Zandt, Timothy Van; Radner, Roy

Working Paper

Real-Time Decentralized Information Processing and Returns to Scale

Discussion Paper, No. 1233

Provided in Cooperation with:
Kellogg School of Management - Center for Mathematical Studies in Economics and Management Science, Northwestern University

Suggested Citation: Zandt, Timothy Van; Radner, Roy (1998) : Real-Time Decentralized Information Processing and Returns to Scale, Discussion Paper, No. 1233, Northwestern University, Kellogg School of Management, Center for Mathematical Studies in Economics and Management Science, Evanston, IL

This Version is available at:
https://hdl.handle.net/10419/221589

Discussion Paper No. 1233

Real-Time Decentralized Information Processing and Returns to Scale

Timothy Van Zandt
Northwestern University
and INSEAD

and

Roy Radner
Stern School of Business
New York University

November 17, 1998

Math Center web site:
http://www.kellogg.nwu.edu/research/math

References

Arrow, K. J. (1974). *The Limits of Organization.* New York: W. W. Norton Co.

Brynjolfsson, E. and Hitt, L. M. (1998). Information technology and organizational design: Evidence from micro data. MIT Sloan School of Management and Wharton School.

Brynjolfsson, E., Malone, T. W., Gurbaxani, V., and Kambil, A. (1994). Does information technology lead to smaller firms? *Management Science*, 40, 1628–1644.

Chandler, A. D. (1966). *Strategy and Structure.* New York: Doubleday.

Chandler, A. D. (1990). *Scale and Scope: The Dynamics of Industrial Capitalism.* Cambridge, MA: Harvard University Press.

Coase, R. (1937). The nature of the firm. *Economica*, 4, 386–405.

Dickinson, H. D. (1939). *Economics of Socialism.* Oxford: Oxford University Press.

Geanakoplos, J. and Milgrom, P. (1991). A theory of hierarchies based on limited managerial attention. *Journal of the Japanese and International Economies*, 5, 205–225.

Hart, O. (1995). *Firms, Contracts, and Financial Structure.* Oxford: Oxford University Press.

Hayek, F. A. v. (1940). Socialist calculation: The competitive 'solution'. *Economica*, 7, 125–149.

Hayek, F. A. v. (1945). The use of knowledge in society. *American Economic Review*, 35, 519–530.

Kaldor, N. (1934). The equilibrium of the firm. *Economic Journal*, 44, 70–71.

Keren, M. and Levhari, D. (1983). The internal organization of the firm and the shape of average costs. *Bell Journal of Economics*, 14, 474–486.

Lange, O. (1936). On the economic theory of socialism: Part I. *Review of Economic Studies*, 4, 53–71.

Lange, O. (1937). On the economic theory of socialism: Part II. *Review of Economic Studies*, 4, 123–142.

Loève, M. (1978). *Probability Theory II.* New York: Springer-Verlag.

Marschak, J. and Radner, R. (1972). *Economic Theory of Teams.* New Haven, CT: Yale Univeristy Press.

Marschak, T. (1972). Computation in organizations: The comparison of price mechanisms and other adjustment processes. In C. B. McGuire and R. Radner (Eds.), *Decision and Organization*, chapter 10, pp. 237–281. Amsterdam: North-Holland. Second edition published in 1986 by University of Minnesota Press.

Marschak, T. (1996). On economies of scope in communication. *Economic Design*, 2, 1–31.

Meagher, K. J. (1996). How to chase the market: An organizational and computational problem in decision making. Australian National University.

NORTHWESTERN UNIVERSITY
CENTER FOR MATHEMATICAL STUDIES IN ECONOMICS AND MANAGEMENT SCIENCE
Discussion Paper No. 1233

Real-Time Decentralized Information Processing and Returns to Scale*

Timothy Van Zandt†
Northwestern University
and INSEAD

Roy Radner
Stern School of Business
New York University

November 17, 1998

Abstract

We use a model of real-time decentralized information processing to understand how constraints on human information processing affect the returns to scale of organizations. We identify three informational (dis)economies of scale: diversification of heterogeneous risks (positive), sharing of information and of costs (positive), and crowding out of recent information due to information processing delay (negative). Because decision rules are endogenous, delay does not inexorably lead to decreasing returns to scale. However, returns are more likely to be decreasing when computation constraints, rather than sampling costs, limit the information upon which decisions are conditioned. The results illustrate how information processing constraints together with the requirement of informational integration cause a breakdown of the replication arguments that have been used to establish nondecreasing technological returns to scale.

JEL Classifications: D83, D23
Keywords: returns to scale, real-time computation, decentralized information processing, organizations, bounded rationality

Authors' addresses:

Timothy Van Zandt
Math Center (CMS-EMS)
2001 Sheridan Road, Room 371
Northwestern University
Evanston, IL 60208-2014
USA

Voice: +1 (847) 491-4414
Fax: +1 (847) 491-2530
Email: tvz@nwu.edu
Web: zandtwerk.kellogg.nwu.edu

Roy Radner
Stern School of Business MEC 9-68
New York University
44 West Fourth Street
New York, NY 10012
USA

Voice: (212) 998-0813
Fax: (212) 995-4228
Email: rradner@stern.nyu.edu

---

*We would like to thank Patrick Bolton, In-Koo Cho, Jacques Cremer, Mathias Dewatripont, George Mailath, Paul Milgrom, two anonymous referees, and numerous seminar participants for helpful comments. Archishman Chakraborty, Sangjoon Kim and Bilge Yilmaz provided valuable research assistance. Previous versions of this paper (which initially had the title "Information Processing and Returns to Scale of a Statistical Decision Problem") contained material that now appears in Van Zandt (1998c).

†This research supported in part by grants SES-9110973, SBR-9223917, and IRI-9711303 from the National Science Foundation, by a CORE Research Fellowship (1993–1994), and by grant 26 of the "Pôle d'Attraction Interuniversitaire" program of the Belgian Government.

v. $\{Y_t\}$ *and* $\{Z_{it}\}$ *are* AR(1) *processes with autoregressive parameters close to* 1 *and with innovation terms whose variances are close to* 2 *and* 1*, respectively; and*

vi. either $\Psi(\epsilon) = \epsilon^2$ *or the stochastic processes* $\left\{\{X_{1t}\}, \{X_{2t}\}, \dots\right\}$ *are Gaussian.*

PROOF OF THEOREM 4B: **Notation:** By assumptions (iv) and (v), we can write $X_{it} = \mu + Y_t + Z_{it}$, where $\{Y_t\}$ and $\{Z_{it}\}$ are mean-zero AR(1) processes. We write these as $Y_t = \gamma Y_{t-1} + V_t$ and $Z_{it} = \beta Z_{i,t-1} + W_{it}$, where $\{V_t\}$ and $\{W_{it}\}$ are noise processes.

Parameter values: Assume that $\gamma = \beta = 1$, that $\mathrm{Var}(V_t) = 2$ and $\mathrm{Var}(W_{it}) = 1$, and that $w = 0$. We will claim that our calculations vary continuously with these parameters, so that the results hold when the values are close to the ones given. In particular, the fact that the processes are not stationary when $\gamma = \beta = 1$ does not invalidate the calculations and results. Assume first that $\Psi(\epsilon) = \epsilon^2$. We will later explain how to adapt the calculations to the Gaussian case.

Intuition: Firm 1 can make a good forecast of X_t^1 simply by observing $X_{1,t-1}$. Firm 1 cannot differentiate Y_t and Z_{it} with this data, but it does not need to. For large n, firm n cannot make as good a forecast of each X_{it} because it cannot use recent data about most of the processes. However, what is mainly needs is to forecast Y_t (a law-of-large-numbers effect diminishes the average loss from errors in forecasting the idiosyncratic terms). For each $s \in \mathbb{N}$ and $i \in \mathbb{N}$, $X_{i,t-s}$ is a noisy observation of Y_{t-s}. Unlike in the sampling problem, the number of these noisy observations used in a forecast is bounded for each s, and so the forecast of Y_t may have a greater expected loss than firm 1's forecast of X_t^1.

Loss for firm 1: By assumption (iii), firm 1 can compute $A_t = X_{1,t-1}$. Since $X_t^1 = X_{1t} = X_{1,t-1} + V_t + W_{1t}$, the expected error is

$$E\left[(V_t + W_{1t})^2\right] \;=\; \mathrm{Var}(V_t) + \mathrm{Var}(W_{1t}) \;=\; 3\ .$$

Loss for firm n: Let $n \geq 2$ and $\pi \in \Pi^n$. Assumption (i) implies that the data from dates $t-1$ and $t-2$ that may be included in H_t^π is at most one of the following:

Case 1 $X_{i,t-1}$ and $X_{i,t-2}$ for some i;
Case 2 $X_{i,t-1}$ and $X_{j,t-2}$ for some i and some $j \neq i$;
Case 3 $X_{i,t-2}$ and $X_{j,t-2}$ for some i and some $j \neq i$.

In addition, H_t^π may include data from periods $t-3$ and earlier.

Consider Case 2. For example, H_t^π includes $X_{1,t-1}$, $X_{2,t-2}$, and data from periods $t-3$ and earlier. To construct a lower bound on the expected loss, we can assume that H_t^π includes $X_{1,t-3}$ and $X_{2,t-3}$. Since A_t^π is a linear decision rule, there are constants $\alpha_1, \alpha_2 \in \mathbb{R}$ such that $A_t^\pi = n\alpha_1 B_1 + n\alpha_2 B_2 + B_3$, where

$$\begin{aligned} B_1 \;&\equiv\; X_{1,t-1} - X_{1,t-3} \;=\; V_{t-1} + W_{1,t-1} + V_{t-2} + W_{1,t-2}\ , \\ B_2 \;&\equiv\; X_{2,t-2} - X_{2,t-3} \;=\; V_{t-2} + W_{2,t-2}\ , \end{aligned}$$

and B_3 is a measurable function of H_{t-3}. Let

$$B \;\equiv\; \frac{X_t^n - X_{t-3}^n}{n} \;=\; V_t + V_{t-1} + V_{t-2} + \frac{1}{n}\sum_{i=1}^{n}(W_{it} + W_{i,t-1} + W_{i,t-2})\ .$$

1 Introduction

1.1 Motivation

The purpose of this paper is to study formally whether and how human information processing constraints can limit the scale of centralized decision making in organizations with endogenous administrative staffs. This abstract question is relevant, for example, to the theory of firms and industrial organization, given that decision making appears to be more centralized when an industry is controlled by a single firm than when output is produced by independent firms. Hence, any advantages to decentralized decision making may limit the scale of firms.

We address this question by characterizing the average cost curve for a statistical decision problem that exhibits centralized decision making, in a model in which decisions are made in real time by an endogenous number of boundedly rational agents. A related model was introduced in Radner and Van Zandt (1992); here we develop a new axiomatic computation model, contrast it with a benchmark sampling problem, and provide more extensive and precise results. In the spirit of the theory of teams, we restrict attention to informational and computational decentralization, leaving aside issues of incentives and governance. The administrative agents in our model are boundedly rational because it takes them time to process and use information. This time represents both managerial wages that must be paid and also, more critically, decision-theoretic delay that constrains the use of recent information. The main theme of this paper is that such delay can lead to decentralization of decision making and bounded firm size—confirming, as stated by Hayek (1945, p. 524), that "we need decentralization because only thus can we ensure that the knowledge of the particular circumstances ... be promptly used".

1.2 Real-time decentralized information processing

We use a real-time computation model—that is, a model where computation constraints are embedded into a temporal decision problem in which data arrive and decisions are made at multiple epochs. Such a model, whose properties are explored in Van Zandt (1998c), captures in a sophisticated way the fact that human information processing constraints limit the use of recent information.[1]

The decision problem we study is the estimation in each period of the sum of n discrete-time stochastic processes. This is one of the control problems faced by a firm or plant that sets its production level centrally in order to meet the uncertain total demand of n sales offices or customers,[2] or by a firm or plant that needs to estimate the average productivity of n workers (machines or shops) based on past individual productivity indices. This decision problem is also part of resource allocation problems—such as allocating capital to n projects or assigning output orders to n production shops—in which one of the steps is aggregating profit, cost or productivity indices in order to calculate a shadow price. The size or scale of the decision problem is n.

[1] Marschak (1972) was the first economic model of real-time processing (that we are aware of). He studied how different price adjustment processes affect delay, but he did not study decentralization of information processing and the effects of problem size.

[2] For example, Benetton's must respond quickly to changing market conditions at its many retail outlets in order to implement just-in-time inventory management practices and thereby reduce inventory costs.

Computation Problem Assumptions 2 and 4 imply that there is $\pi^{n+1} \in \Pi^{n+1}$ such that $A_t^{\pi^{n+1}} = ((n+1)/n)A_t^{\pi^n}$ for $t \in \mathbb{N}$ and such that $C(\pi^{n+1}) = C(\pi^n) \equiv C^*$. For $t \in \mathbb{N}$, let $\ell_t^n \equiv E\big[\psi^n(X_t^n - A_t^{\pi^n})\big]/n$ and $\ell_t^{n+1} \equiv E\big[\psi^{n+1}(X_t^{n+1} - A_t^{\pi^{n+1}})\big]/(n+1)$, so that

$$\begin{aligned} \mathrm{AC}^n(\pi^n) &= \Gamma(\{\ell_t^n\}) + C^*/n \\ \mathrm{AC}^{n+1}(\pi^{n+1}) &= \Gamma(\{\ell_t^{n+1}\}) + C^*/(n+1)\ . \end{aligned}$$

We showed above that $\ell_t^n \geq \ell_t^{n+1}$ for $t \in \mathbb{N}$, and hence that $\Gamma(\{\ell_t^n\}) \geq \Gamma(\{\ell_t^{n+1}\})$ and $\mathrm{AC}^n(\pi^n) \geq \mathrm{AC}^{n+1}(\pi^{n+1})$. We also showed that if either assumption (ii) or (iii) held then $\ell_t^n > L_t^{n+1}$ for $t \in \mathbb{N}$, and hence, by Assumption 10 (part 3), $\Gamma(\{\ell_t^n\}) > \Gamma(\{\ell_t^{n+1}\})$. If instead assumption (i) holds, then $C^* > 0$ and hence $C^*/n > C^*/(n+1)$. In either case, $\mathrm{AC}^n(\pi^n) > \mathrm{AC}^{n+1}(\pi^{n+1})$.

Sampling Problem We let $\pi^{n+1} \in \Pi^{n+1}$ be a sampling procedure such that $\varphi_i^{\pi^{n+1}} = \varphi_i^{\pi}$ for $i \in \{1, \ldots, n\}$ and $\varphi_{n+1}^{\pi^{n+1}} = \varphi_{\text{null}}$. According to Assumption 6, such a procedure π^{n+1} exists and $C(\pi^{n+1}) = C(\pi^n) \equiv C^*$. For $t \in \mathbb{N}$, let $A_t' \equiv ((n+1)/n)A_t^{\pi}$. We showed above that $E\big[\psi^n(X_t^n - A_t^{\pi^n})\big]/n$ is (weakly or strictly) greater than $E\big[\psi^{n+1}(X_t^{n+1} - A_t')\big]/(n+1)$; the latter is, in turn, an upper bound on $E\big[\psi^{n+1}(X_t^{n+1} - A_t^{\pi^{n+1}})\big]/(n+1)$, since A_t' is a function of $H_t^{\pi^{n+1}}$. The rest of the proof is like the one for the computation problem. □

The following assumption ensures that, in Theorem 4A, the diversification effect is present.

Assumption 12 *In the sampling problem, one of the following two conditions holds.*

1. *(a) Ψ is strictly convex. (b) For $i, j \in \mathbb{N}$ such that $i \neq j$ and for $t \in \mathbb{N}$, there are no functions f_i and f_j of H_{t-1} such that $X_{it} - f_i(H_{t-1}) = X_{jt} - f_j(H_{t-1})$ a.e.*
2. *For $i, t \in \mathbb{N}$, if P is a regular conditional probability of X_{it} given $H_t \setminus \{X_{it}\}$, then with strictly positive probability $H_t \setminus \{X_{it}\}$ is such that the conditional probability $P(H_t \setminus \{X_{it}\}, \cdot)\colon \mathcal{B} \to \mathbb{R}$ does not have a support that is bounded above or below.*

PROOF OF THEOREM 4A: **Overview of main step:** The main idea of this proof is that firm kn can achieve lower average costs than firm n by *replicating* the sampling procedure and policy of firm n. Specifically, let $n \in \mathbb{N}$ and $\pi \in \Pi^n$. For $t \in \mathbb{N}$, let ℓ_t be the average period-t expected loss of firm n given π. For $k > 1$, we define a sampling procedure π^k for firm kn, which replicates π, such that $C(\pi^k) = kC(\pi)$. For $t \in \mathbb{N}$, let ℓ_t^k be the average period-t expected loss for firm kn given π^k. We define an upper bound $\bar{\ell}_t^k$ on ℓ_t^k such that $\ell_t > \bar{\ell}_t^k$.

Why this proves the theorem: It follows that $\ell_t > \ell_t^k$ for $t \in \mathbb{N}$ and hence $\Gamma(\{\ell_t\}) > \Gamma(\{\ell_t^k\})$. That is, firm n's average long-run loss given π is greater than firm kn's given π^k. Both firms' average sampling costs are $(1/n)C(\pi)$. Hence, $\mathrm{AC}^n(\pi) > \mathrm{AC}^{kn}(\pi^k)$. By letting π be an optimal sampling procedure for firm n, so that $\mathrm{AC}^n(\pi) = \mathrm{AC}(n)$, we have shown that $\mathrm{AC}(n) > \mathrm{AC}^{kn}(\pi^k) \geq \mathrm{AC}(kn)$.

We can then conclude that firm size is unbounded. Let $\bar{n} \in \mathbb{N}$, let $n' \equiv 1 + \sum_{n=1}^{\bar{n}} n$, and let $n \geq n'$. Any partition of n either has a firm whose size is greater than $\bar{n}$ or has two firms of the same size. In the latter case, these two firms can be combined to reduce average costs and so the partition is not optimal. Hence, the maximum firm size of any optimal partition of n is greater than $\bar{n}$.

large firm into several smaller firms.[5] Thus, diseconomies to centralized decision making may also limit firm size.

We emphasize that by a firm we mean an enterprise—as this term is used in Chandler (1966)—rather than merely a legal entity. For example, in the construction of a large building, many independent contractors work together. During the project, they continue to maintain their independent identities, but they also give up some of their autonomy because of the tight coordination that is required by the project. This paper considers whether there are organizational limits to the scale of such enterprises. As another example,[6] consider two farmers each owning a piece of land and a small tractor. Suppose that they decide to buy a big tractor and cultivate all the land together (not merely share the tractor). Then they have formed an enterprise that did not exist prior to the merger, even if each farmer continues to own his or her own land or to maintain a separate business identity for certain purposes. The farmers would nearly always form a legal partnership after such a merging of operations, but a lack of legal status would not eliminate the economic status of the enterprise. The joint operations involve collective decision making about the cultivation of the land, and the aggregation of information about soil qualities of the two pieces of land and the markets served by the two farmers. Leaving the technological returns aside, such collective decision making may have certain benefits, such as the sharing of information, and certain disadvantages, such as delays in aggregating information. These are precisely the issues we study in this paper.

To capture in a simple and concrete way that decision making is more centralized within a single large firm than among multiple small firms, we identify each decision problem in our model with a single firm. This is consistent with the examples of the decision problem given in Section 1.2, where the decision variable for each problem is a level of output. There is always some centralized control over a firm's total output, but very little coordination of output levels of different firms in the same industry. In these examples, our measure n of scale is proportional to the level of output, which is the usual measure of scale.

Our approach sheds new light on how bounded rationality limits firm size, but it has limitations which could be addressed in future research. First, our identification of a firm with a single centralized decision problem introduces two biases. On the one hand, because there is also decentralized decision making within firms, which is not allowed for by our model, we may underestimate the scale of firms. On the other hand, because there is some coordination among firms—through anonymous market interactions and also through contractual relationships—which is also not captured by our model, we may overestimate the scale of firms.[7]

Second, it would be useful to integrate our complexity-based modeling of organizational decision making with the incentives-based property-rights theory of firms (see Hart (1995)

[5]For example, the subunits could not communicate, coordinate their activities, or allocate resources except as independent firms would do. Even if there continues to exist a common entity that owns the subunits, these subunits would be independent firms, just as the common ownership of the many publicly traded corporations by overlapping sets of stockholders and investments firms does not erase the boundaries between these corporations.

[6]This example is borrowed from comments of an anonymous referee.

[7]While recognizing these limitations, we note that most other models of organizational returns to scale are also based on an ad hoc identification of a firm as some informationally integrated unit. For example, Williamson (1967) defines a firm to be a hierarchy with an exogenous managerial production function. Keren and Levhari (1983) define a firm to be a hierarchy with coordination delay that could be derived from a model of associative computation. Radner (1993, Section 7) defines a firm to be a network for aggregating cohorts of data. Geanakoplos and Milgrom (1991) and Van Zandt (1998f) define a firm to be a group of shops to which resource allocations are coordinated by a hierarchy.

processes $\{X_{it}\}$ and $\{X_{jt}\}$ are mutually independent for $i \neq j$, it follows that $E[X_{it} \,|\, H_t^\pi] = E[X_{it} \,|\, H_{it}^\pi]$. Furthermore,

(1)
$$E\left[(X_t^n - A_t^\pi)^2\right] \;=\; E\left[\left(\sum_{i=1}^n X_{it} - E[X_{it} \,|\, H_{it}^\pi]\right)^2\right] \;=\; \sum_{i=1}^n E\left[\left(X_{it} - E[X_{it} \,|\, H_{it}^\pi]\right)^2\right] .$$

For $i \in \{1, \ldots, n\}$, firm 1 could use a sampling procedure $\pi' \in \Pi^1$ such that $\varphi_1^{\pi'} = \varphi_i^\pi$ (Assumption 6), and would then have the sequence $\{E[(X_{it} - E[X_{it} \,|\, H_{it}^\pi])^2]\}_{t=1}^\infty$ of expected losses. Hence, this sequence belongs to $\mathcal{L}$. Given also the linearity of Γ (Assumption 10) and the additive separability of sampling costs (Assumption 6), the total cost is

$$\mathrm{TC}^n(\pi) \;=\; \sum_{i=1}^n \left(\Gamma\left(\left\{E\left[\left(X_{it} - E[X_{it} \,|\, H_{it}^\pi]\right)^2\right]\right\}_{t=1}^\infty\right) + S(\varphi_i^\pi)\right) .$$

The sampling problem is thus additively separable over the stochastic processes. That is, the problem is to find a single-process information structure $\varphi^* \in \tilde{\varphi}$ that minimizes

(2)
$$\Gamma\left(\left\{E\left[\left(X_{it} - E\left[X_{it} \,\middle|\, \{X_{is} \mid s \in \varphi_t\}\right]\right)^2\right]\right\}_{t=1}^\infty\right) + S(\varphi) ,$$

and then to use a sampling procedure $\pi \in \Pi^n$ such that $\varphi_i^\pi = \varphi^*$ for $i = 1, \ldots, n$. The average cost for any firm is the minimized value of equation (2) and returns to scale are constant.

For the computation problem, we show that the average gain from information processing converges to 0 as $n \to \infty$. For $d \in \mathbb{N}$, let $\lambda_d = E\left[\left(X_{it} - E[X_{it} \,|\, H_{i,t-d}]\right)^2\right]$, which does not depend on i or t because the stochastic processes are exchangeable and stationary. Let $n \in \mathbb{N}$ and $\pi \in \Pi^n$. The right-hand side of equation (1) is a lower bound on the period-t expected loss for the policy $\{A_t^\pi\}$. (This lower bound may not actually be attained, because the decision procedure is not necessarily statistically optimal.) Furthermore, a lower bound on $E\left[(X_{it} - E[X_{it} \,|\, H_{it}^\pi])^2\right]$ is given by $\lambda_{d_{it}^\pi}$, where $d_{it}^\pi = t - \max\{s \mid \langle i, s\rangle \in \Phi_t^\pi\}$ is the minimum lag of the data in H_{it}^π (or $d_{it}^\pi = \infty$ and $E[X_{it} \,|\, H_{i,t-d_{it}}] = E[X_{it}]$ if H_{it}^π is null). Hence, the period-t expected loss for $\{A_t^\pi\}$ is at least $\sum_{i=1}^n \lambda_{d_{it}^\pi}$.

Let $B\colon \mathbb{N} \to \mathbb{N}$ be the bound in Assumption 1, and let $\{d_i\}_{i=1}^\infty$ be the sequence such that $d_i = 1$ for the first $B(1)$ terms, $d_i = 2$ for the next $B(2)$ terms, and so on. This sequence is such that, for $n \in \mathbb{N}$, $\pi \in \Pi^n$, $t \in \mathbb{N}$, and $d \in \mathbb{N}$,

$$\#\left\{i \in \{1, \ldots, n\} \mid d_i \leq d\right\} \;\geq\; \#\left\{i \in \{1, \ldots, n\} \mid d_{it} \leq d\right\} .$$

Hence, because λ_d is decreasing in d, $\sum_{i=1}^n \lambda_{d_{it}^\pi} \geq \sum_{i=1}^n \lambda_{d_i}$. Therefore, $\mathrm{AC}(n) \geq \frac{1}{n}\sum_{i=1}^n \lambda_{d_i}$. Since each stochastic process $\{X_{it}\}$ is regular, $\liminf_{d\to\infty} \lambda_d = \mathrm{Var}(X_{it})$ (Remark 2). Since also $\lim_{i\to\infty} d_i = \infty$, we have $\lim_{i\to\infty} \lambda_{d_i} = \mathrm{Var}(X_{it})$ and $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \lambda_{d_i} = \mathrm{Var}(X_{it})$. Consequently, $\liminf_{n\to\infty} \mathrm{AC}(n) \geq \mathrm{Var}(X_{it})$. Because $\mathrm{Var}(X_{it})$ is the no-information average cost and is an upper bound on $\mathrm{AC}(n)$, $\lim_{n\to\infty} \mathrm{AC}(n) = \mathrm{Var}(X_{it})$.

Suppose also that there is $n \in \mathbb{N}$ such that there is a computation procedure whose average costs are lower than the no-information average cost. Then $\mathrm{AC}(n) < \mathrm{Var}(X_{it})$ and there exists an $\bar{n} \in \mathbb{N}$ such that, for $m \geq \bar{n}$,

$$\lfloor m/n \rfloor n\, \mathrm{AC}(n) + (m \bmod n)\, \mathrm{Var}(X_{it}) \;<\; \mathrm{AC}(m) .$$

The left-hand side of this inequality are the total costs when m processes are partitioned into $\lfloor m/n \rfloor$ firms of size n and $m \bmod n$ firms of size 1. Hence, $\bar{n}$ is a bound on firm size. □

been used to show nondecreasing technological returns to scale, also work in the sampling problem but break down in the computation problem. Furthermore, the proofs link this breakdown to aggregation delay and the informational integration implied by centralized decision making.

Specifically, under the assumptions of Theorem 2, we show that there are constant returns to scale in the sampling problem because a firm should replicate the optimal sampling procedure of a firm of size 1. Under the assumptions of Theorem 4, we show that there are eventually increasing returns to scale in the sampling problem because a firm of size mn can achieve average costs lower than those of a firm of size n by dividing itself into m divisions of equal size that imitate the sampling procedure of the firm of size n. *Such replication strategies do not work in the computation problem because each division would compute only its own forecast.* The aggregation of these forecasts would introduce delay, and so the decision rule would use information that is older than the information used by the smaller firm. Consequently, in the computation problem, there are eventually decreasing returns to scale under the assumptions of Theorem 2 and there may be a firm size that minimizes average costs under the assumptions of Theorem 4.

Empirical research in this area beyond case studies is limited. Brynjolfsson et al. (1994) measure the impact of information technology (IT) on firm size and find that it is linked to smaller firm size. Heuristically, if we claim that firm size is limited in part by managerial delay, then improvements in IT should instead lead to larger firm size (although we do not perform such a comparative statics exercise). However, in a general equilibrium model, improvements in IT also mean that each firm's competitive environment is changing more quickly, and this aggravates the effect of managerial delay. Brynjolfsson and Hitt (1998) find positive correlation between demand for IT and decentralization of decision making within firms. This is a link between hardware and the structure of *human* decision making that our model is not rich enough to capture, but heuristically this might contradict our conclusion that information processing constraints limit centralized decision making. Alternatively, it may mean that firms that operate in rapidly changing environments respond by both decentralizing decision making and improving IT. Further research is needed to resolve these theoretical and empirical issues.

2 The decision problem

We study the real-time computation of a family of forecasting problems that are parameterized by their size or scale n, a strictly positive integer. Our goal is to compare decision problems of different sizes. In the definitions that follow, the exogenous components that vary with n are indexed by n, whereas the endogenous components are not.

Let $\mathbb{Z}$ denote the set of integers and $\mathbb{N}$ the set of strictly positive integers. We fix once and for all a countably infinite set of potential discrete-time stochastic processes, indexed by $i \in \mathbb{N}$, from which the processes that enter into each decision problem are drawn. Process i is denoted by $\{X_{it}\}_{t=-\infty}^{\infty}$ or simply $\{X_{it}\}$. The decision problem of size n involves forecasting the sum $X_t^n \equiv \sum_{i=1}^{n} X_{it}$ of the first n processes at the beginning of each period $t \in \mathbb{N}$, based on their past realizations.[8]

A *forecast* A_t of X_t^n is a random variable measurable with respect to the history $\{X_{1,t-d}, \ldots, X_{n,t-d}\}_{d=1}^{\infty}$. A *policy* is a sequence $\{A_t\}_{t=1}^{\infty}$ of forecasts (also denoted $\{A_t\}$).

[8]Even though the forecasting begins in period 1, we assume a double infinity of time periods for the processes in order to simplify the statement of certain statistical assumptions.

completion of computation. The authors study returns to scale by positing an exogenously given cost function—a function that depends on the scale of the problem, the computational costs, and the delay. Reiter (1996) is also a batch processing model that examines limits to firm size and centralization, but under the postulate that there are bounds on the size of the informational inputs of any organizational unit.

Real-time control is a different, and in some ways richer, methodology for studying the effects of delay on decision making. First, because it is based on a temporal decision problem, we can implicitly derive a "cost of delay" from the degradation of the quality of decisions that are based on old information. Furthermore, because decision rules are endogenous, we do not artificially limit centralization by forcing organizations with large-scale decision problems to bog themselves down with computation and only use old data. Compare this with a benchmark model obtained by embedding a batch processing model into our decision problem. Following Keren and Levhari (1983) and Radner (1993, Section 7), in which all data are collected for a decision at the same point in time and the amount of data is equal to the scale of the firm, we would consider only computation procedures in which the firm calculates the period-t decision from $\{X_{i,t-d}\}_{i=1}^{n}$ for some delay d. The computation constraints require that $n \leq B(d)$ and hence $d \to \infty$ as $n \to \infty$. Consider the assumptions of Theorem 3, with a negligible idiosyncratic component. The problem is then to forecast Y_t from $\{X_{i,t-d}\}_{i=1}^{n}$. Assuming that $\{Y_t\}$ is regular, the average expected loss in the benchmark model is approximately equal in the limit (as $n \to \infty$ and $d \to \infty$) to the average expected loss when there is no information processing. One can thus construct specific examples (see Van Zandt and Radner (1998)) in which firm size is bounded in the benchmark model, whereas Theorem 3 shows that returns to scale are monotonically increasing in our model.

The model by Geanakoplos and Milgrom (1991) is a team-theory model of resource allocation in which an endogenous administrative apparatus hierarchically disaggregates resource allocations. Their model has the advantage of allowing for internal decentralization of decision making, with coordination among the decision-making nodes. Theirs is a static approach that does not explicitly model the hierarchical aggregation of information; rather, there are constraints on information acquisition for individual agents that represent information processing constraints. Hence, their results on returns to scale depend on assumptions about what aggregate information is available exogenously. The assumption under which they conclude that returns to scale are decreasing—that no aggregate information is available—is extreme. However, the notion that aggregate information is less available or of poorer quality than disaggregate information is supported by our model; computational delay means that aggregate information cannot be as recent as disaggregate information. Van Zandt (1998e, 1998f) studies a temporal version of their decision problem, but with real-time information processing.

The work of Orbay (1996) and Meagher (1996) is also related, but with interesting differences. They consider a problem of forecasting a fixed stochastic process without variations in the scale of the decision problem or operations of the firm. However, the amount of data sampled about the process for calculating each decision is endogenous. Because of computational delay, the trade-off is between basing each forecast on a large amount of old information or on a small amount of recent information. The size of the administrative apparatus is roughly proportional to the amount of data incorporated into each decision, so this exercise considers the optimal size of the administrative apparatus for a firm whose scale of production is fixed. They find, for example, that the administrative apparatus tends to be smaller the more quickly the environment is changing.

data. $C(\pi)$ is the long-run costs, including managerial wages, of this information processing. Φ_t^π denotes the indices of the data used to calculate A_t^π.

We model the computation constraints axiomatically rather than constructively. The assumptions allow for decentralized computation, that is, computation performed jointly by many managers or clerks whose numbers and activities are determined endogenously. This property, which is analogous to parallel or distributed processing by networks of machines, cannot be suppressed when studying returns to scale, because managerial resources must be allowed to vary with the scale of the firm. The assumptions stated are satisfied by the computation model in Radner and Van Zandt (1992) and Van Zandt (1998c), which study this same decision problem, and most other distributed processing models, including those that have been used in economics, such as Mount and Reiter (1990) and Reiter (1996).[9]

The fundamental constraint we want to capture is that information processing—which includes the reading and preparation of reports and aggregation of non-numerical information—takes time. To motivate this, the numerical data in our decision problem should be viewed as a proxy for the complex data used by human administrators in actual organizations, or the reader should imagine that the data is not available in a simple numerical format and instead is difficult to understand and substantiate and must be communicated through lengthy reports. We emphasize that our use of a numerical decision problem as a proxy for more realistic human decision problems is standard in economics and derives from the need to impose statistical assumptions, rather than from our need to impose computation constraints. Van Zandt (1998c) explains that the information processing constraints we impose are qualitatively similar to the ones we would impose for more realistic problems.

This time constraint has two effects. First, it adds an administrative cost (reflected in $C(\pi)$) to the calculation of any policy owing to the time managers spend processing information. Second, it restricts the set of feasible policies; in particular, it limits the amount of recent data that can be incorporated into decisions. This second effect is the more important one for this paper, and is captured by the following "iron law of delay".

Assumption 1 *For each lag $d \in \mathbb{N}$, there is a uniform bound on the amount of data whose lag is d or less on which any forecast can depend. Formally, there is a function $B\colon \mathbb{N} \to \mathbb{N}$ such that $\#\{\langle i,s\rangle \in \Phi_t^\pi \mid s \geq t-d\} \leq B(d)$ for $d \in \mathbb{N}$, $\pi \in \Pi$, and $t \in \mathbb{N}$.*

This bound comes from the delay in aggregating information. For example, suppose that policies are computed by having agents perform elementary operations that can have at most k inputs (which can be any previous results, raw data, or constants) and that produce an arbitrary number of outputs. Suppose each operation takes at least δ units of time. Either A_t is a constant or the value of a raw datum, or it is the output of an elementary operation that was begun by time $t-\delta$. Thus, A_t can depend on at most 1 datum that is first available after $t-\delta$. If A_t is the result of an elementary operation begun by time $t-\delta$, then each of the $\leq k$ inputs is either a constant or a raw datum, or is itself the output of an operation begun by time $t-2\delta$. Hence, A_t can depend on at most k data first available after $t-2\delta$. Repeating this argument inductively, A_t can depend on at most k^2 data first available after $t-3\delta$, and on at most $k^{\nu-1}$ data first available after $t-\nu\delta$ for $\nu \in \mathbb{N}$. This implies that, for $d \in \mathbb{N}$, A_t can depend on at most $k^{\lfloor d/\delta \rfloor}$ observations from period $t-d$ or later. The bound would also hold if the delay comes from reading and interpreting raw data and messages; see Van Zandt (1998c) for a discussion.

[9]Kenneth Mount and Stanley Reiter have advocated decentralized information processing as a model of human organizations since 1982. See Van Zandt (1998a, 1998b) for surveys of the use of such models in the economic theory of organizations.

Interestingly, delay *does not* lead to decreasing returns in the computation problem because the amount of data used in the computation is endogenous and, in particular, does not have to increase with the size of the firm. In Section 6.2, we contrast this with the eventually decreasing returns that may obtain in a benchmark batch processing model.

Radner and Van Zandt (1992) characterize the returns to scale for a specific computation model under assumptions (piecewise linear loss and processes that are noisy versions of a common AR(1) process) that are consistent with those of Theorem 3.

5.4 Scalable loss and general processes

The idea behind Theorem 3 is that a larger firm can achieve a lower average loss than a small firm by imitating the decision procedure of a *single* small firm. This is *not* an analog of the principle that leads to nondecreasing technological returns to scale: A large firm can imitate the production processes of *several* small firms whose total size is the size of the large firm. However, in the sampling problem with scalable loss, the analog of this principle—a large firm imitates the sampling procedures of several small firms—does lead to eventually decreasing returns to scale under general statistical assumptions. This is the first part of Theorem 4.

Theorem 4A *Assume that the loss function is scalable and Assumption 12 (stated in the Appendix) holds. In the sampling problem, firm size is unbounded and* $\mathrm{AC}(kn) < \mathrm{AC}(n)$ *for* $n, k \in \mathbb{N}$ *such that* $k > 1$.

There is no such analog for the computation problem. If a large firm imitates the policies of several small firms, it ends up with several forecasts each period. If it attempts to aggregate these forecasts, there is additional delay and so the policy uses information that is older than the information used by the small firms. This does imply that returns to scale are never increasing in the computation problem, as was shown in Theorem 3. However, the second part of Theorem 4 presents a robust example in which firm size is bounded in the computation problem. This result shows how aggregation delay in a centralized decision problem may subvert the Arrow effect.

Theorem 4B *Assume that the loss function is scalable and Assumption 13 (stated in the Appendix) holds. In the computation problem,* $\mathrm{AC}(1) < \mathrm{AC}(n)$ *for* $n \geq 2$ *so* 1 *is a bound on firm size.*

Assumptions 12 and 13 in Theorems 4A and 4B, respectively, are stated in the Appendix because they are rather technical. Assumption 12 is a weak statistical assumption that plays the following role. We obtain the inequality $\mathrm{AC}(kn) \leq \mathrm{AC}(n)$ in the sampling problem by showing that if the firm of size kn replicates the sampling procedure of a firm of size n, then the average sampling cost of the large firm and the small firm are the same, and the average expected loss of the large firm is as low as that of the small firm. To obtain the *strict* inequality $\mathrm{AC}(kn) < \mathrm{AC}(n)$, we appeal to the diversification effect, but this requires, for example, that the processes not be perfectly correlated. Assumption 12 rules out this and similar trivial cases.

For the computation problem, Assumption 13 specifies a detailed but robust example. It assumes, for example, that the processes can be decomposed as $X_{it} = Y_t + Z_{it}$, and that each of the components is a first-order autoregressive processes. When the statistical conditions in Assumption 13 are satisfied, so is Assumption 12; hence, the contrast between

processes. To state this formally, we identify for each $\pi \in \Pi$ and $i \in \mathbb{N}$ the dates of the information about process i provided by π by letting $\varphi_{it}^{\pi} \equiv \{s \in \mathbb{Z} \mid \langle i, s \rangle \in \Phi_t^{\pi}\}$ for $t \in \mathbb{N}$ and $\varphi_i^{\pi} \equiv \{\varphi_{it}^{\pi}\}_{t=1}^{\infty}$. We refer to φ_i^{π} as a *single-process information structure*.[10] (If a process is not sampled at all, then its information structure is $\varphi_{\text{null}} \equiv \{\varnothing\}_{t=1}^{\infty}$.) Our assumption is then that (a) there is a set $\tilde{\varphi}$ of single-process information structures with associated costs, (b) a sampling procedure specifies a single-process information structure in $\tilde{\varphi}$ for each process, and (c) sampling costs are summed over the processes.

Assumption 6 *There is a set $\tilde{\varphi}$ of single-process information structures such that, for $n \in \mathbb{N}$, $\{\varphi_1, \ldots, \varphi_n\} \subset \tilde{\varphi}$ if and only if there is $\pi \in \Pi^n$ such that $\varphi_i^{\pi} = \varphi_i$ for $i \in \{1, \ldots, n\}$. Furthermore, there is $S\colon \tilde{\varphi} \to \mathbb{R}$ such that, for $n \in \mathbb{N}$ and $\pi \in \Pi^n$, $C(\pi) = \sum_{i=1}^{n} S(\varphi_i^{\pi})$. Also, $\varphi_{\text{null}} \in \tilde{\varphi}$ and $S(\varphi_{\text{null}}) = 0$.*

3.4 Comparison of the computation and sampling problems

The policies in the computation problem do not minimize the expected loss conditional on all available information, since they do not even depend on all available information. A weaker notion of statistical optimality of a computation procedure $\pi \in \Pi^n$ is that A_t^{π} minimizes $E[\psi^n(X_t^n - a) \mid H_t^{\pi}]$ almost surely. As discussed in Van Zandt (1998c), a constrained-optimal computation procedure (one that minimizes total costs on Π) need not be statistically optimal in the computation problem because it may be more costly (or impossible) to compute the statistically-optimal decision rule that uses the same information as π. This is one potential difference between the sampling problem and the computation problem.

However, this difference is not relevant to our results. In fact, we never preclude statistical optimality in the computation problem. Instead, the important difference is how much data of a given lag can be used in a forecast. Suppose that, in the sampling problem, the forecast in period t of a firm of size 1 is based on $X_{1,t-2}$. Then, for a firm of size n, it is possible to sample $X_{i,t-2}$ for all $i \in \{1, \ldots, n\}$ with the same average sampling cost faced by the firm of size 1, so that the forecast uses the data surrounded by the solid line in Figure 1. This is not possible in the computation problem because of aggregation delay (Assumption 1). For example, Figure 1 shows the bound on the data of any given lag for the case where $B(d) = 2^{d-1}$. Thus, in the computation problem, aggregation delay creates a negative informational externality among the processes—data of a given lag about one process crowds out data of that lag about other processes.

4 Returns to scale: Assumptions and definitions

4.1 Statistical assumptions

For $t \in \mathbb{Z}$, the vector $\{X_{1t}, X_{2t}, \ldots\}$ is denoted by $\mathbf{X}_t$; then $\{\mathbf{X}_t\}_{t=-\infty}^{\infty}$ or simply $\{\mathbf{X}_t\}$ denotes the vector process. For $t \in \mathbb{Z}$, H_t denotes the history of $\{\mathbf{X}_t\}$ up through period t.

We assume that the processes have finite variance and are stationary and exchangeable.

Assumption 7 *For all $t \in \mathbb{N}$ and $i \in \mathbb{N}$, $0 < \text{Var}(X_{it}) < \infty$.*

[10] The structure φ_{it}^{π} is an element of $2^{\{\ldots, t-2, t-1\}}$, and so formally we define a single-process information structure to be any element of $\prod_{t=1}^{\infty} 2^{\{\ldots, t-2, t-1\}}$.

	Statistical Assumptions	Returns to Scale: Sampling Problem	Returns to Scale: Computation Problem	
Quadratic Loss	mutually dependent	bounded firm size ($\lim_{n\to\infty} \mathrm{AC}(n) = \infty$)	bounded firm size ($\lim_{n\to\infty} \mathrm{AC}(n) = \infty$)	Thm 1
Quadratic Loss	mutually independent	constant (constant per-unit gain)	bounded firm size (per-unit gain $\to 0$)	Thm 2
Scalable Loss	common process plus noise	monotonically increasing	monotonically increasing	Thm 3
Scalable Loss	general	unbounded firm size (replication works)	example with bounded firm size	Thm 4

TABLE 1. Table of results.

independence of the stochastic processes. The quadratic loss function is not favorable to increasing returns because if the average error is constant then the average loss increases linearly with n. Theorem 1 shows that this leads to decreasing returns to scale in both the computation and sampling problems if (heuristically) there is a common component that cannot be perfectly forecasted from past data.

Theorem 1 *Assume the loss is quadratic and that* $E[\mathrm{Cov}(X_{it}, X_{jt} \mid H_{t-1})] > 0$ *for* $i, j \in \mathbb{N}$ *such that* $i \neq j$.[12] *In both the sampling and the computation problems,* $\lim_{n\to\infty} \mathrm{AC}(n) = \infty$ *and firm size is bounded.*

5.2 Quadratic loss and mutually uncorrelated processes

When the loss function is quadratic but the processes are mutually *independent*, a diversification effect counterbalances the curvature of the loss function. This leads to constant returns to scale in the sampling problem. As shown in the proof of Theorem 2, the selection of a sampling procedure is separable over the processes and any firm should replicate an optimal procedure of a firm of size 1.

In the computation problem, such replication is impossible because the firm would compute n forecasts, which must then be aggregated, thereby incurring additional delay. In fact, the aggregation delay implies that the data about "most" processes is "old" in large firms. In Theorem 2, we assume that information becomes useless as it gets older. (Specifically, we assume $\{\mathbf{X}_t\}$ is regular; see Remark 2 immediately after Theorem 2.) Hence, as firm size grows, the average cost converges to the *no-information average cost.* This is defined to be the average cost of the decision procedure that (a) has no administrative cost, (b) makes the same forecast each period, and (c) has an expected loss each period of $\min_{a\in\mathbb{R}} E[\psi^n(X_t^n - a)]$. Such a procedure corresponds to no computation or no sampling.

[12]Recall that $E[\mathrm{Cov}(X_{it}, X_{jt} \mid H_{t-1})] = E[(X_{it} - E[X_{it} \mid H_{t-1}])(X_{jt} - E[X_{jt} \mid H_{t-1}])]$. If the decomposition in Remark 1 holds and if $\{\mathbf{X}_t\}$ or simply $\{Y_t\}$ is regular (see Remark 2), then $E[\mathrm{Cov}(X_{it}, X_{jt} \mid H_{t-1})] > 0$ if and only if the processes are mutually dependent. We conjecture but have not verified that this holds without the decomposition.

where Ψ is a convex function not depending on n such that $\Psi(0) = 0$, $\Psi(\epsilon) > 0$ if $\epsilon \neq 0$, and $E[\Psi(X_{it} - E[X_{it}])] < \infty$. We refer to this as a *scalable* loss function.

A leading example of the scalable loss is when ψ is piecewise linear and does not depend on n:

$$\psi^n(X_t^n - A_t) \;=\; \begin{cases} \gamma_0 \, |X_t^n - A_t| & \text{if } X_t^n - A_t < 0 \,, \\ \gamma_1 \, |X_t^n - A_t| & \text{if } X_t^n - A_t \geq 0 \,. \end{cases}$$

For example, this is the loss when a firm has to make up for excess demand (resp., supply) by buying (resp., selling) output at a price that exceeds (resp., is less than) the firm's unit production cost. The scalable loss also includes the case where a quadratic loss is adjusted for firm size, $\psi^n(X_t^n - A_t) = \frac{1}{n}(X_t^n - A_t)^2$, in which case $\Psi(\epsilon) = \epsilon^2$.

4.3 Long-run loss

As explained in Section 3.1, the function Γ aggregates period-by-period expected losses into a measure of long-run loss. We denote the domain of Γ by $\mathcal{L}$, which must contain the sequence of expected losses for any policy that is generated by a decision procedure (such a policy is said to be *allowable*). Our next assumption restricts the domain $\mathcal{L}$ and assumes that Γ is linear and strictly monotone.

Assumption 10 *If A_t is an allowable policy for a decision problem of size n and $L_t = E[\psi^n(X_t^n - A_t)]$ for $t \in \mathbb{N}$, then $\{L_t\} \in \mathcal{L}$. Furthermore:*

1. *$\mathcal{L}$ is the positive cone of a linear subspace of $\mathbb{R}^{\mathbb{N}}$ containing the constant sequences;*
2. *Γ is a linear functional;*[11]
3. *if $\{L_t\}$ and $\{L'_t\}$ belong to $\mathcal{L}$ and $L_t < L'_t$ for $t \in \mathbb{N}$, then $\Gamma(\{L_t\}) < \Gamma(\{L'_t\})$.*

As a normalization, we also assume that if $\{L_t\}$ is constant then $\Gamma(\{L_t\})$ is equal to the constant value of $\{L_t\}$.

The purpose of the linearity assumption is to make comparisons across problems of different size meaningful (e.g., if the expected loss in each period scales linearly with problem size, then so does the long-run loss). This assumption holds if $\mathcal{L}$ is the set of bounded sequences in $\mathbb{R}^{\mathbb{N}}$ and $\Gamma(\cdot)$ is the discounted present value with respect to a summable sequence of discount factors. It is also consistent with the case where $\Gamma(\{L_t\})$ is the long-run average value of $\{L_t\}$, in which case there is an implicit restriction on the set of decision procedures. Specifically, the set $\mathcal{L}$ must then contain only sequences whose long-run averages are well-defined and $\mathcal{L}$ cannot contain two sequences $\{L_t\}$ and $\{L'_t\}$ such that $L_t < L'_t$ for $t \in \mathbb{N}$ and such that both have the same long-run average (true if $\mathcal{L}$ contains only constant or cyclic sequences). See Van Zandt and Radner (1998) for further discussion and a sketch of how to weaken the monotonicity condition.

4.4 Definitions of returns to scale

For both the computation and sampling problems, we assume that there is a cost-minimizing decision procedure for all n.

[11] That is, Γ can be extended to a linear functional on the subspace spanned by $\mathcal{L}$.