A Theory of Cutoff Formation Under Imperfect Information

Fred M. Feinberg; Joel Huber

Management Science, Vol. 42, No. 1. (Jan., 1996), pp. 65-84.

Stable URL:
http://links.jstor.org/sici?sici=0025-1909%28199601%2942%3A1%3C65%3AATOCFU%3E2.0.CO%3B2-O

Management Science is currently published by INFORMS.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/informs.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org
Tue Feb 13 19:00:56 2007

A Theory of Cutoff Formation Under Imperfect Information

Fred M. Feinberg • Joel Huber
Division of Management and Economics, Scarborough College, University of Toronto, Toronto, Ontario, Canada M1C 1A4
Fuqua School of Business, Duke University, Durham, North Carolina 27708-0120

Numerous models in the Management Science literature contain constructions that are a variant of the following: A decision-maker must choose from a set of alternatives based on imperfect information as to their relative quality, while further evaluation, though costly, provides more accurate information. We examine decision heuristics in which the optimal search policy entails a screening strategy limiting the number of alternatives in the subsequent, costly evaluation. There are two general methods for accomplishing this screening: Quota cutoffs operate by selecting the optimal *number* of alternatives to evaluate; Level cutoffs operate by specifying a minimally-acceptable *level* of the imperfect screening indicator. The present paper has three main objectives. First, to define the Level and Quota cutoff methods, broadly characterize optimal behavior for each and determine what aspects of the decision environment predispose one to be superior to the other; second, to introduce the concomitants of order statistics as a methodology for exploring decision problems when information is imperfectly known; and third, to discuss the pivotal role of default, or fallback, options in a broad class of search problems.

Quota and Level strategies restrict the number of alternatives passing the cutoff-based screen. Because restrictive cutoffs reduce evaluation costs while lowering the expected quality of the item finally selected, changes in the decision environment making the evaluation process less beneficial or increasing its cost drive the optimal cutoff to be more restrictive. In particular, increases in unit evaluation cost, improvement in the quality of a fallback option, decreases in the total number of alternatives available or improvement in the precision of the final evaluation process all lead to more restrictive cutoffs at optimum. These results hold over a remarkably broad range of assumptions and conditions. We also find that a better screening indicator leads to more restrictive screening when evaluation costs are low but, surprisingly, to less restrictive screening when costs are high. Comparing the two strategies, we find the unexpected result that the Quota cutoff strategy is generally superior to the Level, except under one of two fairly uncommon sets of circumstances: when evaluation cost is prohibitively high, or when there is a fallback option of very high quality.
(*Choice Models; Optimal Stopping; Cutoffs; Imperfect Information; Order Statistics; Random Utility Models*)

1. Introduction

Substantial empirical evidence suggests that, when faced with a variety of options and imperfect information about their relative quality, decision-makers employ some type of screening strategy to reduce the number of alternatives requiring subsequent evaluation. Because such an environment—multiple options, imperfect information, non-trivial evaluation costs—typifies a wide variety of real-world decisions, it's not surprising that researchers working in different fields

0025-1909/96/4201/0065$01.25
Copyright © 1996, Institute for Operations Research and the Management Sciences

have posited theoretical bases for its analysis. For instance, Stigler (1961), in a seminal paper, introduced the core concept of search cost and argued that thorough evaluation of all available options is generally suboptimal, while Newell and Simon (1972) argue that rational decision makers may limit search costs by restricting the range of alternatives considered.

The cost of evaluating a number of alternatives, each defined on a number of attributes, can be reduced by a variety of different processes. One can reduce the number of attributes considered (Klein and Bither 1987), the number of alternatives considered, or simplify the analysis for each alternative; further, one can envision combinations of such strategies (Wright 1975). Johnson and Payne (1985) and Payne, Bettman and Johnson (1988) examine the costs and benefits of a number of decision heuristics and show that the appropriate selection strategy typically depends on characteristics of the task and the distribution of alternatives. The present paper focuses on one particular heuristic, the use of a cutoff to limit subsequent evaluations, and asks how this heuristic can be optimally applied.

We examine the tradeoff between costs and benefits in which a choice is made from an optimal subset of alternatives that have been screened by an imperfect indicator. Cost enters the system through the effort to evaluate each alternative that has passed the screen; benefit is measured by the quality of the best of the alternatives found. Clearly, highest quality occurs when all alternatives are evaluated, while minimum effort follows from a very restrictive screen. The question that we address is how to set a cutoff between these two extreme anchors, optimally trading-off evaluation cost against decision quality, and how this optimal cutoff changes with the decision environment.

We concern ourselves here with two ways in which such cutoff-based screening may operate, methods we term "Quota" and "Level" cutoffs. In both cases, an indicator is available to supply ratings of all available alternatives in terms of an imperfect measure of quality. A Quota cutoff specifies the number of alternatives evaluated, while a Level cutoff specifies a minimum score on the indicator that justifies evaluation. A Quota cutoff assures that a certain number will be evaluated, but leaves uncertain the quality of the alternatives within that set. By contrast, a Level cutoff assures a certain level of quality, relative to the screening indicator, but leaves variable the number of alternatives that pass the screen.

To motivate and clarify, a simple example illustrates the parameters of the problem and the nature of the conclusions that can be drawn from an analysis of the Quota and Level problems.

> Suppose that an academic department is trying to hire a doctoral student for an entry-level faculty position based on resumes and recommendations. Faculty score each candidate on a 1–100 scale, and those with the best average scores are brought to campus for a costly two-day visit to determine the best candidate for the position. This rating system has been used with good results for the past several years. The problem facing the faculty lies in determining how many candidates should be brought to campus. One professor reminds his colleagues that they have always invited four candidates to campus and suggests that this same guideline be adopted. Another agrees in principle, but argues that there is a larger pool of graduates this year, and suggests that anyone with an average score of 90 or above be brought in, further pointing out that five candidates meet this criterion this year, but only three would have the previous year. They must make their decisions quickly, as scheduling is difficult, and there will be no opportunity to bring in more candidates if those interviewed are not satisfactory. There is general agreement, however, that it might be acceptable to let the position go unfilled for the time being, as a visiting faculty member on temporary contract, whose performance is considered about average relative to the present pool of candidates, has agreed to stay on for an additional year if needed.

This scenario illustrates key parameters of our framework. Although bringing all available candidates to campus will ensure that the best is identified and inviting none of them will minimize interviewing costs, it is clear that the department must decide on some intermediate policy, trading off its time and monetary resources against the ultimate quality of the candidate chosen. Toward this end, the average rating of each candidate provides an imperfect indicator of quality which can be used to limit the number of costly campus interviews. The professors differ, however, in their opinions on the best way to use this indicator. The first professor feels that specifying a number of candidates, say four, is a good method, even if it's possible that the fifth, excluded candidate might as good as one brought in last year. The second professor argues for some minimally-acceptable level of the ratings indicator (the 1–100 scale) to determine who should be brought in, even though in some years no one, and in other years a very large number, may conceivably pass this screen. This choice of method is one between, respectively, a Quota and a Level cutoff and, although both seem plausible, it is in no way obvious which of the two methods will produce better results in the long-run.

In resolving questions about the relative value of Quota versus Level strategies, our analysis can address a number of related questions. In particular, what happens to the optimal Level or Quota cutoffs if the number of potential candidates drops in a given year? What happens as the screening indicator becomes more precise, perhaps by supplementing the information with telephone calls to the candidates' advisors? Finally, what is the appropriate response to an improvement in the fallback alternative; that is, what happens if the professor on temporary contract is very good, or alternatively very poor?

Before answering these questions, it is useful to frame this research in terms of related work. We make a number of explicit assumptions about the choice process. First we assume irrevocability, that one cutoff defines the set from which the choice must be made. Second, we posit discriminability, the existence of an indicator that permits efficient screening alternatives. Finally, we assume that the act of setting the screen is costless or, at the very least, inexpensive relative to the cost of examining those alternatives that pass it. The discussion of these assumptions serves to define the contexts in which the model applies and to link it with other research.

Irrevocability implies that the choice must be made from the set defined by the cutoff rule. In the academic example, the department must live with the candidates that pass the screen, and cannot repeat the process if dissatisfied with the results. If time is available, an obvious alternative framework involves a sequential strategy in which one observes the quality of the best item according to the indicator and then decides whether to continue. That strategy has the benefit of being able to stop when a superior alternative is discovered, or being able to continue searching if those evaluated to date are unacceptable. There is a vast literature aimed at determining the optimal stopping point as one sequentially evaluates the alternatives; see, for example, Lippman and McCall (1976). Although it may intuitively appear that a sequential strategy will always dominate a fixed choice rule, Morgan and Manning (1985) show that, when faced with time or processing constraints, a non-sequential search strategy can outperform a sequential one.

Irrevocability also poses costs should nothing pass the screen. Suppose, in the faculty recruitment example, that none of the candidates passes the 90% criterion, or none that do pass are acceptable to the faculty. This possibility is accounted for through the value of a fallback option (the part-time faculty member). Largely for technical reasons, we will initially assume that this fallback option has the same value, what we term the *default value*, as choosing at random from the initial pool of candidates. Subsequently, we generalize the model to include both different levels and types of fallback options, and show that the quality of the fallback option has a decisive influence on the profitability of various cutoff strategies.

A second defining aspect of our choice process is discriminability, the availability of an index that can be used to screen alternatives. Clearly, many choices are made without such an indicator. Indeed, a limiting case of our model is that in which the screen is useless, with no correlation between indicator ratings and actual quality. Cutoffs are still valued, but solely through their ability to limit the number of alternatives processed; this, in effect, is Stigler's (1961) original problem.

The final defining aspect of our framework is the assumption that screening is relatively cheap compared with the costs incurred evaluating options that pass the screen. An alternative approach involves costs being assessed sequentially as alternatives are evaluated in the screening process; see, for example, McQueen (1964), and Lippman and McCardle (1991). Grether and Wilde (1984) formulate a model in which screening costs accrue for each alternative evaluated until one is found that satisfies a predetermined cutoff. Their framework is similar to that of a number of problems in which the quality of items passing the screen is traded off against the cost of the screening process itself (e.g., Moskowitz, Plante and Tsia 1993). A screen that is more difficult to pass results in longer search time, higher search costs and a higher quality outcome. By contrast, in our formulation, in which evaluation is not done sequentially, a restrictive screen results in low evaluation costs and relatively low quality, because few alternatives are able to pass into the costly but accurate choice stage. Thus, it is important to differentiate our approach to screening from what is often addressed in the economics of information literature. Critical to our formulation is the relatively low cost of the screening stage compared with the evaluation stage. Such low cost screening opportunities abound. Location often provides an easily-applicable screen when one is considering which homes to inspect, restaurants to visit or different vacation options; explicit tastes ("no local colleges," "only domestic autos") and recommendations from consumer magazines ("best buy," "not acceptable") can offer a reasonable first cut within a large number of consumer product classes.

To summarize, the formal model described below applies when a fallible index is used to limit the number of alternatives that are subsequently subjected to a costly evaluation procedure. This process is neither reversible nor iterative and the parameters are known and stable. Although it may seem that these conditions are sufficiently rare to question the usefulness of the model, consider its use in these seemingly disparate situations:

1. Selection of companies to consider acquiring based on an initial examination of their publicly available information.
2. Qualification of vendors to bid on jobs. Note that in this case the examination cost is not just that of evaluating bids from vendors but also the cost of having to learn about multiple sources.
3. The number and/or minimum qualifications of candidates for Deanships to be submitted to a Provost. Note here that model is appropriate if the Provost's time is the critical resource, while that of the search committee's is less so or is independent of the number of candidates put forward.
4. How many contenders to send to a competition in which the team's score depends on the performance of its best competitor.

2. A Formal Description of the Quota and Level Problems

Suppose the decision-maker has a choice from among n alternatives, and it is possible to use X, a fallible indicator of utility, Y, to limit the number passing to a second stage, where alternatives are evaluated at significant cost. That is, in the first stage the decision-maker limits choice on the basis of indicator ratings, $\{X_i\}$, which imperfectly correspond with unobserved utilities $\{Y_i\}$. Upon inspection, Y_i is assumed to be observed perfectly. The two cutoff methods can be phrased simply as:

Quota cutoff: Choose some number of alternatives to evaluate, n_c.

Level cutoff: Choose some minimally acceptable ratings level, x_c.

The related choice processes can be stated formally as follows:

1. There are a total of n options available. Each alternative is rated, and these ratings (X) correlate imperfectly with utility (Y), with known correlation ρ.
2. One of two methods can be used to eliminate alternatives from costly evaluation. A Quota can be set, and the n_c best-rated alternatives pass to the next stage to be evaluated. Alternatively, a minimally-acceptable ratings Level can be specified, and all alternatives rating at least as well as x_c are evaluated.
3. For each alternative evaluated, a fixed cost, c, will be incurred. For the Quota method, the total evaluation cost will be equal to the Quota selected multiplied by c; for the Level method, the total cost will depend on how many alternatives pass the Level selected, which may vary from one choice occasion to another. If no alternatives pass (a Quota of 0 or a rather high Level), no cost will be incurred.
4. The utilities will be revealed for each alternative passing the screen, and the greatest of these will be awarded to the decision-maker. Should no alternative pass the screen, utility is taken to be 0.

There are several aspects of such a decision process that merit further elaboration. First, both the cost incurred and the payoff in utility associated with evaluating no alternatives (a Quota of $n_c = 0$ or a Level high enough so that no alternatives pass) is initially assumed to be zero. In effect, then, the payoff associated with zero alternatives is identical to the expected return from choosing at random from the original set; it is demonstrated in §5 that this situation can, through the incorporation of *default values*, be directly accounted for within the present framework. Second, the theoretical correlation between the ratings (X) and utility (Y), parameterized by ρ, is assumed to be a known constant. For the class of decision problems for which sequential strategies are impractical or a priori suboptimal, it is reasonable to assume a constant, stable value of ρ. Third, we assume that the evaluation stage is performed without error, that is, the utilities revealed through evaluation are the "real" utilities.[1]

Throughout the paper, the correspondence between ratings X and utility Y is taken to be bivariate normal; this allows the degree of correspondence between the cutoff indicator and utility to be represented by a single parameter, ρ. The use of bivariate normality, while useful, is not essential to the development, which partially parallels those of Stigler (1961) and of Hauser and Wernerfelt (1990). As we will show, there is a natural vantagepoint from which the mathematics of each of the two cutoff methods follows most simply, though some generality is lost, involving rescaling X and Y to have zero mean and unit variance; such an $N(0, 0, 1, 1, \rho)$ distribution has been used to model similar choice scenarios in previous work by Li and Owen (1979), Owen, Li and Chou (1981), Madsen (1982) and by Yeo and David (1984). In this case we define evaluation costs, or simply "cost," as the number of standard deviations of outcome utility the decision maker is willing to give up to avoid evaluating another alternative; standardizing outcome utility as $N[0, 1]$, then, allows evaluation costs to be unambiguously expressed as a real, dimensionless quantity. The resulting loss of generality, involving the values attributed to first- and second-stage default options, is restored in §5.

Objective Function for the Quota Problem: Given n alternatives with ratings values $\{X_i\}$, following the conventions of Yang (1977) and David (1981), we arrange the X-variates in ascending order

$$X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}.$$

This induces an ordering of the paired Y-variates (utilities) called *concomitants* or *induced order statistics,*

$$Y_{[1:n]}, Y_{[2:n]}, \ldots, Y_{[n:n]}.$$

The map from ratings to utility is given by:

$$Y_{[i:n]} = \rho X_{i:n} + \bar{\rho}\epsilon_{[i]}, \tag{1}$$

where $\rho^2 + \bar{\rho}^2 = 1$, and the $\epsilon_{[i]}$ are independent and identically distributed; in the case we consider at length, (X, Y) bivariate normal, the $\epsilon_{[i]}$ are i.i.d. $N[0, 1]$. Detailed discussions of the concomitants and their properties can be found in David (1981), Yang (1977) and Bhattacharya (1984). The appropriate normative decision rule for the Quota cutoff is to increase the number of alternatives selected until the expected value of evaluating one additional alternative no longer compensates for the additional evaluation cost, c (see Lippman and McCall, 1976). This decision rule assures that the optimal Quota cutoff, n_c^*, maximizes the following *expected net gain* function:

$$E[\max\{Y_{[n:n]}, Y_{[n-1:n]}, \ldots, Y_{[n-n_c+1:n]}\}] - cn_c. \tag{2}$$

Calculating the expectation above is a non-trivial exercise, and has been addressed only recently, in the work of Nagaraja and David (1994); the optimal Quota cutoffs arising from an analysis of Eq. (2) are developed more fully in §3.

Objective Function for the Level Problem: The optimal Level cutoff, x_c^*, must maximize an expected net gain function similar to (2). Alternatives surpassing the screen all have a rating higher than x_c, so utility and cost are assessed conditionally on this fact. There will be a set of indices $\{i\}$ for which the ratings $\{X_i\}$ surpass the minimal acceptable level, x_c, with associated utilities $\{Y_i\}$. Because the expected number of alternatives rated above x_c is $n[1 - \Phi(x_c)]$ (with Φ the standard normal cdf), the expected evaluation cost is given by $nc[1 - \Phi(x_c)]$. It follows that the expected net gain for the Level cutoff problem is:

$$E[\max\{Y_i \mid i \ni X_i > x_c\}] - nc[1 - \Phi(x_c)]. \tag{3}$$

In the following two sections, we expand on the objective functions given by (2) and (3), using them to characterize the optimal solutions for the Quota and Level Problems.

[1] The present framework can be modified to accommodate this "second-stage" error. It is not difficult to show that if the (imperfect) second-stage evaluations correlate with "real" utility with correlation ρ' (and if the associated error is uncorrelated with screening error), then this is identical to rescaling the unit evaluation cost to c/ρ'. Thus, for example, if the reliability or correlation of the second-stage is .5, then the cost per evaluation effectively doubles.

3. Optimal Quota Cutoffs

To optimally select a Quota cutoff for a given set of values of ρ, n and c requires that the expected net gain function given by (2) be evaluated or put into a form amenable to numerical integration. Following the derivation of Nagaraja and David (1994), we first seek to find the cdf of the random variable $V_{k,n}$, representing the largest utility value among the k highest-rated alternatives:

$$V_{k,n} \equiv \max[Y_{[n-k+1:n]}, \ldots, Y_{[n:n]}]. \quad (4)$$

In order to obtain this cdf, it is necessary to condition on the ratings values $\{X_i\}$:

$$P(V_{k,n} \le y) = P(Y_{[n-k+1:n]} \le y, \ldots, Y_{[n:n]} \le y)$$
$$= \int_{x_0<x_1<\cdots<x_k} P(Y_{[n-k+1:n]} \le y, \ldots, Y_{[n:n]} \le y$$
$$\mid X_{n-k:n} = x_0, \ldots, X_{n:n} = x_k) f_{X_{n-k:n}}(x_0) k!$$
$$\times \prod_{i=1}^{k} \frac{f_X(x_i)}{1 - F_X(x_0)} dx_i dx_0. \quad (5)$$

Given this conditioning on the order statistics, X_i, the concomitants are independent; therefore:

$$F_{V_{k,n}}(y) = \int_{x_0} \Bigg[k! \int_{x_0<x_1<\cdots<x_k} \prod_{i=1}^{k} P(Y_{[n-i+1:n]} \le y$$
$$\mid X_{n-i+1:n} = x_i) \frac{f_X(x_i)}{1 - F_X(x_0)} dx_i \Bigg] f_{X_{n-k:n}}(x_0) dx_0. \quad (6)$$

The bracketed integrand above is a symmetric function of the variables $\{X_i\}$; thus, it can be simplified as:

$$\left[\int_{x_0<x} P(Y \le y \mid X = x) \frac{f_X(x)}{1 - F_X(x_0)} dx \right]^k. \quad (7)$$

We turn our attention to the specific case of interest, when (X, Y) are bivariate normal, noting first that the pdf for the $(n - k + 1)^{\text{st}}$ order statistic is given by:

$$f_{X_{n-k:n}}(x) = (n-k)\binom{n}{k}[\Phi_X(x)]^{n-k-1}[1 - \Phi_X(x)]^k \phi_X(x), \quad (8)$$

and second that in (7):

$$P(Y \le y \mid X = x) \frac{f_X(x)}{1 - F_X(x_0)} = \Phi\left(\frac{y - \rho x}{\bar{\rho}}\right) \frac{\phi(x)}{1 - \Phi(x_0)}. \quad (9)$$

So, using (8) and (9), we can write the cdf for the maximum of the k concomitants with highest X-values as (note that, for notational consistency, within the inner integral, the variable x in (6) becomes the dummy variable u, while x_0 becomes x):

$$F_{V_{k,n}}(y) = \int_{x=-\infty}^{\infty} \left[\int_{u=x}^{\infty} \Phi\left(\frac{y - \rho u}{\bar{\rho}}\right) \frac{\phi(u)}{1 - \Phi(x)} du \right]^k$$
$$\times (n-k)\binom{n}{k}[\Phi(x)]^{n-k-1}[1 - \Phi(x)]^k \phi(x) dx$$
$$= \binom{n}{k} \int_{x=-\infty}^{\infty} \left[\int_{u=x}^{\infty} \Phi\left(\frac{y - \rho u}{\bar{\rho}}\right) \phi(u) du \right]^k$$
$$\times d[\Phi(x)]^{n-k}. \quad (10)$$

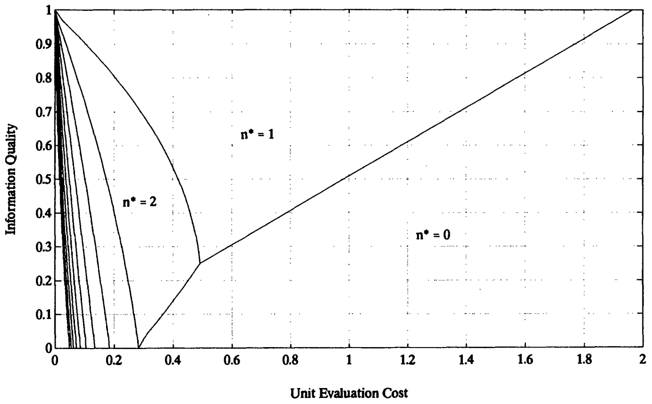
Thus, the formal decision problem for the Quota cutoff involves choosing k to maximize the expected value of $V_{k,n}$ less cost:

$$\max_k \binom{n}{k} \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} y \frac{\partial}{\partial y} \left[\int_{u=x}^{\infty} \Phi\left(\frac{y - \rho u}{\bar{\rho}}\right) \phi(u) du \right]^k$$
$$\times d[\Phi(x)]^{n-k} dy - kc. \quad (11)$$

The expression given by the objective function above can be integrated numerically for specific values of ρ and n, although this method is computationally intensive; for the purpose of obtaining these values, the quantities used in the present paper have been simulated as follows. The simulation proceeds by generating, for each value of ρ, a set of n $N(0, 0, 1, 1, \rho)$ pairs, which are then ordered by X-variate. Averaging the maximum Y for each k over 1,000,000 trials allows computation of the quantity $E[\max\{Y_{[n:n]} \cdots, Y_{[n-k+1:n]}\}]$. A comparison of the results obtained by simulation and those by numerical integration, over a wide range of values of ρ and n, indicates that the values in the simulation and integration differ by no more than one part in 10^3.

The optimal Quota cutoff, n_c^*, is that number of alternatives k for which the cost c of evaluating the $k + 1$st alternative is greater than the expected gain from including it in the set of evaluated alternatives. From such values of n_c^*, it is possible to plot the locus of points that have the same optimal Quota cutoff for fixed n. Figure 1 plots these IsoQuota regions for $n = 25$. In interpreting these graphs, it is important to realize that the cost axis

Figure 1 **IsoQuota Regions (n = 25)**



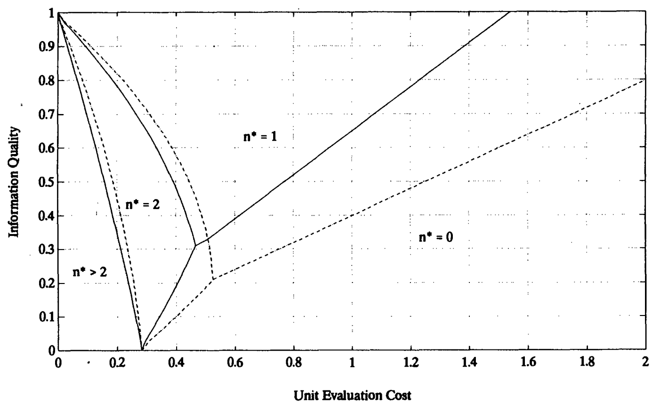
is defined in units of the standard deviation of utility, Y, so that a cost of $c = 1.0$ indicates that the decision maker would give up one standard deviation in the utility of the final result to avoid having to examine another alternative. For most realistic problems, cost will be far smaller, typically less than 0.1.

Sensitivity to c. When cost is high relative to gains due to search, then $n_c^* = 0$; in that region, the optimal strategy is not to evaluate any alternatives at all, as the processing cost outweighs the expected benefit from finding a better alternative. For any point (c, ρ) in the IsoQuota graph, increasing c involves entering a region of smaller n_c^*; as expected, making evaluation more costly entails doing less of it, at optimum. A seemingly unexpected characteristic is that a portion of the $n_c^* = 0$ region borders the $n_c^* = 2$ region, indicating that, as cost decreases in a region of low ρ, it is optimal to change from evaluating no alternatives to evaluating two. The intuition behind this effect is that if screening quality is poor but subsequent evaluations have low cost, it is not particularly valuable to evaluate just one alternative; two are needed to reap the benefits of the much more accurate second stage.

Sensitivity to ρ. The $n = 25$ IsoQuota regions show that, when cost is low ($c < 0.28$), increasing the reliability or quality of available information (i.e., the value of ρ) always entails evaluating fewer alternatives at optimum, and that when cost is high ($c > 0.48$), increasing the quality of information entails evaluating more alternatives. For fairly high evaluation costs (c less than approximately 1.96), choosing the alternative rated best by the indicator ($n_c^* = 1$) eventually proves to be a better strategy than "doing nothing" ($n_c^* = 0$) as the quality of the information, ρ, improves. However, in the low cost regions, this situation is reversed; having a more reliable indicator means there is less benefit from the evaluation stage, and so fewer evaluations are justified. Generally, for any value of cost, highly reliable information dictates that if any alternatives should be evaluated at all, only the best-rated one should be: as ρ nears 1, $n_c^* = 1$. There are numerous examples of this policy, in which choice is essentially dictated by the imperfect indicator; decision-makers may give serious consideration only to the "editor's choice" in a computer magazine, may elect to follow the explicit recommendation of a purchasing committee, or even just repeat a past choice. In all these cases, the indicator provides enough reliability to identify a good alternative, and performing additional evaluation is not deemed to be worth the incremental cost.

Sensitivity to n. For particular choice of ρ and c, the optimal Quota cutoff is quite robust to changes in the number of alternatives available, n. Figure 2 displays the IsoQuota regions for a ten-fold increase in the number of alternatives, moving from $n = 10$ to $n = 100$. This tenfold increase slightly expands the $n_c^* = 1$ region, where it is optimal to pay to evaluate only the best-rated alternative, at the expense of the $n_c^* = 0$ region. This result, that greater number of options leads to greater number to sample, is not intuitively obvious, but it appears to work quite generally. One explanation involves

Figure 2 **Effects of Increasing the Number of Alternatives on the Optimal Quota Cutoff, n = 10 and 100 (– – –)**



the expected difference between the ordered X-variates, which tends to diminish as n increases. Thus, expected difference between, say, the second and third alternative drops as n moves from 10 to 100 and, given one has examined the second alternative, increasing n brings the third alternative closer in and more fully justifies examining it.

4. Optimal Level Cutoffs

Recall that the optimal Level cutoff, x_c^*, must maximize the expected net gain function:

$$E[\max\{Y_i \mid i \ni X_i > x_c\}] - nc[1 - \Phi(x_c)]. \quad (12)$$

Let $p = \Phi(x_c)$, the probability that a certain alternative is rated below x_c, with $q = 1 - p$. Then, the distribution for the number of alternatives surpassing level x_c is binomial, $B(n, q)$. Assuming that, when $k = 0$, no alternative is evaluated and no cost or utility are incurred, we expand the utility component of, $E[\max\{Y_i \mid i \ni X_i > x_c\}]$, as follows:

$$\sum_{k=1}^{n} E[\max\{Y_1, Y_2, \ldots, Y_k\}] \cdot \binom{n}{k} q^k p^{n-k}. \quad (13)$$

It is necessary to calculate the density of Y_i. Because X_i is truncated below at x_c, X_i has pdf $\phi(x)/[1 - \Phi(x_c)]$ on $[x_c, \infty)$. Recalling that $Y_i = \rho X_i + \bar{\rho}\epsilon_i$, its pdf can be calculated as follows:

$$\begin{aligned} f_{Y_i}(u) &= \int_{-\infty}^{\infty} f_{\rho X_i}(x) f_{\bar{\rho}\epsilon_i}(u - x)dx \\ &= [\rho\bar{\rho}(1 - \Phi(x_c))]^{-1} \int_{x_c}^{\infty} \phi\left(\frac{x}{\rho}\right)\phi\left(\frac{u - x}{\bar{\rho}}\right)dx \\ &= \frac{\phi(u)\Phi[(\rho u - x_c)/\bar{\rho}]}{1 - \Phi(x_c)}. \end{aligned} \quad (14)$$

We refer to this pdf for Y_i as f, with $F' = f$. Because the $\{Y_i\}$ are i.i.d., the pdf for $\max\{Y_1, Y_2, \ldots, Y_k\}$ is

$$F_{\max\{Y_1, Y_2, \ldots, Y_k\}} = [F(u)]^k, \quad (15)$$

so that the expected value portion of (13) can be written:

$$E[\max\{Y_1, Y_2, \ldots, Y_k\}] = \int_{-\infty}^{\infty} ud[F(u)]^k. \quad (16)$$

Thus, we can rewrite $E[\max\{Y_i \mid i \ni X_i > x_c\}]$, the utility component for the Level cutoff problem, as follows:

$$\begin{aligned} &E[\max\{Y_i \mid i \ni X_i > x_c\}] \\ &= \sum_{k=1}^{n} \binom{n}{k} q^k p^{n-k} \int_{-\infty}^{\infty} ud[F(u)]^k \\ &= \int_{-\infty}^{\infty} u \frac{\partial}{\partial u} \sum_{k=1}^{n} \binom{n}{k} [qF(u)]^k p^{n-k} du \\ &= \int_{-\infty}^{\infty} u \frac{\partial}{\partial u} [p + qF(u)]^n du \\ &= \int_{-\infty}^{\infty} u \frac{\partial}{\partial u} \left[\Phi(x_c) + \int_{-\infty}^{u} \phi(\nu)\Phi[(\rho\nu - x_c)/\bar{\rho}]d\nu\right]^n du. \end{aligned} \quad (17)$$

The second-to-last integral expression above is valid for (X, Y) from an arbitrary jointly continuous bivariate distribution, provided the integral exists; the choice of the bivariate normal merely facilitates computation of $f(u)$. Denoting expected utility (17) by $\mathrm{EU}[x_c, \rho, n]$, the optimal Level cutoff, x_c^*, maximizes expected net gain, implying a first-order condition for local, interior extrema:

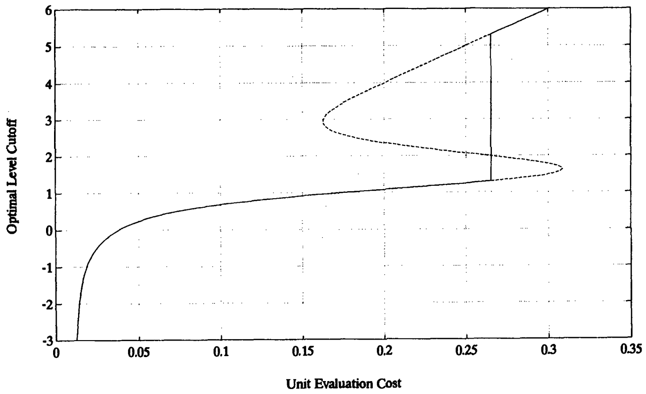
$$\begin{aligned} &\mathrm{EU}[x_c, \rho, n] - nc[1 - \Phi(x_c)] \Rightarrow \\ &\partial(\mathrm{EU}[x_c, \rho, n])/\partial x_c + nc\phi(x_c) = 0; \\ &c = -\frac{1}{n\phi(x_c)} \frac{\partial}{\partial x_c} \int_{-\infty}^{\infty} u \frac{\partial}{\partial u} \left[\Phi(x_c) + \int_{-\infty}^{u} \phi(\nu)\Phi[(\rho\nu - x_c)/\bar{\rho}]d\nu\right]^n du. \end{aligned} \quad (18)$$

It is demonstrated in the appendix that this implies the following relationship, at optimum, between c and x_c^*:

$$c = -\int_{-\infty}^{\infty} u \frac{\partial}{\partial u} \left[\Phi[(\rho x_c^* - u)/\bar{\rho}] \times \left(\Phi(x_c^*) + \int_{-\infty}^{u} \phi(\nu)\Phi[(\rho\nu - x_c^*)/\bar{\rho}]d\nu\right)^{n-1}\right]du. \quad (19)$$

Equation (19) offers a description of all local interior extrema for given values of ρ and n. While complex, this expression can be evaluated precisely using numerical methods. For most values of c, ρ, and n, (19) is satisfied by a single value of x_c^*. However, for some

Figure 3 **Points Satisfying First-Order Condition (– – –) and Actual Global Optima; $\rho = 0.05$, $n = 25$**



combinations of c, ρ, and n, (19) is satisfied by several values of x_c^*, some of which can represent minima, plateaus or non-optimal local maxima; any such value of x_c^* must therefore be checked for global optimality.[2]

One such situation, for $\rho = 0.05$ and $n = 25$, is shown in Figure 3, where an evaluation cost of 0.1 leads to an optimal Level cutoff of about 0.75 on the standardized X scale. As one might expect, lowering the cost leads to less restrictive cutoffs, while raising cost leads to more restrictive cutoffs. Note that the optimal cutoff is discontinuous near $c = 0.27$, jumping from a little more than 1.0 to greater than 5.0. Although such a discontinuity does not exist in all regions of the parameter space, this fundamental shift in strategy does; the resulting "contingent action" strategies, addressed at length in §6, will be seen to be one of two distinct regions of the parameter space where the Level cutoff method proves superior to the Quota.

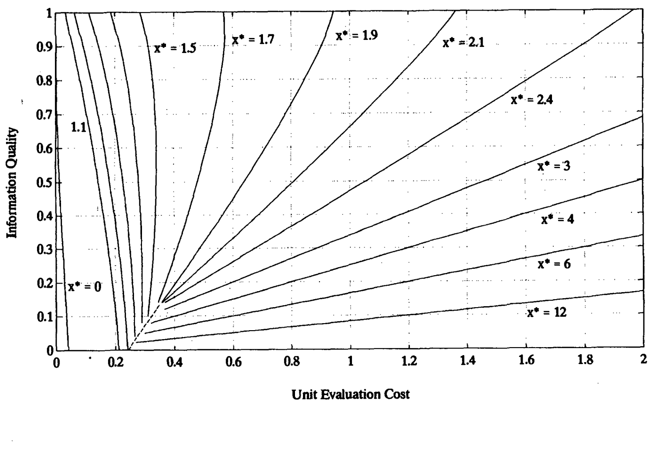
Analogous to the IsoQuota regions of Figure 1, it is possible to construct IsoLevel contours according to the curves of constant x_c^* in the (c, ρ) plane, allowing an assessment of the impact of the environmental parameters c, ρ and n. Note that, for a fixed value of n, specifying a constant Level cutoff x_c^* is equivalent to specifying some constant expected number of alternatives to be evaluated; for example, a Level cutoff of $x_c^* = 0$ corresponds to

$$n[1 - \Phi(x_c)] = (25)[1 - \Phi(0)] = 12.5$$

alternatives, on average, passing the $x_c^* = 0$ cutoff. These IsoLevel contours, for $n = 25$, are pictured in Figure 4.

Sensitivity to c. Although it is obvious that expected net gain cannot increase with evaluation cost c, it is less apparent whether an analogous statement can be made about the relationship between c and the cutoff x_c^*; indeed, as Figure 3 demonstrates for $\rho = 0.05$, the first-order condition (19) for c is not monotonic in x_c. However, from the IsoLevel contours of Figure 4 it is clear that, for any point (c, ρ), increasing c involves moving to an IsoLevel of higher x_c^*; this indicates that, just as in the Quota case, increasing c always requires increasing x_c^*, implying less evaluation at optimum. Also analogous to the Quota case is the "wall" of discontinuities displayed in the lower left, the "low c, low ρ" area, of Figure 4 (see footnote 2). In contrast to the Quota case, consider the region where the unit evaluation cost is large; recall that, for large c, the optimal Quota cutoff is $n_c^* = 0$ and, as this entails no evaluation, there is no cost, no utility and no expected net gain. Unlike the situation for Quota cutoffs, no matter how high the cost, there is some Level cutoff guaranteed to supply a positive expected net gain; in Figure 4, even in the region of high

Figure 4 **IsoLevel Contours ($n = 25$)**



[2] Such jump discontinuities are similar to those in the Quota case involving "discontinuous" jumps from $n_c^* = 0$ to $n_c^* = 2$. Figure 4 displays this graphically, where curves of low IsoLevel (x_c^* constant) meet curves of high IsoLevel. Were x_c^* visualized as height above the (c, ρ) plane, the curve made up of such meetings would appear as a sheer wall of discontinuities, widening to infinity as ρ approached 0.

c and low ρ (where the Quota cutoff implies that no evaluation is optimal), there are IsoLevel contours, albeit of exceptionally large value.

Sensitivity to ρ. As is demonstrated by the IsoLevel contours of Figure 4, increasing ρ does not have a clear directional impact on the optimal Level cutoff. For low levels of processing cost, increasing ρ leads to higher optimal Level cutoffs (less evaluation), while for high cost the opposite holds; there is even a range of costs (near $c = 0.25$), where raising ρ from 0 through 1 leads first to lower (more evaluation), then to higher (less evaluation), IsoLevels. All these results are analogous to those for Quota cutoffs.

Sensitivity to n. In analogy with Figure 2, which details the effects on increasing n tenfold from $n = 10$ to $n = 100$, it is possible to construct two different graphs. The first would depict different, fixed IsoLevel contours (values of x_c^*) for both $n = 10$ and $n = 100$; the second, rather than fix values of x_c^*, would instead choose to fix the values for the expected number of alternatives passing the screen, comparing a lower level of x_c^* for $n = 10$ to its corresponding higher value for $n = 100$. These IsoLevel graphs can both be summarized verbally. For the first of these, increasing n always leads to an increase in x_c^*. Unlike Quota cutoffs, which are rather insensitive to changes in n, Level cutoffs must change rather decisively to avoid evaluating an excessive number of alternatives; this may indicate that, from a behavioral standpoint, the act of adjusting Level cutoffs requires far more attention and processing than adjusting the corresponding Quota cutoff. For the second of these IsoLevel graphs, if one were to examine the *expected* number of alternatives passing the optimal Level cutoff as a function of n, the impact of increasing n is essentially the same as in the Quota case: more alternatives should be evaluated. Thus, although the value of the optimal cutoff, x_c^*, increases with n, it is nevertheless reasonable to claim that, as n increases, the optimal response for the Level cutoff (as well as for the Quota) is to become *less* restrictive.

Summary. There is a remarkable parallel in the sensitivities of the Level and the Quota cutoff methods, at optimum, to the parameters of the decision environment. First, for both, increasing c always results in more restrictive cutoffs. Second, increasing ρ leads to more restrictive cutoffs when costs are low, while the reverse is true when costs are high. Third, when information is poor, discontinuities arise in the optimal strategy. There are, however, aspects in which the two cutoff methods markedly differ. While Quota cutoffs are remarkably robust to changes in the size of the choice set, n, Level cutoffs are particularly sensitive to n. Doubling the size of the choice set entails twice as many alternatives, on average, surpassing a fixed Level cutoff, requiring drastic increases in x_c^*; a fixed Quota cutoff, by its nature, implicitly adjusts to such changes. Finally, while escalating evaluation costs will eventually price the Quota cutoff out of the market ($n_c^* = 0$), this is not the case for a Level cutoff, for which the optimal value x_c^* is always finite, a difference which plays a crucial role in §6, where the relative efficiency of the two methods is compared (this asymptotic relationship is demonstrated in the appendix).

5. First- and Second-stage Defaults

One of the goals of the present paper is to illustrate the role of default, or fallback, options in optimal selection problems. Recall the example of deciding which of a large pool of job applicants to interview based on the quality of their resumes, all of which have been rated on some scale of overall suitability. It is possible that none of the applicants, based on these resume ratings, seems appropriate for the job, and none is interviewed; in terms of the cutoff methods in question, this could be accomplished by setting a Quota of 0 or a resume-ratings Level high enough so that none manages to surpass it. Regardless of the cutoff method used to bring it about, such a decision would only be plausible if it were *acceptable* to let the job go unfilled, meaning that there must be some clearly defined payoff or cost associated with keeping the job open. This fallback option may come about in a variety of ways, for instance, if the person presently doing the job has agreed to stay on until a satisfactory replacement can be located, if one can hire a temporary person to fill in, or if the position can go unfilled for some specified length of time during which a new search for appropriate candidates can be undertaken. In any of these situations, the associated costs of electing *not* to carry out any evaluation at the present time can be accounted for and explicitly factored into the search process. As both the Quota and Level models have been formulated, the payoff associated with evaluating no alternatives, what we will refer to as a *first-stage default* (D_1), has been assumed to be zero. This assumption of a zero first-stage default, given the zero-mean, unit-variance scaling used in our analysis, is equivalent to assuming that, rather than engage in the costly evaluation stage, one chooses *at random* from the original set of alternatives. Should the true fallback option have a greater or lesser value than the average of the original set, the model must be modified to allow a different first stage default value.

By contrast, the *second-stage default* (D_2) comes into play after one has conducted the costly search and identified no satisfactory alternatives. For example, suppose that a broad search is undertaken for the position but none of the applicants, upon being interviewed, is perceived as being right for the job. The second-stage default is the value of the option one can fall back on *after* a set of detailed evaluations has been performed. As both the Quota and Level methods have been analyzed thus far, the value of the second-stage default has been negative infinity, that is, the decision maker had to choose from among the alternatives that passed the screen, no matter how poor they appeared after evaluation.

It should be clear that, conceptually, the first- and second-stage defaults are quite distinct, and can take on different values. For example, it is possible that one's assistant will stay on indefinitely so long as one chooses not to interview (i.e., evaluate) possible replacements, but will give notice to look for another job if one decides to bring in even a single potential replacement for an interview. Such a situation, involving a first-stage default of zero and a second-stage default of negative infinity, is the type we have examined thus far. An argument can be made to the effect that, if there is any nonzero default in either stage, parsimony dictates they have the same value, and there is merit to this view. As we shall see, while setting both defaults to zero ($D_1 = D_2 = 0$) is not very different from the situation which we have analyzed ($D_1 = 0, D_2 = -\infty$), setting both defaults to negative infinity ($D_1 = D_2 = -\infty$), capturing a situation in which one absolutely must wind up with an acceptable option, is somewhat of a disaster for the Level cutoff, necessitating an infinitely negative cutoff and all alternatives being evaluated, no matter the cost. In any case, although the mathematical development will treat D_1 and D_2 independently, graphical illustrations will, in the interest of brevity, take them to have equal value.

5.1. Quota Cutoffs with Defaults

First-Stage Default: The Quota method requires no additional analytics to accommodate a first-stage default option. At any (c, ρ)-point in the IsoQuota regions of Figure 1, one need only compare the value of expected net gain to the first-stage default value; if the default is of greater value (i.e., $n_c^* = 0$), it is chosen. In this light, the entire $n_c^* = 0$ region in Figure 1 can be seen to represent a surface of constant height, where expected net gain is zero; any other value for the first-stage default would function similarly, with larger (i.e., positive) values causing the $n_c^* = 0$ region to "subsume" a greater portion of the regions to its left, and smaller (i.e., negative) values of the default causing the $n_c^* = 0$ region to retreat to the right.

Second-Stage Default: Should none of the n_c alternatives specified by the Quota cutoff be associated with Y-values greater than the value of the second-stage default, D_2, the payoff incurred is not the maximum of the Y-values but D_2. Thus, the expected net gain function becomes:

$$E[\max\{Y_{[n:n]}, \ldots, Y_{[n-n_c+1:n]}, D_2\}] - cn_c. \quad (20)$$

The resulting derivation of the numerically-integrable net gain function is exactly analogous to the $D_2 = -\infty$ case (§3, (11)), with the exception that the minimally-acceptable payoff is now D_2:

$$\max_k \binom{n}{k} \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} \max(y, D_2) \frac{\partial}{\partial y} \left[\int_{u=x}^{\infty} \Phi\left(\frac{y - \rho u}{\bar{\rho}}\right) \phi(u) du \right]^k d[\Phi(x)]^{n-k} dy - kc. \quad (21)$$

This maximization problem should be viewed with the caveat that the first-stage default is not accounted for, and can take on any value, positive or negative. Calculation of IsoQuota regions is analogous to the previous situation, depicted in Figure 1, where the defaults were unequal, with $D_1 = 0$ and $D_2 = -\infty$; for the purposes of graphical representation, we will assume that $D_1 = D_2$. Figure 5 shows IsoQuota regions for different values of the common first- and second-stage default. As might be expected, as the value of default increases, there is a general tendency toward evaluating fewer options at optimality.

5.2. Level Cutoffs with Defaults

First-Stage Default: We consider the case in which, if no alternatives are evaluated (due to a high Level cutoff), the decision-maker has available a default option, of value D_1. Thus, there is an additional payoff of D_1, accrued with probability $[\Phi(x_c)]^n$, implying an expected net gain and first-order condition, respectively:

$$\mathrm{EU}[x_c, \rho, n] - nc[1 - \Phi(x_c)] + D_1[\Phi(x_c)]^n;$$

$$\partial(\mathrm{EU}[x_c, \rho, n])/\partial x_c + nc\phi(x_c) + nD_1\phi(x_c)[\Phi(x_c)]^{n-1} = 0. \tag{22}$$

The role of unit cost, c, has therefore been replaced by the quantity $c + D_1[\Phi(x_c)]^{n-1}$, and the relationship between x_c^* and c, at optimum can be stated simply as an analog of (19):

$$c = -D_1[\Phi(x_c^*)]^{n-1} - \int_{-\infty}^{\infty} u \frac{\partial}{\partial u}\left[\Phi\left(\frac{\rho x_c^* - u}{\bar{\rho}}\right)\left(\Phi(x_c^*) + \int_{-\infty}^{u} \phi(\nu)\Phi[(\rho\nu - x_c^*)/\bar{\rho}]d\nu\right)^{n-1}\right]du. \tag{23}$$

Thus, for a fixed Level cutoff, x_c^*, an increase in the value of the first-stage default D_1 will cause the cutoff to be optimal for a smaller value of c (i.e., put simply, increasing D_1 for a fixed c will raise the optimal cutoff, x_c^*). In effect, "subsidizing" the situation in which nothing passes the cutoff causes the cutoff to be raised, at optimum. Note as well, though, that this lower cutoff value will correspond to a different (higher) level of profit; replacing, in the profit formulation (22) above, the value of c associated with no first-stage default with the one given by (23), shows that the (positive) change in profit between a first-stage default of $D_1 > 0$ and one of $D_1 = 0$ is given by:

$$nc[1 - \Phi(x_c)] - n(c - D_1[\Phi(x_c)]^{n-1}) \times [1 - \Phi(x_c)] + D_1[\Phi(x_c)]^n = [n(1 - \Phi(x_c)) + \Phi(x_c)]D_1[\Phi(x_c)]^{n-1} > 0. \tag{24}$$

Second-Stage Default: We consider the case in which, after evaluation, the decision-maker "prefers" the default option, of value D_2, to any of the Y-values turned up by the evaluation process. The density of Y_i is the same as in the $D_2 = -\infty$ case (see (14)). In the sequence of equations ((16)–(19)) leading to the relationship between x_c^* and c, the only change needed is in the variable used for expected value calculations, u, which has been replaced by $\max\{u, D_2\}$. Thus, we have the following relationship at optimum:

$$c = -\int_{-\infty}^{\infty} \max\{u, D_2\} \frac{\partial}{\partial u}\left[\Phi[(\rho x_c^* - u)/\bar{\rho}]\left(\Phi(x_c^*) + \int_{-\infty}^{u} \phi(\nu)\Phi[(\rho\nu - x_c^*)/\bar{\rho}]d\nu\right)^{n-1}\right]du. \tag{25}$$

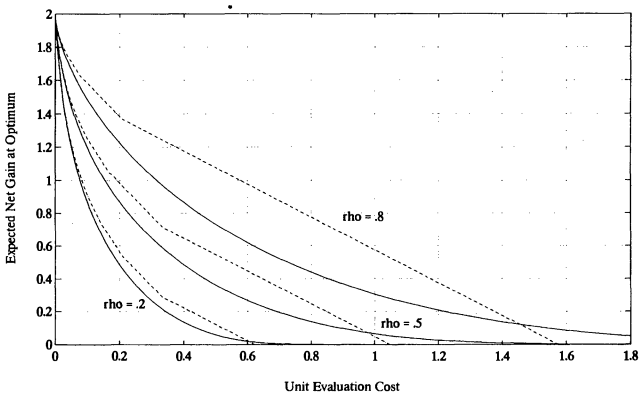
If, in analogy to Figure 5, the IsoLevel contours of Figure 4 were amended in accordance with (23) and (25), for a common value of $D_1 = D_2$, the resulting contours would all relocate leftward as the common default increased. Simply stated, increasing the common default (in fact, increasing either default alone) causes the optimal cutoff x_c^* to increase. Note that this situation is similar to that of Figure 5, where increasing the value of the common default caused a leftward migration of each of the IsoQuota boundary lines.

6. The Relative Effectiveness of the Quota and Level Cutoff Methods

As the Quota and Level cutoff methods both allow large sets of alternatives to be restricted with little effort, it is useful to assess their relative effectiveness. Although the original purpose of this study was to determine, and subsequently prove, which of the two cutoff methods was best, no such proofs are possible; each of the methods has unique qualities allowing it to dominate the other for certain values of the parameters specifying the choice environment, ρ, n, c, D_1 and D_2. Thus, the best one can hope for is a taxonomy of sorts as to which regions of the parameter space respond best to which of the cutoff methods. We note at the outset, however, that the economic intuition suggesting that the Quota method, due to its ability to fix cost (coupled with a vague application of Jensen's inequality), should dominate the Level method, is largely borne out: for the parametric values arguably most typical of the majority of choice problems, the Quota method will turn out to be superior to the Level. However, for two specific regions of the parameter space, describing seemingly less common choice scenarios, Level cutoffs can outperform Quota cutoffs. The goal at hand, then, is to provide a description of these different regions of the parameter space and relate them where possible to actual behavior.

Figure 6a depicts the expected net gain, as a function of cost, for Quota and Level strategies for three different values of ρ, a "low" value ($\rho = 0.2$), a "moderate" value ($\rho = 0.5$), and a "high" value ($\rho = 0.8$), for $n = 25$ and $D_1 = D_2 = 0$; qualitative features of this figure are broadly representative of choice scenarios for any value of $n > 1$ and values of $D_1 = D_2$ from negative infinity to not too much more than 0 (that is, values which the

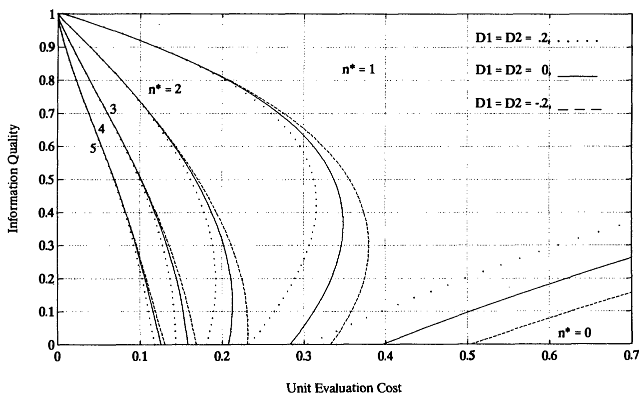
Figure 6a **Expected Net Gain for Quota (– – –) and Level, $D_1 = D_2 = 0$; $n = 25$**



best of the set of n alternatives is very likely to surpass). Note that, for all three values of ρ, there is some value of cost (e.g., $c \approx 0.97$ for $\rho = 0.5$) above which the Level strategy is superior to the Quota; it is important to understand that, so long as ρ is non-zero, this must happen; that is, there is *always* some processing cost c above which the Level cutoff strategy is superior.[3] However, the resulting expected net gain is often minuscule (e.g., for $c = 3.8$, $\rho = 0.80$ and $n = 25$, the optimal cutoff is $x_c^* = 4.75$, with a resulting net gain of 4×10^{-6}), simply because the probability of any alternative passing the cutoff is very small ($\approx 10^{-6}$). We term these seldom-invoked Level cutoffs *contingent action strategies*, as they require action to be taken only a small proportion of the time, usually under uncommon circumstances.

Thus, while Level cutoffs may be superior for some decision environments, the expected net gain involved can be small enough so that the strategy may have little practical importance unless the cost of a poor decision is very high. In "everyday" situations, where one

Figure 5 **IsoQuota Regions with $D_1 = D_2 = \{-0.2, 0, 0.2\}$**

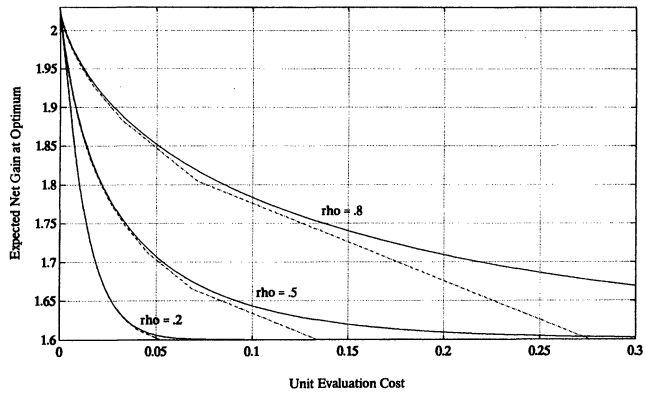


[3] This occurs for the following reason. For very high evaluation cost, specifying even a single alternative with a Quota strategy will be prohibitively costly. However, with a Level strategy, one can set a cutoff so high that only a single alternative is ever likely to surpass it, an alternative of such quality that it is all but certain to offset even an extremely high evaluation cost. As shown in the appendix, the asymptotic relationship $x_c^* = c/\rho$ ensures that no matter how high the unit evaluation cost, there is a finite Level cutoff guaranteed to provide a positive expected net gain.

typically expects to perform detailed evaluation on two or more alternatives, Quota cutoffs generally perform better. This is not to say that the high-cost regions where Level cutoffs flourish are without relevance to decision-making; indeed, from an evolutionary vantagepoint, rarely-used cutoff rules assume a vital role. Consider, for instance, how a cutoff is set as to when an environmental stimulus warrants some type of protective motion or detailed attention. The overwhelming number of such stimuli being received at any given moment dictate that some type of perceptual screening take place, screening that, by its nature, could be neither sequential ("order all impinging stimuli according to how threatening they are and investigate in that order") nor Quota-based ("investigate the two most threatening stimuli"). The human perceptual apparatus seems to be armed with a remarkable variety of such rarely-enacted Level-based contingent action strategies, from deciding which noises in the night may indicate an intruder to when one should move one's hand to avoid being badly burned, even if one has never directly encountered these situations before.

It would be incorrect to conclude that the contingent action region, where costs are high and practically everything is screened out, is the only one in which Level cutoffs can outperform Quota, as there is an additional region of the parameter space where this occurs. This other region is the one where the default option (assuming $D_1 = D_2$) is quite good, that is, its value is a significant proportion of the expected payoff associated with evaluating all possible options. Figure 6b depicts such a situation, for $\rho = \{0.2, 0.5, 0.8\}$, $n = 25$ and $D_1 = D_2 = 1.6$. Put in terms of the $\{\rho, n, c, D_1, D_2\}$ parameter space, whereas the previous region of Level superiority could be described in terms of cost only, perhaps as "high c," the present region focuses not on cost but on the default values, and might be termed "high D_1 and high D_2," where "high" denotes a value that is a large proportion of the maximum payoff possible from the entire set of alternatives; for $n = 25$, this payoff is roughly 1.96, the expected maximum of a set of 25 independent $N[0, 1]$ variables. That this "high D_1 and high D_2" region should emerge as one where the Level strategy is superior offers a hint about how the Level cutoff uses contingent action to its advantage. Level cutoffs offer a particularly effective way to exploit a lucrative

Figure 6b **Expected Net Gain for Quota (– – –) and Level, $D_1 = D_2 = 1.6$; $n = 25$**



first-stage default, because a Level cutoff can be set so that one is unlikely to pay evaluation costs unless one actually exceeds the default value, while a Quota cutoff does not allow this flexibility. Conversely, a particularly poor first-stage default hangs like Damocles' sword over the Level cutoff—one can never be totally sure of avoiding it—while any Quota cutoff above 1 is certain to do so.

Considering the excessive evaluation costs and low returns that accrue in the contingent action region, and the relative rarity of exceptionally high default values, it is safe to assert that the Quota strategy is more effective than the Level strategy for "typical" choice problems. The Quota strategy has other advantages as well. As discussed earlier, a Quota cutoff is less affected by the total number of available alternatives than the corresponding Level cutoff. Further, being based on a rank ordering of the best, the Quota cutoff is also less sensitive to errors in assessing these distributions. Apart from being more robust to changes or uncertainty in the environment, the Quota strategy is easier to calculate intuitively and to adjust when one is wrong, simply because it is integer valued, and does not require one to estimate the number of alternatives that can be expected to surpass a given level.

Given its apparent dominance, one could ask why the Quota strategy is not more in evidence in behavior. One possible explanation is that pure opportunities to use Quota strategies occur rarely in nature. Alternatives are typically not ranked by some criterion and available in sets of varying sizes to decision-makers. Another possibility is that a choice made using a Level cutoff offers a measure of defensibility, owing to its having passed some numerical criterion, something a Quota cutoff cannot provide; it speaks better of one's standards and year-to-year consistency to consider only those applicants scoring above a prespecified Level on an entrance exam, rather than to consider some fixed number of top scorers, whatever the minimum score in that particular group happens to be. However, it appears that a large region of the IsoQuota graph, that with $n_c^* = 1$, is consistent with a number of decision heuristics. These heuristics include choosing the brand one chose last time, or choosing a consumer magazine's best-rated product. These heuristics generate an imperfect indicator of quality, and in these cases reflect a Quota strategy with $n_c^* = 1$.

In contrast to the scarcity of environments in which Quota strategies are available, opportunities for Level cutoffs abound. Consider decisions only to examine cars in a given price range, or from a given country. Buyer guides and brand shelves are organized to facilitate Level cutoffs. Further, if we relax our assumption of irrevocability, then a Level cutoff can efficiently be used to emulate a Quota strategy. Suppose that a thorough evaluation takes time, but there is very little cost to noting the number passing the screen and then reentering the screening process. In that case, a cutoff that is too restrictive may be is eased, or one too loose made tighter. Empirical evidence for this includes a experiment of Widing and Talarzyk (1991), in which minimal levels of attributes could be iteratively increased or decreased; respondents employed Level cutoffs to derive a small set of alternatives from which to choose, in effect using successive Level cutoffs to approximate a Quota one.

That the essentially general superiority of the Quota over the Level method can be nullified for certain extreme values of the parameters c, D_1 and D_2 should not be taken to imply that the other "primitive" parameters of the choice environment, ρ and n, have no effect on the two methods' relative effectiveness. Their role is more one of amplification rather than parity; generally speaking, a Level/Quota dominance relation holding in some region of the $\{c, D_1, D_2\}$ parameter subspace for particular values of ρ and n will tend to hold more strongly, or over a wider range of parametric values, as either ρ or n increase. Given the numerous potential interactions, even taken pairwise, among the parameters necessary to specify the choice environment for either the Quota or Level methods, it is rather remarkable to be able to summarize, albeit sketchily, the dominance relationships between the two methods so simply as follows: Except when evaluation cost is very high or the default options are exceptionally good, the Quota method is preferred to the Level, and the degree of preference for either increases with the quality of prior information and the number of alternatives from which to choose.

7. Extensions and Directions for Further Research

It is important to keep in mind that the results found here relate to the choice environment defined at the outset—one in which screening by an imperfect indicator is inexpensive, where subsequent evaluation is costly, and where one cannot come back to the well if dissatisfied. Within such a choice environment, we detailed the characteristics of two cutoff strategies, Quota and Level, with respect to five parameters of the decision environment: the cost per evaluation, the quality of the screening indicator, the number of alternatives, and the (default) values of the two fallback options. This section examines the contributions of this research in terms of providing direct guidance for decision makers, and for the general development of theory in such complex decision problems.

Perhaps the best way to discuss the relevance of the cutoff theory to decision makers is to return to the questions raised initially in the academic hiring scenario. The first question involved whether the number of candidates brought in should be based on a fixed, predetermined number (Quota) or on the faculty rating scores (Level). Given that the number of candidates expected is appropriately in the three-to-six range, the Quota cutoff turns out to be superior to the Level, not only because it offers greater expected return at optimum, but also due to its greater robustness to uncertainty in choice environment parameters. This robustness is highlighted in the answer to the second question, concerning the change in cutoffs given a decrease in the number of available candidates. Even a moderate decrease in the number of alternatives leads the optimal Level cutoff to drop sharply, to avoid having too few candidates pass the screen; in contrast, fewer alternatives causes the optimal Quota strategy to become only slightly more restrictive or, because it is integer-valued, to not change at all. The third question concerns the faculty rating system; what if it were made more reliable, perhaps through the incorporation of additional information? In this case, the answer is that cutoffs at optimum should become more restrictive with improvements in quality of the indicator. The final question asks what happens, at optimum, to either type of cutoff as the fallback option improves. The answer here is that cutoffs should generally get more restrictive and, if the quality of the fallback option is sufficiently strong, one might then enter the region in which no candidate should be brought in, and a contingent action strategy is appropriate; it is also possible, if the fallback option is exceptionally good, for the Level method to outperform the Quota.

There is a general principle that accounts for most of these results, applicable in choice environments not justifying a contingent action strategy. Because cost resides in the evaluation stage, rather than the screening stage, **any change in the choice environment that increases the relative benefit of the evaluations, or decreases their cost, produces a less restrictive screen at optimum.** A number of conclusions follow directly from this principle. For example, reducing processing costs increases the optimal number passing the screen, as does increasing the reliability of the evaluation process. Further, the availability of a better fallback option also makes the optimal screen more restrictive, because the incremental value of each alternative evaluated is limited by the probability that it may not be superior to the fallback option.

Less obviously, decreasing the quality or reliability of the screening indicator (ρ) means that there is greater relative benefit from the evaluations, so they will be used more extensively. Put differently, decreasing the quality of the screening indicator increases the variance resolved in the evaluations, thereby justifying more of them. In a parallel way, an increase in the number of original alternatives also leads to an increase in the optimal number passing. In terms of expected value, allowing more alternatives to enter can only decrease the range of the X-values (ratings) among the k best-rated alternatives, meaning that one's ability to make choices based on these X-values alone will be impaired; simply put, one's ability to *distinguish* the alternatives, based on their ratings, is impaired as more alternatives are made available. Thus, in accordance with the principle above, the evaluation stage becomes more useful, in that it allows one to better distinguish the alternatives, and so more evaluations are performed at optimum.[4]

Focusing on the relative costs and benefits in the evaluation stage permits one to deduce the impact of changes in the choice environment on the restrictiveness of optimal cutoffs. Specifically, one can predict that cutoffs at optimum will be more restrictive with (1) greater processing costs, (2) greater accuracy in the cutoff indicator, (3) fewer total alternatives, and (4) better fallback options. One intriguing aspect of these findings is that they hold in parallel fashion for both Quota and Level strategies, reflected in the remarkably close correspondence and parametric sensitivities of the IsoQuota regions and IsoLevel contours. It is important to recall that this parallelism requires that restrictiveness be defined as inversely related to the *number* passing the screen in the Quota case and to the *expected number* passing in the Level case.

If the parallelism found between the characteristics of the Level and the Quota is suggestive, this is even more the case in the general dominance of Quota over Level in regions where one expects to evaluate one or more alternatives. There is no theoretical reason of which we are aware that would predict this dominance. Further, the Quota method succeeds even though the integer responses required of it hamper it relative to the continuous response available to the Level method.

[4] Notice that this is borne out by the asymptotic result for high cost, $c = \rho x_c^*[\Phi(x_c^*)]^{n-1}$, where the optimal Level cutoff is quite robust to changes in n (see Appendix). The extent to which this is true for distributions other than bivariate normal has not been addressed, though we have verified it graphically when the $N(0, 1)$ X distribution is replaced by a uniform distribution of zero mean and unit variance. It is not known what conditions on the joint distribution of (X, Y) would be sufficient for this to be proved.

The greater efficiency of Quota over Level is an important practical result because Quota cutoffs typically have to be generated, rather than being offered up by nature. By contrast, Level cutoffs abound; witness the widespread use of distance from work in apartment selection, price in the choice of durables, or standardized exam scores in student admission decisions. The normative implication is that decisions requiring significantly costly evaluations can be streamlined if one can set up a relatively cheap screening mechanism, and that this screen will be even more efficient if based on the number of alternatives evaluated, rather than a fixed attribute level. This screening can be performed by a helper, a committee, or a computer. The results here broadly indicate that, when used optimally, a simple rule like "Find the three best candidates" proves superior to a rule like "Evaluate candidates with GMAT scores over 700."

By contrast, the Level strategy emerges as better than the Quota when evaluation costs are high or when the fallback option is extremely favorable. The Level cutoff is then used to select alternatives with very high X-variates; this is termed a "contingent action strategy" because explicit action (evaluation or choice) is taken only rarely. Despite being generated from a decision process that appears to be very similar to the standard cutoff problem, the contingent action strategy is importantly different in two respects. First, items are expected to pass the optimal cutoff very rarely, typically on the order of one time in a thousand or less, so that the high processing cost is ordinarily incurred for only the one item passing the screen. Second, in a contingent action environment, a more effective screen (higher ρ) justifies a *less* restrictive cutoff, effectively reversing the intuition given earlier about the inverse relationship between the quality of the indicator and the number of items that should be examined. While this kind of decision rule appears to occur in a number of contexts, it is sufficiently different from the standard use of cutoffs to warrant a separate exploration of its properties.

In addition to providing guidance in choice scenarios explicitly requiring cutoffs to be made, the framework introduced in this paper has implications for those modeling such behavior. Although there has been some work along these lines in statistics and process control (Schilling 1982, Yeo and David 1989) and in the management sciences (Gensch 1987, Hauser and Wernerfelt 1990, Roberts and Lattin 1991), this is the first time, to our knowledge, that any comparison has been made between the effectiveness of Quota and Level cutoffs generated through imperfect indicators. These comparisons would not have been feasible without appealing to the concomitants of order statistics, a general methodology which seems to have been thusfar underutilized in the vast literature on random utility models.

Another contribution of the present work is in its incorporation of default, or fallback, options. We distinguish two rather different types—a first-stage default in the event that no alternatives pass the screen, and a second-stage default in the event that all options passing the screen are unacceptable—and discuss their effects on the optimal properties of cutoff selection. It is important to realize that such decision problems really cannot be framed in generality without the incorporation of defaults. We anticipate that default values will play a central role in extending the present choice scenario, in which one makes a single, irrevocable cutoff, to more complex ones in which several successive cutoffs are made. In a multi-cutoff or sequential-choice setting, a default value precisely corresponds to the expected value of being able to "return" to the original set of alternatives to restrict or derestrict it in order to improve the quality of the final choice.

The comparative statics given above are most useful in providing directional guidance to a decision maker who uses a cutoff in the face of an altered decision environment, similar to the situation presented by the faculty screening questions. One valuable use of such directional guidance is in the evaluation and calibration of cutoff heuristics. For example, Huber and Klein (1991) showed that reducing indicator accuracy led to less restrictive cutoffs among naive respondents, although they were unable to estimate how close these cutoffs were to an optimal policy. Of course, if suitable values of ρ and evaluation cost are known, or can be reliably estimated, then the methods developed here permit an optimal policy to be determined; being able to specify an optimal or near-optimal policy thus allows decision makers to assess the normative effectiveness of their cutoff heuristics. In assessing an optimal normative policy, external estimates of ρ might be preferable to subjective ones, given the notorious instability inherent in subjective estimation (Sanbonmatsu, Shavett and Gibson 1994). Such "objective" estimates can come from a variety of sources, pretests, correlations calculated from past data or from closely related evaluative situations (Gokhale and Press 1982, Jennings, Amable and Ross 1982).

In terms of theoretical generalizations, our major finding may be that so few exist. In studying the cutoff problem, our original intent was, naively in retrospect, to be able to offer analytical proof or exhaustive evidence for statements of the "Quota is always better than Level" or "Cutoffs get more restrictive with increases in ρ" variety. While such statements appear to hold in vast regions of the parameter space, they tend not to hold in all of it, making sweeping generalizations tempting, but demonstrably inaccurate. One is inclined to say then that the chief results of this study are less main effects than interaction effects; generalizations need to be specified on the basis of more than a single parameter at a time. This is not to say that there are no implications that hold across the board. Several dictated by economic principle are generally borne out, for instance, that less evaluation is performed and net gain is poorer for higher unit evaluation cost. Further, the result that evaluation should be more extensive when more alternatives are available, which was not readily apparent from economic theory or other theoretical work on choice behavior, also seems to hold in general.

Even if the two-stage, single irrevocable-cutoff choice framework adopted here is kept intact, there are several assumptions of the model, made in the name of parsimony and tractability, which might well be relaxed to accommodate known realities of various decision environments. Several of these involve scaling assumptions, mainly linearity. For example, the expected return in utility for an additional unit of many numerical indicators—test scores in candidates, mpg in automobiles, current profits for companies—tends to exhibit diminishing returns. This problem can be ameliorated by combining different attributes into a score, perhaps in a compensatory fashion or with the aid of a technique like factor analysis, and then rescaling that score to be in constant units of utility. A second linearity assumption that could be relaxed is that of constant cost per evaluation. One can easily envision marginal processing costs increasing or decreasing as a function of the number evaluated. Increasing costs could derive from fatigue, boredom or the increased recording or memory costs as multiple alternatives are compared. Decreasing costs can come from learning, standardization and process improvements. A variable evaluation cost specification is not much of an impediment to obtaining closed-form solutions for either the Level or Quota problems, as it can be accommodated peripherally in formulating a first-order condition; that is, it affects only the cost portion of the net gain function, not the utility portion. However, any such changes would tend to limit the generality of claimed properties of the resulting optimal strategies.

In conclusion, both the Level and Quota strategies are valuable heuristics because they are simple to use and adaptable. Their simplicity derives from the ease with which they can be implemented or transferred to other situations, their adaptability from the relatively simple informational basis for changing the cutoff rule. This adaptation is easily enacted with either strategy and does not require knowledge of the calculations shown here. Decision makers may note, for example, that they are examining too many "long shot" alternatives and subsequently tighten the screen or, conversely, may uncover an outstanding alternative that did not clear the screen and resolve to cast a broader net the next time around. Thus, while exploration into other heuristics, such as those involving incremental or multiattribute selection, will certainly prove important, it is the simplicity and adaptability of the Level and Quota strategies which ultimately justifies both their study and use.[5]

[5] The authors wish to thank Professors H. A. David and H. N. Nagaraja for sharing their expertise with order statistics, S. Lippman, K. McCardle, two anonymous reviewers and the area editor for their numerous suggestions, as well as the Social Sciences and Humanities Research Council of Canada for their continuing support.

Appendix: Scaling, Asymptotics, Proof of (19)

Scaling of Problem Parameters

Throughout the mathematical development, ratings, X, and utility, Y, were assumed to have been standardized to have joint distribution $N(0, 0, 1, 1, \rho)$. Here we assume that ratings, utilities, unit evaluation cost and default values are given in dimensioned units, and show how X and Y are defined relative to these. Assume the following notation for the nonstandardized parameters: R = ratings, U = utilities, k = unit evaluation cost, Δ_1 = first-stage default, Δ_2 = second-stage default; the joint distribution (R, U) is now $N(\mu_R, \mu_U, \sigma_R^2, \sigma_U^2, \rho)$. Then, the expected net gain for, for example, the Quota problem is given by:

$$E[\max\{U_{[n:n]}, U_{[n-1:n]}, \ldots, U_{[n-n_c+1:n]}, \Delta_2\}] - Kn_c. \quad \text{(A.1)}$$

Recalling that $U_{[i:n]} = \mu_U + \sigma_U Y_{[i:n]}$, (A.1) is given by:

$$\mu_U + \sigma_U\left[E\left[\max\left\{Y_{[n:n]}, Y_{[n-1:n]}, \ldots, Y_{[n-n_c+1:n]}, \frac{\Delta_2 - \mu_U}{\sigma_U}\right\}\right] - \frac{K}{\sigma_U} n_c\right]. \quad \text{(A.2)}$$

Hence, the maximizing (A.2) is equivalent to maximizing (2), and the standardized parameters are seen to be, as expected, $X = (R - \mu_R)/\sigma_R$, $Y = (U - \mu_U)/\sigma_U$, $c = K/\sigma_U$, $D_2 = (\Delta_2 - \mu_U)/\sigma_U$ and, analogously, $D_1 = (\Delta_1 - \mu_U)/\sigma_U$. The derivation for the Level cutoff objective function is analogous.

Asymptotics for Level Cutoff with High Cost

We consider the optimal strategies for the Quota and Level methods as cost, c, tends towards infinity. If the local extremal condition (19) is to be satisfied as c increases, x_c^* must increase as well; if not, the total cost of evaluation $nc[1 - \Phi(x_c)]$ would dominate expected utility, yielding negative expected net gain. Then, considering [19] as $x_c \to \infty$, because $\Phi[(\rho u - x_c)/\bar{\rho}]$ vanishes more quickly than $\Phi[(u - \rho x_c)/\bar{\rho}]$, the first-order condition reduces to:

$$c = -[\Phi(x_c^*)]^{n-1} \int_{-\infty}^{\infty} u \frac{\partial}{\partial u}(1 - \Phi[(u - \rho x_c^*)/\bar{\rho}])du. \quad \text{(A.3)}$$

The integral represents the mean of a random variable with cdf $\Phi[(u - \rho x_c)/\bar{\rho}]$. Thus, a local extremum for the Level cutoff as $c \to \infty$ is given by

$$c = \rho x_c^*[\Phi(x_c)]^{n-1}. \quad \text{(A.4)}$$

Because $\Phi(x_c) \to 1$, (A.4) supplies first-order asymptotics for high unit evaluation cost; these can be used to validate graphical solutions for the general Level cutoff problem:

$$\begin{aligned} &x_c^* \to c/\rho > 0, \\ &\partial x_c^*/\partial c \to 1/\rho > 0, \\ &\partial x_c^*/\partial n \to (c/\rho)^2 \phi(c/\rho) > 0, \\ &\partial x_c^*/\partial \rho \to -c/\rho^2 < 0. \end{aligned} \quad \text{(A.5)}$$

Proof of (19), The Relationship Between c and x_c^* at Optimum for the Level Cutoff

We have the following first-order condition for local, interior extrema:

$$c = -\frac{1}{n\phi(x_c)} \frac{\partial}{\partial x_c} \int_{-\infty}^{\infty} u \frac{\partial}{\partial u}\left[\Phi(x_c) + \int_{-\infty}^{u} \phi(\nu)\Phi[(\rho\nu - x_c)/\bar{\rho}]d\nu\right]^n du. \quad \text{(A.6)}$$

Once the order of the derivatives is switched (so that the x_c derivative is taken inside the integral), the large bracketed expression can be differentiated with respect to x_c:

$$\frac{\partial}{\partial x_c}\left(\Phi(x_c) + \int_{-\infty}^{u} \phi(\nu)\Phi[(\rho\nu - x_c)/\bar{\rho}]d\nu\right) = \phi(x_c) - \int_{-\infty}^{u} \phi(\nu)\phi[(\rho\nu - x_c)/\bar{\rho}]d\nu/\bar{\rho}. \quad \text{(A.7)}$$

The integrand on the right-hand side can be rearranged using an identity based on the normal pdf:

$$\phi(\nu)\phi[(\rho\nu - x_c)/\bar{\rho}] = \phi(x_c)\phi[(\nu - \rho x_c)/\bar{\rho}]. \quad \text{(A.8)}$$

Substituting the above into the right-hand side of (A.7) allows the integral to be taken:

$$\begin{aligned} &\phi(x_c) - \int_{-\infty}^{u} \phi(x_c)\phi[(\nu - \rho x_c)/\bar{\rho}]d\nu/\bar{\rho} \\ &= \phi(x_c)[1 - \Phi[(u - \rho x_c)/\bar{\rho}]] \\ &= \phi(x_c)\Phi[(\rho x_c - u)/\bar{\rho}]. \end{aligned} \quad \text{(A.9)}$$

Finally, substitution, after cancellation of like terms, yields the relationship between c and x_c^* at optimum:

$$c = -\int_{-\infty}^{\infty} u \frac{\partial}{\partial u}\left[\Phi[(\rho x_c^* - u)/\bar{\rho}]\left(\Phi(x_c^*) + \int_{-\infty}^{u} \phi(\nu)\Phi[(\rho\nu - x_c^*)/\bar{\rho}]d\nu\right)^{n-1}\right]du. \quad \text{(A.10)}$$

References

Bhattacharya, P. K., "Induced Order Statistics: Theory and Applications," in P. R. Krishnaia and P. K. Sen, (Eds.), *Handbook of Statistics*, Volume 4, North Holland, Amsterdam, pp. 383–403, 1984.

David, H. A., *Order Statistics*, Second Edition, John Wiley, New York, 1981.

Gensch, D. H., "A Two-Stage Disaggregate Attribute Choice Model," *Marketing Sci.*, 6, 3 (1987), 223–239.

Gokhale, D. V. and J. Press, "Assessment of a Prior Distribution for a Correlation in a Bivariate Normal Distribution," *J. Royal Statistical Society* (B), 145, 2 (1982), 237–249.

Grether, D. and L. Wilde, "An Analysis of Conjunctive Choice," *J. Consumer Res.*, 10 (March 1984), 373–385.

Hauser, J. R. and B. Wernerfelt, "An Evaluation Cost Model of Consideration Sets," *J. Consumer Res.*, 16 (March 1990), 393–408.

Huber, J. and N. M. Klein, "Adapting Cutoffs to the Choice Environment, The Effects of Attribute Correlations and Reliability," *J. Consumer Res.*, 18 (Dec. 1991), 346–357.

Jennings, D. L., T. M. Amable, and L. Ross, "Informational Covariation Assessments: Data-based vs. Theory-based Judgments," in D. Kahnemann, P. Slovic and A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases*, pp. 211–230, Cambridge University Press, Cambridge, 1982.

Johnson, E. and J. L. Payne, "Effort and Accuracy in Choice," *Management Sci.*, 31 (1985), 395–414.

Klein, N. M. and S. W. Bither, "An Investigation of Utility-Directed Cutoff Selection," *J. Consumer Res.*, 14 (Sept. 1987), 240–256.

Li, L. and D. B. Owen, "Two-Sided Screening Procedures in the Bivariate Case," *Technometrics*, 21, 1 (1979), 79–85.

Lippman, S. A. and R. McCall, "The Economics of Job Search: A Survey (Part I)," *Economic Inquiry*, 14 (1976), 155–189.

—— and K. McCardle, "Uncertain Search: A Model of Search Among Technologies of Uncertain Values," *Management Sci.*, 37 (1991), 1474–1490.

Madsen, R. W., "A Selection Procedure Using a Screening Variable," *Technometrics*, 24, 4 (1982), 301–306.

McQueen, J. B., "Optimal Policies for a Class of Search and Evaluation Problems," *Management Sci.*, 10 (1964), 746–759.

Morgan, P. and R. Manning, "Optimal Search," *Econometrica*, 53, 4 (1985), 923–944.

Moskowitz, H., R. Plante, and H. S. Tsia, "A Multistage Screening Model for Evaluation and Control of Misclassification Error in Detection of Hypertension," *Management Sci.*, 39, 3 (1993), 307–321.

Nagaraja, H. N. and H. A. David, "Distribution of the Maximum of Concomitants of Selected Order Statistics," *The Annals of Statistics*, 22, 1 (March 1994), 478–494.

Newell, A. and H. A. Simon, *Human Problem Solving*, Prentice Hall, Englewood Cliffs, NJ, 1972.

Owen, D. B., L. Li, and Y. M. Chou, "Prediction Intervals for Screening Using a Measured Correlated Variable," *Technometrics*, 23, 2 (1981), 165–170.

Payne, J. L., J. R. Bettman, and E. Johnson, "Adaptive Strategy Selection in Decision Making," *J. of Experimental Psychology: Learning, Memory and Cognition*, 9 (1988), 676–696.

Roberts, J. H. and J. M. Lattin, "Development and Testing of a Model of Consideration Set Composition," *J. Marketing Res.*, 28 (November 1991), 429–440.

Sanbonmatsu, D. M., S. Shavett, and B. D. Gibson, "Salience, Set Size and Illusory Correlation: Making Moderate Assumptions about Extreme Targets," *J. Personality and Social Psychology*, 66, 4 (1994), 1020–1033.

Schilling, E. G., *Acceptance Sampling in Quality Control*, Marcel Dekker, New York, 1982.

Stigler, G. J., "The Economics of Information," *J. Political Economy*, 69 (June 1961), 213–225.

Widing, R. E. and W. W. Talarzyk, "Electronic Information Systems for Consumers: An Evaluation of Computer Assisted Format in Multiple Decision Environments," Working Paper, Case Western Reserve University, Cleveland, OH, 1991.

Wright, P., "Consumer Choice Strategies: Simplifying vs. Optimizing," *J. Marketing Res.*, 12 (February 1975), 60–67.

Yang, S. S., "General Distribution Theory of the Concomitants of Order Statistics," *The Annals of Statistics*, 5, 5 (1977), 996–1002.

Yeo, W. B. and H. A. David, "Selection Through an Associated Characteristic with Applications to the Random Effects Model," *J. American Statistical Association*, 79, 6 (1984), 399–405.

Accepted by Robert T. Clemen; received August 26, 1993. *This paper has been with the authors* $10\frac{1}{2}$ *months for* 2 *revisions.*