\{z, z'\}; p) + \mathcal{C}(\sup\{z, z'\}; p) \leq \mathcal{C}(z; p) + \mathcal{C}(z'; p)$ for all z, z' and p. The function $\mathcal{C}$ has increasing (resp., decreasing) differences in z and p if $\mathcal{C}(z'; p) - \mathcal{C}(z; p)$ is an increasing (resp., decreasing) function of p for $z' \geq z$. If $\mathcal{C}$ is submodular in z and has increasing (resp., decreasing) differences in z and p, then ψ is decreasing (resp., increasing) in p.

For each result that relies on numerical tests, we replace the label "proof" by "numerical test" and describe the test we have run. If the test is systematic and exhaustive, we label the result as a "proposition"; if the test only calculates a limited number of examples, we label the result as a "conjecture".

5 Computation procedures: PRAM

The policies outlined in Section 3 involve mainly addition, so we include only this operation explicitly into our computation models. We consider first a simple model without communication costs, called the PRAM, that has indeterminate organization structure; then we study a variant of the hierarchical associative computation model of Radner (1993).

Remark 2 Because, as is common in economic theory, we have adopted a simple stylized decision problem for actual complex management problems, we must adopt a correspondingly simple computation model as a proxy for actual complex human information processing. See Van Zandt (1999).

5.1 Computation model

The PRAM (parallel random access machine) is the simplest model of distributed computation. We first define the possible operations on data and the time each one takes. Computation then involves performing such operations on raw data and on partial results from previous operations. It is a model of distributed computation because operations can be performed concurrently. The model implicitly suppresses communication costs and delays, as if all managers had instant access to all data. As a consequence, the assignment of managers to operations and the communication of information among managers is indeterminate. It is the structure of such information flows that typically is used to characterize organizational structure in the decentralized information processing literature. However, with the PRAM we can still use the amount of computation performed each period as a measure of the size of the administrative staff.

We assume there is one operation, adding two numbers, which takes one period. Let w be the managerial wage rate or cost per operation.

Remark 3 The assumption that an operation takes one period is merely a modeling decision that allows the calendar length of the discrete time unit to be the amount of time it takes to perform an operation. An increase in processing speed corresponds to a reduction in the calendar length of a period and therefore to a reduction in the per-period volatility of the environment.

A PRAM can sum n numbers in $\lceil \log_2 n \rceil$ periods as follows: The data are divided into $n/2$ pairs, which are concurrently summed in one period. The $(n/2)$ partial results are then divided into pairs and summed in the next period, and so on. The number of data or partial results is cut in half each period, and hence it takes $\lceil \log_2 n \rceil$ periods to have a single partial result left, which is the sum of the n numbers. This requires $n-1$ operations and hence has a managerial cost of $(n-1)w$. If the PRAM adds the numbers in more than $\lceil \log_2 n \rceil$ periods, then fewer operations need to be performed concurrently but the total number of operations remains $n-1$. Since there are no communication costs, this increase in delay does not reduce the managerial cost and hence is not efficient.

5.2 Decision procedures without recall

In a policy without recall, the computation task is to add n numbers every k periods. As explained in Section 5.1, each sample can and should be summed in $\lceil \log_2 n \rceil$ periods. Hence, we can eliminate d as a parameter of the policies. Recycling notation, we denote the loss by

$$\mathcal{L}^{\mathrm{nr}}(n,k) \;=\; \frac{\sigma_\epsilon^2}{n} + \left(\lceil \log_2 n \rceil + \frac{k-1}{2}\right)\sigma_v^2 \,. \tag{7}$$

The managerial cost is $(n-1)w$ for each sample and hence the average managerial cost over the planning cycle is $(n-1)w/k$. The overall cost function is thus

$$\mathcal{C}^{\mathrm{nr}}(n,k) \;=\; \frac{\sigma_\epsilon^2}{n} + \left(\lceil \log_2 n \rceil + \frac{k-1}{2}\right)\sigma_v^2 + \frac{(n-1)w}{k} \,. \tag{8}$$

In the benchmark case in which $w=0$, it is optimal to sample each period ($k=1$). The resulting design problem is to choose n so as to minimize $\mathcal{L}^{\mathrm{nr}}(n,1) = \sigma_\epsilon^2/n + \lceil \log_2 n \rceil \sigma_v^2$. We obtain the following conclusion: as the environment changes more quickly, it is better to use fewer managers, thereby basing decisions on fewer, but more recent, data.

Proposition 1 *Consider PRAMs with no recall. Assume $w=0$. Then $\mathcal{L}^{\mathrm{nr}}(n,1)$ has increasing differences in n and σ_v^2 but has decreasing differences in n and σ_ϵ^2. Therefore, optimal sample size is decreasing in environmental volatility σ_v^2 and is increasing in the noisiness σ_ϵ^2 of observations. Hence, the size of the managerial staff decreases as the environment changes more quickly or observations become less noisy.*

Proof: That the cost function has increasing differences in σ_v^2 but decreasing differences in σ_ϵ^2 can be seen by the fact that $\partial\mathcal{C}(n,k)/\partial n$ is increasing in σ_v^2 and decreasing in σ_ϵ^2. Since the number of operations performed per period is $n-1$, a corollary is that the size of the managerial staff increases with sample size. □

When $w>0$, the sampling interval k is also a relevant decision variable.

Proposition 2 *Consider PRAMs with no recall. Assume $w>0$. Then $\mathcal{C}^{\mathrm{nr}}(n,k)$ is submodular in (n,k). Furthermore, $\mathcal{C}^{\mathrm{nr}}(n,k)$ has decreasing differences in (n,k) and σ_ϵ^2 but increasing differences in (n,k) and σ_v^2. Therefore, optimal (n,k) is increasing in σ_ϵ^2 and decreasing in σ_v^2.*

Proof: For testing submodularity of $\mathcal{C}^{\mathrm{nr}}(n,k)$, we can omit any additive terms that do not involve both n and k. This leaves only nw/k, which is easily seen to be submodular.

For checking increasing or decreasing differences for a given parameter, we can omit any additive terms that do not involve both the choice variables and the parameter. In the case of σ_ϵ^2, this leaves σ_ϵ^2/n, which clearly has decreasing differences in n and σ_ϵ^2. In the case of σ_v^2, this leaves $(\lceil \log_2 n \rceil + k/2)\,\sigma_v^2$, which clearly has increasing differences in (n,k) and σ_v^2.
□

Proposition 2 does not tell us what happens to the size of the administrative staff as the environment becomes more volatile. Recall that our proxy for the size of the staff is the average amount $(n-1)/k$ of managerial time used to calculate predictions. As σ_v^2 rises, both n and k fall. Consider the extreme values. As $\sigma_v^2 \downarrow 0$, it is possible to achieve approximately zero loss with very few managers by choosing large n (to get an accurate estimate of the state) but an even larger k (which reduces managerial costs but has little effect on the loss because the environment is changing slowly). Therefore, as $\sigma_v^2 \downarrow 0$, the managerial costs and hence the size of the administrative staff must go to 0. On the other hand, as the environment changes ever more quickly, sample size decreases to 1—at which point the administrative cost is 0 (a single observation is drawn each period and this becomes the estimate; there is no information processing).

Conjecture 1 *Consider PRAMs without recall. Assume $w > 0$. The relationship between σ_v^2 and managerial size is approximately an inverted* U.

NUMERICAL TEST: We calculate the optimal n and k as follows. The functional form for $\mathcal{C}(n,k)$ is convex in k, so the optimal k for fixed n is the solution to the first-order condition $\partial\mathcal{C}/\partial k = 0$ (rounded either up or down). We thereby obtain cost as a function of n and perform an exhaustive search over n, up to an upper bound determined by the condition that a decrease in n from the optimum must cause the loss to increase. For several values of w, we calculate optimal sample size n, sampling interval k, and managerial size $(n-1)/k$ for $-\log\sigma_v^2$ ranging from 1 to 10 in 400 increments of .025. We find a relationship similar to that illustrated in Figure 1. □

Figure 1 shows an example of n, k, and n/k as a function of $-\log\sigma_v^2$ for $\sigma_\epsilon^2 = 1$ and $w = .001$. Consistent with Proposition 2, n and k decrease monotonically with volatility. Managerial size n/k varies nonmonotonically. Although it has an approximate U shape, there are several local maxima. This is probably because of the rounding in delay $\lceil \log_2 n \rceil$.

5.3 Decision procedures with recall

Compared to a policy without recall, the computation of a policy with recall adds two components that are due to the need to average each newly calculated sample average with the current estimate, as in equation (5). Delay is increased by some amount d_r, to $\lceil \log_2 n \rceil + d_r$, and the managerial time per sample is increased by some amount m_r, to $n - 1 + m_r$. It is for this reason that the decision procedures with recall do not always dominate decision procedures without recall, even though $\mathcal{L}^{\mathrm{r}}(n,d,k) < \mathcal{L}^{\mathrm{nr}}(n,d,k)$ for all n, d, and k.

The long-run loss $\mathcal{L}^{\mathrm{r}}(n,d,k)$ from equation (6) plus the long-run managerial cost can be expressed as a function of n and k:

$$(9) \qquad \mathcal{C}^{\mathrm{r}}(n,k) \;=\; \left(\sqrt{k^2/4 + k\sigma_\epsilon^2/(n\sigma_v^2)} + \lceil \log_2 n \rceil + d_r - 1/2\right)\sigma_v^2 + (n-1+m_r)w/k\,.$$

Consider the polar case in which $w = 0$. As with no recall, it is optimal to sample every period so that $k = 1$. Cost as a function of n becomes

$$(10) \qquad \mathcal{L}^{\mathrm{r}}(n,1) \;=\; \left(\sqrt{(1/4 + \sigma_\epsilon^2/(n\sigma_v^2)} + \lceil \log_2 n \rceil + d_r - 1/2\right)\sigma_v^2\,.$$

Proposition 3 *Consider PRAMs with recall. The solution correspondence to* $\min_n \mathcal{L}^{\mathrm{r}}(n,1)$ *is increasing in σ_ϵ^2 and decreasing in σ_v^2.*

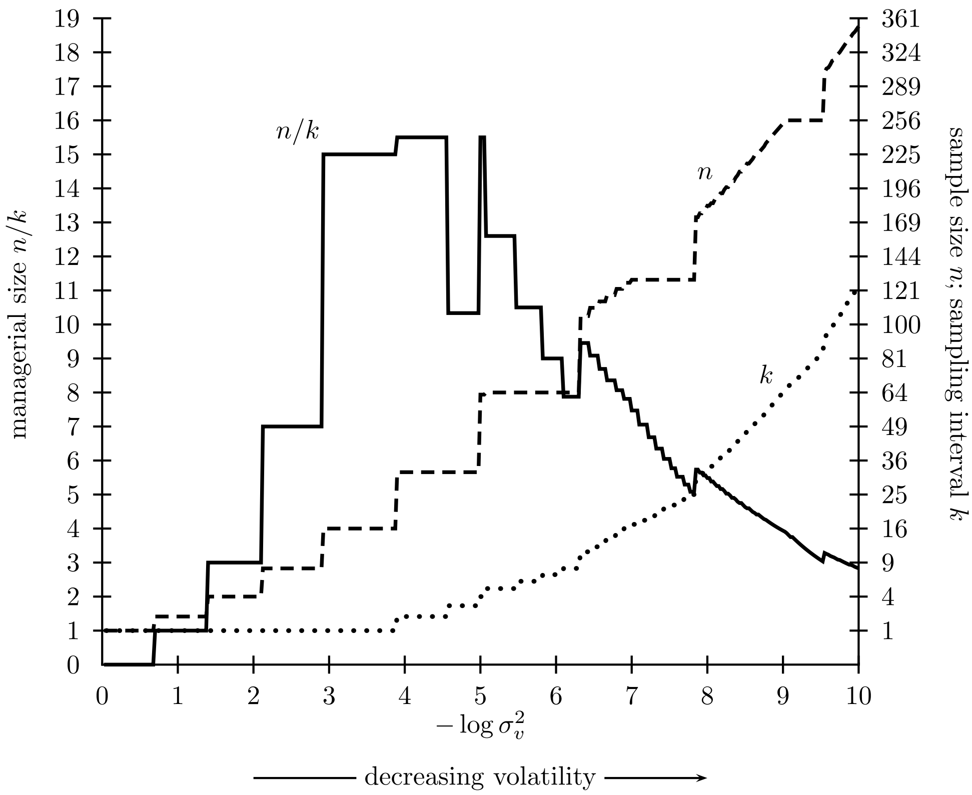

FIGURE 1. Optimal sample size n (dashed), sampling interval k (dotted), and managerial size n/k (solid) as a function of environmental volatility σ_v^2, where $\sigma_\epsilon^2 = 1$ and $w = .001$. Scale: $\sigma_v^2 \to -\log \sigma_v^2$; $n \to \sqrt{n}$; $k \to \sqrt{k}$; $n/k \to$ natural.

PROOF: Minimizing $\mathcal{L}^{\mathrm{r}}(n,1)$ is equivalent to minimizing $\sqrt{(1/4+\sigma_\epsilon^2/(n\sigma_v^2)} + \lceil \log_2 n \rceil$. A solution must always be a power of 2, because otherwise increasing n by 1 causes the first term to decrease but does not increase the second term (delay). Therefore, we restrict $n = 2^z$ for $z \in \{0, 1, \ldots\}$ and rephrase the design problem as $\min_z f(z;\lambda)$, where

$$f(z;\lambda) \;=\; \sqrt{1/4+\lambda 2^{-z}} + z$$

and $\lambda = \sigma_\epsilon^2/\sigma_v^2$. Observe that

$$\frac{\partial f}{\partial \lambda} \;=\; \frac{1}{2^{z+1}\sqrt{1/4+\lambda 2^{-z}}} \;=\; \frac{1}{\sqrt{2^{2z}+\lambda 2^{z+2}}}\,,$$

which is decreasing in z. Hence $\partial^2 f/\partial z \partial \lambda < 0$, and f has decreasing differences in z and λ. It follows that the optimal z is an increasing correspondence of λ. (Submodularity holds vacuously because the choice variable is one-dimensional.) Therefore, optimal n is increasing in σ_ϵ^2 and decreasing in σ_v^2. $\square$

Now consider the case of a positive managerial wage. We can see from equation (9) that, when the environment is nearly stable ($\sigma_v^2 \approx 0$), all procedures achieve nearly zero loss. This is different from policies without recall, because with recall any procedure is effectively using an infinite amount of data. As the environment changes more slowly, many data from one period and the same number of accumulated data from previous periods yield nearly the same loss. Therefore, if managers are costly, then the optimal procedure should be the one that has very low cost and hence low managerial size. On the other hand, if the environment is extremely volatile, then the cost of delay dominates and optimal sample size must again be 1.

We thus have the following proposition.

Proposition 4 *Consider PRAMs with recall. Fix $w > 0$ and $\sigma_\epsilon^2 \geq 0$. There is a $\bar{\sigma}_v^2 >$ such that, for $\sigma_v^2 > \bar{\sigma}_v^2$, sample size is $n = 1$ and hence managerial size is at most m_r. As $\sigma_v^2 \downarrow 0$, managerial size converges to* 0.

PROOF: Consider a procedure with $k = n = 1$. The cost is then

$$(\sqrt{1/4+\sigma_\epsilon^2/\sigma_v^2} + d_r - 1/2)\sigma_v^2 + m_r w\,.$$

Any procedure with $n > 1$ has $d \geq 1$ and hence a cost of at least $(1+d_r)\sigma_v^2$. There is a $\bar{\sigma}_v^2$ such that, for $\sigma_v^2 > \bar{\sigma}_v^2$, this cost is larger than the cost for $k = n = 1$. Hence, for $\sigma_v^2 > \bar{\sigma}_v^2$, an optimal procedure has $n = 1$.

For any fixed n and k, the loss converges to 0 as $\sigma_v^2 \downarrow 0$, so the total cost converges to $(n-1+m_r)w/k$. Hence, $(n-1+m_r)w/k$ is an approximate upper bound on total cost for small σ_v^2. This holds for all n and k; by letting $k \to \infty$, this upper bound converges to 0. The managerial cost of an optimal procedure must therefore converge to 0 as $\sigma_v^2 \downarrow 0$, which means that managerial size also converges to 0. $\square$

6 Computation procedures: Stationary hierarchies

As an alternative computation model, we consider a variant of the model of associative information processing in Radner (1993). Each manager can read and aggregate one report (a raw datum or a partial result from another manager) in one period. For example, two managers who wish to calculate the prediction $\hat{x}_t$ without recall from a sample of size 10 can do so by each sequentially reading and aggregating five data and then having one of the managers aggregate the partial result of the other manager, setting the estimate equal to the sum of the data. This takes 6 periods, so those data must have been collected in period $t-6$.

This model has an implicit communication cost, because it takes one period to read any report. Decentralizing information process therefore involves a trade-off: it reduces delay but increases managerial costs. Hence, unlike in the PRAM model—where it was always optimal to use the minimal delay for a given sample size—delay re-enters as an independent design variable.

The communication cost also makes the communication structure determinate. If one manager receives a message from another manager and adds it to his current partial result, the reading and aggregation of the message takes one period. If instead two managers send their partial results to a third manager, then the latter needs two periods to read and aggregate the messages. Hence, the first communication flow is typically better.

The computation task for policies without recall is identical to the periodic processing problem in Radner (1993), and the task for policies with recall is similar. As shown in Van Zandt (1998), efficient organizations in Radner's model are not hierarchical, because communication costs can be reduced and throughput increased by having the cohorts processed by different managers (and any one manager does not always process cohorts with the same managers). Following Orbay (2001), we impose a *stationarity* assumption, which requires that each sample be processed in the same way by the same hierarchy. One motivation for this assumption is that it captures the unmodeled communication costs that occur when managers are constantly shifting their channels of communication. The administrative staff thus consists of a fixed hierarchy of managers. We assume further that managers are paid a salary and hence that the managerial cost is mw, where m is the number of managers in the hierarchy and w is the per-period salary.

The stationarity assumption limits the frequency with which a hierarchy may collect data from the environment. If the maximum amount of time any manager spends per sample is k consecutive periods, then a stationary hierarchy can start processing a new sample and compute a prediction every k periods but not more frequently. It is inefficient to collect samples less frequently, because there is no impact on managerial costs: managers end up with more idle time but are paid salaries anyway.

With this computation model, a computation procedure is not merely an abstract algorithm but rather a hierarchical organization; the parameters of the policy that a hierarchy computes can be interpreted as parameters of the hierarchy. We introduce new terminology that makes use of this interpretation. Since the time k between samples is the maximum number of inputs any manager can read per sample, we call k the hierarchy's *span limit.* The sample size n is the amount of data the hierarchy can process per cohort, so we call it the hierarchy's *capacity.* The number m of managers is the *size* of the hierarchy. The delay d of a policy is the hierarchy's delay. Orbay (2001) characterizes the efficient frontier of such stationary hierarchies with respect to these four parameters and provides an algorithm for determining n as a function of (m, d, k) for points on the frontier; see Orbay (2001, Section 4) for details.

As in the preceding section, we consider a benchmark case in which managers are costless. Hierarchies should then have the maximum capacity n for fixed d and k. Such hierarchies are said to be *maximal.* Orbay (2001) shows that the number $M(d, k)$ of managers and the capacity $N(d, k)$ of a maximal hierarchy with delay d and span limit k is given recursively by

$$N(0,k)=1\,,\quad N(d,k)\;=\;\sum_{i=1}^{\min(d,k)} N(d-i,k)\,; \tag{11}$$

$$M(0,k)=0\,,\quad M(d,k)\;=\;1+\sum_{i=1}^{\min(d,k)} M(d-i,k)\,. \tag{12}$$

For the recursion, $N(d,k)$ and $M(d,k)$ are defined for all positive values of d and k, but a hierarchy must have $d \geq k$.

6.1 Hierarchies without recall

Consider stationary hierarchies that compute policies without recall. Suppose first that $w=0$, so that optimal hierarchies are maximal. The reduced-form loss to be minimized is $\mathcal{L}^{\mathrm{nr}}(d,k)=\sigma_\epsilon^2/N(d,k)+(d+(k-1)/2)\sigma_v^2$.

We obtain an analog to Proposition 1 as follows.

Proposition 5 *Consider hierarchies without recall. Assume $w=0$. Optimal size and delay decrease with environmental volatility σ_v^2 and increase with observation error σ_ϵ^2. The same is approximately true for span limit.*

Numerical test: The solution to $\min_{d,k}\mathcal{L}^{\mathrm{nr}}(d,k)$ depends only on the ratio $\sigma_\epsilon^2/\sigma_v^2$. Therefore, we can normalize $\sigma_\epsilon^2=1$ and test how d, k, and $M(d,k)$ depend on σ_v^2. For each σ_v^2, we numerically minimize $\mathcal{L}^{\mathrm{nr}}(d,k)$. We do this for initial increments of 0.1 in the value of $\log\sigma_v^2$; near values of σ_v^2 at which d or k changes, we calculate the solution for smaller increments. In the numerical results, d and $M(d,k)$ are monotone; k is nearly monotone except at a few values of σ_v^2. Figure 2 shows d, k, and $M(d,k)$ as a function of σ_v^2. □

Proposition 6 *Consider hierarchies without recall. Assume $w=0$. Optimal hierarchy size grows to infinity as the environmental volatility σ_v^2 converges to 0.*

Proof: In the limit, $\mathcal{L}^{\mathrm{nr}}(d,k)=\sigma_\epsilon^2/N(d,k)$ when $\sigma_v^2=0$. Thus, the loss decreases with the hierarchy's capacity. For maximal hierarchies, the number of managers increases without bound as a function of capacity. □

When managers are costly, we write cost as

$$\mathcal{C}^{\mathrm{nr}}(m,n,d,k)=\sigma_\epsilon^2/n+(d+(k-1)/2)\sigma_v^2+mw\,,$$

with parameters restricted to the feasible values for stationary hierarchies. We once again obtain a nonmonotone relationship between volatility and managerial size, and managerial size converges to 1 as $\sigma_v^2\downarrow 0$.

Proposition 7 *Consider hierarchies with no recall. Assume $w>0$ and $\sigma_\epsilon^2>0$. If $\sigma_v^2>\sigma_\epsilon^2$ then the optimal hierarchy consists of a single manager who processes one observation each period, so $k=d=n=1$. For small σ_v^2, the optimal hierarchy also has one manager and hence $k=d=n$, but $n\to\infty$ as $\sigma_v^2\to 0$.*

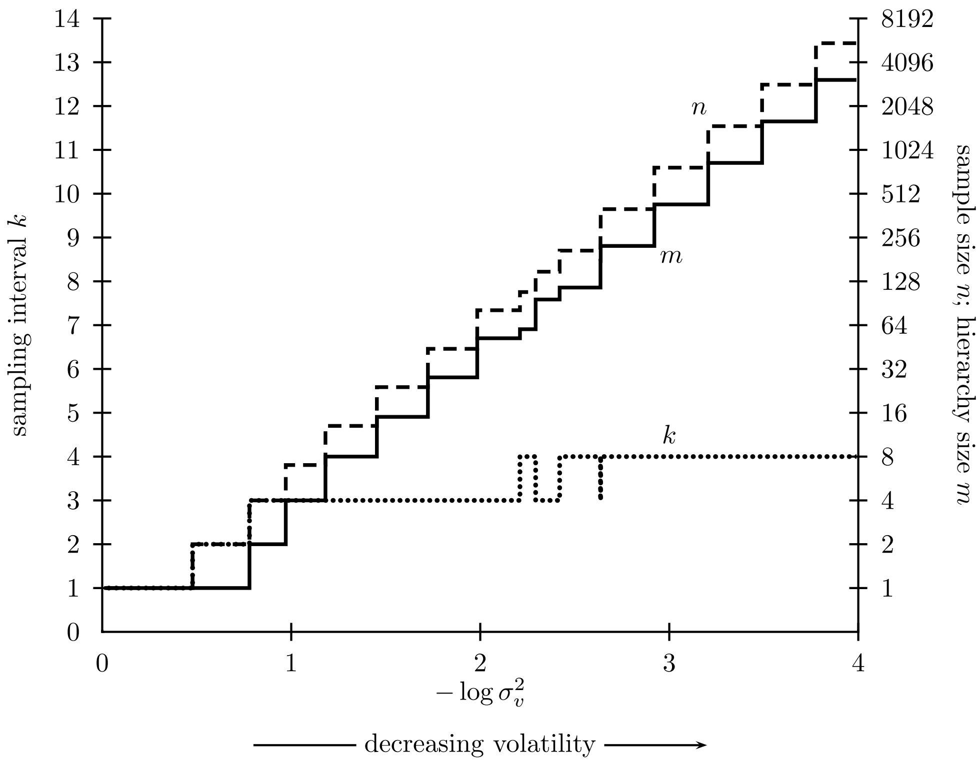

FIGURE 2. Optimal hierarchies without recall as a function of σ_v^2 when managers are costless. Graph shows sample size n (dashed, $\log_2$ scale), sampling interval k (dotted, natural scale), and hierarchy size m (solid, $\log_2$ scale). Sampling error σ_ϵ^2 is normalized to 1.

PROOF: An efficient hierarchy with a single manager processes a cohort of size n in n periods and processes a new sample every n periods. Thus, once $m = 1$, we have $k = d = n$.

Consider such a one-manager hierarchy. If the sample size is 1, then the cost is $\sigma_\epsilon^2 + \sigma_v^2 + w$. Any other hierarchy has a delay of at least 2 (and $k \geq 1$ and $m \geq 1$), so its cost is at least $2\sigma_v^2 + w$. The first hierarchy has a lower cost if $\sigma_v^2 > \sigma_\epsilon^2$.

Consider a one-manager hierarchy that processes samples of size n. The cost for this policy converges to $\sigma_\epsilon^2/n + w$ as $\sigma_v^2 \downarrow 0$. Therefore, $\sigma_\epsilon^2/n + w$ is an approximate upper bound on total cost for small σ_v^2. This holds for all n; by letting $n \to \infty$, this upper bound converges to w. Hence the managerial cost of optimal procedures must converge to w as $\sigma_v^2 \downarrow 0$. This is possible only if $n \to \infty$ and $m \to 1$. □

Conjecture 2 *Consider hierarchies with no recall. Assume $w > 0$. Optimal hierarchy size has the following approximate relationship to environment volatility σ_v^2: hierarchy size first increases and then decreases as σ_v^2 increases from zero.*

NUMERICAL TEST: We test this for several values of w and σ_ϵ^2 by solving for the optimal hierarchy for 100 evenly spaced values of $-\log \sigma_v^2$. The calculation of the optimal hierarchy uses the characterization of the efficiency frontier in Orbay (2001). The optimization finds only local optima, but for each case we start the search from four different points and verify that it yields the same solution. Figure 3 shows d, k, and $\log m$ as a function of σ_v^2 for $\sigma_\epsilon^2 = 1$ and $w = 0.001$. □

6.2 Optimal hierarchies with recall

As with the PRAM, we give a reduced-form treatment of the change in the computation task that results from recall. The updating of the estimate from newly calculated sample means can always be delegated to a new manager who performs no other operations. We thus assume that the number of managers increases by 1 and that the updating introduces a delay d_r. This introduces a loss $d_r\sigma_v^2$ and a managerial cost w, which are independent of the other design variables. These costs therefore affect the comparison between hierarchies with recall and hierarchies without recall, but not the ranking withing the class of hierarchies with recall. To simplify notation, we suppress such costs. The overall cost is thus

$$\mathcal{C}^{\mathrm{r}}(m,n,d,k) \;\equiv\; \left(\sqrt{(k^2/4 + k\sigma_\epsilon^2/(n\sigma_v^2)} + d - 1/2\right)\sigma_v^2 + wm\,. \tag{13}$$

In the benchmark case of $w = 0$, hierarchies are maximal and the cost is only the loss, which we express as a function of d and k:

$$\mathcal{L}^{\mathrm{r}}(d,k) \;\equiv\; \left(\sqrt{(k^2/4 + k\sigma_\epsilon^2/(N(d,k)\sigma_v^2)} + d - 1/2\right)\sigma_v^2\,. \tag{14}$$

Proposition 8 *Consider hierarchies with recall. Assume $w = 0$. Optimal size, delay, and span limit will decrease with environmental volatility σ_v^2.*

NUMERICAL TEST: The solution to $\min_{d,k} \mathcal{L}^{\mathrm{r}}(d,k)$ depends only on the ratio $\sigma_\epsilon^2/\sigma_v^2$. Therefore, we can normalize $\sigma_\epsilon^2 = 1$ and test how d, k, and $M(d,k)$ depend on σ_v^2. The calculations proceed as described for Proposition 5. Figure 4 shows d, k, and $M(d,k)$ as a function of σ_v^2. □

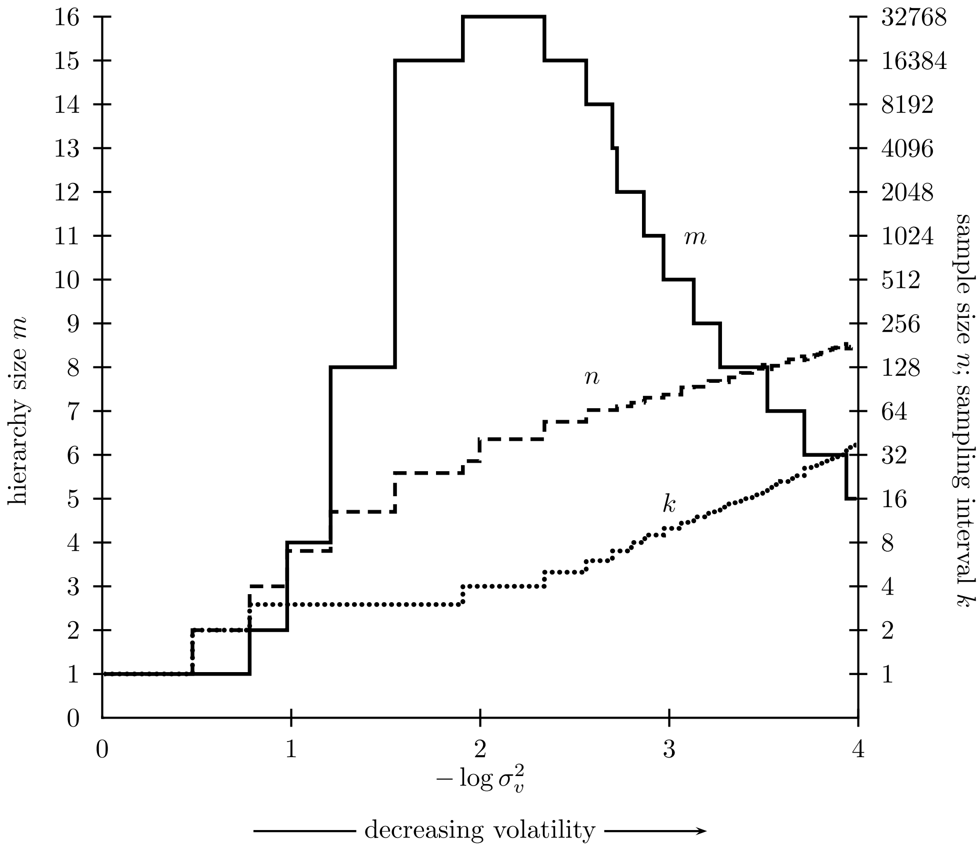

FIGURE 3. Optimal hierarchies without recall as a function of σ_v^2 for the case $w = 0.001$ and $\sigma_\epsilon^2 = 1$. Graph shows sample size n (dashed, $\log_2$ scale), sampling interval k (dotted, $\log_2$ scale), and hierarchy size m (solid, natural scale).

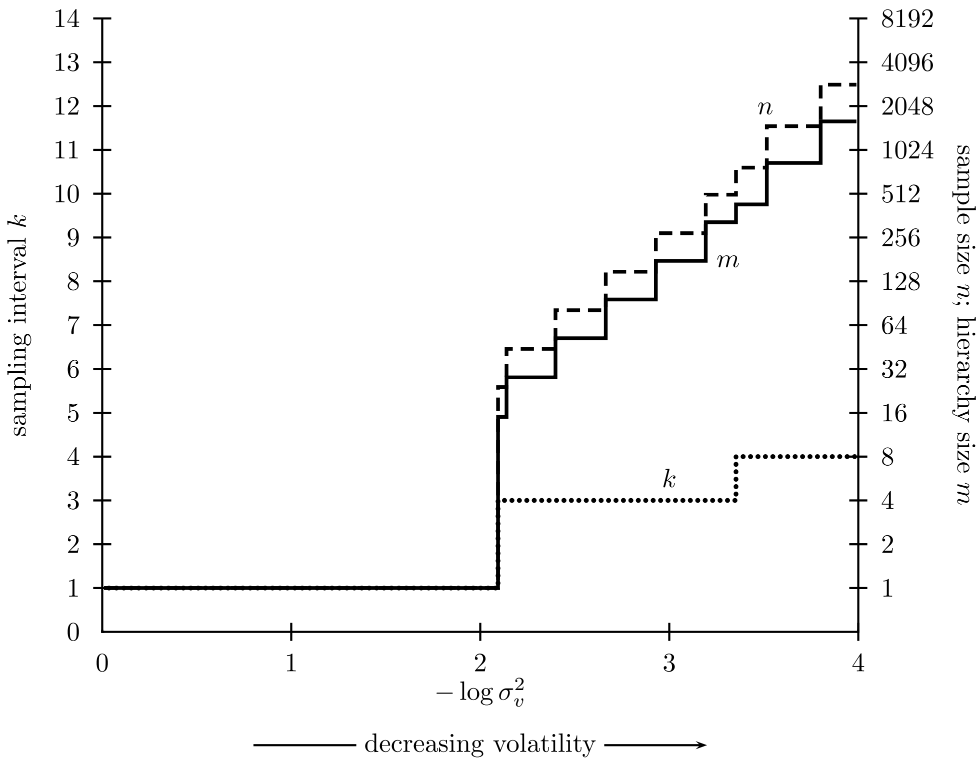

FIGURE 4. Optimal hierarchies with recall as a function of σ_v^2 when managers are costless. Graph shows sample size n (dashed, $\log_2$ scale), sampling interval k (dotted, natural scale), and hierarchy size m (solid, $\log_2$ scale). Sampling error σ_ϵ^2 is normalized to 1.

Proposition 9 *Consider hierarchies with recall. Assume* $w = 0$*. If* $\sigma_v^2 > \sigma_\epsilon^2/2$*, then the optimal hierarchy has one manager who processes one observation each period so that* $n = d = k = 1$*.*

PROOF: Minimizing $\mathcal{L}^{\mathrm{r}}(n,d,k)$ is equivalent to minimizing

$$\hat{\mathcal{L}}(n,d,k) \;=\; \sqrt{(k^2/4 + k\sigma_\epsilon^2/(N(d,k)\sigma_v^2)} + d\,.$$

Consider a hierarchy with one manager who processes one observation each period, so that $n = d = k = 1$. Then $\hat{\mathcal{L}}(1,1,1) = \sqrt{1/4 + \sigma_\epsilon^2/\sigma_v^2} + 1$. Any other hierarchy has delay of at least 2, so $\hat{\mathcal{L}}(\cdot)$ is at least $\sqrt{1/4} + 2 = 5/2$. The first hierarchy is thus optimal if $\sqrt{1/4 + \sigma_\epsilon^2/\sigma_v^2} + 1 < 5/2$. This simplifies to $\sigma_v^2 > \sigma_\epsilon^2/2$. □

Now consider the case of a positive managerial wage. We obtain an analog to Proposition 4. When the environment is very stable, all hierarchies achieve nearly zero loss and the optimal hierarchy should be that which minimizes cost; it has just one manager. When the environment is highly volatile, the cost of delay dominates and optimal hierarchy must again consist of a single manager.

Proposition 10 *Consider hierarchies with recall. Assume* $w > 0$ *and* $\sigma_\epsilon^2 > 0$*. If* $\sigma_v^2 > \sigma_\epsilon^2/2$*, then the optimal hierarchy has size 1 and consists of a single manager who processes one observation each period* ($n = d = k = 1$)*. For small* σ_v^2*, hierarchy size is 1 and* $n = d = k$*.*

PROOF: If $\sigma_v^2 > \sigma_\epsilon^2/2$ then, according to Proposition 9, a single-manager hierarchy that processes one observation each period has strictly lower loss than any other hierarchy. Since it also has the lowest managerial cost, it is optimal.

By equation (13), for all hierarchies the loss converges to 0 as $\sigma_v^2 \downarrow 0$. Thus, for small σ_v^2, the optimal hierarchy must have the lowest managerial cost and hence must have size 1. □

This suggests an inverted-U relationship between volatility and managerial size, which we confirm in several examples.

Conjecture 3 *Consider hierarchies with recall. Assume* $w > 0$*. As environmental volatility increases from zero, optimal hierarchy size first increases and then decreases.*

NUMERICAL TEST: We test this for several values of w and σ_ϵ^2 by solving for the optimal hierarchy for 100 evenly spaced values of $-\log \sigma_v^2$; see Conjecture 2 for details. Figure 5 shows d, k, and $\log m$ as functions of σ_v^2 for $\sigma_\epsilon^2 = 1$ and $w = 10^{-5}$. The general inverted-U shape is preserved across simulations. □

7 Summary

We have found several relationships between managerial size and environmental volatility. For the benchmark case in which $w = 0$, managerial size decreases monotonically as volatility rises. This is due to the following trade-off: higher sample size increases the precision of estimates for when the sample was taken but also increases delay before the estimate is used; the loss due to this delay increases when the environment is more volatile. However, when $w > 0$, the relationship between volatility and managerial size is nonmonotonic. When

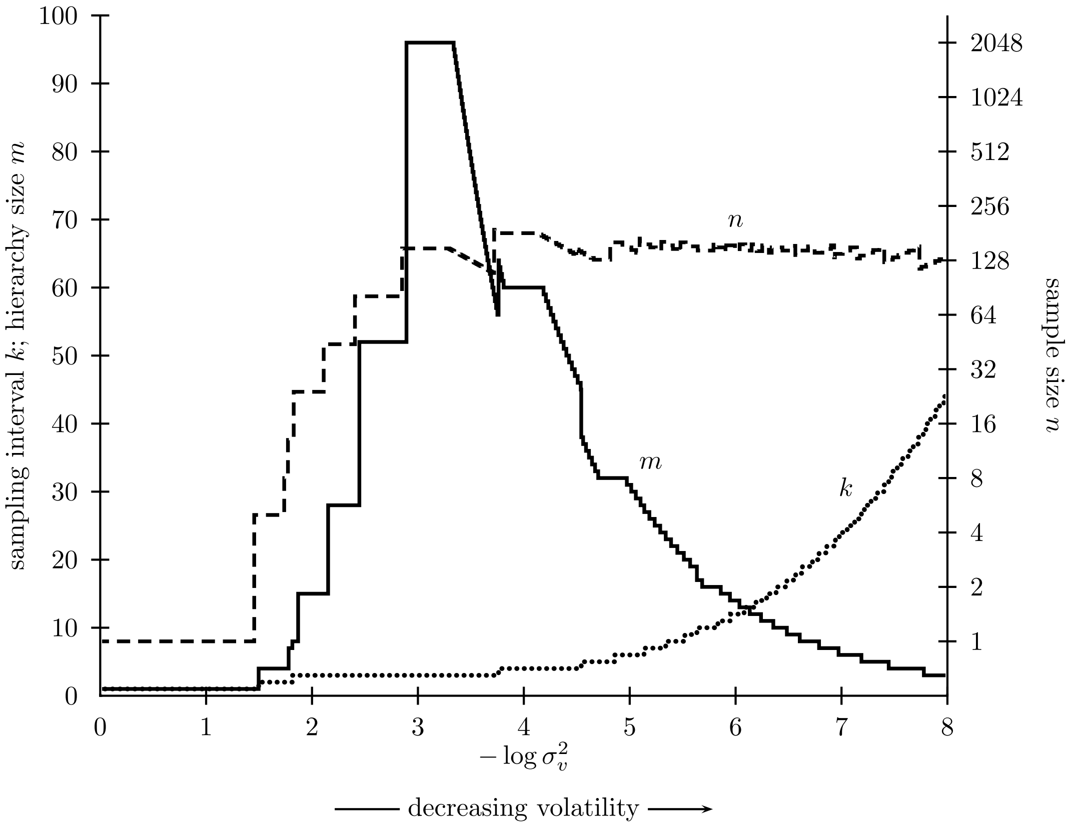

FIGURE 5. Optimal hierarchies with recall as a function of σ_v^2 for the case of $\sigma_\epsilon^2 = 1$ and $w = 0.00001$. Graph shows sample size n (dashed, $\log_2$ scale), sampling interval k (dotted, natural scale), and hierarchy size m (solid, natural scale).

volatility is very high and hence delay incurs a high loss, sample size and thus managerial size is small. As volatility initially decreases, delays incurs a lower cost, so sample size and managerial size increase. However, increasing environmental stability also causes the value of frequent sampling to decrease, which pushes in the direction of smaller managerial size. For sufficiently low volatility, this effect dominates and optimal managerial size begins to decrease as the environment becomes more stable.

These results are confirmed both for policies with recall and for policies without recall, as well as for the PRAM and stationary hierarchies computation models.

References

Milgrom, P. and Roberts, J. (1990). Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica*, 58, 1255–1277.

Milgrom, P. and Shannon, C. (1994). Monotone comparative statics. *Econometrica*, 62, 157–180.

Orbay, H. (2001). Information processing hierarchies. *Journal of Economic Theory*. Forthcoming.

Radner, R. (1993). The organization of decentralized information processing. *Econometrica*, 62, 1109–1146.

Topkis, D. (1979). Equilibrium points in nonzero-sum n-person submodular games. *SIAM Journal of Control and Optimization*, 17, 773–787.

Van Zandt, T. (1998). The scheduling and organization of periodic associative computation: Efficient networks. *Economic Design*, 3, 93–127.

Van Zandt, T. (1999). Real-time decentralized information processing as a model of organizations with boundedly rational agents. *Review of Economic Studies*, 66, 633–658.

Van Zandt, T. and Radner, R. (2001). Real-time decentralized information processing and returns to scale. *Economic Theory*. Forthcoming.

Vives, X. (1990). Nash equilibrium with strategic complementarities. *Journal of Mathematical Economics*, 19, 305–321.