View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Institutional Knowledge at Singapore Management University

Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of Business

Lee Kong Chian School of Business

9-2007

Combining predictors to achieve optimal trade-offs between selection quality and adverse impact

Wilfried DE CORTE
Ghent University

Filip LIEVENS
Singapore Management University, filiplievens@smu.edu.sg

Paul R. SACKETT
University of Minnesota - Twin Cities

DOI: https://doi.org/10.1037/0021-9010.92.5.1380

Follow this and additional works at: https://ink.library.smu.edu.sg/lkcsb_research

Part of the Human Resources Management Commons, and the Organizational Behavior and Theory Commons

Citation

DE CORTE, Wilfried; LIEVENS, Filip; and SACKETT, Paul R.. Combining predictors to achieve optimal trade-offs between selection quality and adverse impact. (2007). *Journal of Applied Psychology*. 92, (5), 1380-1393. Research Collection Lee Kong Chian School Of Business.
Available at: https://ink.library.smu.edu.sg/lkcsb_research/5587

This Journal Article is brought to you for free and open access by the Lee Kong Chian School of Business at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Lee Kong Chian School Of Business by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.

Published in Journal of Applied
Psychology, 2007 September, Volume 92,
Issue 5, Pages 1380-1393
http://doi.org/10.1037/0021-9010.92.5.1380

Combining Predictors to Achieve Optimal Trade-Offs Between Selection Quality and Adverse Impact

Wilfried De Corte and Filip Lievens
Ghent University

Paul R. Sackett
University of Minnesota, Twin Cities Campus

The authors propose a procedure to determine (a) predictor composites that result in a Pareto-optimal trade-off between the often competing goals in personnel selection of quality and adverse impact and (b) the relative importance of the quality and impact objectives that correspond to each of these trade-offs. They also investigated whether the obtained Pareto-optimal composites continue to perform well under variability of the selection parameters that characterize the intended selection decision. The results of this investigation indicate that this is indeed the case. The authors suggest that the procedure be used as one of a number of potential strategies for addressing the quality–adverse impact problem in settings where estimates of the selection parameters (e.g., validity estimates, predictor intercorrelations, subgroup mean differences on the predictors and criteria) are available from either a local validation study or meta-analytic research.

Keywords: personnel selection, adverse impact, optimal trade-off, selection quality, predictor composites

In a recent article, Campion et al. (2001) observed that the quandary posed by the often competing goals of valid selection and a diverse workforce "may constitute the most perplexing problem facing the practice of personnel selection today" (p. 158). This quandary, henceforth referred to as the *selection quality–adverse impact problem*, emerges because many valid preemployment tests (e.g., cognitive ability tests) show substantial effect sizes (i.e, standardized mean differences) by race or ethnicity, resulting in selection rates that vary substantially for applicant groups that differ in terms of these characteristics (e.g., Sackett, Schmitt, Ellingson, & Kabin, 2001).

To address this quandary, researchers have considered a number of alternatives, including banding, adaptation of the presentation format or the content of tests, modification of the test taker's attitudes, and within-group norming (cf. Sackett et al., 2001; Sackett & Wilk, 1994). Another option is to use a composite of selection predictors that have different effect sizes to obtain a better trade-off between the goals of selection quality and adverse impact than would otherwise be the case. This strategy addresses the common scenario in which an employer plans to use a composite of a set of predictors (e.g., cognitive tests, personality tests, interviews, work samples). In this scenario, selection practitioners know how to maximize the mean criterion score of selected applicants, namely, by inputting all predictors into a regression equation and using the resulting weights. But often the employer asks, "Is there an alternative way of using the predictors that comes close to this optimal solution in terms of the level of criterion performance achieved but does so with less adverse impact?" At this point, selection practitioners do not know how to respond to such a request other than by trial and error with various predictor weights (Hattrup, Rock, & Scalia, 1997; Pulakos & Schmitt, 1996; Sackett & Ellingson, 1997; Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997). In fact, this is what one also commonly sees in technical reports: an examination of a series of alternative models that use varying combinations of available predictors, weighted in differing ways. The current article provides an elegant solution to this problem. Note that in the United States, the legal system calls for an investigation of whether there are alternatives with substantially equal validity but less adverse impact. We view combining available predictors in differing ways as an example of such "alternatives." Note that the question of "substantially equal" is a value judgment: Some may say this means "no more than a 1% reduction from the utility gain from the performance maximizing weights," others might accept a 5% reduction, and still others a 10% reduction. The method described in this article permits one to directly answer the question of how much of an improvement in minority hiring can be achieved within whatever constraint (1%, 5%, 10%, and so forth, reduction in the criterion) the researcher finds acceptable.

Prior Research on Predictor Composite Alternatives

Over the last 10 years, supplementing valid, cognitive predictors with noncognitive predictors has emerged as a potentially useful strategy for addressing the selection quality–adverse impact trade-off. If noncognitive predictors are relevant to the job of interest, these predictors might increase the validity coefficients obtained. In addition, if subgroup differences for these noncognitive tests are smaller, a composite of a traditional cognitive test with these noncognitive tests will often lead to smaller subgroup differences than use of the traditional cognitive test alone (Sackett et al., 2001).

Wilfried De Corte, Department of Data Analysis, Ghent University, Ghent, Belgium; Filip Lievens, Department of Personnel Management and Work and Organizational Psychology, Ghent University; Paul R. Sackett, Department of Psychology, University of Minnesota, Twin Cities Campus.

Correspondence concerning this article should be addressed to Wilfried De Corte, Department of Data Analysis, Ghent University, H. Dunantlaan 1, 9000, Ghent, Belgium. E-mail: wilfried.decorte@ugent.be

Various studies have examined the effectiveness of using composite alternatives. For instance, Pulakos and Schmitt (1996) compared a traditional verbal ability measure with three alternative predictors: biodata, a situational judgment test, and a structured interview. Use of the verbal ability test alone resulted in a standardized effect size, *d*, of 1.03, whereas use of a composite of the four measures produced an effect size of 0.63, and a composite of the three alternative predictors produced an effect size of 0.23 (see also Ryan, Ployhart, & Friedel, 1998). Other studies have explored more generally the effects of multiple selection parameters on adverse impact. For example, Schmitt et al. (1997) examined the effect of the number of predictors, predictor intercorrelations, validities, and level of predictor subgroup differences on adverse impact associated with the predictor composite (see also Doverspike, Winter, Healy, & Barrett, 1996). Sackett and Ellingson (1997) developed a set of helpful implications associated with estimating the effect of predictor composites on adverse impact. Hattrup et al. (1997) investigated the effect of varying the importance of different criterion dimensions of job performance on adverse impact.

These previous studies share several limitations. First, prior research has used a trial-and-error strategy for determining various predictor weights to find a composite alternative that comes closest to meeting the two objectives. Second, in most studies, including the recent study by Potosky, Bobko, and Roth (2005), the weights of the predictors have been determined by regressing the composite performance criterion on the predictors. Although this regression-based approach maximizes validity, it does not answer the question of whether there is an alternative way of using the predictors that comes close to the regression-based weighting in terms of predictive efficiency but does so with less adverse impact. Only De Corte (1999) showed how nonlinear programming can be used to combine job performance predictors into a predictor composite such that the average quality of the composite-selected employees is maximized and the adverse impact ratio remains within acceptable bounds. However, an important limitation of De Corte (1999) was that only one of the goals (i.e., selection quality) was optimized, whereas the other objective (adverse impact) was dealt with as a constraint. In other words, De Corte (1999) did not address how available predictors should best be combined (i.e., weighed) to achieve certain desired levels of both quality and adverse impact. Yet, in cases in which both these goals are valued, this seems to be the primary issue that selection practitioners are confronted with when they consider the alternative of using predictor composites to address the selection quality–adverse impact problem in a particular application (Outtz, 2002).

To overcome the limitations of previous research, we developed a general procedure for determining values for the predictor weights such that the resulting predictor composites provide an optimal balance or trade-off between the quality and adverse impact objectives. The procedure applies optimization methods from the field of operations research and results in the determination of so-called Pareto-optimal predictor composites. The next section presents the procedure. First, we provide a basic introduction to single- and multi-objective optimization and indicate the relevance of the approach for the purpose at hand. Next, we clarify the notion of Pareto-optimal predictor composites and describe the application context and the details of the optimization method used to obtain these composites.

Procedure for Obtaining Pareto-Optimal Predictor Composites

Single- and Multi-Objective Optimization

Optimization is one of the primary tools used in *operations research*, which is the interdisciplinary science of the usage of quantitative methods to assist decision makers in designing, analyzing, and improving the performance of scientific, financial, organizational, or engineering systems. Basically, optimization focuses on the minimization or maximization, subject to certain constraints, of an objective. The objective as well as the constraints are typically functions of certain variables that are usually referred to as *decision variables*.

As an example, suppose that a farmer has a piece of land, 210 acres (849,843.33 m^2) large, to be planted by either Crop 1 or Crop 2 or some combination of both. The farmer has limited permissible amounts of fertilizer (i.e., 20,000 lb; 9,071.9 kg) and pesticide (i.e., 550 lb; 249.5 kg) that can be used, each of which is required in different amounts per unit area for Crop 1 and Crop 2. Crop 1 requires 125 lb (56.7 kg) of fertilizer and 2 lb (0.9 kg) of pesticide per acre, whereas Crop 2 requires 70 lb (31.8 kg) of fertilizer and 3 lb (1.4 kg) of pesticide per acre. Let the selling price for Crop 1 and Crop 2 be \$625 per acre and \$495 per acre, respectively, and if the area planted with Crop 1 and Crop 2 is denoted as x_1 and x_2, respectively, the objective of the optimization can be expressed as

$$\text{maximize } 625x_1 + 495x_2 \text{ (i.e., maximize the revenue).}$$

The decision variables are the areas x_1 and x_2 to be planted with Crops 1 and 2, whereas the constraints of the optimization problem are

$$x_1 + x_2 \leq 210 \text{ (limit on the total area to be planted)}$$

$$125x_1 + 70x_2 \leq 20{,}000 \text{ (limit on the fertilizer)}$$

$$2x_1 + 3x_2 \leq 550 \text{ (limit on the pesticide)}$$

$$x_1 \geq 0 \text{ and } x_2 \geq 0 \text{ (no negative areas).}$$

The example illustrates a so-called *linear* programming (or optimization) problem because the objective and the contraints are all linear functions of the decision variables. The solution of the program shows values of 96.4 and 113.6 for x_1 and x_2, respectively, meaning that the farmer should plant 96.4 acres (390,118.6 m^2) with Crop 1 and 113.6 acres (459,724.8 m^2) with Crop 2. Next, looking at the values of the constraints for the solution, it can be observed that the first two restrictions are met exactly (i.e., the value of the constraint is equal to the upper limit value at the solution), whereas the final constraint is not. Thus, $96.4 + 113.6 = 210$ and $125 \times 96.4 + 70 \times 113.6 = 20{,}000$ (within rounding precision), whereas $2 \times 96.4 + 3 \times 113.6 = 533.6 < 550$. Constraints that are met exactly (which will always be the case for equality constraints) are usually referred to as *binding constraints*.

When either the objective of an optimization problem or any one of its constraints is nonlinear in the decision variables, the problem is called a *nonlinear* programming problem. Another distinction is between constrained (i.e., at least one constraint is present) and unconstrained programs. Still further distinctions are made, and we refer the reader to, for example, Nash and Sofer (1996) for a more

detailed description of these problem types and the different techniques used to solve them.

Because only one objective must be optimized in the example just given, the problem is a single-objective programming problem. Such problems are usually characterized by a single optimum solution (i.e., a single optimum value for the decision variables). This is no longer the case when more than one objective is of interest, however. The presence of multiple objectives implies that in such problems there is rarely a single solution that is optimal according to every objective. Rather, of the possible solutions that produce a given level of one objective (e.g., a given level of mean performance achieved via selection), there is one that produces a higher value on a second objective (e.g., rate of minority selection) than any other. That solution is referred to as *Pareto optimal*; other terms used in the literature for this are *efficient*, *nondominated*, and *noninferior*. There is a Pareto-optimal solution for every attainable level of each outcome. For example, for any attainable level of mean performance, there is a Pareto-optimal solution, namely, the one weighting of the predictors out of many possible that would produce the highest rate of minority selection for that level of mean performance.

Relevance of the Optimization Approach

The principal issue addressed in this article corresponds to a multi-objective programming (MOP) problem with two objectives. The first objective is to maximize the selection quality, and the second is to minimize the adverse impact. Also, both objectives are typically in conflict, as many valid selection predictors show substantial effect sizes. Furthermore, with a given set of predictor variables, both objectives are a function of the same decision variables. More specifically, both objectives depend on (a) the weights with which the predictors are combined in the predictor composite and (b) the cutoff value used for the composite.

Methods to solve MOP problems are usually classified in terms of when they elicit preference information from the decision maker in order to choose between the Pareto-optimal solutions (Marler & Arora, 2004). Preferences may be elicited before the method is applied (*a priori articulation methods*), during the application (*interactive methods*), or after the method has been applied (*a posteriori articulation methods*), or no preference information may be used at all. The most popular a priori articulation method is the weighted sum method in which the different objectives are assigned weights and then summed to a single objective. Goal programming, where goals (i.e., specific values for each objective) are specified for each objective, is another a priori approach.

None of the a priori approaches is well suited to address the selection quality–adverse impact MOP problem, however. The main reason for this (and a reason that also applies to the interactive and the no-preference methods, which are essentially basic variants of the a priori methods) is that their application requires a number of difficult ad hoc decisions, such as, for example, the choice of the relative weight or the goal value to be assigned to the objectives. For one to make these decisions effectively, it is helpful to first have a clear picture of the palette of possible Pareto-optimal solutions that can be achieved for a given set of selection predictors. A posteriori methods focus explicitly on such a picture and are therefore considered hereafter to study the selection quality–adverse impact MOP problem. Before discussing the actual implementation of the chosen a posteriori method, we first elaborate the notion of Pareto-optimal predictor composites.

Definition of Pareto-Optimal Composites and Trade-Offs

A predictor composite is called *Pareto optimal* when it results in a Pareto-optimal trade-off between the goals of quality and adverse impact, given the details of the intended selection scenario (e.g., selection ratio) and the available selection predictors. There is a Pareto-optimal trade-off for any given level of selection quality. For example, there may be multiple combinations of predictors and predictor weights that would produce a mean standardized performance level of 1.0 among those selected. However, of those multiple solutions, the Pareto-optimal one would be the one that maximizes the rate of minority selection (i.e., maximizes the adverse impact ratio). Similarly, there may be many combinations of predictors and predictor weights that would result in a given level of the adverse impact ratio (e.g., .80). The Pareto-optimal one would be the one that maximizes selection quality. More generally, a trade-off between selection quality and adverse impact is called *Pareto optimal* when any weighing scheme for the predictors that differs from the one that is associated with the optimal trade-off results in either a decrease in the selection quality or an increase in the level of adverse impact. The entire set of Pareto-optimal trade-offs is referred to as the *Pareto-optimal trade-off function* or the *Pareto surface* (Keeney & Raiffa, 1993). The present method is aimed at a representative collection of Pareto points (i.e., of Pareto-optimal trade-offs) where each point corresponds to a specific weighing of the predictors and is characterized by an optimal pair of values for the selection quality and the adverse impact objective.

Application Context

The new procedure for determining Pareto-optimal predictor composites is indicated whenever the goals of selection quality and diversity are in conflict. This is hardly limiting, however, because these goals are in conflict whenever they cannot both be optimized by exactly the same weighting of the available selection predictors. Thus, even in the case in which the most valid predictors would have the smallest effect sizes, the goals of quality and diversity will most likely still be in conflict because the predictor weights that maximize, for example, the composite validity (i.e., the regression weights) will almost surely differ from the weights that minimize the composite effect size.

There is also little or no restriction on the variety of planned, "fixed applicant pool" (cf. *Standards for Educational and Psychological Testing*, American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1999, p. 152) selection scenarios that can be handled by the method. Both probationary and nonprobationary selections, as well as situations in which the applicants come from several different minority populations, can be addressed. Thus, the method can cope, for example, with decisions in which the applicant pool comprises White, Black, and Hispanic candidates and the latter two applicant groups are treated as separate minority groups. The method can also deal with selection scenarios where candidates who are successful on an initial screen are admitted to a job training program and where it is later decided on the basis of the

training performance which of these initially selected candidates will be retained.

For reasons of simplicity, the following description of the procedure focuses on nonprobationary scenarios involving only one minority group. Also, we assume that representative estimates are available for the selection parameters (i.e., the selection rate and the proportional representation of the minority and the majority candidates in the applicant pool, as well as for the effect size, validity, and intercorrelations of the available predictors) that characterize the selection scenario and the available predictors. These estimates may be derived either from a local past or current validation study or from the findings reported in the constantly growing number of meta-analytic studies on the characteristics of selection predictors and their relationship to the most important performance dimensions (e.g., Bobko, Roth, & Potosky, 1999; Hough, Oswald, & Ployhart, 2001; McKay & McDaniel, 2006; Potosky, Bobko, & Roth, 2005; Salgado, Anderson, Moscoso, Bertua, & De Fruyt, 2003; Schmidt & Hunter, 1998; Schmitt, Clause, & Pulakos, 1996). In a later section we address the issue of uncertainty in estimation of selection parameters.

In addition to representative estimates for the selection parameters, the present procedure also requires an assumption about the distribution of the predictor scores in the majority and the minority population. In particular, it is assumed that the predictors and the criterion dimension(s) have a joint multivariate normal distribution with the same variance–covariance matrix but differing means in the majority and the minority applicant populations. The assumption is also invoked by De Corte (1999), and it is essentially equivalent to the one that underlies previous study results on the effect of different predictor combinations on minority hiring and adverse impact (e.g., Hattrup et al., 1997; Schmitt et al., 1997). As shown in Appendix A, this assumption permits expressing the selection outcomes of adverse impact (as indicated by the adverse impact ratio) and quality (see later discussion) in terms of the selection parameter data, on the one hand, and the values of the decision variables, on the other hand. As noted earlier, these decision variables correspond for the present selection scenarios to the weights assigned to the predictors in forming the predictor composite and to the cutoff score that must be applied to the predictor composite to ensure that the intended selection rate is achieved.

With the aforementioned specifications, the problem addressed by the present method can be summarized as follows: Given the selection parameter data that characterize the selection scenario, find values for the decision variables such that the resulting values of the quality and impact objectives represent a Pareto-optimal trade-off.

Method

As noted above, the problem addressed by our procedure is a typical example of a multi-objective optimization problem. There it was also argued that our problem is best handled by an a posteriori method that has the aim of represent the entire set of Pareto-optimal trade-offs. Among these methods (see, e.g., Marler & Arora, 2004, for an overview), we propose to use the technique of normal-boundary intersection, developed by Das and Dennis (1998), because it enjoys a number of advantages that are particularly relevant in the present context. First, the technique generates Pareto-optimal trade-offs between adverse impact and selection quality that are uniformly spread over the entire Pareto surface. As a consequence, a representative subset of all Pareto-optimal composite predictors is obtained as well. Second, for each Pareto-optimal trade-off, it is possible to determine the associated relative importance attached to the two selection objectives when the trade-off function is convex as is usually the case. Finally, the generated set of Pareto points is independent of the scaling of the individual selection objectives. This feature is of particular relevance when the selection quality objective is defined in terms of the utility of the selection because this utility can be expressed in different monetary or nonmonetary units.

The implementation of the normal-boundary intersection technique consists of two stages. In the first stage, two constrained nonlinear programming problems (cf. Appendix B) are solved to obtain the predictor weighing schemes, $\mathbf{b}_q$ and $\mathbf{b}_a$, that result in the maximum possible value for the selection quality objective (denoted as q_{max}) and the maximum possible value of the adverse impact objective (denoted as a_{max}), respectively. These weighing schemes characterize two optimal trade-off points, henceforth referred to as $\mathbf{t}_q$ and $\mathbf{t}_a$, respectively. The first optimal trade-off, $\mathbf{t}_q = (q_{max}, a_q)'$ (with a_q the value of the adverse impact objective associated with the weighing $\mathbf{b}_q$ of the predictors), corresponds to the situation in which only the quality objective is judged to be of importance. In turn, the second optimal trade-off, $\mathbf{t}_a = (q_a, a_{max})'$ (with q_a the value of the selection quality objective associated with the weighing $\mathbf{b}_a$ of the predictors), represents the situation in which only the adverse impact objective is of concern.

As detailed in Appendix B, the optimal trade-offs $\mathbf{t}_q$ and $\mathbf{t}_a$ permit the determination of a payoff matrix $\mathbf{\Phi}$ that is subsequently used in the formulation of the nonlinear programs that are solved in the second stage of the normal-boundary intersection method. These second-stage nonlinear programs result in new optimal trade-offs, each of which corresponds to a particular valuation (expressed in terms of relative importance, see later discussion) of the two selection objectives. Together with the earlier obtained trade-offs $\mathbf{t}_q$ and $\mathbf{t}_a$, these new trade-off points provide an evenly spaced and, therefore, representative characterization of the entire Pareto-optimal trade-off curve.

Several approaches may be adopted to solve the nonlinear programming problems in the two stages of the normal-boundary intersection method. Bazaraa, Sherali, and Shetty (2006) provided a detailed discussion of these approaches, observing that all successful approaches require that the values of the objective function (i.e., the quantity that is to be optimized) and of the eventual constraints of the nonlinear program can be computed analytically (instead of through simulation) from the values of the decision variables. As noted earlier and further detailed in Appendix A, this requirement is met in the present context. Thus, the solution of the different nonlinear programs poses no particular problems.

Implementation

We wrote a computer program to implement the normal-boundary intersection method. The program runs on a personal computer under the Windows 95/98, NT, XP, and 2000 Professional operating systems. To execute the program, the user must prepare only a single input file that details the nature of the studied selection decision and summarizes the characteristics of the pre-

dictor and the criterion variables. The executable code and a manual that describes the preparation of the input file and the actual usage of the program can be downloaded from the Internet at http://users.ugent.be/~wdecorte/software.html. The documentation also contains an example application and provides further details on the output generated and the way in which this output can be further processed by means of freely available graphical software (e.g., the graphical procedures of the R programming environment available at http://lib.stat.cmu.edu/R/CRAN).

To enhance the applicability of the program and to ensure its relevance, we added several control options. A first option permits the user to choose the way in which the selection quality objective is operationalized. Whereas the adverse impact ratio (i.e., the ratio between the selection rate in the minority applicant group and that in the majority group) always represents the adverse impact objective, the selection quality objective can be expressed in terms of either the validity of the predictor composite, the average criterion score of the selected applicants, or the utility of the selection (see Appendix A).

Still other program options can be used (a) to determine the number of computed trade-off points; (b) to specify whether a nonprobationary or a probationary selection decision is intended, in which case the composite criterion cutoff can eventually be fixed at a given value that expresses the current standard of acceptable job performance; (d) to constrain the predictor weights; and (e) to bound from below and above, or to fix the proportion of initially hired employees for probationary selections. The option to constrain the predictor weights provides the means to prevent some of the solution weights from being too small (large) or even having negative values. Such negative weights are usually unacceptable because they lead to a composite predictor in which valued job-related attributes are counted against the applicants.

Example Application

The example application relates to a situation in which four predictors are available to select a given proportion of selectees from a heterogeneous applicant group where 75% of the candidates are members of the majority population (i.e., Whites) and the remaining 25% belong to the Black minority population (cf. Hattrup et al., 1997). The four predictors are cognitive ability (CA), structured interview (SI), conscientiousness (CO), and biodata (BI). The overall performance criterion is a weighted sum of the two dimensions, with weights 3 and 1 for the task performance (TP) and the contextual performance (CP) dimensions, respectively. The latter weight values were chosen to conform to those used in other related studies (cf. Hattrup et al., 1997). In general, the specification of the weights of the criterion dimensions is essentially a matter of organizational policy (cf. Murphy & Shiarella, 1997).

Table 1 summarizes the values used for the different characteristics of both the predictors and the criterion dimensions. The predictor effect size (i.e., Black–White mean difference), validity (with respect to TP), and intercorrelation data were borrowed from the meta-analytic study of Bobko et al. (1999), and the validity values for the CP dimension correspond to results presented by Hattrup et al. (1997) and McManus and Kelly (1999). The Hattrup et al. study provided also the value of the correlation between the two criterion dimensions, whereas the criterion effect size values were copied from McKay and McDaniel (2006). All used values correspond to uncorrected estimates to "consider the operational use of potential sets of predictors" (Bobko et al., 1999, p. 563).

Table 1
Effect Sizes and Intercorrelations Between the Performance Predictors and the Performance Criteria

Variable	*d*	Intercorrelation matrix 1	2	3	4	5	6
Predictors							
1. Cognitive ability	1.00	—					
2. Structured interview	0.23	.24	—				
3. Conscientiousness	0.09	.00	.12	—			
4. Biodata	0.33	.19	.16	.51	—		
Criterion dimensions							
5. Task (job) performance	0.21	.30	.30	.18	.28	—	
6. Contextual performance	0.13	.16	.26	.20	.25	.17	—

Although the example first and foremost focuses on the merits of the method described earlier, we preferred to use predictor and criteria data that reflect the results of previous summary studies to ensure a representative application of the procedure. We emphasize, though, that this is simply an example of how the methods developed here can be used. That a reader may question a particular chosen value (e.g., the correlation between two specific predictors) is not an impediment to the article's fundamental goal of developing and illustrating the use of these decision-making techniques in the personnel selection context.

Results

Given the aforementioned values for the selection parameters, the present procedure was used to analyze the selection quality–adverse impact trade-off surface for an intended nonprobationary selection system with a selection rate of 15%. In this illustration, selection quality is indexed by the mean criterion score obtained by those selected. Table 2 and Figure 1 summarize the obtained results.

The solid line in the left part of Figure 1 displays the entire set of Pareto-optimal trade-offs for the quality and adverse impact objectives, whereas the four lines in the right part of the figure detail the predictor weighing schemes that correspond to the different optimal trade-offs. Alternatively, Table 2 provides details on a selected number of these optimal trade-offs. The selected trade-offs are indicated by a bullet symbol in Figure 1, and for each of these trade-offs Table 2 indicates the value of the selection quality (i.e., expected criterion score of a selected applicant) and the adverse impact ratio objective, as well as the value of the four predictor weights (scaled to have unit sum) that correspond to the optimal trade-off. Thus, using Figure 1 and Table 2 it can be verified that, for example, Optimal Trade-Off Point 5 shows values 0.48 and 0.74 for the adverse impact and the expected criterion score objective, respectively, and that the optimal trade-off is attained when using weights 0.09, 0.45, 0.15, and 0.31 for the CA, the SI, the CO, and the BI predictor, respectively. Also, because the point represents an optimal trade-off, no predictor composite can do at least as well on either one of the objectives and, at the same time, do better on the other objective. Thus, for this selection decision, it is not possible to combine the four predictors into a

Table 2
Selected Pareto-Optimal Selection Quality–Adverse Impact Trade-Offs for a Selection From a Black and White Applicant Group

Optimal trade-off	Adverse impact	Expected criterion score	Predictor weight			
			CA	SI	CO	BI
1	.30	.78	.28	.34	.12	.26
2	.35	.77	.21	.38	.13	.28
3	.39	.76	.16	.41	.14	.29
4	.43	.75	.12	.43	.15	.30
5	.48	.74	.09	.45	.15	.31
6	.52	.73	.05	.47	.16	.32
7	.55	.71	.02	.49	.16	.32
8	.59	.70	.00	.51	.20	.29
9	.62	.68	.00	.53	.28	.19
10	.65	.66	.00	.54	.35	.12
11	.68	.63	.00	.55	.40	.05
12	.70	.61	.00	.54	.46	.00
13	.72	.58	.00	.44	.56	.00
14	.74	.56	.00	.38	.62	.00
15	.76	.53	.00	.32	.68	.00
16	.78	.50	.00	.26	.74	.00
17	.80	.47	.00	.21	.79	.00
18	.82	.44	.00	.16	.84	.00
19	.83	.41	.00	.11	.89	.00
20	.85	.38	.00	.06	.94	.00
21	.87	.35	.00	.00	1.00	.00

Note. CA = cognitive ability; SI = structured interview; CO = conscientiousness; BI = biodata.

composite such that, for example, the expected criterion score of the selected applicants is at least equal to 0.74 while the adverse impact ratio exceeds .48.

Apart from the quality and adverse impact ratio values that characterize each optimal trade-off point, Figure 2 also portrays two additional features of these optimal trade-offs. First, the dashed line on the graph provides the details of the worst possible trade-off that corresponds to each of the optimal trade-offs. Second, the dot-dashed curve of the graph represents the relative importance of the adverse impact objective (as compared with the relative importance of the quality objective) that corresponds to the different optimal trade-off points. In fact, each Pareto-optimal trade-off between selection quality and adverse impact is characterized by a different weighting of these two objectives. Thus, attaching a weight of, for example, 3 and 1 to the quality and adverse impact objectives will lead to a different optimal trade-off than attaching equal weights to both objectives. For each optimal trade-off, these weights can be scaled to have unit sum in order to express the relative importance attached to the two objectives.

To illustrate these worst possible and importance curves, reconsider Pareto-Optimal Trade-Off Point 5. For a value of 0.48 for the adverse impact ratio objective, this optimal trade-off shows a best possible quality value of 0.74. The corresponding symbol on the dashed curve of worst possible trade-offs indicates that for the same adverse impact ratio value the worst possible weighting of the selection predictors will lead to a value of only 0.52 for the selection quality objective. Next, from the position of the matching bullet symbol on the dot-dashed line with respect to the right vertical axis (labeled "Relative importance AI [adverse impact] objective"), it can be verified that the optimal trade-off with values 0.48 and 0.74 for the adverse impact and the quality objectives corresponds to the situation in which the relative importance attached to the objectives equals 0.24 and 0.76, respectively.

To compare the results of the present procedure with those obtained by De Corte (1999), as well as to further illustrate the interpretation of the optimal and worst possible trade-off curves and the importance curve, we look more closely at the trade-offs, numbered 1, 17, and 21 in Figure 2. As indicated by the importance curve, Point 1 represents the optimal trade-off in the case in which the importance attached to the adverse impact objective is equal to zero (no importance), whereas Point 21 corresponds to the optimal trade-off when the adverse impact objective receives exclusive importance. From the graph and Table 2 it can further be seen that Trade-Off 1 is characterized by a mean criterion score of the successfully selected applicants of 0.78 and an associated adverse impact ratio of 0.30; whereas Trade-Off 21 has corresponding values of 0.35 and 0.87, respectively. As expected, Optimal Trade-Off Point 21, which attaches maximum importance to the adverse impact objective, is associated with a predictor weighing scheme in which all but the predictor with the smallest effect size (i.e., predictor CO) receive a zero weight in the composite selection predictor (cf. the last row of Table 2). Alternatively, as all predictors showed an effect size in favor of the majority group, the optimal trade-off when maximum importance is attached to the quality objective (i.e., Trade-Off 1) corresponds to a regression-based weighting of the predictors (cf. the first row of Table 2).

Next, consider Point 17 of the optimal trade-off surface curve. This trade-off is characterized by an average criterion score of 0.47 and a value of 0.80 for the adverse impact ratio. Observe that this trade-off point is in fact identical to the solution of the selection quality–adverse impact problem as proposed by De Corte (1999) because the latter solution corresponds to the maximum possible selection quality under the binding constraint that the adverse impact ratio be equal to 0.80 (i.e., meet the 80% rule). Thus, compared with the method proposed by De Corte (1999), the present procedure does not merely result in one particular point of the quality–adverse impact optimal trade-off surface but instead approximates the entire set of such optimal trade-off points. In addition, for each generated optimal trade-off, it is possible to determine the corresponding importance attached to the objectives as well as the associated predictor weights. Thus, from the relative importance curve on the graph it can be verified that the present optimal trade-off with a value of .80 for the adverse impact objective corresponds to the assignment of relative importance values of .62 and .38 to the adverse impact and the quality objective, respectively.

Finally, we illustrate how this technique can be used to answer the question posed at the beginning of this article, namely, "Is there a different weighting of predictors that will come close (i.e., within a specified distance) to the maximum mean quality attainable but with less adverse impact?" To address this question, one must specify one's definition of *close*; once a given decision maker defines it (e.g., "anything within 95% of the maximum mean quality attainable"), then Figure 1 permits this question to be answered. As noted earlier, the maximum mean quality attainable with these predictors at this selection ratio is 0.78. Thus, we can move down the optimal trade-off curve to the point where mean quality is 0.74 (i.e., 95% of 0.78); we find that the Pareto-optimal

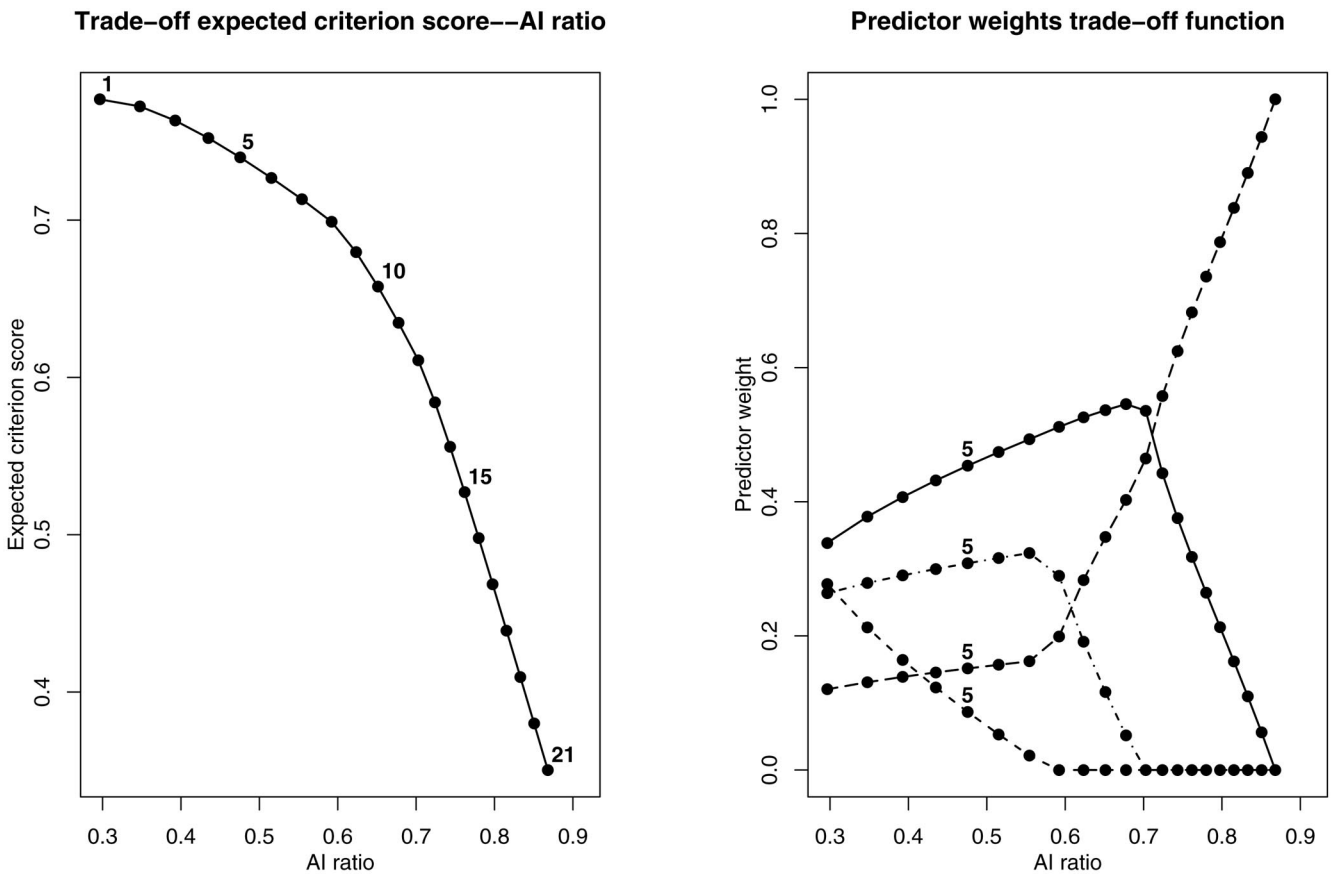

Figure 1. Left panel: Pareto-optimal trade-off curve, with the selected Pareto-optimal trade-offs detailed in Table 2, for a nonprobationary selection with a .15 selection rate. Right panel: Predictor weighing systems that correspond to the Pareto-optimal trade-offs (cf. Table 2; right). Short-dashed line = Predictor 1 weight function; solid line = Predictor 2 weight function; long-dashed line = Predictor 3 weight function; dot-dashed line = Predictor 4 weight function. AI = adverse impact.

weighting of predictors at this point produces an adverse impact ratio of 0.48, compared with the value of 0.30 for the weighting that maximizes quality. And we can turn to the right side of the figure and see the predictor weights that would be used to obtain this result. Also, using the next-to-last equation of Appendix A, the gain in terms of the adverse impact objective can be expressed in terms of the percentage of increase in minority hiring as compared with the minority hiring under the maximum quality condition. It is then obtained that the gain in adverse impact from 0.30 to 0.48 corresponds to a 51.4% improvement in minority hiring. Thus, the results reveal that accepting a 5% reduction in the selection quality can increase the minority hiring rate by more than 50%.

Discussion

From the optimal and the worst possible trade-off curves depicted in Figure 2, it can be verified that the conflict between the goals of attaining a high selection quality and a high adverse impact ratio is not a simple one. In all but one case (i.e., Trade-Off Point 21), a given target value for one of the objectives corresponds to a broad range of possible values for the other objective. Yet, only one of the latter values is Pareto optimal, whereas the others correspond to predictor weighing schemes that can be bettered. As an example, consider Optimal Trade-Off Point 5 and the corresponding worst possible trade-off marked with the bullet symbol on the worst possible trade-off line. As indicated by the dotted vertical line that connects these two trade-offs, there are many other nonoptimal trade-offs, all with the same value for the adverse impact objective but with a smaller value for the selection quality than that associated with the corresponding optimal trade-off. In the language of decision theory, these alternatives are "dominated" by the Pareto-optimal trade-off. Also, the quality difference between the optimal trade-off and the corresponding worst possible trade-off is no less than 0.74 – 0.52 = 0.22 standard units. This substantial difference (as compared with the maximum possible quality difference of 0.78 – 0.35 = 0.43 between any two predictor weighing systems) clearly emphasizes the importance of choosing appropriate predictor weights if the selection practitioner wants to achieve a trade-off with a given adverse impact value.

From the horizontal dotted line on the same plot (i.e., the dotted line that ends at Optimal Trade-Off Point 10), a similar conclusion can be drawn in the case in which the practitioner desires a given value for the quality objective: To achieve an average criterion score

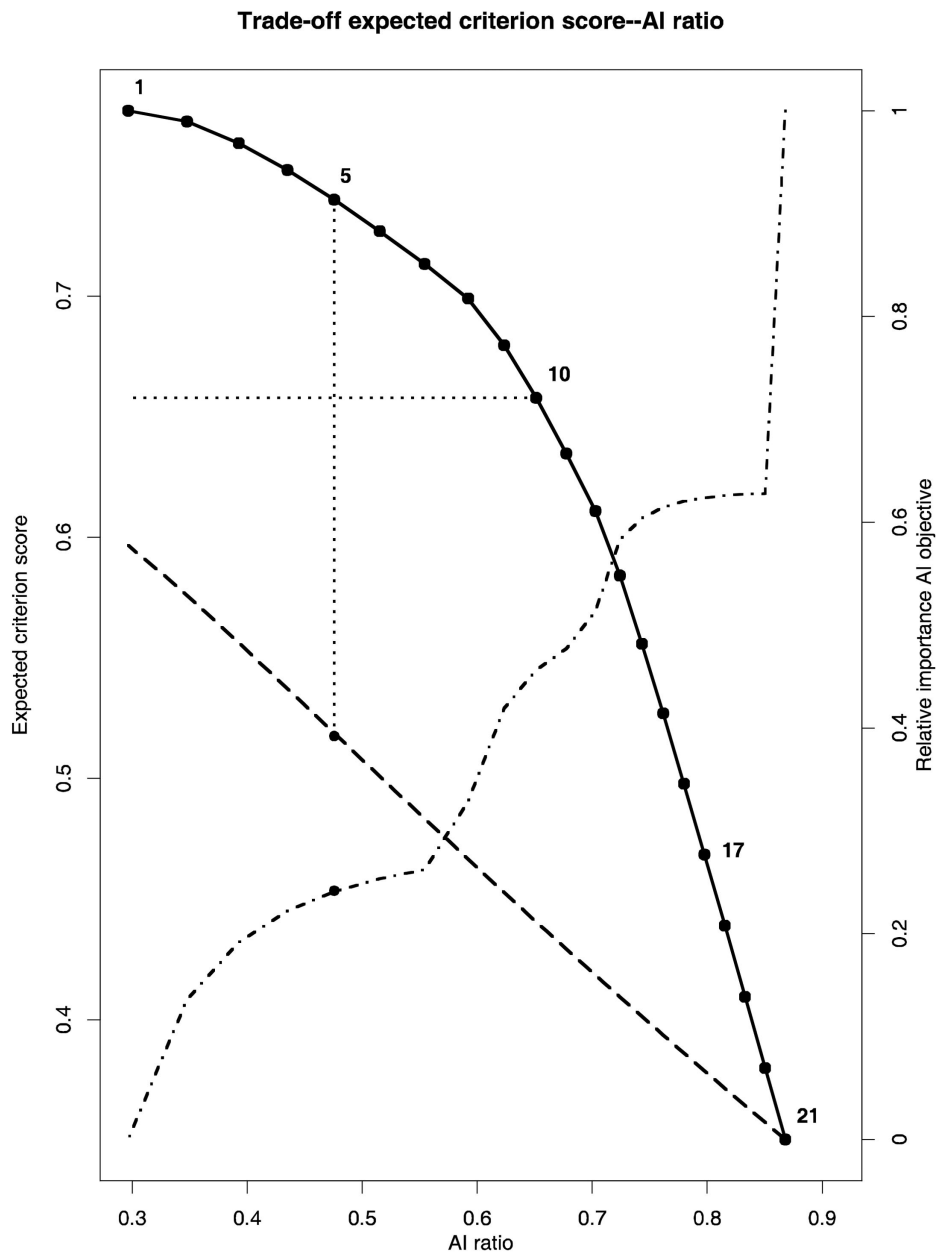

Figure 2. Pareto-optimal trade-off curve (solid line), worst possible trade-off curve (dashed line), and relative importance of the adverse impact (AI) objective (dot-dashed line) for a nonprobationary selection with a .15 selection rate.

of .66, again many weighing schemes are possible, with the worst weighing resulting in an adverse impact value of 0.30, whereas the best weighing corresponds to a value of .65 for this objective. Thus, although the adverse impact and the quality objectives are typically in conflict, this does not at all mean that they cannot be balanced.

As a final observation, note that the present method does not tell the decision maker which combination of predictor weights should be used. Rather, it presents the decision maker with the entire set of Pareto-optimal trade-offs. This feature should not be regarded as a drawback, however, because it simply reflects the reality that the resolution of competing goals typically requires a decision as to the relative importance of these objectives. Although the latter decision cannot be resolved analytically, it is no doubt helpful that the method allows the decision maker to visualize the relationship between the relative importance attached to the two objectives and the resulting optimal trade-off.

Robustness of Pareto-Optimal Predictor Weighing Schemes

As is true for any proposal to study the consequences of a selection decision, the present method derives results that depend

on the particular values used for the selection rate, the selection predictor effect sizes, validities, and so on. Although this dependency is of no consequence when analyzing hypothetical selection scenarios, it becomes a major concern when the method is used as a decision tool when planning a future selection system because in that case the values of the selection parameters will typically not be known with certainty. It is therefore important to assess whether Pareto-optimal predictor weighing schemes, as computed from a given set of parameter values, are robust in the sense that they also perform well when the selection parameters have somewhat different values.

Monte Carlo Method

To settle the robustness issue, we perform sensitivity analysis using Monte Carlo simulation methods. These methods are often used for analyzing uncertainty (sensitivity), where the goal is to determine how variation, lack of knowledge, or error in certain input parameters affects the performance or adequacy of the outcome quantities that are of concern (Saltelli, Chan, & Scott, 2001). The basic Monte Carlo scheme consists of two phases. In the first phase, appropriate distributions are chosen to represent the uncertainty or variability for each of the problem input parameters (here represented by the predictor validities, effect sizes, etc.), whereas in the second phase, the relevant output quantities (here the selection quality and adverse impact values) are repeatedly calculated. More specifically, in each repetition, a value for each input parameter is randomly sampled from the distribution of that parameter, and the combination of the thus-obtained input parameter values is used to compute the outcome quantities. These outcome values are stored and subsequently analyzed to study the effects of the uncertainty.

In a first, preliminary step, the available selection parameter data are used as input to our method to determine a sample of optimal trade-off points, and the predictor weighing schemes that correspond to these optimal trade-offs are stored. These predictor weighing schemes are henceforth referred to as the *optimal weighing schemes*. For each selected optimal trade-off, we also sample a number of trade-offs that are dominated by the optimal trade-off (i.e., produce lower values of the quality and adverse impact objectives) and determine the predictor weighing systems that lead to these dominated trade-offs. The resulting weighing systems are henceforth called the *set of dominated weighing systems* associated with the corresponding optimal weighing system. In the case in which an optimal predictor weighing scheme is fairly robust, we expect to find that its associated trade-off for the selection quality and the adverse impact objectives, as computed from somewhat different values for the selection parameters, will usually maintain dominance over the similarly computed trade-offs associated with the weighing systems that belong to its set of dominated weighing systems.

Next, the Monte Carlo scheme is implemented. At each replication, a value is sampled from the chosen distribution of each selection parameter. Using these values, we apply both the optimal weighing system and its associated set of dominated weighing systems to calculate the selection quality and the adverse impact ratio of the selection. Finally, we count within each Monte Carlo replication and for each optimal weighing system the number of times that both objectives have a more favorable (i.e., larger) value as compared with the corresponding values associated with the weighing systems that it dominates and aggregate these counts over the total number of replications. Although a further aggregation over the different optimal weighing systems is possible as well, such an aggregation is not preferred because it eliminates the possibility of verifying whether the different optimal weighing systems perform homogeneously.

In addition to our documenting the frequency with which the optimal predictor weighing schemes perform better than the corresponding dominated weighing system on both selection objectives, two further counts are registered. The first is the total number of times that dominated weight systems outperform the corresponding optimal scheme on both quality and adverse impact, whereas the second is the frequency with which the optimal weights, compared with their respective dominated weight systems, lead to a better performance on an appropriately weighted combination of the two selection objectives. In particular, the weights assigned to the selection objectives are for each optimal weighing system chosen identical to the importance of these objectives as implied by the optimal trade-off that is associated with the optimal weighing scheme.

By expressing these frequencies as a proportion of the maximum possible frequency, we obtain three sensitivity indices for each sampled optimal predictor weighing scheme. The first and the third index summarize the extent to which the optimal weighing scheme outperforms its associated dominated weighing systems over the studied variability of the predictor parameter values, whereas the second index indicates the proportion with which the dominated weighing schemes are better than the optimal weighing on both quality and adverse impact. Thus, if a particular optimal weighing scheme would be completely robust for variability in the predictor parameter values, this scheme would be characterized by values of one for the first and the third index, and a value of zero for the second index.

Application

To illustrate, we applied the uncertainty analysis to the earlier discussed selection problem. Because the analysis focuses on the robustness of the optimal predictor weighing schemes for variability in the predictor parameter values, fairly broad uniform distributions were chosen to represent the gamut of possible sample values for the predictor parameters. The distribution chosen for each parameter was centered on the effect size, validity, and intercorrelation values reported in Table 1 with range equal to the value of the estimate. Thus, the distribution used for, for example, the majority–minority difference for the cognitive ability predictor was rectangular with endpoints of 0.5 and 1.5, respectively. The analysis focused on the performance of the predictor composites associated with the earlier selected Optimal Trade-Off Points 2 to 20. Optimal Trade-Offs 1 and 21 are not considered because they do not have dominated trade-offs. As shown in Figure 3, the 19 selected trade-offs are equally spaced over the entire trade-off surface, which ensures that the results are representative for the entire set of optimal predictor weighing schemes. Also, for one of the selected optimal trade-offs (i.e., Trade-Off 10), the figure depicts the set of all trade-offs that are dominated by this optimal trade-off and indicates which of the dominated trade-offs within the set were actually sampled to assess the performance under

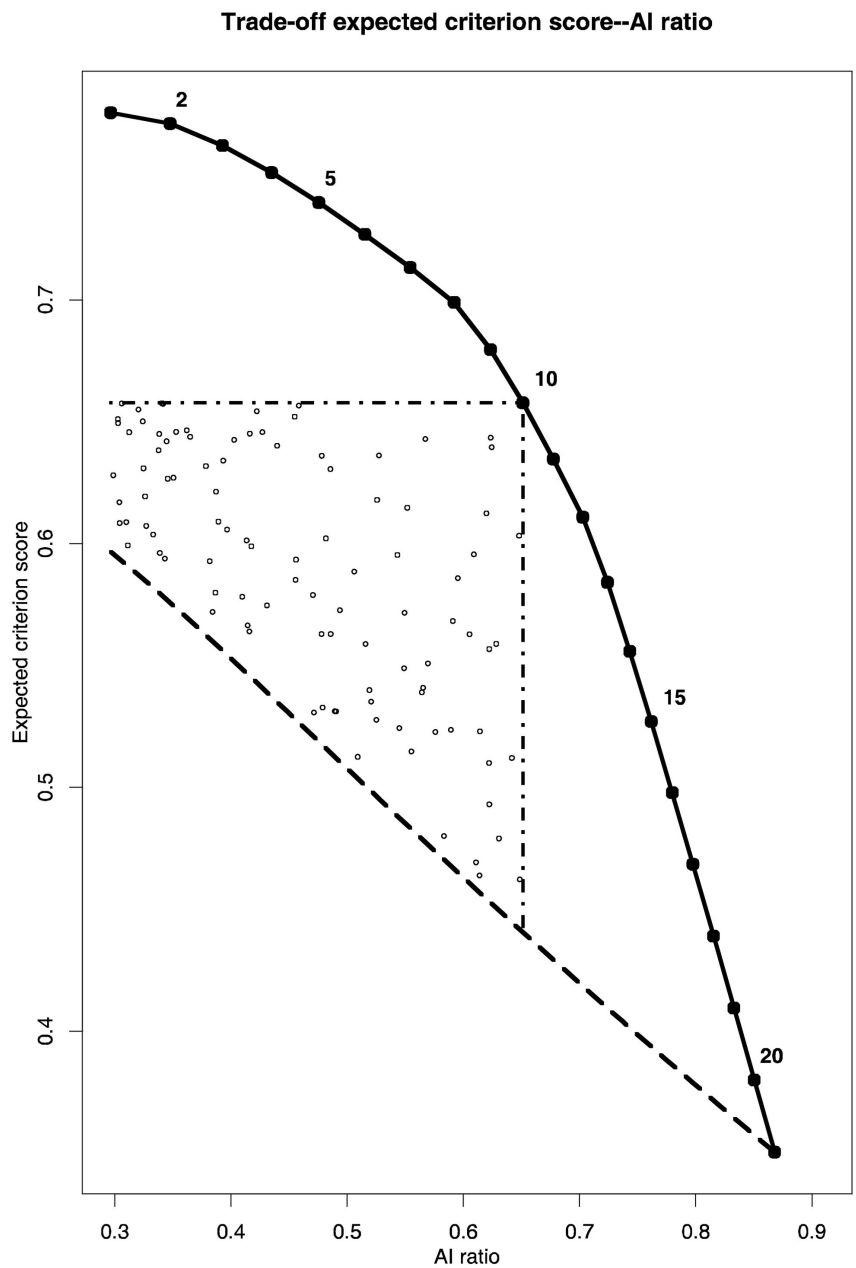

Figure 3. Pareto-optimal trade-offs used in the Monte Carlo simulation to study the robustness of the corresponding optimal predictor weighing schemes. The dots represent the set of dominated trade-offs used to evaluate the robustness of the Pareto-optimal predictor weighing scheme associated with Pareto-Optimal Trade-Off 10. AI = adverse impact.

uncertainty of this optimal predictor composite. The results of the assessment for this as well as for the other 18 selected optimal predictor weighing schemes are summarized in Table 3.

The rows of the table correspond to the 19 studied optimal composites, whereas the proportions in the columns indicate the obtained values for the sensitivity indices. Thus, the numbers in the second column show the proportion in instances in which the optimal composite continues to dominate the sampled nonoptimal composites over the studied variability in the selection parameter data, whereas the proportions in the third column pertain to the reverse. Finally, the proportions in the fourth column show the rate with which the optimal composite outperforms the nonoptimal composites on the weighted combination of the selection quality and the adverse impact objectives, using the importance weights corresponding to the optimal trade-off.

By and large, the obtained results show that optimal predictor composites that are based on representative estimates of the selection parameters continue to perform well under the studied vari-

Table 3
Performance Under Uncertainty of Optimal Predictor Composites

Optimal trade-off	Dominates	Is dominated	Outperforms
2	.62	.03	.87
3	.65	.02	.87
4	.70	.02	.88
5	.70	.01	.89
6	.69	.02	.87
7	.67	.01	.88
8	.68	.01	.91
9	.66	.02	.92
10	.65	.01	.93
11	.66	.01	.96
12	.61	.02	.93
13	.62	.01	.96
14	.63	.01	.97
15	.65	.01	.97
16	.63	.01	.97
17	.64	.01	.98
18	.65	.01	.98
19	.66	.01	.97
20	.63	.01	.98

ability of the selection parameters. The proportions reported in column 2, and especially those of column 4, are well above the values that can be expected if the optimal composites would not be robust for uncertainty in the selection parameters, whereas the rates in column 3 show that optimal composites become only very rarely dominated by the corresponding nonoptimal composites. Also, similar value patterns were found when studying several other selection scenarios (both probationary and nonprobationary) that involved different collections of predictors. The results of the uncertainty analyses therefore offer substantial support to the proposition that a decision-making approach to the formation of predictor composites is indeed feasible.

Discussion

This article has presented a procedure for determining the set of Pareto-optimal trade-offs between selection quality and adverse impact that can be achieved through differential weighting of the available predictors. The procedure addresses a wide variety of selection scenarios, including both probationary and nonprobationary decisions as well as situations in which the applicants come from several different minority populations. As compared with previous related proposals (e.g., De Corte, 1999; Hattrup et al., 1997; Schmitt et al., 1997), the procedure provides a more general as well as a more detailed analysis of the selection quality–adverse impact problem because it offers an attractive tool to visualize the entire set of trade-off alternatives. In addition, the optimal predictor weighing schemes that correspond to the different possible optimal trade-offs are obtained as well.

It was also noted that each optimal trade-off corresponds to a particular relative importance of the selection objectives. This means that each optimal trade-off can be obtained by solving an optimization problem in which the objective is a particular weighted sum (i.e., a sum with weights identical to the relative importance of the objectives) of the expected quality and adverse impact of the selection. Thus, each optimal trade-off corresponds to the maximization of an expected utility that combines the quality and diversity goals with a specific relative importance, and the individual optimal trade-offs differ in the actual expected utility (i.e., in the actual relative importance with which the two goals are combined) that is maximized. All this shows that the present procedure and its results are consistent with the decision-theoretic framework advocated by Petersen and Novick (1976) in their discussion of models of test fairness. Moreover, the procedure reflects top-down selection using the same composite cutoff score for the different applicant populations (cf. Appendix A) and in no way assumes subgroup-based adjustment of the predictor scores. It therefore meets one of the key provisions in the 1991 Civil Rights Act, which states that "it shall be an unlawful employment practice for a respondent, in connection with the selection or referral of applicants or candidates for employment or promotion, to adjust the scores of, use different cutoff scores for, or otherwise alter the results of, employment related tests on the basis of race, color, religion, sex, or national origin." For the same reasons, it is clear that the procedure does not involve using race, religion, and so on as a determinant of the decision to accept or reject candidates. Instead of such a practice, the procedure simply admits the inclusion of workforce diversity as an additional objective to be met by the selection system.

Similar to earlier methods (e.g., Hattrup et al., 1997; Schmitt et al., 1997), the present procedure can be applied as a research tool to study the effects of, for example, the validity, effect size, and intercorrelations of the selection predictors on both selection quality and adverse impact. The article did not focus on this type of effect study, however, but investigated the potential of the procedure as a decision tool to assist industrial–organizational psychologists in choosing between alternative predictor composites when planning a selection in a context in which both the goals of workforce quality and diversity are of importance. We believe that the latter approach, as compared with the first type of study, is of more immediate relevance to industrial–organizational psychologists, especially in these frequent situations where they have no other option than to use readily available predictors to implement a selection system without local evidence.

Even given representative estimates of the selection parameter values, it is still possible that the present method leads to predictor composites that behave rather poorly under somewhat different conditions. Although this possibility is equally present for composites derived in other ways, such as regression-based composites, it cannot be neglected. Thus, we also investigated the impact of uncertainty in the selection parameter values. The results indicated that the optimal predictor composites, derived from representative estimates of the selection parameters, continue to perform well under considerable variability of these parameters. It is therefore suggested that selection practitioners should consider using this procedure to determine optimal predictor weighing schemes, especially when no other alternative to address the quality–adverse impact problem can be applied and representative estimates of the selection parameters are indeed available.

Although the present proposal can address a broad range of situations, hurdle-based decisions still remain out of scope. The extension to multistage selections will not be easy, however, because for these situations an analytical method to link adverse impact and selection quality to the relevant selection parameters is

not yet available. One could eventually resolve this issue by adapting the approach outlined by De Corte (1998) to the case in which the total applicant group is a mixture of majority and minority candidates.

The extension to multistage scenarios could also help to address the eventuality that some of the selected candidates refuse the job offer. Following a suggestion by Murphy (1986), a two-stage approach could then be used in which the likelihood of accepting a job offer is considered as the second-stage predictor. However, this approach requires fairly detailed information about the underlying process of refusal, and it is doubtful that this information will usually be available (Murphy, 1986; Ryan, Sacco, McFarland, & Kriska, 2000; Schmit & Ryan, 1997). Alternatively, in case of random job refusal, the present procedure can be applied without any modification because then only the selection rate must be adjusted to account for the estimated probability of job refusal.

The results also depend on an assumption about the distribution of the predictors in the different applicant populations. As noted earlier, this assumption is essentially identical to the one invoked in all other studies on the effects of predictor composites on the level of adverse impact and the average criterion score of the selected applicants. Also, the assumption is consistent with the limited information that is presently available on the distribution of predictor scores in applicant populations (e.g., Schmidt, Hunter, McKenzie, & Muldrow, 1979). The assumption is not always required, however. In particular, the assumption can be dropped in the case in which one chooses the effect size and the validity of the predictor composite to represent the goals of diversity and quality, respectively. This choice also reduces the data requirements of the present procedure because then data on the composition of the applicant pool as well as on the selection rate are no longer needed.

In conclusion, this article focused on using predictor composites to address the adverse impact–selection quality problem, but we emphasize that other routes to workforce diversity, such as banding and the development of new, low-impact predictors, have important merits as well and that it is often preferable to use a combination of these alternatives. Also, such combinations, with optimal predictor weighing as one of the alternatives, will usually be possible. Thus, either by itself or in combination with other alternatives, the present method to determine Pareto-optimal predictor composites may offer a substantial contribution to the adverse impact–selection quality conflict. Compared with earlier procedures, the method provides valuable trade-off information that is otherwise unavailable, and it presents this information in a simple and understandable fashion. Because of these unique features, we hope that the proposal will find routine application in the design of selection systems.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: Author.

Bazaraa, M. S., Sherali, H. D., & Shetty, C. M. (2006). *Nonlinear programming: Theory and methods* (3rd ed.). New York: Wiley.

Bobko, P., Roth, P. L., & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. *Personnel Psychology, 52,* 561–589.

Boudreau, J. W. (1991). Utility analysis for decisions in human resource management. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 2, 2nd ed., pp. 621–745). Palo Alto, CA: Consulting Psychologists Press.

Campion, M. A., Outtz, J. L., Zedeck, S., Schmidt, F. L., Kehoe, J. F., Murphy, K. R., & Guion, R. M. (2001). The controversy over score banding in personnel selection: Answers to 10 key questions. *Personnel Psychology, 54,* 149–185.

Das, I., & Dennis, J. E. (1998). Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems. *SIAM Journal on Optimization, 8,* 631–657.

De Corte, W. (1998). Estimating and maximizing the utility of sequential selection decisions with a probationary period. *British Journal of Mathematical and Statistical Psychology, 51,* 101–121.

De Corte, W. (1999). Weighing job performance predictors to both maximize the quality of the selected work force and control the level of adverse impact. *Journal of Applied Psychology, 84,* 695–702.

Doverspike, D., Winter, J. L., Healy, M. C., & Barrett, G. V. (1996). Simulations as a method of illustrating the impact of differential weights on personnel selection outcomes. *Human Performance, 9,* 259–273.

Hattrup, K., Rock, J., & Scalia, C. (1997). The effects of varying conceptualizations of job performance on adverse impact, minority hiring, and predictor performance. *Journal of Applied Psychology, 82,* 656–664.

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment, 9,* 152–194.

Keeney, R. L., & Raiffa, H. (1993). *Decisions with multiple objectives.* Cambridge, England: Cambridge University Press.

Marler, R. T., & Arora, J. S. (2004). Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization, 26,* 369–395.

McKay, P., & McDaniel, M. (2006). A re-examination of Black–White mean differences in work performance: More data, more moderators. *Journal of Applied Psychology, 91,* 538–554.

McManus, M. A., & Kelly, M. L. (1999). Personality measures and biodata: Evidence regarding their incremental predictive value in the life insurance industry. *Personnel Psychology, 52,* 137–148.

Murphy, K. R. (1986). When your top choice turns you down: Effect of rejected offers on the utility of selection tests. *Psychological Bulletin, 99,* 133–138.

Murphy, K. R., & Shiarella, A. H. (1997). Implications of the multidimensional nature of job performance for the validity of selection tests: Multivariate framework for studying test validity. *Personnel Psychology, 50,* 823–854.

Nash, S., & Sofer, A. (1996). *Linear and nonlinear programming.* New York: McGraw-Hill.

Outtz, J. L. (2002). The role of cognitive ability tests in employment selection. *Human Performance, 15,* 161–171.

Petersen, N. S., & Novick, M. R. (1976). An evaluation of some models for culture-fair selection. *Journal of Educational Measurement, 13,* 3–29.

Potosky, D., Bobko, P., & Roth, P. L. (2005). Forming composites of cognitive ability and alternative measures to predict job performance and reduce adverse impact: Corrected estimates and realistic expectations. *International Journal of Selection and Assessment, 13,* 304–315.

Pulakos, E. D., & Schmitt, N. (1996). An evaluation of two strategies for reducing adverse impact and their effects on criterion-related validity. *Human Performance, 9,* 241–258.

Raju, N. S., Burke, M. J., & Normand, J. (1990). A new approach for utility analysis. *Journal of Applied Psychology, 75,* 3–12.

Ryan, A. M., Ployhart, R. E., & Friedel, L. (1998). Using personality testing to reduce adverse impact: A cautionary note. *Journal of Applied Psychology, 83,* 298–307.

Ryan, A. M., Sacco, J. M., McFarland, L. A., & Kriska, S. D. (2000).

Applicant self-selection: Correlates of withdrawal from a multiple hurdle process. *Journal of Applied Psychology, 85,* 163–179.

Sackett, P. R., & Ellingson, J. E. (1997). The effects of forming multipredictor composites on group differences and adverse impact. *Personnel Psychology, 50,* 707–721.

Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing and higher education. *American Psychologist, 56,* 302–318.

Sackett, P. R., & Wilk, S. L. (1994). Within-group norming and other forms of score adjustment in preemployment testing. *American Psychologist, 49,* 929–954.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & De Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities as predictors of work behaviors: A European meta-analysis. *Personnel Psychology, 56,* 573–605.

Saltelli, A., Chan, K., & Scott, E. M. (2001). *Sensitivity analysis.* West Sussex, England: John Wiley.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124,* 262–274.

Schmidt, F. L., Hunter, J. E., McKenzie, R. C., & Muldrow, T. W. (1979). Impact of valid selection procedures on workforce productivity. *Journal of Applied Psychology, 64,* 609–626.

Schmit, M. J., & Ryan, A. M. (1997). Applicant withdrawal: The role of test-taking attitudes and racial differences. *Personnel Psychology, 50,* 855–876.

Schmitt, N., Clause, C. S., & Pulakos, E. D. (1996). Subgroup differences associated with different measures of some common job-relevant constructs. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 11, pp. 115–139). New York: Wiley.

Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology, 82,* 719–730.

Tallis, G. M. (1961). The moment generating function of the truncated multinormal distribution. *Journal of the Royal Statistical Society, Series B, 23,* 223–229.

Appendix A

Computation of the Selection Utility and the Adverse Impact Ratio

This appendix details the computation of the selection utility, ΔU, and the adverse impact ratio (or ratios in the case of more than one minority group) as a function of the decision variables $\mathbf{b} = (b_1 \ldots b_P)'$, x_c, and y_c (with x_c and y_c the composite predictor and the criterion cutoff values, respectively) and the selection parameter data. Because nonprobationary selection is the special case of probationary selection for which $y_c = -\infty$, only the latter variant is discussed. Also, as ΔU is equal to $U_p - U_0$, with U_p and U_0 the payoffs of the predictor-based and the corresponding random selection, respectively, and U_0 can be obtained in a similar way as U_p, the discussion focuses on the derivation of the payoff U_p. According to Boudreau (1991), this payoff can be expressed as the benefits of the selection minus the testing, the separation, and the training costs. Thus,

$$U_p = \sum_g \{TN_g S_g^{(s)} \bar{V}_g^{(s)} - N_g c_t - N_g(S_g - S_g^{(s)})c_s - N_g S_g c_f\}$$

where T corresponds to the number of time periods that a successfully selected applicant remains on the job; N_g indicates the number of applicants from group g; S_g and $S_g^{(s)}$ refer to the proportion of applicants from group g that are selected and successfully selected, respectively; $\bar{V}_g^{(s)}$ denotes the average money valued criterion performance of a successfully selected applicant from group g; and c_s, c_f, and c_t are the costs per individual of separation, training, and testing, respectively.

Given the earlier introduced assumptions, the averages $\bar{V}_g^{(s)}$ (with $g = 1, \ldots, G$) can be determined as follows:

$$\bar{V}_g^{(s)} = \bar{V} + \sigma_V r_{YV} \bar{Y}_{sg}^{(s)}$$

with $\bar{V}$ and σ_V the average and the standard deviation of the money valued criterion performance in the total applicant population, respectively, r_{YV} the correlation between the rated and the money valued job performance (cf. Raju, Burke, & Normand, 1990), and $\bar{Y}_{sg}^{(s)}$ the globally standardized (i.e., with respect to the total applicant population) average criterion score of the successfully selected applicants from group g.

To determine the quantities $\bar{Y}_{sg}^{(s)}$ as well as the probabilities of (successful) selection, we invoke the earlier introduced assumption that the predictors $\mathbf{u} = (U_1 \ldots U_P)'$ and the criterion dimensions $\mathbf{w} = (W_1 \ldots W_C)'$ follow a joint $P + C$-variate normal distribution with the same variance–covariance matrix but a different mean vector in the applicant populations. Assuming further, without loss of generality, that the predictors and the criterion dimensions have unit variance in the different populations, it then follows that the raw composite predictor $X_r = \mathbf{b}'\mathbf{u}$ and the raw global criterion score $Y_r = \mathbf{a}'\mathbf{w}$ (with $\mathbf{a} = (a_1 \ldots a_C)'$, the vector of preassigned weights to the separate criterion dimensions) have a joint bivariate normal distribution with the same covariance matrix but a different mean vector in the applicant populations. Because X_r and Y_r have the same covariance matrix in the different populations, a common rescaling can be applied to the raw scores X_r and Y_r such that the rescaled composite predictor scores, X, and the rescaled global criterion scores, Y, have unit variance in each applicant population. The correlation between X and Y, r_{XY}, will be equal to the following:

$$r_{XY} = \frac{(\mathbf{b}',\mathbf{0}')\mathbf{R}(\mathbf{0}',\mathbf{a}')'}{\sqrt{\mathbf{b}'\mathbf{R}_u\mathbf{b}}\sqrt{\mathbf{a}'\mathbf{R}_w\mathbf{a}}}$$

where $\mathbf{0}$ is a zero vector of appropriate order, $\mathbf{R}$ is the joint correlation matrix of $\mathbf{u}$ and $\mathbf{w}$, and $\mathbf{R}_u$ and $\mathbf{R}_w$ are the correlation matrices of $\mathbf{u}$ and $\mathbf{w}$, respectively. Also, adopting the convention to equate the predictor averages and the criterion averages in one of the minority groups (by convention, the first) to zero, it follows that the average scores in this minority population on both the composite predictor and the global criterion, $\bar{X}_1$ and $\bar{Y}_1$, also equal zero, whereas the corresponding averages in the other populations,

$\bar{X}_g$ and $\bar{Y}_g$ (with $g = 2, \ldots, G$), are equal to $\mathbf{b}'\mathbf{d}_{ug}/\sqrt{\mathbf{b}'\mathbf{R}_u\mathbf{b}}$ and $\mathbf{a}'\mathbf{d}_{wg}/\sqrt{\mathbf{a}'\mathbf{R}_w\mathbf{a}}$, respectively, where $\mathbf{d}_{ug}$ and $\mathbf{d}_{wg}$ correspond to the vector of effect sizes of the predictors $\mathbf{u}$ and the criterion dimensions $\mathbf{w}$ in population g.

The preceding results completely specify the joint distribution of X and Y in the subpopulations such that the formulas on the value and the mean vector of truncated binormal distributions (cf. Tallis, 1961) can subsequently be used to evaluate the proportions S_g and $S_g^{(s)}$, as well as the average global criterion scores of the successfully selected employees from the different applicant groups. From these averages, denoted as $\bar{Y}_g^{(s)}$, the corresponding globally standardized criterion averages, $\bar{Y}_{sg}^{(s)}$, are then obtained as follows:

$$\bar{Y}_{sg}^{(s)} = \frac{\bar{Y}_g^{(s)} - \bar{Y}}{\sigma_Y}$$

where $\bar{Y} = \Sigma_g p_g \bar{Y}_g$, $\sigma_Y = 1 + \Sigma_g p_g(\bar{Y}_g - \bar{Y})^2$, and p_g denotes the proportion of applicants from group g in the total candidate population.

Finally, letting the subscript $g = G$ indicate the majority applicant population, the value of the adverse impact ratio for the minority groups (i.e., the groups $g = 1, \ldots, G - 1$), a_g, can be computed as follows:

$$a_g = \frac{S_g}{S_G}.$$

In the case of probationary selection, the ratio can be computed as follows:

$$a_g = \frac{S_g^{(s)}}{S_G^{(s)}}.$$

Appendix B

Solution of Nonlinear Programming Problems

This appendix details the nonlinear programming problems that are solved in the two stages of the normal-boundary intersection method. Using the earlier introduced notation, we obtain the optimal predictor weighing scheme that maximizes the selection quality objective by solving the following nonlinear program.

Maximize selection quality over $\mathbf{z} = (\mathbf{b}', x_c, y_c)'$ subject to the following constraints:

1. $\Sigma_g p_g S_g^{(s)} = s$.
2. $\mathbf{b}'\mathbf{R}_u\mathbf{b} = 1$.
3. Application-specific constraints.

In the preceding formulation, the selection quality objective is equal to either the selection utility, the average quality of the selected applicants, or the validity of the composite predictor. Constraint 1 expresses the requirement to obtain the intended (successful) selection rate s. Also, Constraint 2, which fixes the variance of the composite predictor to the arbitrarily chosen value of 1, is added to ensure a unique solution of the nonlinear program.

The second nonlinear program of the first stage is identical to the preceding program, except that this time the adverse impact ratio is maximized. When the total applicant population comprises members from several different minority groups, the second nonlinear program is solved separately for each minority group.

In the second stage of the normal-boundary intersection method, the nonlinear programs have, in case of a single minority group, the following format.

Maximize r over $\mathbf{z}$ and r subject to the aforementioned detailed constraints (Constraints 1, 2, and 3) and

4. $\mathbf{\Phi\beta} - r\mathbf{\Phi 1} = \mathbf{v}_z$.

In the last constraint, $\mathbf{1}$ is the column vector of all ones, $\boldsymbol{\beta}=(\beta,1-\beta)'$, and $\mathbf{v}_z = \mathbf{t}_z - \mathbf{t}_o$, with $\mathbf{t}_o = (q_{max}, a_{max})'$ and $\mathbf{t}_z = (q_z, a_z)'$, where q_z and a_z are the values of the selection quality and the adverse impact objectives associated with the values $\mathbf{z}$ of the predictor weights and the composite predictor and criterion cutoffs. Also, the trade-off matrix $\mathbf{\Phi}$ is as follows:

$$\mathbf{\Phi} = \begin{pmatrix} 0 & q_a - q_{max} \\ a_q - a_{max} & 0 \end{pmatrix}$$

where a_q, q_a, q_{max}, and a_{max} are as defined in the text. The above nonlinear program is solved repeatedly for different but equally spaced values of β between 0 and 1.

Received March 6, 2006
Revision received December 13, 2006
Accepted January 5, 2007 ■