MANCHESTER 1824
The University of Manchester

The University of Manchester Research

On the use of spline regression in the study of congruence in organizational research.

DOI:
10.1177/1094428117715067

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):
Edwards, J., & Parry, M. (2018). On the use of spline regression in the study of congruence in organizational research. *Organizational Research Methods*, *21*(1), 68-110. https://doi.org/10.1177/1094428117715067

Published in:
Organizational Research Methods

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.

Download date:26. Aug. 2022

On the Use of Spline Regression in the Study of Congruence in Organizational Research

Jeffrey R. Edwards
Kenan-Flagler Business School
University of North Carolina
Chapel Hill, NC 27599-3490
919-962-3144
jredwards@unc.edu

Mark E. Parry
Henry W. Bloch School of Management
University of Missouri-Kansas City
5110 Cherry Street
Kansas City, MO 64110-2499
816-235-6724
parryma@umkc.edu

Previous versions of this paper were presented at the annual meeting of the Society of Industrial and Organizational Psychology, Atlanta, GA, May, 2006, and the annual meeting of the Academy of Management, Vancouver, BC, August, 2015.

The authors thank Clyde B. Schechter, M.D., for his technical assistance with Stata.

In press, *Organizational Research Methods.*

Abstract

The study of congruence is central to organizational research. Congruence refers to the fit, match, similarity, or agreement between two constructs and is typically framed as a predictor of outcomes relevant to individuals and organizations. Previous studies often operationalized congruence as the algebraic, absolute, or squared difference between two component variables. Difference scores suffer from numerous methodological problems, which stimulated the development of alternative procedures. For algebraic and squared difference scores, the primary alternatives involve linear and quadratic regression equations. For absolute difference scores, the extant alternative is piecewise regression, which avoids certain problems with absolute difference scores but relies on untested assumptions that are central to congruence research. In this article, we develop an alternative to absolute difference scores based on spline regression, yielding a comprehensive approach for testing hypotheses that underlie absolute difference scores while avoiding the shortcomings of piecewise regression analysis. We demonstrate the advantages of spline regression over absolute difference scores and piecewise regression using an empirical example.

The study of congruence is central to organizational research (Edwards, 1994). Congruence refers to the fit, match, or similarity between two constructs, such as the demands of the job and the abilities of the person, the needs of the person and the rewards provided by the job, the values of the person and the organization, or the actual and ideal structure of a firm (Edwards & Shipp, 2007; Kristof, 1996; Venkatraman, 1989). Congruence is usually treated as a predictor of outcomes relevant to individuals and organizations, such as job satisfaction, job performance, organizational commitment, psychological and physical well-being, and firm performance (Boyd, Haynes, Hitt, Bergh, & Ketchen, 2012; Edwards, 1991; Hoffman & Woehr, 2006; Kristof-Brown, Zimmerman, & Johnson, 2005).

During much of its history, congruence research relied on difference scores, which are prone to numerous methodological problems (Cronbach, 1958; Edwards, 1994; Johns, 1981). Many of these problems can be overcome by analyses that use the component variables that constitute difference scores as joint predictors, supplemented by higher-order terms as needed to represent the functional form implied by the difference score in question. This approach was developed by Edwards (1994), who presented alternatives to algebraic, absolute, and squared difference scores along with procedures for testing their associated constraints, which can be viewed as hypotheses that motivate the use of difference scores from a conceptual standpoint.

As shown by Edwards (1994), algebraic and squared difference scores can be expressed as restricted versions of linear and quadratic regression equations, respectively. However, the expression for an absolute difference score is less straightforward, because computing the score involves a decision rule such that, if the difference is positive, its sign is unaltered, whereas if the difference is negative, its sign is reversed. To represent this decision rule, Edwards (1994) formed a dummy variable that indicated whether the difference was positive or negative and incorporated this variable as a moderator in a piecewise regression equation. This approach provides tests of whether the slopes of the component variables are equal in magnitude and opposite in sign for positive and negative differences, as implied by an absolute difference score. However, this approach has one crucial shortcoming: *it does not allow the researcher to verify*

whether, in fact, the slopes relating the component variables to the outcome change where the component variables are equal. Rather, this assumption is taken for granted, given that the dummy variable indicating whether the difference is positive or negative must be coded prior to analysis. Methodological work that followed Edwards (1994) has focused primarily on the quadratic regression equation that serves as an alternative to squared difference scores (Edwards, 2002; Edwards & Parry, 1993). A general approach that permits comprehensive tests of congruence effects implied by absolute difference scores has yet to be developed.

Developing a viable alternative to absolute difference scores is important, for several reasons. First, theories of congruence often frame its effects in terms of an absolute difference between component constructs. For instance, the theory of job satisfaction developed by Locke (1969) indicated that, for most job attributes, satisfaction results from the absolute difference between perceived and valued (i.e., desired) amounts. Similarly, the theory of stress and behavior in organizations proposed by McGrath (1976) asserts that experienced stress is a function of the absolute difference between the task demands and individual abilities. Likewise, in strategic management research, the concept of fit between the firm and its environment has been portrayed as an absolute difference (Venkatraman, 1989). Beyond these explicit examples, most congruence theories assert that outcomes increase or decrease as the components of congruence deviate from one another in either direction, and a parsimonious interpretation of this assertion is that the effects of incongruence are linear, corresponding to an absolute difference. Second, absolute difference scores have been used in numerous studies of congruence, spanning topics such as leader-member exchange, met expectations, job satisfaction, work-related stress, person-job fit, value congruence, interpersonal similarity, organizational climate, and strategic alignment. As such, procedures that avoid problems with absolute difference scores would prove useful for many domains of management research. Third, developing procedures that address the full set of assumptions underlying absolute difference scores would bring closure to methodological issues that were made evident decades ago (Edwards, 1994) but have yet to be fully resolved.

In this article, we present spline regression as an alternative to absolute difference scores. Spline regression comprises a family of procedures for analyzing functions that change slope at specific points whose number and location can be treated as parameters to be estimated (Marsh & Corimer, 2002; Smith, 1979; Suits, Mason, & Chan, 1978). Although spline regression is typically applied to two-dimensional functions, it can be extended to three-dimensional surfaces in which the slope changes along a fold in the surface, as implied by an absolute difference score relating two component variables to an outcome (Edwards, 1994).

This article is organized as follows. We begin with an introduction to spline regression, describing two-dimensional functions that are typically analyzed with this method. We then extend spline regression to three-dimensional surfaces relevant to congruence research. We explain how features of these surfaces correspond to parameters in spline regression equations and how these equations are estimated and interpreted. Next, we present an empirical example that compares spline regression to absolute difference scores and the piecewise regression approach presented by Edwards (1994). We conclude by discussing the strengths and limitations of spline regression and how it complements polynomial regression in congruence research.

Before proceeding, it is important to position spline regression relative to polynomial regression. As noted earlier, polynomial regression was developed as an alternative to algebraic and squared difference scores. In contrast, spline regression is an advancement over piecewise regression, which itself is an alternative to absolute difference scores. Spline regression and polynomial regression can both be used to investigate congruence hypotheses and, as such, the methods should not be viewed as competitors. The methods differ in that polynomial regression is suited to hypothesized surfaces that are curvilinear and symmetric, whereas spline regression applies to surfaces that are linear, potentially asymmetric, and can have more than one line along which the surface changes in slope. Ultimately, we believe that polynomial regression and spline regression should be viewed as complementary methods for congruence research, with the choice between these methods determined by the particulars of the hypotheses that the researcher wishes to investigate. The distinctions between polynomial regression and spline regression are

elaborated in the Discussion section of this article, where we also describe how the methods can be integrated into a unified analytical framework.

Overview of Spline Regression

As noted earlier, spline regression is a method for estimating functions that change slope at one or more points (Marsh & Corimer, 2002; Smith, 1979; Suits et al., 1978). Spline regression differs from piecewise regression (Neter, Wasserman, & Kutner, 1989; von Eye & Schuster, 1998) in two ways. First, piecewise regression requires the researcher to specify *a priori* the point at which a function changes slope without being able to test this specification, whereas spline regression can treat this point as a parameter to be estimated. Second, although piecewise regression can estimate functions that are discontinuous at the point where the slope is specified to change, indicating a vertical displacement of the function, the functions estimated by spline regression are continuous. By specifying functions as continuous, the point or points at which the slope changes can be estimated. This tradeoff is worthwhile for congruence research because, from a conceptual standpoint, there is little basis for predicting that functions abruptly jump where two component variables are equal. In contrast, the premise that outcomes are minimized or maximized along the line of congruence, as opposed to some other line, is central to congruence research and should therefore be treated as a core hypothesis to be tested empirically. To clarify the correspondence between piecewise regression and spline regression, we start with a piecewise regression function and then modify it to yield a spline regression function. We then extend these methods to three dimensions.[1]

Piecewise Regression Function

A basic piecewise regression equation can be written as follows:

$$Y = a_0 + a_1X + a_2W + a_3XW + e. \quad (1)$$

In Equation 1, X is an independent variable, Y is the dependent variable, and W is a dummy variable that equals 0 when X is less than or equal to some critical value, which we designate c, and equals 1 when X is greater than c.[2] The terms W and XW capture the difference in the intercept and slope, respectively, of the function relating X to Y on either side of the point c.

Finally, the a_i in Equation 1 are unstandardized regression coefficients (in this article, we use a_i to represent coefficients in piecewise regression equations and b_i to indicate coefficients in spline regression equations).

The interpretation of Equation 1 is clarified by substituting 0 or 1 for W to recover the intercept and slope of the function relating X to Y on either side of the point c. When $X \leq c$, $W = 0$, and Equation 1 simplifies to:

$$Y = a_0 + a_1X + e. \quad (2)$$

Hence, a_0 and a_1 are the intercept and slope, respectively, of the function relating X to Y when $X \leq c$. When $X > c$, $W = 1$, and Equation 1 becomes:

$$Y = (a_0 + a_2) + (a_1 + a_3)X + e. \quad (3)$$

Thus, the intercept and slope of the function when $X > c$ are represented by the compound terms $(a_0 + a_2)$ and $(a_1 + a_3)$, respectively.

Figure 1 provides examples of hypothetical piecewise regression functions based on Equation 1. Figure 1a shows a function that resembles an inverted-V, which might occur when an employee is dissatisfied by too much or too little contact with coworkers (Harrison, 1978). The function in Figure 1b is positively sloped on the right side and flat on the left side, as when an employee experiences increased stress when job demands exceed his or her abilities but not when demands fall short of abilities (French, Rodgers, & Cobb, 1974). Finally, Figure 1c has a positive slope on the left and negative but less steep slope on the right, which could represent a person's pay satisfaction when compared to a referent other, indicating that underpayment is more dissatisfying than overpayment (Adams, 1965).

As illustrated in Figure 1, a piecewise regression function can be discontinuous, depicted by the vertical jump in the functions at point c. This discontinuity can occur because Equation 1 does not constrain the intercepts of the segments of the function relating X to Y shown in Equations 2 and 3. The constraint that yields a continuous function can be derived by substituting c for X in Equations 2 and 3, thereby obtaining the predicted value of Y when X equals c, and setting these two expressions equal to one another. For Equation 2, the predicted

value of Y when X equals c is $a_0 + a_1c$, and for Equation 3, the predicted value of Y when X equals c is $(a_0 + a_2) + (a_1 + a_3)c$. Setting these expressions equal to one another yields:

$$a_0 + a_1c = (a_0 + a_2) + (a_1 + a_3)c. \quad (4)$$

We simplify this equation to obtain:

$$a_2 = -a_3c. \quad (5)$$

Imposing the constraint indicated by Equation 5 on Equation 1 yields the following equation:

$$\begin{aligned} Y &= a_0 + a_1X - a_3cW + a_3XW + e \\ &= a_0 + a_1X + a_3(X - c)W + e. \end{aligned} \quad (6)$$

Equation 6 can be rewritten as follows:

$$Y = a_0^* + a_1^*X + a_2^*(X - c)W + e. \quad (7)$$

The asterisks in this expression are intended to convey that the a_i^* estimated from Equation 7 will generally differ from the a_i estimated from Equation 1, given that these equations use different independent variables as predictors. Specifically, whereas Equation 1 has three predictors, X, W, and XW, Equation 6 has two predictors, X and $(X - c)W$. Hence, modifying Equation 1 by dropping W and replacing XW with $(X - c)W$ forces the two segments in Equations 2 and 3 to meet at the point c (Neter et al., 1989).

As with Equation 1, the interpretation of Equation 7 becomes apparent when W is replaced with 0 or 1. When $W = 0$, such that $X \leq c$, Equation 7 reduces to:

$$Y = a_0^* + a_1^*X + e. \quad (8)$$

Alternately, when $W = 1$, meaning that $X > c$, Equation 7 becomes:

$$Y = (a_0^* - a_2^*c) + (a_1^* + a_2^*)X + e. \quad (9)$$

Figure 2 contains plots of continuous piecewise regression functions that correspond to Equation 7. These plots are consistent with those in Figure 1, except that the functions are now continuous at the point where $X = c$.

Spline Regression

Equation 7 can be translated into a spline regression function with two modifications. The first is merely a matter of notation, whereby we relabel the coefficients from a_i^* to b_i:

$$Y = b_0 + b_1X + b_2(X - c)W + e. \quad (10)$$

The second modification involves replacing W with the term $(X > c)$:

$$Y = b_0 + b_1X + b_2(X - c)(X > c) + e. \quad (11)$$

In Equation 11, $(X > c)$ is a logical expression that equals 1 when the condition described by the expression is satisfied and equals 0 otherwise. Hence, when $X > c$, $(X > c)$ equals 1, and when $X \leq c$, $(X > c)$ equals 0. The key difference between the expression $(X > c)$ in Equation 11 and the dummy variable W in Equation 10 is that the value at which $(X > c)$ equals 0 or 1 is estimated as the parameter c, whereas the value at which W equals 0 or 1 is fixed *a priori*. In spline regression terminology, the parameter c that represents where the slope of the function changes is a *knot*. Estimating c along with b_0, b_1, and b_2 can be accomplished with nonlinear regression procedures (Bates & Watts, 1988; Seber & Wild, 2003), which we discuss later in this article.

The interpretation of the spline function in Equation 11 is facilitated by examining the intercepts and slopes of the functions on either side of c. When $X \leq c$, $(X > c)$ in Equation 11 equals 0, and the equation reduces to:

$$Y = b_0 + b_1X + e. \quad (12)$$

Conversely, when $X > c$, the term $(X > c)$ equals 1, and Equation 11 becomes:

$$\begin{aligned} Y &= b_0 + b_1X + b_2(X - c) + e \\ &= b_0 + b_1X + b_2X - b_2c + e \\ &= (b_0 - b_2c) + (b_1 + b_2)X + e. \end{aligned} \quad (13)$$

Hence, when $X \leq c$, the intercept and slope of the function relating X to Y are b_0 and b_1, whereas when $X > c$, the intercept and slope are $(b_0 - b_2c)$ and $(b_1 + b_2)$, respectively.

Figure 3 displays functions that parallel those in Figure 2. However, c no longer signifies a fixed constant, but instead represents an estimated parameter, corresponding to Equation 11. Figure 3a again shows that coworker satisfaction decreases when an employee has too much or too little contact with coworkers, but introduces the notion that satisfaction peaks when actual contact with coworkers is slightly greater than desired, reflecting the premise that excess contact can bring instrumental and emotional support that have salutary effects beyond fulfilling the

person's desire for contact itself (House, 1981). In Figure 3b, the knot is also shifted to the right, indicating that stress does not increase until demands exceed abilities by some threshold, such that a slight excess of demands over abilities is experienced not as stress but instead as challenge (Caplan, 1983). Finally, Figure 3c suggest that pay satisfaction does not decrease until the focal person's pay exceeds that of a referent other by an amount that represents a tolerance of overpayment for the self relative to others (Adams, 1965).

Spline Regression With Multiple Knots

Equations for spline functions with one knot can be extended to include additional knots. For instance, a spline function with two knot locations labeled c_1 and c_2 can be estimated using the following equation:

$$Y = b_0 + b_1X + b_2(X - c_1)(X > c_1) + b_3(X - c_2)(X > c_2) + e. \quad (14)$$

In Equation 14, the expression $(X > c_1)$ equals 0 when $X \leq c_1$ and equals 1 when $X > c_1$, and the expression $(X > c_2)$ equals 0 when $X \leq c_2$ and equals 1 when $X > c_2$. Equation 14 can be further extended to estimate functions with more than two knots.

The interpretation of Equation 14 is clarified by examining the intercepts and slopes of the spline function on either side of c_1 and c_2. For this illustration, we assume that $c_1 < c_2$ (if $c_1 > c_2$, then the subscripts on c_1 and c_2 can be reversed to conform to the equations given here). When $X \leq c_1$, the expressions $(X > c_1)$ and $(X > c_2)$ equal 0. Hence, Equation 14 reduces to:

$$Y = b_0 + b_1X + e. \quad (15)$$

When $c_1 < X \leq c_2$, $(X > c_1)$ equals 1 and $(X > c_2)$ equals 0, such that Equation 14 becomes:

$$\begin{aligned} Y &= b_0 + b_1X + b_2(X - c_1) + e \\ &= b_0 + b_1X + b_2X - b_2c_1 + e \\ &= (b_0 - b_2c_1) + (b_1 + b_2)X + e. \end{aligned} \quad (16)$$

Finally, when $X > c_2$, $(X > c_1)$ and $(X > c_2)$ both equal 1, and Equation 14 is therefore:

$$\begin{aligned} Y &= b_0 + b_1X + b_2(X - c_1) + b_3(X - c_2) + e \\ &= b_0 + b_1X + b_2X - b_2c_1 + b_3X - b_3c_2 + e \\ &= (b_0 - b_2c_1 - b_3c_2) + (b_1 + b_2 + b_3)X + e. \end{aligned} \quad (17)$$

Taken together, Equations 15, 16, and 17 give the intercepts and slopes for the three segments of the spline functions indicated by Equation 14.

Drawing from Equation 14, Figure 4 shows illustrative spline functions with two knots that are substantively similar to those in Figure 3, with the exception that each function has a flat region near the middle. In Figure 4a, the employee does not experience dissatisfaction until contact with coworkers falls outside of what might be considered a zone of indifference (Kulka, 1979). In Figure 4b, stress increases not only when demands exceed abilities, but also increases when demands fall well short of abilities, perhaps due to boredom or the underutilization of valued skills. Finally, Figure 4c indicates that pay satisfaction is not affected until the focal person's pay deviates from that of a referent other by an amount large enough to be viewed as meaningful.

Extending Spline Regression to Three Dimensions

The principles underlying two-dimensional spline functions can be extended to three-dimensional spline surfaces. These surfaces can be viewed as extensions of the piecewise regression surfaces considered by Edwards (1994). Here, we review the principles of piecewise regression surfaces and then move to spline regression surfaces.

Piecewise Regression Surface

Edwards (1994) applied piecewise regression as an alternative to absolute difference scores. We begin by considering an equation that uses an absolute difference as a predictor:[3]

$$Z = d_0 + d_1|Y - X| + e. \quad (18)$$

Note that we use d_i to signify coefficients from an equation that uses an absolute difference score as a predictor. Following Edwards (1994), an alternative expression for the absolute difference function can be written as follows:

$$Z = d_0 + d_1[(1 - 2W)(Y - X)] + e. \quad (19)$$

In Equation 19, W is a dummy variable coded such that $W = 0$ when $Y \geq X$ and $W = 1$ when $Y < X$. Thus, when the $(Y - X)$ difference is positive, (1 – 2W) reduces to 1, whereas when the $(Y - X)$ difference is negative, (1 – 2W) becomes –1, thereby reversing the sign of the difference. As

such, the term $(1 - 2W)$ produces the same result as an absolute value transformation. When the $(Y - X)$ difference equals zero, the term $[(1 - 2W)(Y - X)]$ becomes zero regardless of whether W is coded 0 or 1. Expanding Equation 19 and rearranging terms yields the following expression:

$$Z = d_0 - d_1X + d_1Y + 2d_1XW - 2d_1YW + e. \quad (20)$$

Equation 20 is a constrained version of the following piecewise linear equation:

$$Z = a_0 + a_1X + a_2Y + a_3W + a_4XW + a_5YW + e. \quad (21)$$

Comparing these two equations shows that Equation 20 imposes the following constraints on the coefficients in Equation 21: (a) $a_1 = -a_2$; (b) $a_4 = -a_5$; (c) $a_4 = -2a_1$; and (d) $a_3 = 0$. These constraints should be viewed as hypotheses to be tested such that, if the constraints hold, the surface relating X and Y to Z is consistent with that implied by an absolute difference score. Additional guidelines for using Equation 21 to conduct confirmatory tests of absolute difference scores are provided by Edwards (1994, 2002).

Equation 21 can be used to derive the intercepts and slopes of two triangular sections of the surface relating X and Y to Z, with the boundary separating the sections running along the $Y = X$ line. These terms can be recovered by substituting 0 or 1 for W in Equation 21. When $W = 0$, Equation 21 simplifies to:

$$Z = a_0 + a_1X + a_2Y + e. \quad (22)$$

Alternately, when $W = 1$, Equation 21 becomes:

$$\begin{aligned} Z &= a_0 + a_1X + a_2Y + a_3 + a_4X + a_5Y + e \\ &= (a_0 + a_3) + (a_1 + a_4)X + (a_2 + a_5)Y + e. \end{aligned} \quad (23)$$

Hence, when $Y \geq X$, the surface has an intercept of a_0 and slopes of a_1 and a_2, respectively, for X and Y. When $Y < X$, the intercept of the surface is $(a_0 + a_3)$, and slopes of X and Y are $(a_1 + a_4)$ and $(a_2 + a_5)$.

Illustrative surfaces corresponding to Equations 20 and 21 are shown in Figures 5 and 6, respectively. The surfaces in Figure 5 are symmetric on either side of the $Y = X$ line and are flat and continuous along the $Y = X$ line, consistent an absolute difference score. Substantively, Figure 5a indicates that coworker satisfaction decreases symmetrically as actual and desired

contact with coworkers differ in either direction and remains constant when actual and desired coworker contact are equal, regardless of their absolute levels. Analogously, Figure 5b shows that experienced stress increases symmetrically as job demands and employee abilities deviate from one another and is constant as demands and abilities jointly increase or decrease. Finally, Figure 5c indicates that pay satisfaction decreases to the same degree as received and referent other pay differ in either direction and remains constant for any fixed difference between received and referent other pay. These surfaces are extensions of the two-dimensional functions in Figure 2 but differ in two ways. First, the two-dimensional functions treat desired coworker contact, employee abilities, and referent other pay as constants, whereas the surfaces allow these standards of comparison to vary. Second, unlike the functions in Figure 2, the surfaces in Figure 5 impose the symmetry constraints associated with an absolute difference score.

Unlike the surfaces in Figure 5, the surfaces in Figure 6 are asymmetric about the $Y = X$ line and are sloped and discontinuous along the $Y = X$ line. The surface in Figure 6a is equivalent to that in Figure 5a but adds a discontinuity along the $Y = X$ line such that, for a given difference between X and Y, coworker satisfaction is lower for excess rather than deficient contact with coworkers. Compared to Figure 5b, the surface in Figure 6b indicates that, along the $Y = X$ line, experienced stress is higher when job demands and employee abilities are both high than when both are low and that stress is higher when demands exceed abilities than when demands fall short of abilities. The surface in Figure 6c differs from that in Figure 5c in that pay satisfaction is higher when received and referent other pay are both high than when both are low and also when self pay exceeds referent other pay. Further examples that compare surfaces corresponding to absolute difference scores and piecewise regression equations are provided by Edwards (1994, 2002).

Equation 21 places no restrictions on the sections of the surface for which W is coded 0 or 1. Because of this, the surface can be discontinuous along the $Y = X$ line, as illustrated in Figure 6. The surface can be made continuous by imposing constraints that force the sections of the surface to meet along the $Y = X$ line. To derive these constraints, we begin by rewriting

Equations 22 and 23 to describe the shape of each surface along the $Y = X$ line. The shape of a surface along a particular line can be found by substituting the expression for the line into the equation for the surface (Edwards & Parry, 1993). We apply this principle to Equations 22 and 23. Substituting the expression $Y = X$ into Equation 22 yields:

$$\begin{aligned} Z &= a_0 + a_1X + a_2X + e \\ &= a_0 + (a_1 + a_2)X + e. \end{aligned} \tag{24}$$

Similarly, substituting $Y = X$ in Equation 23 gives:

$$\begin{aligned} Z &= (a_0 + a_3) + (a_1 + a_4)X + (a_2 + a_5)X + e \\ &= (a_0 + a_3) + (a_1 + a_2 + a_4 + a_5)X + e. \end{aligned} \tag{25}$$

If the two sections of the surface described by Equations 24 and 25 meet along the $Y = X$ line, then the intercept and slope in Equation 24 must equal the intercept and slope in Equation 25. Setting equal the intercepts from Equations 24 and 25 and simplifying gives:

$$\begin{aligned} a_0 &= a_0 + a_3 \\ a_3 &= 0. \end{aligned} \tag{26}$$

Similarly, setting equal the slopes from Equations 24 and 25 and simplifying produces:

$$\begin{aligned} a_1 + a_2 &= a_1 + a_2 + a_4 + a_5 \\ a_4 &= -a_5. \end{aligned} \tag{27}$$

Substituting Equations 26 and 27 into Equation 21 produces a piecewise regression equation in which the surfaces for $W = 0$ and $W = 1$ meet along the $Y = X$ line. This substitution yields:

$$\begin{aligned} Z &= a_0 + a_1X + a_2Y - a_5XW + a_5YW + e \\ &= a_0 + a_1X + a_2Y + a_5(Y - X)W + e. \end{aligned} \tag{28}$$

We now rewrite this constrained equation by replacing the a_i with a_i^* and listing the coefficient subscripts as integers in ascending order:

$$Z = a_0^* + a_1^*X + a_2^*Y + a_3^*(Y - X)W + e. \tag{29}$$

Again, replacing the a_i with a_i^* is meant to convey that the coefficients estimated using Equation 29 will generally differ from the corresponding coefficients in Equation 21 (e.g., a_1 will differ from a_1^*) given that the two equations use different independent variables.

Equation 29 can be used to write expressions for the surface on either side of the $Y = X$ line that defines the location of the seam. When $Y \geq X$, $W = 0$, and Equation 29 reduces to:

$$Z = a_0^* + a_1^* X + a_2^* Y + e \tag{30}$$

Alternately, when $Y < X$, $W = 1$, and Equation 29 becomes:

$$Z = a_0^* + a_1^* X + a_2^* Y + a_3^* (Y - X) + e$$

$$Z = a_0^* + a_1^* X + a_2^* Y + a_3^* Y - a_3^* X + e$$

$$Z = a_0^* + (a_1^* - a_3^*) X + (a_2^* + a_3^*) Y + e. \tag{31}$$

Figure 7 shows surfaces that parallel those in Figure 6, with the caveat that the surfaces now meet along the seams. As before, the seams are constrained to run along the $Y = X$ line, based on how W is coded prior to estimation. This constraint can be treated as an assumption to be tested using spline regression, as described below.

Spline Regression Surface

We now translate Equation 29 into an equation corresponding to a spline regression surface. First, we replace the equation for the seam associated with Equation 29, which is $Y = X$, with the general expression $Y = c_0 + c_1 X$. This replacement allows the intercept and slope of the seam, which we notate c_0 and c_1, respectively, to take on values other than 0 and 1. Replacing $Y = X$ with $Y = c_0 + c_1 X$ in turn replaces the term $(Y - X)$ in Equation 29 with the term $(Y - c_0 - c_1 X)$, which can be seen as follows. Recall that the term $(Y - X)$ in Equation 29 originated by substituting $Y = X$ into Equations 22 and 23 to obtain the intercepts and slopes of the two sections of the surface on either side of the $Y = X$ line. Analogously, we can obtain the intercepts and slopes of the two sections on either side of the $Y = c_0 + c_1 X$ line by substituting the expression for this line into Equations 22 and 23. For Equation 22, we have:

$$\begin{aligned} Z &= a_0 + a_1 X + a_2 (c_0 + c_1 X) + e \\ &= a_0 + a_1 X + a_2 c_0 + a_2 c_1 X + e \\ &= (a_0 + a_2 c_0) + (a_1 + a_2 c_1) X + e. \end{aligned} \tag{32}$$

For Equation 23, we obtain:

$$Z = (a_0 + a_3) + (a_1 + a_4) X + (a_2 + a_5)(c_0 + c_1 X) + e$$

$$= (a_0 + a_3) + (a_1 + a_4)X + a_2c_0 + a_2c_1X + a_5c_0 + a_5c_1X + e$$

$$= (a_0 + a_3 + a_2c_0 + a_5c_0) + (a_1 + a_4 + a_2c_1 + a_5c_1)X + e. \quad (33)$$

Next, we set the intercepts and the slopes for the two sections of the surface on either side of the $Y = c_0 + c_1X$ line to be equal. For the intercepts, we have:

$$a_0 + a_2c_0 = a_0 + a_3 + a_2c_0 + a_5c_0$$

$$a_3 = -a_5c_0. \quad (34)$$

For the slopes, we obtain:

$$a_1 + a_2c_1 = a_1 + a_4 + a_2c_1 + a_5c_1$$

$$a_4 = -a_5c_1 \quad (35)$$

We now substitute Equations 34 and 35 into Equation 21, which yields:

$$Z = a_0 + a_1X + a_2Y - a_5c_0W - a_5c_1XW + a_5YW + e$$

$$= a_0 + a_1X + a_2Y + a_5(Y - c_0 - c_1X)W + e. \quad (36)$$

We then rewrite Equation 36 by replacing the a_i with a_i^* and listing the subscripts in ascending order to obtain:

$$Z = a_0^* + a_1^*X + a_2^*Y + a_3^*(Y - c_0 - c_1X)W + e. \quad (37)$$

As before, we replace the a_i with a_i^* and list the subscripts in ascending order to indicate that the coefficient estimates from Equation 37 will generally differ from those obtained using Equation 21. Comparing Equation 29 to Equation 37 shows that the term $(Y - X)$ has been replaced with $(Y - c_0 - c_1X)$.

Second, we replace W with the logical expression $(Y < c_0 + c_1X)$, which distinguishes between the sections of the surface on either side of the $Y = c_0 + c_1X$ line:

$$Z = a_0^* + a_1^*X + a_2^*Y + a_3^*(Y - c_0 - c_1X)(Y < c_0 + c_1X) + e. \quad (38)$$

The expression $(Y < c_0 + c_1X)$ resolves to 1 when the condition it describes is satisfied and equals 0 otherwise. Finally, we change the a_i^* to b_i, which we use to distinguish spline regression from piecewise regression. The resulting equation is:

$$Z = b_0 + b_1X + b_2Y + b_3(Y - c_0 - c_1X)(Y < c_0 + c_1X) + e. \quad (39)$$

The interpretation of the spline surface for Equation 39 is clarified by computing the

intercepts and slopes of the portions of the surface on either side of the seam described by $Y = c_0 + c_1X$. When $Y \geq c_0 + c_1X$, $(Y < c_0 + c_1X)$ equals 0, and Equation 39 simplifies to:

$$Z = b_0 + b_1X + b_2Y + e. \tag{40}$$

When $Y < c_0 + c_1X$, $(Y < c_0 + c_1X)$ equal 1, and Equation 39 becomes:

$$\begin{aligned} Z &= b_0 + b_1X + b_2Y + b_3(Y - c_0 - c_1X) + e \\ &= b_0 + b_1X + b_2Y + b_3Y - b_3c_0 - b_3c_1X + e \\ &= (b_0 - b_3c_0) + (b_1 - b_3c_1)X + (b_2 + b_3)Y + e. \end{aligned} \tag{41}$$

Thus, when $Y \geq c_0 + c_1X$, the surface relating X and Y to Z has an intercept of b_0 and slopes for X and Y of b_1 and b_2, whereas when $Y < c_0 + c_1X$, the intercept and slopes for X and Y are $(b_0 - b_3c_0)$, $(b_1 - b_3c_1)$, and $(b_2 + b_3)$.

The terms in Equations 40 and 41 can be further analyzed to examine relevant properties of the surface. For instance, in studies of congruence, it is often useful to determine whether the coefficients on X and Y are equal in magnitude but opposite in sign, as implied when an absolute difference score is used to represent a congruence hypothesis. This hypothesis involves two tests that compare the coefficients on X and Y within Equations 40 and 41, as follows: (a) $b_1 = -b_2$, or $b_1 + b_2 = 0$; and (b) $b_1 - b_3c_1 = -(b_2 + b_3)$, or $b_1 + b_2 + b_3(1 - c_1) = 0$. If these two conditions are satisfied, then the slopes for X and Y do not differ from being equal in magnitude but opposite in sign on both sides of the seam.

Another property of the surface often relevant to congruence research is whether the surface is symmetric on either side of the seam. Symmetry implies that the slopes for X and Y on one side of the seam are equal in magnitude but opposite in sign when compared to their slopes on the other side of the seam. When treated as a hypothesis, symmetry involves two tests that again use the coefficients on X and Y from Equations 40 and 41: (a) $b_1 = -(b_1 - b_3c_1)$, or $2b_1 - b_3c_1 = 0$; and (b) $b_2 = -(b_2 + b_3)$, or $2b_2 + b_3 = 0$. If both of these conditions are satisfied, then the surface does not deviate from being symmetric about the seam.

A third property of the surface that merits attention is the slope of the surface along the seam. This property is relevant because most congruence hypotheses rely on the assumption that

an outcome is maximized or minimized along the $Y = X$ line and has the same value along this line regardless of whether X and Y are both low or high. Support for this prediction would be evidenced by a surface with a seam running along the $Y = X$ line, such that $c_0 = 0$ and $c_1 = 1$, and a slope of zero along the seam. The estimated values of c_0 and c_1 can be tested directly, and the intercept and slope of a surface along its seam can be computed by substituting the expression for the seam (i.e., $Y = c_0 + c_1X$) into Equation 40:

$$\begin{aligned} Z &= b_0 + b_1X + b_2(c_0 + c_1X) + e \\ &= b_0 + b_1X + b_2c_0 + b_2c_1X + e \\ &= (b_0 + b_2c_0) + (b_1 + b_2c_1)X + e. \end{aligned} \quad (42)$$

Hence, the intercept and slope of the surface along its seam are $(b_0 + b_2c_0)$ and $(b_1 + b_2c_1)$. As might be expected, the same result is obtained when $Y = c_0 + c_1X$ is substituted into Equation 41 given that, by construction, the portions of the spline surface indicated by Equations 40 and 41 have the same intercept and slope along the seam.

Figure 7 shows surfaces that are analogous to those in Figure 6 but relax the constraint that the seams run along the $Y = X$ line. In Figure 7a, the seam is shifted to the right, such that employees are more satisfied when they have more contact with coworkers than would meet their stated desires, which again could reflect supplemental benefits of social support. Figure 7b indicates that stress begins to increase when demands are greater than abilities and that these effects are more pronounced when abilities are low, given that the seam is shifted to the right and rotated counterclockwise relative to the $Y = X$ line. Finally, Figure 7c again shows that pay satisfaction begins to decrease when the person's pay is greater than a referent other but also indicates that this decrease begins further to the right (i.e., into the region of excess pay) when self and other pay are both high than when both are low.

Spline Regression Surface With Multiple Seams

Equations for spline surfaces with one seam can be extended to surfaces with two or more seams. Consider a surface with two seams described by the lines $Y = c_{10} + c_{11}X$ and $Y = c_{20} + c_{21}X$. An equation for this surface can be written by extending Equation 39 as follows:

$$Z = b_0 + b_1X + b_2Y + b_3(Y - c_{10} - c_{11}X)(Y < c_{10} + c_{11}X) + b_4(Y - c_{20} - c_{21}X)(Y < c_{20} + c_{21}X) + e. \quad (43)$$

As before, the expression $(Y < c_{10} + c_{11}X)$ equals 0 when $Y \geq c_{10} + c_{11}X$ and equals 1 when $Y < c_{10} + c_{11}X$. Similarly, the expression $(Y < c_{20} + c_{21}X)$ equals 0 when $Y \geq c_{20} + c_{21}X$ and equals 1 when $Y < c_{20} + c_{21}X$. Equation 43 can be extended for surfaces with more than two seams.

Surfaces corresponding to Equation 43 can have three or four sections, depending on whether the two seams cross within the joint range of X and Y in the X,Y plane. For simplicity, we consider a surface for which the seams denoted by $Y = c_{10} + c_{11}X$ and $Y = c_{20} + c_{21}X$ do not cross within the range of X and Y. Consequently, the two seams separate the surface into three portions. When the two seams cross within the range of X and Y, the surface has four portions, which makes the interpretation of the surface more complex but nonetheless tractable. For this illustration, the $Y = c_{10} + c_{11}X$ seam is located above the $Y = c_{20} + c_{21}X$ seam in the X,Y plane such that, for each value of X, the value of Y described by $c_{10} + c_{11}X$ is greater than the value of Y indicated by $c_{20} + c_{21}X$. Under these conditions, when $Y \geq c_{10} + c_{11}X$, the expressions $(Y < c_{10} + c_{11}X)$ and $(Y < c_{20} + c_{21}X)$ both equal 0, and Equation 43 reduces to:

$$Z = b_0 + b_1X + b_2Y + e. \quad (44)$$

When $c_{10} + c_{11}X > Y \geq c_{20} + c_{21}X$, such that Y lies between the two seams, Equation 43 becomes:

$$\begin{aligned} Z &= b_0 + b_1X + b_2Y + b_3(Y - c_{10} - c_{11}X) + e \\ &= b_0 + b_1X + b_2Y + b_3Y - b_3c_{10} - b_3c_{11}X + e \\ &= (b_0 - b_3c_{10}) + (b_1 - b_3c_{11})X + (b_2 + b_3)Y + e. \end{aligned} \quad (45)$$

Finally, when $Y < c_{20} + c_{21}X$, Equation 43 is as follows:

$$\begin{aligned} Z &= b_0 + b_1X + b_2Y + b_3(Y - c_{10} - c_{11}X) + b_4(Y - c_{20} - c_{21}X) + e \\ &= b_0 + b_1X + b_2Y + b_3Y - b_3c_{10} - b_3c_{11}X + b_4Y - b_4c_{20} - b_4c_{21}X + e \\ &= (b_0 - b_3c_{10} - b_4c_{20}) + (b_1 - b_3c_{11} - b_4c_{21})X + (b_2 + b_3 + b_4)Y + e. \end{aligned} \quad (46)$$

As before, the equations that describe the various portions of the surface can be used to analyze properties relevant to congruence research. For instance, the coefficients on X and Y for each section of the surface can be compared to determine whether X and Y have slopes that are

equal in magnitude but opposite in sign. This property can be evaluated by testing the following constraints: (a) $b_1 = -b_2$, or $b_1 + b_2 = 0$; (b) $b_1 - b_3c_{11} = -(b_2 + b_3)$, or $b_1 + b_2 + b_3(1 - c_{11}) = 0$; and (c) $b_1 - b_3c_{11} - b_4c_{21} = -(b_2 + b_3 + b_4)$, or $b_1 + b_2 + b_3(1 - c_{11}) + b_4(1 - c_{21}) = 0$. Testing these three constraints indicates whether the slopes for X and Y are equal in magnitude but opposite in sign for Equations 44, 45, and 46, respectively. Analogously, slopes for X can be compared across the three sections of the surface by testing differences between b_1, $b_1 - b_3c_{11}$, and $b_1 - b_3c_{11} - b_4c_{21}$, which are the coefficients on X from Equations 44, 45, and 46. The corresponding coefficients on Y are b_2, $b_2 + b_3$, and $b_2 + b_3 + b_4$, as again shown in Equations 44, 45, and 46.

Finally, the slope of the surface along each seam can be computed by substituting the expression for a seam line into an equation that describes the surface on either side of the seam. For instance, substituting the expression $Y = c_{10} + c_{11}X$ into Equation 44 yields:

$$\begin{aligned} Z &= b_0 + b_1X + b_2(c_{10} + c_{11}X) + e \\ &= b_0 + b_1X + b_2c_{10} + b_2c_{11}X + e \\ &= (b_0 + b_2c_{10}) + (b_1 + b_2c_{11})X + e. \end{aligned} \tag{47}$$

Thus, the intercept and slope of the surface along the seam described by $Y = c_{10} + c_{11}X$ are $(b_0 + b_2c_{10})$ and $(b_1 + b_2c_{11})$, respectively. Analogously, substituting the expression $Y = c_{20} + c_{21}X$ into Equation 45 gives:

$$\begin{aligned} Z &= (b_0 - b_3c_{10}) + (b_1 - b_3c_{11})X + (b_2 + b_3)(c_{20} + c_{21}X) + e \\ &= b_0 - b_3c_{10} + b_1X - b_3c_{11}X + b_2c_{20} + b_2c_{21}X + b_3c_{20} + b_3c_{21}X + e \\ &= (b_0 - b_3c_{10} + b_2c_{20} + b_3c_{20}) + (b_1 - b_3c_{11} + b_2c_{21} + b_3c_{21})X + e. \end{aligned} \tag{48}$$

Hence, the intercept and slope of the surface along the seam $Y = c_{20} + c_{21}X$ are $(b_0 - b_3c_{10} + b_2c_{20} + b_3c_{20})$ and $(b_1 - b_3c_{11} + b_2c_{21} + b_3c_{21})$, respectively. The compound coefficients on X in Equations 47 and 48 can be tested to determine whether the surface is flat or sloped along the seams.

Figure 8 displays spline surfaces with two seams that build on the functions in Figure 4 and the surfaces in Figure 7. Figure 8a shows that coworker satisfaction decreases when actual contact with coworkers deviates from desired contact by some threshold, again representing a

zone of indifference. The surface in Figure 8b indicates that stress increases when demands exceed abilities and, to a lesser extent, when demands fall short of abilities, and that, for a given difference between demands and abilities, stress is higher when demands and abilities are both high than when both are low. This surface also shows that stress is unaffected by small deviations of demands from abilities when both are low rather than high, as indicated by the distance between the seams at the near corner of the surface. Finally, Figure 8c shows that pay satisfaction decreases as the focal person is underpaid, and to a lesser extent, overpaid relative to a referent other. The surface adds that deviations of the focal person's pay relative to a referent other is more tolerable when pay received by the self and other are both high than when both are low, as indicated by the difference in the spread between the seams at the far and near corners of the surface.

Estimating Spline Regression Equations

As noted previously, nonlinear regression can be used to analyze spline regression equations that treat knot and seam locations as parameters to be estimated (Bates & Watts, 1988; Motulsky & Ransnas, 1987; Seber & Wild, 2003). We strongly recommend that researchers estimate the locations of knots and seams, even when theory indicates where they should be, because doing so treats knot and seam locations as hypotheses to be tested, as opposed to assumptions that are taken for granted. This recommendation is based on the premise that theoretically motivated model specifications should be examined empirically, thereby exposing theories to falsification or corroboration. Accordingly, in this section we provide an overview of nonlinear regression to help organizational researchers understand its basic principles, make informed decisions when applying this method, and organize and interpret its results.

Estimation Algorithms

Nonlinear regression uses iterative procedures that modify parameter estimates at each step, with the goal of producing estimates that minimize some type of loss function. A loss function commonly used in nonlinear regression is the sum of squared residuals, consistent with the least-squares criterion of OLS regression. Beginning with an initial set of starting values, the

loss function is computed and new parameters values are generated, with this process repeating until the estimation algorithm has converged, as indicated by a decrease in the loss function or changes in the parameter estimates that are small enough to be considered negligible.

Nonlinear regression can be implemented using various algorithms (Bates & Watts, 1988; Motulsky & Ransnas, 1987; Seber & Wild, 2003). Direct-search algorithms, such as the simplex method, generate $k + 1$ sets of parameter estimates, where k is the number of parameters in the equation, and compute the loss function for each set. The set with the highest value of the loss function is replaced with a new set that blends the best of the remaining k sets, and the process repeats until convergence is reached. Other methods derive parameter estimates by seeking the gradient vector of the loss function, along which the function decreases most rapidly. For instance, the method of steepest descent computes the slope of the gradient vector at each iteration and generates parameter estimates that yield the greatest reduction in the loss function. With this method, initial iterations tend to move rapidly down the gradient vector, but later iterations can zigzag around the minimum of the loss function. The Newton-Raphson method guides iterations using the first and second derivatives of the gradient vector to find the path along which the loss function decreases at the fastest rate. Unfortunately, computing the second derivatives of the gradient vector can be problematic, leading to local rather than global convergence and sometimes causing the loss function to increase rather than decrease. These problems can be ameliorated by methods that approximate the second derivatives rather than computing them in their entirety. Among these methods, the Gauss-Newton, quasi-Newton, and Levenberg-Marquardt algorithms have proven useful, although their performance depends on the extent to which their approximations of the second derivatives disregard information that would be needed to accurately track the gradient vector of the loss function.

The algorithms described here are available in major statistical programs such as SPSS, Stata, SAS, SYSTAT, and R. Each algorithm has its own strengths and weaknesses, and no algorithm can be deemed superior in all situations. Therefore, it is prudent to implement several algorithms and compare their results. If the parameter estimates are similar, then it is reasonable

to conclude that the results are not specific to the algorithm used. If estimates differ, then we recommend trying different sets of starting values and ensuring that the programs are using the same convergence criteria. If these efforts fail, then alternative specifications of the regression equation should be considered, perhaps by increasing or decreasing the number of seams in the surface. Doing so can reveal whether the hypothesized surface is too discrepant with the data to yield stable estimates across estimation algorithms.

Specifying Starting Values

The effectiveness of nonlinear estimation algorithms depends on the starting values specified for the parameters prior to the first iteration. A carefully selected set of starting values can increase the likelihood that the algorithm will successfully converge, require fewer iterations to reach convergence, and avoid local minima in the iteration process (Seber & Wild, 2003). For spline surfaces, we describe three approaches for specifying starting values. First, starting values can be chosen that match the hypothesized surface. Although this approach is straightforward, its effectiveness depends on the extent to which the hypothesized surface is consistent with the data, which is unknown prior to analysis. A second approach is to use the hypothesized seam locations to code dummy variables, apply piecewise regression with the portions of the surface constrained to meet along the seams, and use the resulting coefficient estimates along with the intercepts and slopes of the hypothesized seams as starting values. For instance, if a surface was hypothesized to have one seam with an intercept of $c_0 = 0$ and a slope of $c_1 = 1$, such that the seam runs along the $Y = X$ line, then a dummy variable W would be coded 0 or 1 to distinguish cases where Y is greater than or less than X, and Equation 29 would be estimated. The obtained estimates of a_0^*, a_1^*, a_2^*, and a_3^* would serve as starting values for b_0, b_1, b_2, and b_3 in Equation 37, along with starting values of 0 and 1 for c_0 and c_1. A third approach starts by estimating an unconstrained piecewise regression equation that does not force the sections of the surface to meet and uses the resulting coefficient estimates to solve for the lines along which the sections of the surface would intersect if they were projected into one another. The intercepts and slopes of these lines of intersection are used as starting values for the seams, and the coefficient estimates

from the unconstrained piecewise equation are used as starting values for the remaining parameters. To illustrate, for a surface with one seam, Equation 21 would be estimated, and the line along which these two sections of the surface intersect would be found by solving Equations 34 and 35 for c_0 and c_1, respectively, which yields $c_0 = -a_3/a_5$ and $c_1 = -a_4/a_5$. These terms would serve as starting values for c_0 and c_1, and a_0, a_1, a_2, and a_5 estimated using Equation 21 would serve as starting values for b_0, b_1, b_2, and b_3, respectively. Of these three approaches, we consider the second and third approaches superior to the first, and we have found that these two approaches tend to yield very similar results. In practice, the third approach uses more information from the data, whereas the second approach is simpler because it does not require the researcher to compute where the sections of the surface intersect. Moreover, there is no guarantee that the starting values for seams derived using the third approach will fall within the range of the data. In the example that follows, we applied the second approach and verified that the results were effectively the same when the third approach was used.

Convergence Criteria

Statistical programs that implement nonlinear estimation rely on various convergence criteria to determine when the iterations will cease. Typical convergence criteria include the change in the loss function, changes in the parameter estimates, and the maximum number of iterations. Default values for these criteria vary widely across programs. For instance, default values used by SPSS, Stata, SAS, and SYSTAT range from .0001 to .00000001 for changes in the loss function, from .0001 to .00000001 for changes in parameter estimates, and from 20 to 10,000 for the maximum number of iterations (convergence criteria in R depend on the package used). Differences in these criteria can cause parameter estimates to differ even when programs use the same estimation algorithm. Thus, it is important to specify the same convergence criteria when comparing results across programs. Convergence criteria can also be modified when an estimation algorithm fails to converge or produces estimates that fall outside reasonable bounds.[4]

Model Evaluation

Nonlinear regression models can be evaluated using various criteria. For instance, when

the loss function used for estimation minimizes the sum of squared residuals, the R^2 for the equation can be used to assess the proportion of variance explained by the model, analogous to the R^2 provided by OLS regression. In addition, the R^2 values from nested equations, such as those that specify different numbers of seams, can be compared using the conventional F-test (Motulsky & Ransnas, 1987). In rare instances, the R^2 can actually decrease when a seam is added to the equation, which is a symptom that the data have been overfitted. When this occurs, the F-test can be disregarded, and the equation without the additional seam should be preferred. Most statistical packages that perform nonlinear estimation also report estimates of the variances and covariances of parameters, which can be used to construct confidence intervals for individual parameters and linear combinations of parameters, and some packages (e.g., Stata) can test linear and nonlinear combinations of parameters, from which confidence intervals can be formed. However, such analyses rely on normal theory to derive sampling distributions of parameters and R^2 values, which rests on assumptions that might not hold for nonlinear estimation (Motulsky & Ransnas, 1987; Seber & Wild, 2003), particularly when nonlinear combinations of parameters are tested, such as the products b_3c_0 and b_3c_1 in Equation 41. These problems can be addressed by applying nonparametric procedures that do not rely on distributional assumptions, such as the bootstrap (Efron & Tibshirani, 1993; Julious, 2001; Mooney & Duval, 1993; Stine, 1989). In the following example, we demonstrate methods that rely on normal theory and also apply the bootstrap, using bias-corrected confidence intervals derived with the percentile method.

Empirical Example

Sample and Measures

We illustrate spline regression and compare it to absolute difference scores and piecewise linear regression using data from a larger study of person-environment fit in organizations (Cable & Edwards, 2004; Edwards & Cable, 2009). These data involve 950 respondents who completed measures of the actual and desired amounts of various attributes of jobs and overall satisfaction with the job. For illustration, we focus on the following four job attributes, which we list here along with a sample item for each attribute: (a) authority (e.g., having definite lines of authority);

(b) relationships (e.g., forming relationships with coworkers); (c) variety (e.g., doing a variety of things); and (d) autonomy (e.g., doing my work in my own way). All measures consisted of three items that were rated on 5-point scales ranging from "none" to "a very great amount" for both actual amount and desired amounts. Descriptive statistics, correlations, and reliability estimates for the measures used for analysis are reported in Table 1.

Analyses

The analyses reported here were conducted using Stata 14. Analyses of absolute difference scores and piecewise linear equations were conducted using OLS regression invoked with the *regress* command, and spline regression equations were estimated with nonlinear least-squares regression using the *nl* command. Absolute difference scores were estimated using Equation 18. We applied two versions of piecewise regression analysis, one in which the surfaces on either side of the $Y = X$ line were not forced to meet (Edwards, 1994), and another that constrained the surfaces to meet along the $Y = X$ line. These analyses use Equations 21 and 29, respectively. For both equations, W was coded such that $W = 0$ when $Y \geq X$ and $W = 1$ when $Y < X$. The spline regression analyses specified surfaces with one seam, corresponding to Equation 39.

The nonlinear regression procedure in Stata implements the Gauss-Newton approach which, as noted earlier, approximates the second derivatives used to guide iterations that track the loss function along the gradient vector, which can help avoid estimation problems. We relied on the default maximum number of iterations, which was 10,000, and the convergence criteria for successive parameter estimates and for the residual sum of squares, both of which were .00001. To examine the stability of the nonlinear regression results across statistical packages, we repeated the analyses using SPSS 23, SAS 9.0, and SYSTAT 12, using the Gauss-Newton and quasi-Newton algorithms implemented by these programs and the same maximum number of iterations and convergence criteria. The results from these programs were virtually identical to those reported here and yielded the same substantive conclusions. Stata syntax used for these analyses, as well as syntax for SPSS, SAS, and SYSTAT, the data used in our example, and

Excel files that compute surface features and bootstrap confidence intervals, are available on the website of the first author (http://public.kenan-flagler.unc.edu/faculty/edwardsj/).

Our analyses were guided by the hypothesis that job satisfaction is greatest when the actual and desired amounts of job attributes are equal and decreases as actual and desired amounts deviate from each other in either direction. This hypothesis follows from theories that specify job satisfaction as an outcome of the fit between actual and desired job attributes (e.g., Dawis, 1992; Edwards, Caplan, & Harrison, 1998; Locke, 1976; Rice, McFarlin, Hunt, & Near, 1985) and was explicitly expressed as an absolute difference function by Locke (1969). This hypothesis is also implied by the many studies that have used the absolute difference between actual and desired job attributes to predict job satisfaction.

Starting values for the spline regression equations were derived using the second of the three approaches described earlier. To recap, Equation 29 was estimated with W coded 0 or 1 depending on whether Y was greater than or less than X. The obtained estimates were used to specify starting values for Equation 39, whereby $b_0 = a_0^*$, $b_1 = a_1^*$, $b_2 = a_2^*$, $b_3 = a_3^*$, $c_0 = 0$, and $c_1 = 1$. We compared the coefficient estimates from Equation 39 using these starting values to estimates obtained using the third approach discussed earlier, in which results from Equation 21 were used to specify starting values for Equation 39 as $b_0 = a_0$, $b_1 = a_1$, $b_2 = a_2$, $b_3 = a_5$, $c_0 = -a_3/a_5$, and $c_1 = -a_4/a_5$. The coefficient estimates from Equation 39 using these two approaches to setting starting values were virtually identical, with no coefficients showing differences above the second decimal place and 17 of the 20 coefficients equal to three or more decimal places. As such, the choice between these two approaches to specify starting values had no material impact on the results reported here.

Statistical tests were conducted by constructing confidence intervals based on normal theory for absolute difference and piecewise regression equations using OLS estimation and both normal theory and the bootstrap for spline regression equations using nonlinear estimation. Confidence intervals were constructed for individual coefficients as well as linear and nonlinear combinations of coefficients required to assess features of the estimated spline surfaces. Normal

theory confidence intervals were constructed for individual coefficients using the standard errors reported by Stata. For expressions that involved linear and nonlinear combinations of coefficients, we used the *nlcom* postestimation command, which reports standard errors and confidence intervals for the computed quantities. To construct bootstrap confidence intervals, we applied the *boot* function in Stata to draw 10,000 bootstrap samples, estimate the focal spline regression equation for each sample, and use the resulting coefficients individually and in combination to compute the quantities required to test the spline surface. For these quantities, we constructed bias-corrected confidence intervals using the percentile method (Stine, 1989). For illustrative purposes, we report both normal theory and bootstrap confidence intervals for the spline regression equations estimated with nonlinear regression, but we base our interpretation on the bootstrap confidence intervals, which are more appropriate for these analyses.

In addition to the primary analyses described here, we also conducted supplemental analyses in which we estimated spline regression equations that incorporated two seams. Surfaces with two seams are conceptually relevant for the variables involved because, as illustrated by our hypothetical examples, such surfaces can contain a zone of indifference in which satisfaction remains constant when the actual and desired amounts of a job attribute differ to a limited degree and decreases when the difference between actual and desired amounts reaches some critical threshold (Kroeger, 1995; Kulka, 1979).

Results

Descriptive statistics. Table 1 reports descriptive statistics, correlations, and reliability estimates for the measures used in our example. Means were near or slightly above the midpoint of the 1-5 scale used for each measure, and minima and maxima in conjunction with bivariate plots showed that the paired actual and desired measures for each job dimension exhibited adequate dispersion for analyzing the shapes of the surfaces relating these measures to job satisfaction. Prior to analysis, the actual and desired measures were centered at the scale midpoint of 3, such that the measures could range from –2 to +2 and had a midpoint of zero. This rescaling facilitates interpretation such that the $Y = X$ and $Y = -X$ lines, which are key

reference lines in congruence research (Edwards, 1994), running diagonally across the X,Y plane under the surface and intersecting at the center of the surface, which is the point $X = 0$, $Y = 0$.

Absolute Difference Scores. Table 2 gives results for regressions of job satisfaction on absolute difference scores computed for each of the four job dimensions, following Equation 18. Coefficient estimates are supplemented by 95% confidence intervals computed using the reported standard errors, which give lower and upper limits based on normal theory (these limits are labeled LL_N and UL_N, respectively). As can be seen, the coefficients on the absolute difference scores were negative, and their confidence intervals excluded zero, inviting the interpretation that job satisfaction is maximized when actual and desired amounts of each job dimension are equal and decreases symmetrically as actual amounts deviate from desired amounts in either direction. Surfaces corresponding to these results are displayed in Figure 9. Using Equation 20, the coefficients for the absolute difference scores in Table 2 were translated into coefficients for actual and desired job dimensions (i.e., X and Y, respectively) on either side of the $Y = X$ line, as shown in Table 3. These coefficients make evident the constraints imposed by an absolute difference score, such that the slopes for X and Y are equal in magnitude but opposite in sign on either side of the $Y = X$ line. Naturally, these results suffer from the problems with difference scores (Cronbach, 1958; Edwards, 1994; Johns, 1981) and are reported here for comparison with the results that follow.

Unconstrained piecewise regression. Table 4 contain results from analyses using the unconstrained piecewise regression equations in Equations 21. As indicated by comparing Equations 20 and 21, the results in Table 3 would be consistent with the functional form implied by an absolute difference score if $a_1 = -a_2$, $a_4 = -a_5$, $a_4 = -2a_1$, and $a_3 = 0$, with the further condition that a_1, a_2, a_4, and a_5 differ from zero (Edwards, 1994). This pattern did not hold for any of the four job dimensions, as shown by the coefficients in Table 4. In addition, the four constraints imposed by the absolute difference scores were tested as a set, which is equivalent to testing the increase in R^2 for the unconstrained piecewise regression equation over the absolute difference score equation. F statistics for these tests were 17.30, 25.55, 18.41, and 23.67,

respectively, for authority, relationships, variety, and autonomy (all numerator and denominator degrees of freedom were 4 and 944, respectively, and all $p < .05$). These tests are equivalent to statistically comparing the R^2_{AD} and R^2_{UP} values in Table 12, which refer to the R^2 values estimated using the absolute difference and unconstrained piecewise regression equations, respectively. For all four job dimensions, the R^2_{UP} values were significantly larger than the corresponding R^2_{AD} values, meaning the constraints imposed by the absolute difference scores were rejected. These results are sufficient to reject the functional form implied by the absolute difference score. Differences between the surfaces for the absolute difference scores and the unconstrained piecewise regression equations can be seen by comparing Figures 9 and 10. In addition, the coefficients in Table 4 were used to compute the coefficients for X and Y on either side of the $Y = X$ line, using Equations 22 and 23. These coefficients are given in Table 5 and can be compared with those in Table 3 to further probe the differences in results for the absolute difference scores and the unconstrained piecewise regression equations. In particular, as shown in Table 3, the results for the absolute difference scores suggested that actual and desired amounts of each job attribute had equal but opposite relationships with job satisfaction for both negative and positive discrepancies. In contrast, the results in Table 5 indicate: (a) when actual amounts are less than desired amounts (i.e., left of the $Y = X$ line), job satisfaction is positively related to actual amounts and, to a lesser extent, negatively related to desired amounts; and (b) when actual amounts are greater than desired amounts (i.e., right of the $Y = X$ line), job satisfaction is positively related to desired amounts and, to a lesser extent, negatively related to actual amounts. These results further disconfirm the simple symmetric surfaces implied by the absolute difference scores.

Constrained piecewise regression. Table 6 reports results for constrained piecewise regression equations, corresponding to Equation 29, and the coefficients in Table 6 are translated into coefficients for X and Y on either side of the $Y = X$ line in Table 7, using Equations 30 and 31. Results for the constrained equations were similar to those for the unconstrained equations, as can be seen by comparing the corresponding coefficients in Tables 5 and 7. Omnibus tests of

the difference between the unconstrained and constrained equations can be obtained by comparing their respective R^2 values (i.e., R^2_{UP} and R^2_{CP}) using an F-test, as reported in Table 12. As can be seen, the R^2 values did not statistically differ for any of the four job dimensions. The similarities of the results for the unconstrained and constrained equations are further evidenced by comparing the corresponding surfaces in Figures 10 and 11. As can be seen, the surfaces for the constrained equations eliminate the slight discontinuities along the surfaces for the unconstrained equations, but otherwise, the surfaces are quite similar and yield the same substantive interpretations.

Spline regression. Results for the spline regression analyses are reported in Table 8. As noted earlier, these analyses used coefficient estimates from the constrained piecewise regression equations as starting values, such that b_0, b_1, b_2, and b_3 were initiated at a_0^*, a_1^*, a_2^*, and a_3^*, whose values are reported Table 6, and c_0 and c_1 were started at 0 and 1, which are the intercept and slope of the $Y = X$ line. Coefficients from Table 8 are translated into slopes for X and Y on either side of each seam in Table 9, and the spline regression surfaces are plotted in Figure 12.

Comparing the results in Tables 6 and 8 shows that the estimates of b_0, b_1, b_2, and b_3 deviated from the starting values of a_0^*, a_1^*, a_2^*, and a_3^*. However, the key difference in the results is that, in Table 8, the intercepts and slopes of the seams are not constrained to 0 and 1, respectively, but instead are estimated parameters. As shown in Table 8, the intercepts differed from 0 for authority and relationships, and the slopes differed from 1 for relationships and variety. Only autonomy yielded intercept and slope estimates that did not statistically differ from 0 and 1. Thus, the spline regression results revealed that, for three of the four surfaces, the assumption that the seam ran along the $Y = X$ line was rejected, whereas for one surface, the seam did not deviate statistically from the $Y = X$ line. These findings are central to congruence hypotheses and cannot be obtained from piecewise regression analysis.

Surfaces corresponding to these results are shown in Figure 12. As is apparent, the seams for the authority, relationships, and variety surfaces deviate from the $Y = X$ line. The locations of the seams relative to the $Y = X$ line are further examined in Table 10, which reports the points at

which the seams of the surfaces cross the $Y = -2 - X$, $Y = -X$, and $Y = 2 - X$ lines, which correspond to X and Y coordinates of (–1, –1), (0, 0), and (1, 1), respectively, along the $Y = X$ line (details regarding the computation and interpretation of these values are provided in the Appendix). To facilitate interpretation, we designate the coordinates of (–1, –1), (0, 0), and (1, 1) as low, medium, and high levels, respectively, of X and Y. For authority, the seam is shifted to the right of the $Y = X$ line across all three levels of X and Y, which indicates that job satisfaction is consistently greater when actual authority exceeds desired authority. For relationships, the seam is rotated counterclockwise and shifted to the left of the $Y = X$ at medium and high levels of X and Y, meaning that job satisfaction is greater when actual relationships are less than desired relationship when both variables are above their midpoints. For variety, the seam is rotated counterclockwise such that, when actual and desired variety are both low, job satisfaction is greater when actual exceeds desired, whereas when actual and desired variety are both high, job satisfaction is greater when actual falls short of desired. Finally, for autonomy, job satisfaction was greater when actual autonomy was less than desired autonomy when both variables were high, but this finding should be interpreted in light of the fact that the seam for autonomy did not differ statistically from the $Y = X$ line.

In addition to examining the locations of the seams, the interpretations of the surfaces are facilitated by using the coefficient estimates in Table 8 to assess the slopes of the surface on either side of the seam, as given in Table 9. These results indicate that: (a) on the left of the seams, job satisfaction increases as actual amounts increase and, to a lesser extent, as desired amounts decrease; and (b) on the right of the seams, job satisfaction increases as desired amounts increase and, to a lesser extent, and actual amounts decrease. Thus, using the seams as reference lines, the increases in job satisfaction when actual amounts increased toward desired amounts were greater than the decreases in job satisfaction when actual amounts exceeded desired amounts.

A final set of results is reported in Table 11, which gives the intercepts and slopes of each surface along its seam. For all job dimensions, the surface was positively sloped along the seam,

indicating that job satisfaction increased as actual and desired amounts jointly increased. Taking into account the locations of the seams within the X,Y plane (see Figure 12), we conclude that job satisfaction was maximized when actual and desired authority were approximately equal, actual relationships were less than desired relationships by about two units, actual variety was less than desired variety by about one unit, and when actual and desired autonomy were about equal.

We also compared the R^2 values from the spline regression to those obtained using absolute difference scores, unconstrained piecewise regression, and constrained piecewise regression. As indicated in Table 12, the variance explained by the spline regression equations was statistically greater than that explained by the absolute difference, unconstrained piecewise, and constrained piecewise equations. Thus, spline regression was superior not only in terms of the substantive information yielded, but also in terms of the proportion of variance explained.

As a last step, we compared the one-seam spline regression equations to two-seam equations, using starting values from constrained piecewise regression equations that changed slopes along the $Y = 1 + X$ and $Y = -1 + X$ lines. For relationships, variety, and autonomy, only one seam remained within the range of the data. In contrast, both seams were within the range of the data for authority, although the increment in variance explained by the two-seam equation relative to the one-seam equation was only .002, which was trivial in substantive terms and did not reach statistical significance, $F(3, 941) = 0.60$, $p > .05$. For illustrative purposes, the surface for the two-seam equation for authority is shown in Figure 14, which indicates a small zone of indifference when actual and desired authority are both high. However, comparing this surface to the surface for the one-seam equation in Figure 13a shows that the surfaces are quite similar, which is consistent with the finding that the equations for both surfaces explained nearly the same amount of variance in job satisfaction.

Discussion

This article has presented spline regression as a useful analytical approach for the study of congruence in organizational research. Spline regression represents a key step forward in the development of alternatives to absolute difference scores, which have yet to progress beyond the

piecewise regression approach proposed by Edwards (1994). Although this approach avoids certain problems with absolute difference scores, it does not allow the researcher to determine whether outcomes are maximized or minimized along the $Y = X$ line. This assumption is central to congruence research, and the inability to corroborate or refute this assumption renders tests of congruence hypotheses incomplete. This assumption can be rigorously examined with spline regression, thereby allowing researchers to comprehensively test congruence hypotheses cast in terms of the absolute difference between two component variables, as set forth by Locke (1969), McGrath (1976), Venkatraman (1989), and implied by numerous empirical studies that have used absolute difference scores to represent the effects of congruence. Spline regression also enables researchers to investigate congruence hypotheses that go beyond those represented by absolute difference scores, such as surfaces with seams that are shifted or rotated from the $Y = X$ line and with multiple seams, as exemplified by two-seam surfaces that depict a zone of indifference that spans the $Y = X$ line. As such, spline regression constitutes an important contribution to the set of analytical tools available in congruence research.

With spline regression, congruence researchers now have viable alternatives to algebraic, squared, and absolute difference scores, which historically represented the three dominant ways that congruence has been operationalized (Edwards, 1994). For algebraic difference scores, a viable alternative is linear regression using the component variables as predictors, supplemented by analyses to determine whether the hypothesis represented by an algebraic difference is supported. For squared difference scores, researchers can rely on polynomial regression along with response surface methodology, in which the regression equation is a quadratic function of the component variables (Edwards, 2002; Edwards & Parry, 1993). For absolute difference scores, we now have spline regression, which overcomes limitations with piecewise regression and rounds out the alternatives to the three primary difference scores that historically dominated congruence research.

Spline regression bears certain similarities to polynomial regression, in that both methods provide alternatives to difference scores that depict functions in which an outcome is maximized

or minimized along the $Y = X$ line and decreases or increases symmetrically as the component variables differ in either direction. However, these two methods have some relevant differences. For instance, polynomial regression specifies the effects of congruence as curvilinear, whereas spline regression treats these effects as linear. In addition, polynomial regression requires that the effects of incongruence are symmetric, given that any cross-section of a quadratic surface is a parabolic function with equal but opposite curvature on either side of a vertical line termed the *axis of symmetry* (Kuang & Kase, 2012). In contrast, spline regression surfaces can have slopes that differ on either size of the seam. Moreover, the estimation of polynomial regression coefficients is exact, meaning there is only one solution for a given set of data and variables, whereas spline regression coefficient estimates can differ depending on the starting values and estimation algorithms used. In our experience, we have found that polynomial regression and spline regression tend to yield similar substantive interpretations when applied to the same data, with the primary differences involving the asymmetries and zones of indifference that spline regression can depict.

We see several worthwhile opportunities for the further development of spline regression. One opportunity involves the integration of spline regression and polynomial regression under a single unified analytical framework. This integration would begin by supplementing the linear terms used in spline regression with the three quadratic terms included in polynomial regression. The resulting equation would allow curvature of the surface on either side of the seam, which in turn would follow a quadratic rather than linear function through the X,Y plane. An equation of this form is as follows:

$$Z = b_0 + b_1X + b_2Y + b_3X^2 + b_4XY + b_5Y^2 + b_8(Y - c_0 - c_1X - c_2X^2)(Y < c_0 + c_1X + c_2X^2) + \text{e}. \quad (47)$$

Equation 47 subsumes spline regression and polynomial regression as special cases. Specifically, if b_8 equals zero, Equation 47 becomes the quadratic regression equation typically used in polynomial regression analysis. Alternately, if b_3, b_4, b_5, and c_2 equal zero, Equation 47 reduces to Equation 39, which is a spline regression equation with one linear seam. Thus far, our

limited attempts to estimate Equation 47 have brought challenges involving convergence and interpretation. As such, further work is needed to determine whether Equation 47 will provide the basis for generating a useful and practical method for integrating spline and polynomial regression.

Although the application of spline regression presented here has several strengths, it also has certain limitations that signify areas for further development. First, like OLS regression, spline regression relies on the assumption that the variables involved in the analysis are measured without error. Measurement error in Z will attenuate R^2 estimates, whereas measurement error in X and Y can bias coefficient estimates upward or downward. We have explored the estimation of spline regression estimation using structural equation modeling with latent variables but have yet to devise a viable procedure. Future work along these lines would be very beneficial.

Second, when applying spline regression, obtaining useful starting values can be difficult. The approaches we illustrated relied on theoretical assumptions about seam locations, which were used to code dummy variables for piecewise regression equations, the results of which were used to derive starting values for nonlinear estimation. If theoretical assumptions such as those we employed are grossly incorrect, starting values can differ markedly from values that best fit the data, and nonlinear estimation procedures can encounter local minima or fail to converge. In general, it is prudent to try different sets of starting values to determine whether parameter estimates are stable. If available theory does not suggest seam locations, then data can be plotted to visually identify where the dependent variable changes in slope relative to the independent variables, or systematic combinations of lines passing through the X,Y plane could be examined (e.g., all lines one unit apart that run parallel or perpendicular to the $Y = X$ line).

Third, because the estimation algorithms used to implement nonlinear regression are iterative, they can fail to converge in practice. Moreover, differ statistical packages rely on different algorithms and incorporate different convergence criteria, which means that the same data can produce different results depending on which package is used. For our example, we

applied four statistical packages (i.e., SPSS, Stata, SAS, and SYSTAT), and although the obtained results were very close, they were not identical. As such, the nonlinear estimation procedures used for spline regression involve some degree of instability that should be kept in mind when drawing conclusions from nonlinear regression.

In addition to these limitations, researchers who apply spline regression should keep various practical issues in mind. First, although we consider spline regression a method for testing *a priori* congruence hypotheses, it could also be used for exploratory research, as when knots or seams are progressively added to spline regression equations and the increment in R^2 is tested. Although such analyses are possible, they would rarely be useful, because congruence research has progressed beyond the point where exploratory analyses would be necessary or informative. For example, virtually all theories of congruence discuss how the fit between two constructs relates to outcomes, and such discussions are sufficient to hypothesize the location of a seam and slopes on either side of the seam, and testing hypotheses such as these calls for analyses that are confirmatory rather than exploratory. Second, the use of spline regression raises the usual considerations of statistical power. Although statistical power has received scant attention in the spline regression literature, it stands to reason that the statistical power associated with spline regression analysis is a function of sample size, effect size, and the chosen Type I error rate. In spline regression, the effects of interest include the location of the seam, the slopes of the surface on either side of the seam, the slope of the surface along the seam, and the change in slope of the surface at the seam. These effects can be more readily detected when samples are large in absolute terms and also when the data are distributed throughout the X,Y plane. If the data are clustered on one side of the hypothesized seam location, then a change in the slope of the surface along the seam, as well as the location of the seam itself, will be difficult to detect.

Summary and Conclusion

Spline regression holds great promise for the study of congruence in organizational research. In this article, we reviewed the fundamentals of spline regression and distinguished it from piecewise regression, explained methods for specifying and interpreting two-dimensional

spline regression functions, and extended these methods to three-dimensional surfaces. The empirical example we presented demonstrates how spline regression can be used to conduct detailed tests of congruence hypotheses and generate substantively meaningful results. We encourage researchers to consider spline regression as the culmination of alternatives to absolute difference scores and a useful companion to polynomial regression in the arsenal of analytical methods for congruence research.

References

Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 267-299). New York: Academic Press.

Bates, D. M., & Watts, D. G. (1988). *Nonlinear regression and its applications*. New York: Wiley.

Berry, W. D. (1993). *Understanding regression assumptions*. Newbury Park, CA: Sage.

Caplan, R. D. (1983). Person-environment fit: Past, present, and future. In C. L. Cooper (Ed.), *Stress research* (pp. 35-78). New York: Wiley.

Cronbach, L. J. (1958). Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior* (pp. 353-379). Stanford, CA: Stanford University Press.

Dawis, R. V. (1992). Person-environment fit and job satisfaction. In C. J. Cranny, P. C. Smith, & E. F. Stone (Eds.), *Job satisfaction* (pp. 69-88). New York: Lexington.

Edwards, J. R. (1991). Person-job fit: A conceptual integration, literature review, and methodological critique. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (vol. 6, pp. 283-357). New York: Wiley.

Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Processes, 58,* 51-100 (erratum, *58,* 323-325).

Edwards, J. R., Cable, D. M., Williamson, I. O., Lambert, L. S., & Shipp, A. J. (2006). The phenomenology of fit: Linking the person and environment to the subjective experience of fit. *Journal of Applied Psychology.*

Edwards, J. R., Caplan, R. D., & Harrison, R. V. (1998). Person-environment fit theory: Conceptual foundations, empirical evidence, and directions for future research. In C. L. Cooper (Ed.), *Theories of organizational stress* (pp. 28-67). Oxford: Oxford University Press.

Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an

alternative to difference scores in organizational research. *Academy of Management Journal, 36,* 1577-1613.

Edwards, J. R., & Shipp. A. J. (in press). The relationship between person-environment fit and outcomes: An integrative theoretical framework. In C. Ostroff and T. A. Judge (Eds.), *Perspectives on organizational fit*. San Francisco: Jossey-Bass.

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

French, J. R. P., Jr., Rodgers, W. L., & Cobb, S. (1974). Adjustment as person-environment fit. In G. Coelho, D. Hamburg, & J. Adams (Eds.), *Coping and adaptation* (pp. 316-333). New York: Basic Books.

Greene, W. H. (2003). *Econometric analysis* (5th Ed.). Prentice Hall.

Harrison, R. V. (1978). Person-environment fit and job stress. In C. L. Cooper & R. Payne (Eds.), *Stress at work* (pp. 175-205). New York: Wiley.

House, J. S. (1981). *Work, stress, and social support*. Reading, MA: Addison-Wesley Publishing Company.

Johns, G. (1981). Difference score measures of organizational behavior variables: A critique. *Organizational Behavior and Human Performance, 27,* 443-463.

Julious, S. A. (2001). Inference and estimation in a changepoint regression problem. *The Statistician, 50,* 51-61.

Kristof, A. L. (1996). Person-organization fit: An integrative review of its conceptualization, measurement, and implications. *Personnel Psychology, 49,* 1-49.

Kristof-Brown, A. L., Zimmerman, R. D., & Johnson, E. C. (2005). Consequences of individual's fit at work: A meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Personnel Psychology, 58,* 281-342.

Kroeger, N. W. (1995). Person-environment fit in the final jobs of retirees. *Journal of Social Psychology, 135,* 545-551.

Kuang, Y., & Kase, E. (2012). *Pre-calculus for dummies* (2nd ed.) New York: Wiley.

Kulka, R. A. (1979). Interaction as person-environment fit. In L. R. Kahle (Ed.), *New directions for methodology of behavioral science* (pp. 55-71). San Francisco: Jossey-Bass.

Locke, E. A. (1976). The nature and causes of job satisfaction. In M. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 1297-1350). Chicago: Rand McNally.

Marsh, L. C., & Cormier, D. R. (2002). *Spline regression models*. Thousand Oaks, CA: Sage.

Motulsky, H. J., & Ransnas, L. A. (1987). Fitting curves to data using nonlinear regression: A practical and nonmathematical review. *The FASEB Journal, 1,* 365-374.

Neter, J., Wasserman, W., & Kutner, M. H. (1989). *Applied linear regression models* (2nd ed.). Homewood, IL: Irwin.

Popper, K. R. (1959). *The logic of scientific discovery*. New York: Basic Books.

Rice, R. W., McFarlin, D. B., Hunt, R. G., & Near, J. P. (1985). Organizational work and the perceived quality of life: Toward a conceptual model. *Academy of Management Review, 10,* 296-310.

Seber, G. A. F., & Wild, C. J. (2003). *Nonlinear regression*. New York: Wiley.

Smith, P. L. (1979). Splines as a useful and convenient statistical tool. *The American Statistician, 33,* 57-62.

Stine, R. (1989). An introduction to bootstrap methods. *Sociological Methods & Research, 18,* 243-291.

Suits, D., Mason, A., & Chan, L. (1978). Spline functions fitted by standard regression methods. *The American Statistician, 60,* 132-139.

Wold, S. (1974). Spline functions in data analysis. *Technometrics, 16,* 1-11.

Footnotes

[1] The derivations that follow are intentionally thorough and take the reader through every step that leads from two-dimensional piecewise and spline functions to three-dimensional functions for absolute difference scores, piecewise regression, and spline regression. For readers who wish to skip these derivations, the key equations that result are as follows: (1) Equation 11 is a two-dimensional spline function with one knot; (2) Equation 14 is a two-dimensional spline function with two knots; (3) Equation 39 is a three-dimensional spline surface with one seam; and (4) Equation 43 is a three-dimensional spline surface with two seams.

[2] In practice, W can be coded either 0 or 1 with for cases where $X = c$. For this article, we chose to code W as 0 for these cases, acknowledging that this choice is arbitrary. Fortunately, when the piecewise function is constrained to be continuous, the obtained results are the same regardless of whether W equals 0 or 1 for cases where $X = c$. This condition applies to work presented here, which focuses on continuous functions.

[3] Although Edwards (1994) expressed the absolute difference between X and Y as $|X - Y|$, we use $|Y - X|$ because, in the derivations that follow, the dummy variable W will be coded 0 on the left side of the surface and 1 on the right side of the surface. This coding parallels that used for two-dimensional piecewise functions, for which W was coded 0 for the left portion of the function and 1 for the right portion of the function (see Equation 1).

[4] We suspect some researchers might feel uneasy about the notion that spline regression results can vary depending on the estimation algorithm, starting values, and convergence criteria used. However, these nuances are shared by other methods of analysis, such as structural equation modeling with maximum likelihood estimation. These issues underscore the fact that analytical choices can influence the obtained results, and it is prudent to consider different options and compare how they influence the conclusions drawn from an investigation.

Appendix

Determining Where Seams Cross Lines of Interest

To determine where a seam of a spline surface crosses a line of interest, we begin with an equation for the seam (for these derivations, we consider a surface with one seam):

$$Y = c_0 + c_1X \tag{A1}$$

We start with the $Y = -X$ line, which runs perpendicular to the $Y = X$ line and intersects at the point $X = 0$, $Y = 0$. To find where the seam crosses the $Y = -X$ line, we substitute the equation for this line into Equation A1, replacing Y with $-X$:

$$-X = c_0 + c_1X. \tag{A2}$$

Solving for X gives the point on the X axis where the seam crosses the $Y = -X$ line:

$$X = -c_0/(c_1 + 1). \tag{A3}$$

Note that Equation A3 is express in terms of X. To recover the corresponding value of Y, we note that Y is the negative of X along the $Y = -X$ line, which means the Y coordinate where the seam crosses the $Y = -X$ line is the negative of the right side of Equation A3, which is:

$$Y = c_0/(c_1 + 1). \tag{A4}$$

Hence, the point at which the seam crosses the $Y = -X$ line is $X = -c_0/(c_1 + 1)$, $Y = c_0/(c_1 + 1)$.

We now derive the distance of this point from the $Y = X$ line, which indicates how far the seam is shifted from the line of congruence along the $Y = -X$ line. This distance is the difference between the point $X = 0$, $Y = 0$, where the $Y = X$ line intersects the $Y = -X$ line, and the point $X = -c_0/(c_1 + 1)$, $Y = c_0/(c_1 + 1)$, which is the point where the seam crosses the $Y = -X$ line. Because the X and Y axes are perpendicular, the distance between these points along the $Y = -X$ line can be conceived as the diagonal of a right triangle, where the side with respect to the X axis is the difference between the points $X = 0$ and $X = -c_0/(c_1 + 1)$, which is $-c_0/(c_1 + 1)$, and the side with respect to the Y axis is the difference between the points $Y = 0$ and $Y = c_0/(c_1 + 1)$, or $c_0/(c_1 + 1)$. Using the Pythagorean theorem, the length of the diagonal can be computed as $c = \sqrt{a^2 + b^2}$, where c is the length of the diagonal along the $Y = -X$ line and a and b are the sides with respect to the X and Y axes, which are $-c_0/(c_1 + 1)$ and $c_0/(c_1 + 1)$, respectively. Because the length of

one side is the negative of the other, a = –b, the expression under the radical $a^2 + b^2$ simplifies to $2a^2$ or, equivalently, $2b^2$, and therefore $c = \sqrt{2}a = \sqrt{2}b$. When computing c, we recommend using $\sqrt{2}b$, because b refers to the shift of the seam from the $Y = X$ line with respect to the Y axis, such that negative values of c indicate a downward shift and positive values signify an upward shift. Thus, we compute the shift of the seam from the $Y = X$ line as $c_{Y=-X} = \sqrt{2}c_0/(c_1 + 1)$, where the subscript $Y = -X$ on c refers to shift along the $Y = -X$ line.

The interpretation of $c_{Y=-X}$ can be clarified by referring to Figure A1, which plots the seam for the surface relating actual and desired relationships to job satisfaction, corresponding to the floor of Figure 12b. Drawing from Table 8, the estimates of c_0 and c_1 are 1.54 and 3.32, respectively. Substituting these values into the equation $c_{Y=-X} = \sqrt{2}c_0/(c_1 + 1)$ yields a value of 0.51. This value is reported in Table 10 and shown in Figure A1 near the center of the graph. As indicated in Table 10, this value differs from zero, meaning that the upward shift of the seam from the $Y = X$ line along the $Y = -X$ line can be deemed statistically significant.

Other lines that run parallel $Y = -X$ line can also be of interest. For instance, consider the $Y = 2 - X$ line, which crosses the $Y = X$ line at the point $X = 1$, $Y = 1$. Substituting this expression into the equation for the seam gives:

$$2 - X = c_0 + c_1X. \tag{A5}$$

Solving for X yields:

$$X = (2 - c_0)/(c_1 + 1). \tag{A6}$$

This is the point where the seam crosses the $Y = 2 - X$ line, as measured on the X axis. The point on the Y axis is found by substituting Equation A5 into $Y = 2 - X$, which gives:

$$Y = 2 - (2 - c_0)/(c_1 + 1). \tag{A7}$$

Thus, the seam crosses the $2 - X$ line at the point $X = (2 - c_0)/(c_1 + 1)$, $Y = 2 - (2 - c_0)/(c_1 + 1)$. The distance of this point from the point where the $2 - X$ line intersects the $Y = X$ line is found by subtracting the X and Y values for this intersection, which are $X = 1$, $Y = 1$. Subtracting these values and rearranging terms yields $X = (1 - c_0 - c_1)/(c_1 + 1)$, $Y = -(1 - c_0 - c_1)/(c_1 + 1)$. Again, these values for X and Y serve as a and b in the equation for the Pythagorean theorem, and

applying the expression $c = \sqrt{2}b$ derived earlier gives $c_{Y=2-X} = -\sqrt{2}(1 - c_0 - c_1)/(c_1 + 1)$. Using the estimates 1.54 and 3.32 for c_0 and c_1, we obtain $c_{Y=2-X} = 1.26$, as reported in Table 10 and shown in the upper right corner of Figure A1. This value differs statistically from zero, as shown in Table 10.

As a third example, we address the $Y = -2 - X$ line, which crosses the $Y = X$ line at the point $X = -1$, $Y = -1$. Substituting $Y = -2 - X$ into the seam equation yields:

$$-2 - X = c_0 + c_1X. \quad \text{(A8)}$$

Solving for X gives:

$$X = -(2 + c_0)/(c_1 + 1). \quad \text{(A9)}$$

Hence, the seam crosses the $Y = -2 - X$ line at the point $(-2 - c_0)/(c_1 + 1)$ on the X axis. The corresponding point on the Y axis, which is found by substituting Equation A9 into $Y = -2 - X$, is:

$$Y = -2 + (2 + c_0)/(c_1 + 1). \quad \text{(A10)}$$

The distance from this point to the point where the $-2 - X$ line crosses the $Y = X$ line is found by subtracting the coordinates $X = -1$, $Y = -1$. Performing these operations and rearranging terms yields $X = -(1 + c_0 - c_1)/(c_1 + 1)$, $Y = (1 + c_0 - c_1)/(c_1 + 1)$. Substituting the Y coordinate for b into the expression $c = \sqrt{2}b$ yields $c = \sqrt{2}(1 + c_0 - c_1)/(c_1 + 1) = -.0.25$, which is the distance from the $Y = X$ line to the seam along the $Y = -2 - X$ line, as depicted in the lower left corner of Figure A1. As shown in Table 10, this value does not statistically differ from zero.

The procedures developed here can be applied to other lines of interest to the researcher. For congruence hypotheses, we believe it is worthwhile to consider lines that run parallel to the $Y = -X$ line that cross the $Y = X$ line at points that represent unit increases and decreases of both X and Y within the range of the data. In the present example, these points were $X = 1$, $Y = 1$ and $X = -1$, $Y = -1$. If X and Y had been measured on 7-point scales, such that their scale-centered counterparts ranged from –3 to 3, then we would add the points $X = 2$, $Y = 2$ and $X = -2$, $Y = -2$, such that the lines of interest would be $Y = -4 - X$, $Y = -2 - X$, $Y = -X$, $Y = 2 - X$, and $Y = 4 - X$. Of course, these are general guidelines, and the researcher could choose other lines as dictated by

the hypothesis under consideration. Again, tests of where the seam crosses these lines should only be conducted with the range of the data, as represented by the bivariate scatterplot of the observed values of *X* and *Y*.

Table 1

Means, Standard Deviations, Reliabilities, and Correlations of Study Variables

Variable	*M*	*SD*	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. Actual Authority	3.31	0.98	(.90)								
2. Desired Authority	3.57	0.72	.02	(.87)							
3. Actual Relationships	2.85	0.81	.28	.18	(.90)						
4. Desired Relationships	3.27	0.79	.12	.28	.39	(.91)					
5. Actual Variety	3.16	0.90	.23	.20	.42	.17	(.90)				
6. Desired Variety	3.67	0.70	.19	.25	.21	.32	.39	(.88)			
7. Actual Autonomy	3.10	0.82	.30	.20	.47	.16	.63	.28	(.82)		
8. Desired Autonomy	3.74	0.64	.13	.29	.17	.28	.26	.57	.35	(.79)	
9. Job Satisfaction	3.68	0.92	.22	.18	.36	.13	.35	.13	.44	.12	(.90)

Note: $N = 950$. Reliability estimates (Cronbach's alpha) are reported along the diagonal. Correlations greater than .06 in absolute magnitude are statistically significant ($p < .05$).

Table 2

Results for Absolute Difference Scores Predicting Job Satisfaction

	d_0	d_1	R^2
Authority	3.97**	–0.32**	.09**
LL_N	3.89	–0.38	
UL_N	4.05	–0.25	
Relationships	3.88**	–0.29**	.05**
LL_N	3.80	–0.37	
UL_N	3.96	–0.21	
Variety	3.97**	–0.40**	.10**
LL_N	3.90	–0.48	
UL_N	4.05	–0.33	
Autonomy	4.03**	–0.46**	.14**
LL_N	3.95	–0.54	
UL_N	4.11	–0.39	

Note: $N = 950$. For columns labeled d_0 and d_1, table entries are unstandardized regression coefficients estimated using Equation 18, followed by the lower and upper limits of 95% confidence intervals (LL_N and UL_N, respectively), with the subscript N indicating that the intervals are based on normal theory, computed from the standard errors reported in the regression output. For the column labeled R^2, table entries are squared multiple correlations.

* $p < .05$ ** $p < .01$

Table 3

Absolute Difference Score Surfaces on Left and Right of Y = X Line

	Left of $Y = X$ Line			Right of $Y = X$ Line		
	INT	X	Y	INT	X	Y
	d_0	$-d_1$	d_1	d_0	d_1	$-d_1$
Authority	3.97**	0.32**	–0.32**	3.97**	–0.32**	0.32**
LL_N	3.89	0.24	–0.38	3.89	–0.38	0.25
UL_N	4.05	0.38	–0.25	4.05	–0.25	0.38
Relationships	3.88**	0.29**	–0.29**	3.88**	–0.29**	0.29**
LL_N	3.80	0.21	–0.37	3.80	–0.37	0.21
UL_N	3.96	0.37	–0.21	3.96	–0.21	0.37
Variety	3.97**	0.40**	–0.40**	3.97**	–0.40**	0.40**
LL_N	3.90	0.33	–0.48	3.90	–0.48	0.33
UL_N	4.05	0.48	–0.33	4.05	–0.33	0.48
Autonomy	4.03**	0.46**	–0.46**	4.03**	–0.46**	0.46**
LL_N	3.95	0.39	–0.54	3.95	–0.54	0.39
UL_N	4.11	0.54	–0.39	4.11	–0.39	0.54

Note: $N = 950$. Table entries are based on unstandardized regression coefficients estimated using Equation 18, followed by the lower and upper limits of 95% confidence intervals (LL_N and UL_N, respectively), with the subscript N indicating that the intervals are based on normal theory, computed from the standard errors reported in the regression output.

* $p < .05$ ** $p < .01$

Table 4

Results for Unconstrained Piecewise Regression Equations Predicting Job Satisfaction

	a_0	a_1	a_2	a_3	a_4	a_5	R^2
Authority	3.71**	0.42**	0.02	0.12	–0.66**	0.52**	.15**
LL_N	3.59	0.34	–0.10	–0.07	–0.82	0.34	
UL_N	3.84	0.51	0.14	0.31	–0.51	0.70	
Relationships	3.84**	0.51**	–0.09	–0.02	–0.36**	0.29**	.14**
LL_N	3.74	0.42	–0.19	–0.18	–0.58	0.08	
UL_N	3.93	0.61	0.01	0.14	–0.14	0.50	
Variety	3.81**	0.56**	–0.15**	–0.03	–0.74**	0.57**	.17**
LL_N	3.70	0.47	–0.26	–0.21	–0.96	0.33	
UL_N	3.91	0.65	–0.05	0.15	–0.51	0.82	
Autonomy	3.79**	0.61**	–0.15**	0.01	–0.40**	0.73**	.22**
LL_N	3.68	0.53	–0.25	–0.19	–1.14	0.42	
UL_N	3.89	0.70	–0.04	0.21	–0.53	1.05	

Note: $N = 950$. For columns labeled a_0 through a_5, table entries are unstandardized regression coefficients estimated using Equation 21. Below the coefficients are the lower and upper limits of 95% confidence intervals (LL_N and UL_N, respectively), with the subscript N indicating that the intervals are based on normal theory, computed from the standard errors reported in the regression output. For the column labeled R^2, table entries are squared multiple correlations.

* $p < .05$ ** $p < .01$

Table 5

Unconstrained Piecewise Regression Surfaces on Left and Right of Y = X Line

	Left of $Y = X$ Line			Right of $Y = X$ Line		
	INT	X	Y	INT	X	Y
	a_0	a_1	a_2	$a_0 + a_3$	$a_1 + a_4$	$a_2 + a_5$
Authority	3.71**	0.42**	0.02	3.83**	–0.24**	0.54**
LL_N	3.59	0.34	–0.09	3.69	–0.37	0.40
UL_N	3.84	0.51	0.14	3.98	–0.11	0.68
Relationships	3.84**	0.51**	–0.09	3.82	0.15	0.20*
LL_N	3.74	0.42	–0.19	3.69	–0.04	0.02
UL_N	3.93	0.61	0.01	3.94	–0.35	0.38
Variety	3.80**	0.56**	–0.15**	3.78**	–0.18	0.42**
LL_N	3.70	0.47	–0.26	3.63	–0.38	0.20
UL_N	3.91	0.65	–0.05	3.92	0.03	0.64
Autonomy	3.79**	0.61**	–0.15**	3.80**	–0.23	0.58**
LL_N	3.68	0.53	–0.25	3.62	–0.52	0.29
UL_N	3.89	0.69	–0.04	3.97	0.07	0.88

Note: $N = 950$. Table entries are based on unstandardized regression coefficients estimated using Equation 21. Below the coefficients are the lower and upper limits of 95% confidence intervals (LL_N and UL_N, respectively), with the subscript N indicating that the intervals are based on normal theory, computed from the standard errors reported in the regression output. For the column labeled R^2, table entries are squared multiple correlations.

* $p < .05$ ** $p < .01$

Table 6

Results for Constrained Piecewise Regression Equations Predicting Job Satisfaction

	a_0^*	a_1^*	a_2^*	a_3^*	R^2
Authority	3.72**	0.42**	–0.03	0.60**	.15**
LL_N	3.67	0.34	–0.13	0.46	
UL_N	3.86	0.49	0.07	0.74	
Relationships	3.83**	0.50**	–0.10*	0.33**	.14**
LL_N	3.75	0.41	–0.19	0.14	
UL_N	3.90	0.58	–0.01	0.52	
Variety	3.79**	0.51**	–0.17**	0.72**	.16**
LL_N	3.70	0.43	–0.26	0.51	
UL_N	3.88	0.59	–0.07	0.93	
Autonomy	3.79**	0.59**	–0.15**	0.82**	.22**
LL_N	3.70	0.52	–0.25	0.55	
UL_N	3.87	0.67	–0.06	1.10	

Note: $N = 950$. For columns labeled a_0^* through a_3^*, table entries are unstandardized regression coefficients estimated using Equation 29. Below the coefficients are the lower and upper limits of 95% confidence intervals (LL_N and UL_N, respectively), with the subscript N indicating that the intervals are based on normal theory, computed from the standard errors reported in the regression output. For the column labeled R^2, table entries are squared multiple correlations.

* $p < .05$ ** $p < .01$

Table 7

Constrained Piecewise Regression Surfaces on Left and Right of Y = X Line

	Left of $Y = X$ Line			Right of $Y = X$ Line		
	INT	X	Y	INT	X	Y
	a_0^*	a_1^*	a_2^*	a_0^*	$a_1^* - a_3^*$	$a_2^* + a_3^*$
Authority	3.76**	0.42**	–0.03	3.76**	–0.19**	0.57**
LL_N	3.67	0.34	–0.12	3.67	–0.29	0.46
UL_N	3.86	0.49	0.07	3.86	–0.08	0.68
Relationships	3.83**	0.50**	–0.10*	3.83**	0.17*	0.23**
LL_N	3.75	0.41	–0.19	3.75	0.00	0.07
UL_N	3.90	0.58	–0.01	3.90	0.33	0.39
Variety	3.79**	0.51**	–0.17**	3.79**	–0.21*	0.56**
LL_N	3.70	0.43	–0.26	3.70	–0.39	0.37
UL_N	3.88	0.59	–0.07	3.88	–0.03	0.74
Autonomy	3.78**	0.59**	–0.15**	3.78**	–0.23	0.67**
LL_N	3.70	0.52	–0.25	3.70	–0.48	0.41
UL_N	3.87	0.67	–0.06	3.87	0.02	0.92

Note: $N = 950$. Table entries are based on unstandardized regression coefficients estimated using Equation 29. Below the coefficients are the lower and upper limits of 95% confidence intervals (LL_N and UL_N, respectively), with the subscript N indicating that the intervals are based on normal theory, computed from the standard errors reported in the regression output. For the column labeled R^2, table entries are squared multiple correlations.

* $p < .05$ ** $p < .01$

Table 8

Results for Spline Regression Equations Predicting Job Satisfaction

	b_0	b_1	b_2	b_3	c_0	c_1	R^2
Authority	3.69**	0.39**	0.02	0.73**	–0.72**	1.20**	.16**
LL_N	3.59	0.32	–0.07	0.50	–1.17	0.82	
UL_N	3.78	0.46	0.12	0.97	–0.28	1.59	
LL_B	3.59	0.30	–0.08	0.46	–1.83	0.81	
UL_B	3.78	0.47	0.13	1.03	–0.29	1.86	
Relationships	4.14**	0.84**	–0.10	0.21**	1.54*	3.32**†	.16**
LL_N	3.90	0.61	–0.21	0.06	0.18	1.11	
UL_N	4.38	1.06	0.01	0.37	2.91	5.52	
LL_B	3.85	0.54	–0.23	0.04	–0.01	1.80	
UL_B	4.51	1.17	0.03	0.37	6.18	11.98	
Variety	3.88**	0.69**	–0.15**	0.55**	0.12	1.59**†	.18**
LL_N	3.75	0.56	–0.27	0.35	–0.20	1.14	
UL_N	4.02	0.81	–0.04	0.76	0.45	2.04	
LL_B	3.76	0.53	–0.29	0.28	–0.03	1.18	
UL_B	4.08	0.88	–0.02	0.81	0.71	2.69	
Autonomy	3.85**	0.69**	–0.17**	0.73**	0.09	1.23**	.23**
LL_N	3.72	0.58	–0.28	0.44	–0.16	0.94	
UL_N	3.97	0.79	–0.05	1.01	0.34	1.53	
LL_B	3.73	0.58	–0.30	0.37	–0.26	0.99	
UL_B	3.98	0.80	–0.04	1.14	0.35	1.67	

Note: $N = 950$. For columns labeled b_0 through c_1, table entries are unstandardized spline regression coefficients estimated using Equation 37. Below the coefficients, the rows labeled

LL_N and UL_N contain the lower and upper limits, respectively, of 95% confidence intervals based on normal theory, and the rows labeled LL_B and UL_B contain the lower and upper limits, respectively, of 95% bias-corrected percentile based confidence intervals derived using the bootstrap. The column labeled R^2 contains squared multiple correlations.

* $p < .05$ ** $p < .01$

† The 95% confidence interval for c_1 excluded 1.00.

Table 9

Spline Regression Surfaces on Left and Right of Seam

	Left of Seam			Right of Seam		
	INT	*X*	*Y*	INT	*X*	*Y*
	b_0	b_1	b_2	$b_0 - b_3c_0$	$b_1 - b_3c_1$	$b_2 + b_3$
Authority	3.69**	0.39**	0.02	4.22**	–0.49**	0.76**
LL_N	3.59	0.32	–0.07	3.90	–0.72	0.54
UL_N	3.78	0.46	0.12	4.53	–0.27	0.97
LL_B	3.59	0.30	–0.08	3.90	–0.85	0.52
UL_B	3.78	0.47	0.13	4.76	–0.25	1.03
Relationships	4.14**	0.84**	–0.10	3.82**	0.14	0.11*
LL_N	3.90	0.61	–0.21	3.74	–0.01	–0.00
UL_N	4.38	1.06	0.01	3.90	0.29	0.22
LL_B	3.85	0.54	–0.23	3.74	–0.06	0.00
UL_B	4.51	1.17	0.03	3.92	0.28	0.24
Variety	3.88**	0.69**	–0.15**	3.82**	–0.19	0.40**
LL_N	3.75	0.56	–0.27	3.69	–0.38	0.23
UL_N	4.02	0.81	–0.04	3.94	–0.01	0.57
LL_B	3.76	0.53	–0.29	3.71	–0.40	0.19
UL_B	4.08	0.88	–0.02	3.92	0.02	0.63
Autonomy	3.85**	0.69**	–0.17**	3.78**	–0.21	0.56**
LL_N	3.72	0.58	–0.28	3.65	–0.48	0.30
UL_N	3.97	0.79	–0.05	3.92	0.06	0.82
LL_B	3.73	0.58	–0.30	3.67	–0.64	0.24

UL_B	3.98	0.80	–0.04	3.96	0.14	0.97

Note: $N = 950$. Table entries are based on unstandardized spline regression coefficients estimated using Equation 37. Below the coefficients, the rows labeled LL_N and UL_N contain the lower and upper limits, respectively, of 95% confidence intervals based on normal theory, and the rows labeled LL_B and UL_B contain the lower and upper limits, respectively, of 95% bias-corrected percentile based confidence intervals derived using the bootstrap.

* $p < .05$ ** $p < .01$

Table 10

Shifts of Seams Along Three Lines of Interest

	Shift Along $Y = -2 - X$ line	Shift Along $Y = -X$ line	Shift Along $Y = 2 - X$ line
	$\frac{\sqrt{2}(c_0 - 2c_1)}{c_1 + 1}$	$\frac{\sqrt{2}c_0}{c_1 + 1}$	$\frac{\sqrt{2}(c_0 + 2c_1)}{c_1 + 1}$
Authority	–0.59*	–0.46**	–0.33*
LL_N	–1.00	–0.69	–0.53
UL_N	–0.19	–0.23	–0.13
LL_B	–1.31	–0.85	–0.52
UL_B	–0.16	–0.21	–0.07
Relationships	–0.26	0.50*	1.26**
LL_N	–1.00	0.20	0.76
UL_N	0.19	0.81	1.77
LL_B	–0.65	0.00	0.41
UL_B	0.14	0.98	2.02
Variety	–0.26*	0.07	0.39*
LL_N	–0.56	–0.11	0.18
UL_N	0.05	0.25	0.60
LL_B	–0.62	–0.02	0.13
UL_B	–0.00	0.36	0.80
Autonomy	–0.09	0.06	0.20*
LL_N	–0.38	–0.11	0.03
UL_N	0.19	0.22	0.37
LL_B	–0.48	–0.15	0.04

UL_B	0.17	0.23	0.34

Note: $N = 950$. Table entries are based on unstandardized spline regression coefficients estimated using Equation 37. Below the coefficients, the rows labeled LL_N and UL_N contain the lower and upper limits, respectively, of 95% confidence intervals based on normal theory, and the rows labeled LL_B and UL_B contain the lower and upper limits, respectively, of 95% bias-corrected percentile based confidence intervals derived using the bootstrap.

* $p < .05$ ** $p < .01$

Table 11

Intercepts and Slopes Along Seams

	Intercept	Slope
	$b_0 + b_2c_0$	$b_1 + b_2c_1$
Authority	3.67**	0.42**
LL_N	3.51	0.30
UL_N	3.83	0.54
LL_B	3.43	0.30
UL_B	3.81	0.60
Relationships	3.98**	0.50**
LL_N	3.84	0.26
UL_N	4.13	0.73
LL_B	3.82	0.25
UL_B	4.24	0.91
Variety	3.87**	0.44**
LL_N	3.77	0.31
UL_N	3.96	0.58
LL_B	3.76	0.30
UL_B	3.96	0.64
Autonomy	3.83**	0.48**
LL_N	3.74	0.37
UL_N	3.93	0.59
LL_B	3.73	0.38
UL_B	3.92	0.61

Note: N = 950. Table entries are based on unstandardized spline regression coefficients estimated using Equation 37. Below the coefficients, the rows labeled LL_N and UL_N contain the lower and upper limits, respectively, of 95% confidence intervals based on normal theory, and the rows labeled LL_B and UL_B contain the lower and upper limits, respectively, of 95% bias-corrected percentile based confidence intervals derived using the bootstrap.

* $p < .05$ ** $p < .01$

Table 12

Results for Constrained Piecewise Regression Equations Predicting Job Satisfaction

	R^2_{AD}	R^2_{UP}	R^2_{CP}	R^2_{SP}
Authority				
Relationships	$.08^{**}$	$.15^{**}_{a}$	$.15^{**}_{a}$	$.16^{**}_{a,b,c}$
Variety	$.05^{**}$	$.14^{**}_{a}$	$.14^{**}_{a}$	$.16^{**}_{a,b,c}$
Autonomy	$.10^{**}$	$.17^{**}_{a}$	$.16^{**}_{a}$	$.23^{**}_{a,b,c}$
	$.14^{**}$	$.22^{**}_{a}$	$.22^{**}_{a}$	$.23^{**}_{a,b,c}$

Note: $N = 950$. Table entries are square multiple correlations obtained from Equations 18, 21, 29, and 37, which correspond to the absolute difference score, unconstrained piecewise regression, constrained piecewise regression, and spline regression, respectively. The subscripts a, b, and c indicate that the squared multiple correlation was larger than that obtained from Equations 18, 21, and 29, respectively.

* $p < .05$ ** $p < .01$

Figure Captions

Figure 1. Two-dimensional unconstrained piecewise regression functions.

Figure 2. Two-dimensional constrained piecewise regression functions.

Figure 3. Two-dimensional spline regression functions with one knot.

Figure 4. Two-dimensional spline regression functions with two knots.

Figure 5. Three-dimensional absolute difference functions.

Figure 6. Three-dimensional unconstrained piecewise regression functions.

Figure 7. Three-dimensional constrained piecewise regression functions.

Figure 8. Three-dimensional spline regression functions with one seam.

Figure 9. Three-dimensional spline regression functions with two seams.

Figure 10. Estimated surfaces for absolute difference scores.

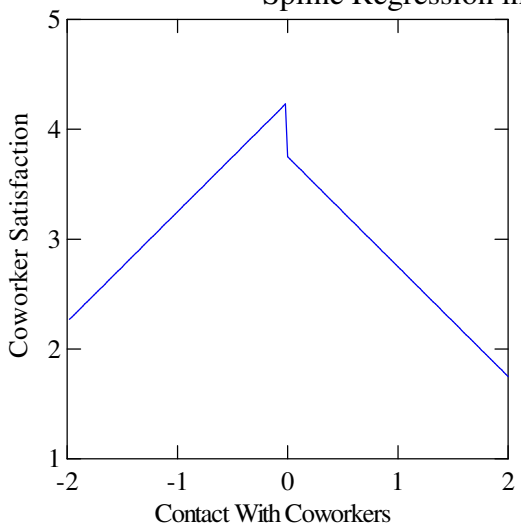
Figure 11. Estimated surfaces for unconstrained piecewise regression equations.

Figure 12. Estimated surfaces for constrained piecewise regression equations.

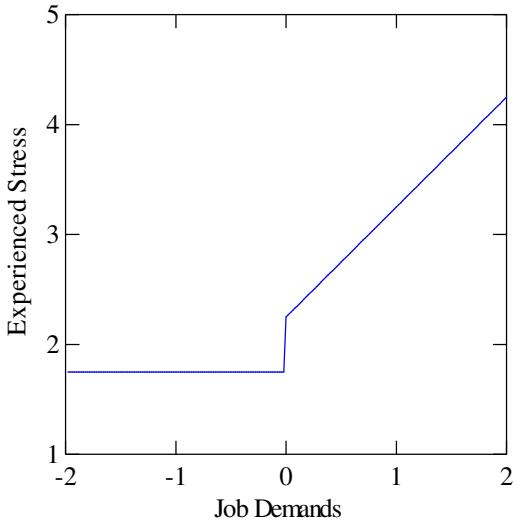
Figure 13. Estimated surfaces for spline piecewise regression equations with one seam.

Figure 14. Estimated spline regression surface with two seams for authority.

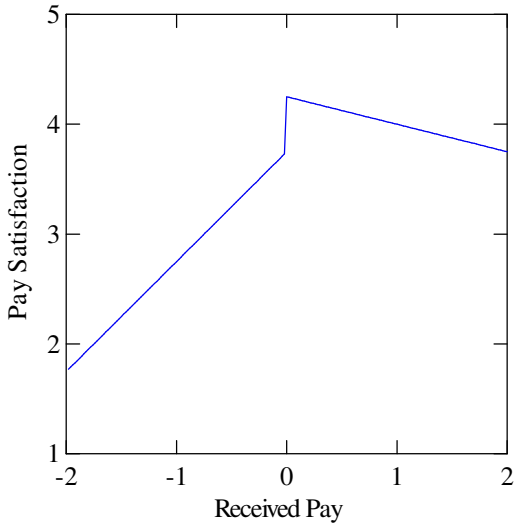
Figure A1. Illustration of seam crossing three lines of interest for relationships.



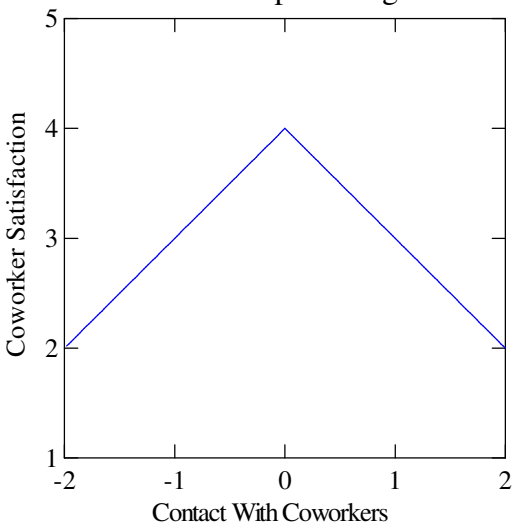
a. Contact with coworkers and coworker satisfaction



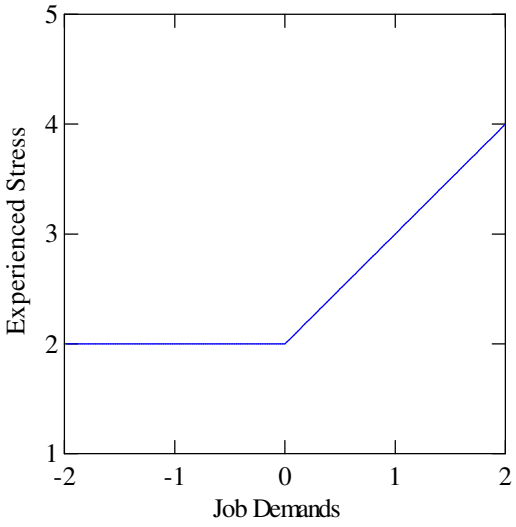
b. Job demands and experienced stress



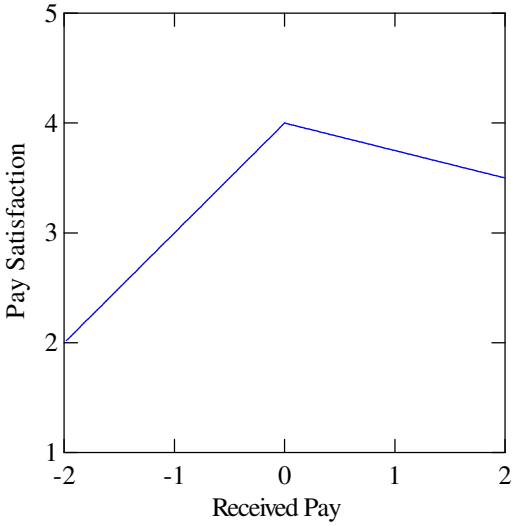
c. Received pay and pay satisfaction



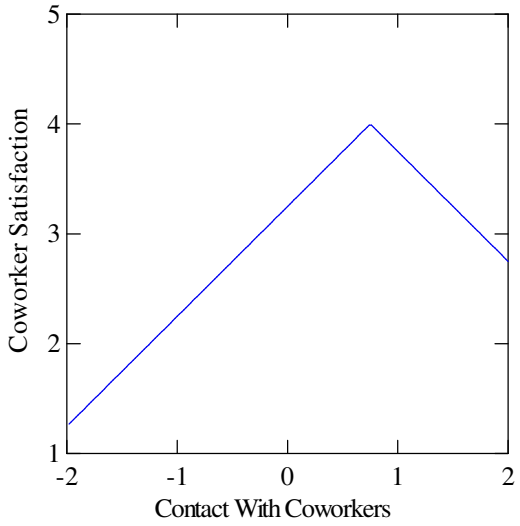
a. Contact with coworkers and coworker satisfaction



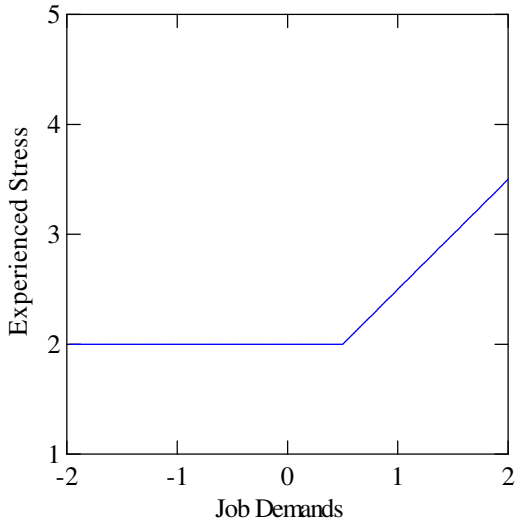
b. Job demands and experienced stress



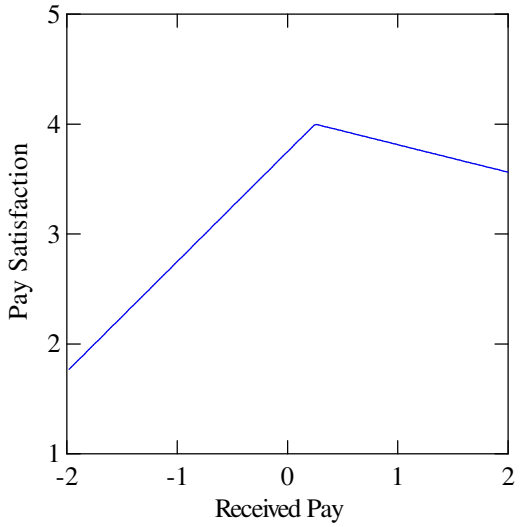
c. Received pay and pay satisfaction



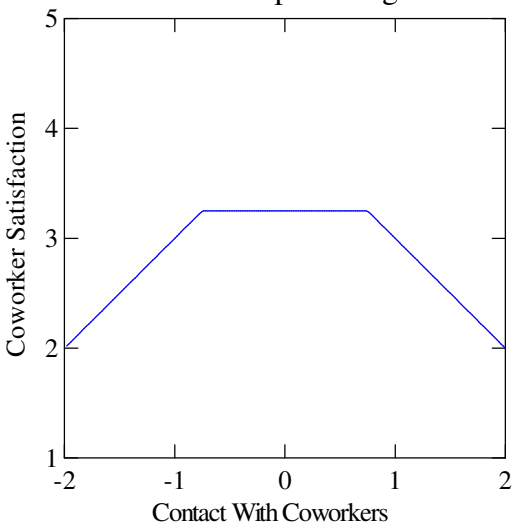
a. Contact with coworkers and coworker satisfaction



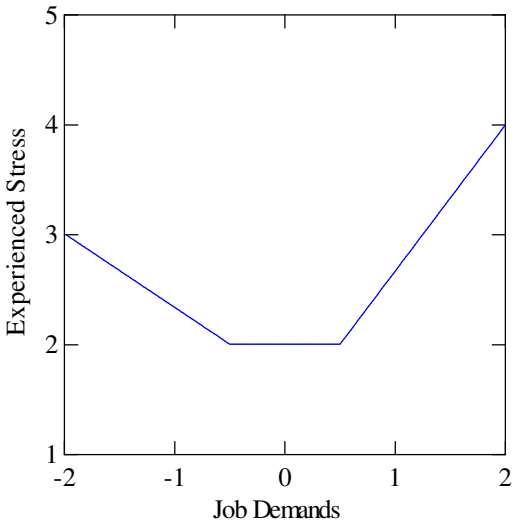
b. Job demands and experienced stress



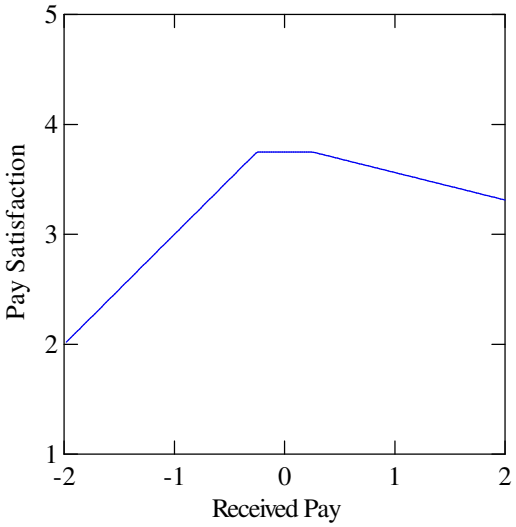
c. Received pay and pay satisfaction



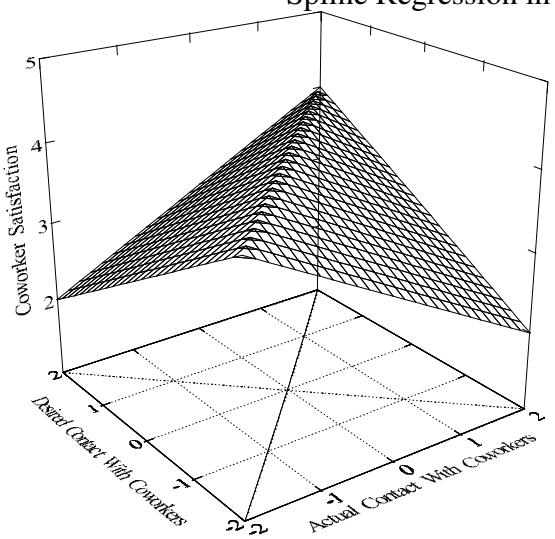
a. Contact with coworkers and coworker satisfaction



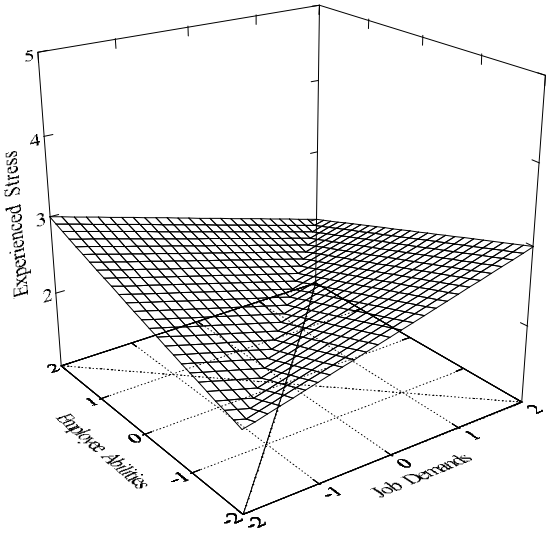
b. Job demands and experienced stress



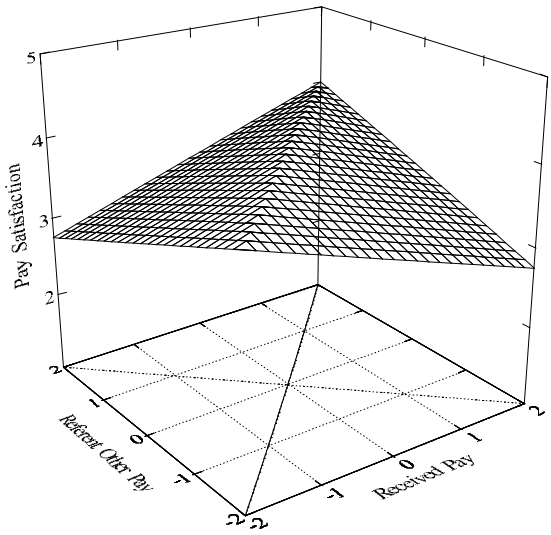
c. Received pay and pay satisfaction



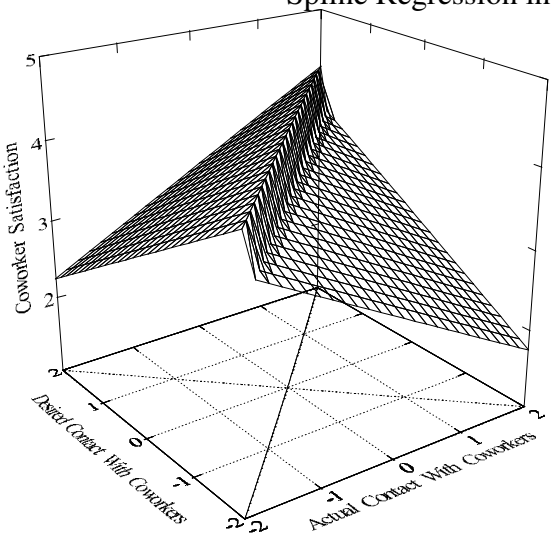
a. Actual and desired contact with coworkers and coworker satisfaction



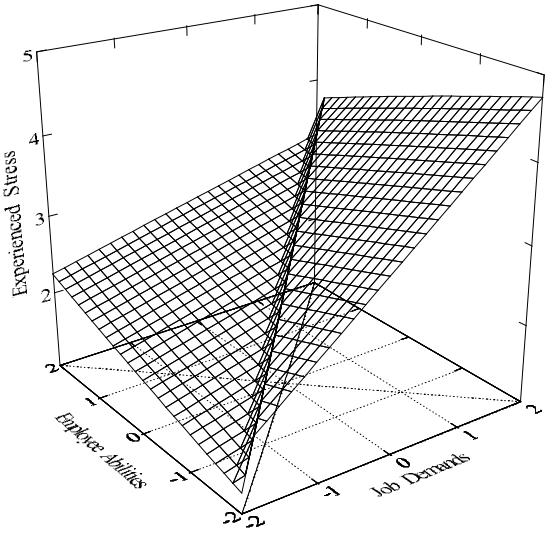
b. Job demands, employee abilities, and experienced stress



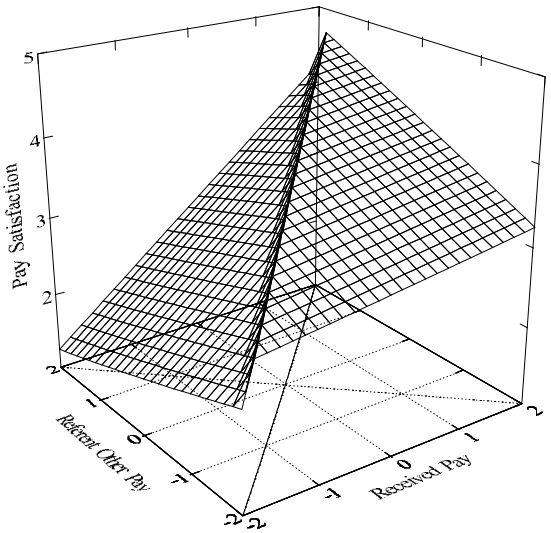
c. Received and referent other pay and pay satisfaction



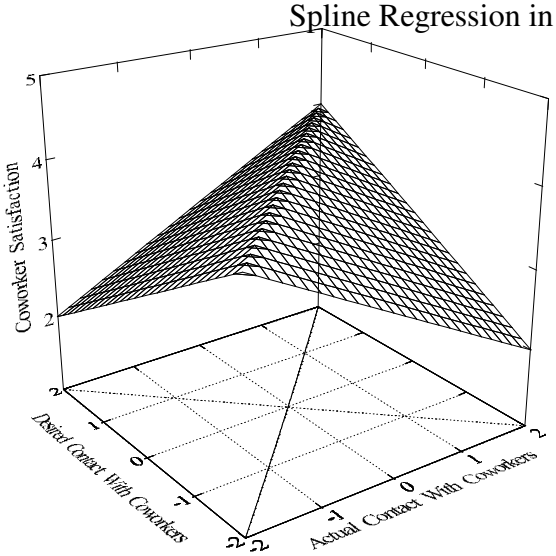
a. Actual and desired contact with coworkers and coworker satisfaction



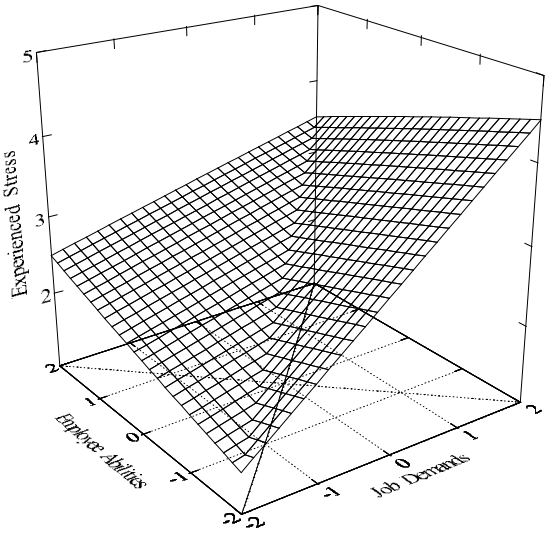
b. Job demands, employee abilities, and experienced stress



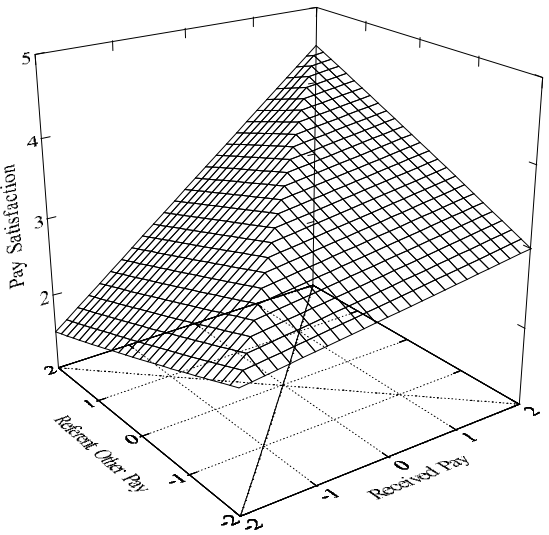
c. Received and referent other pay and pay satisfaction



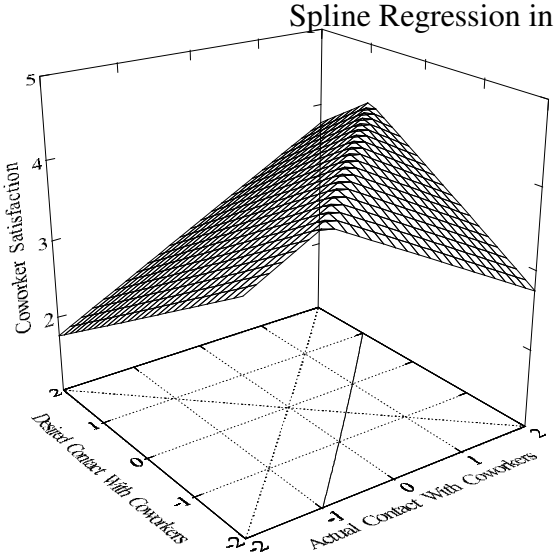
a. Actual and desired contact with coworkers and coworker satisfaction



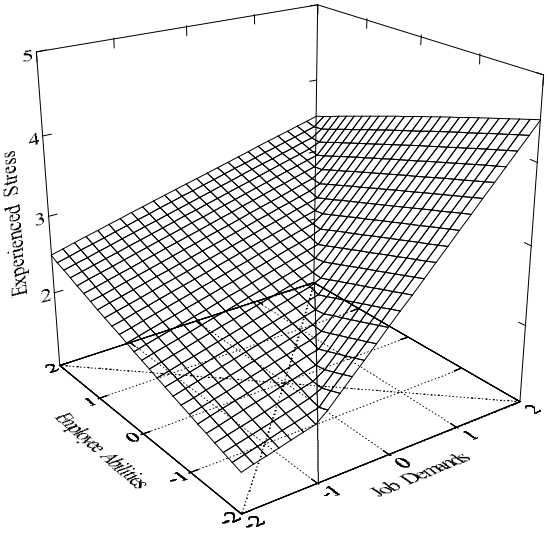
b. Job demands, employee abilities, and experienced stress



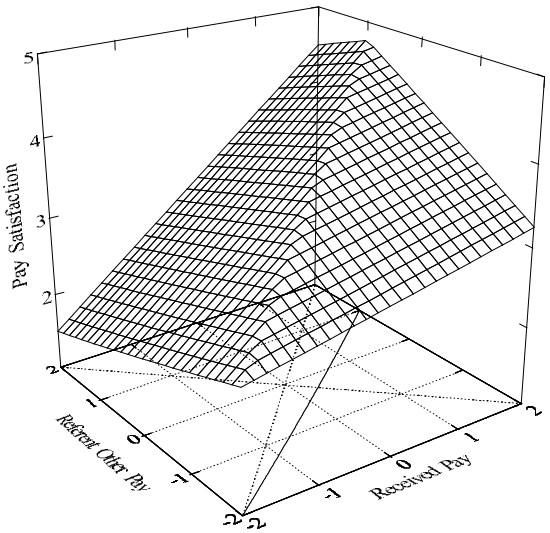
c. Received and referent other pay and pay satisfaction



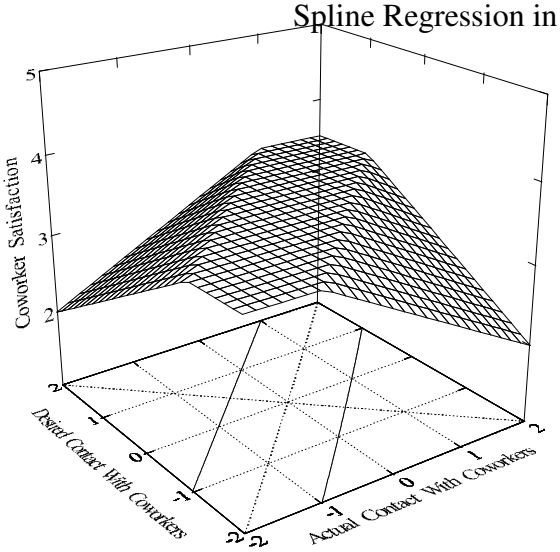
a. Actual and desired contact with coworkers and coworker satisfaction



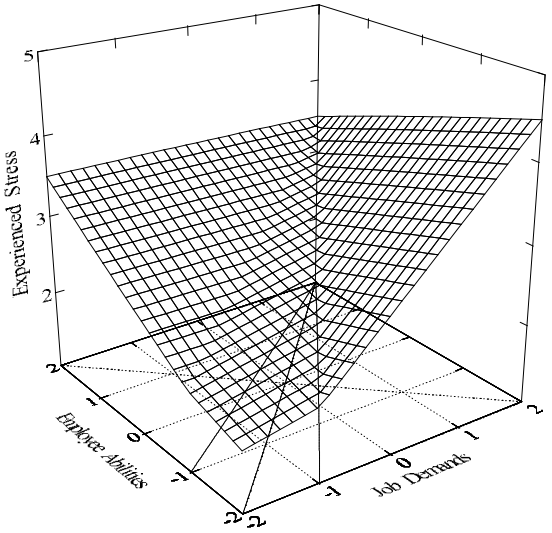
b. Job demands, employee abilities, and experienced stress



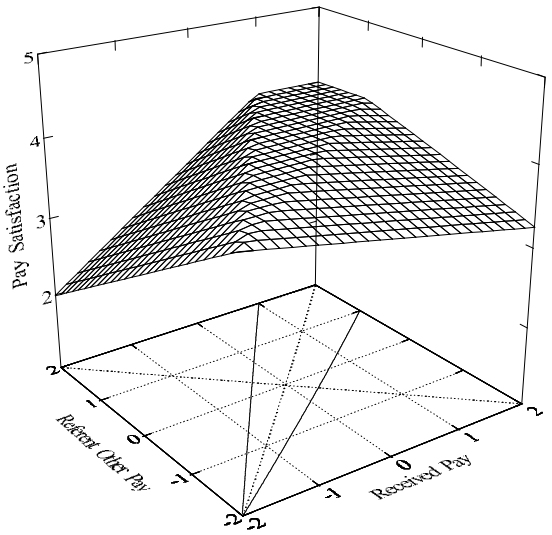
c. Received and referent other pay and pay satisfaction



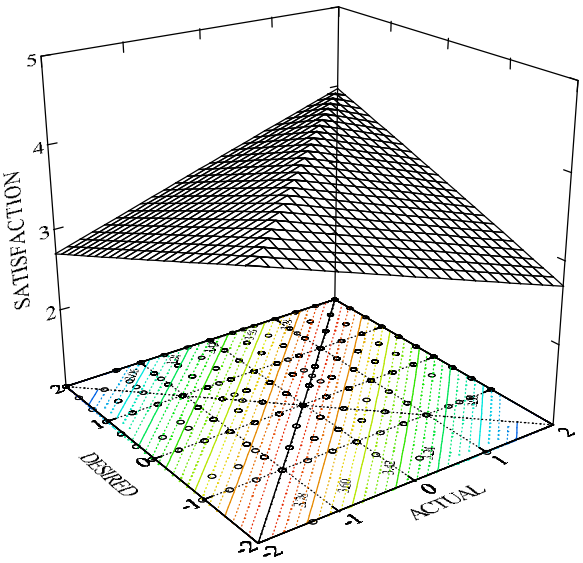
a. Actual and desired contact with coworkers and coworker satisfaction



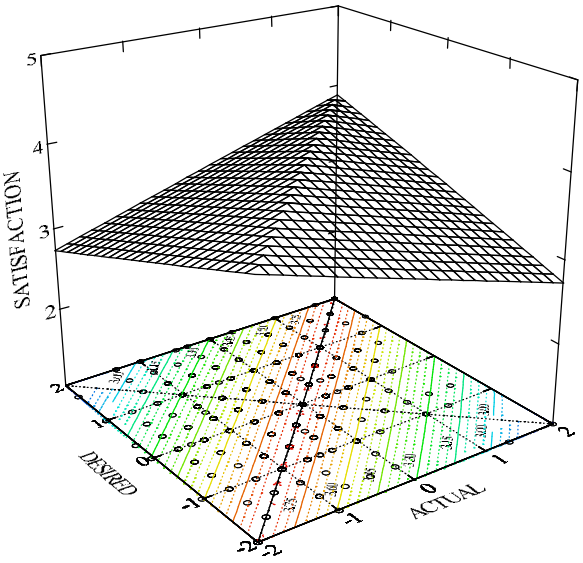
b. Job demands, employee abilities, and experienced stress



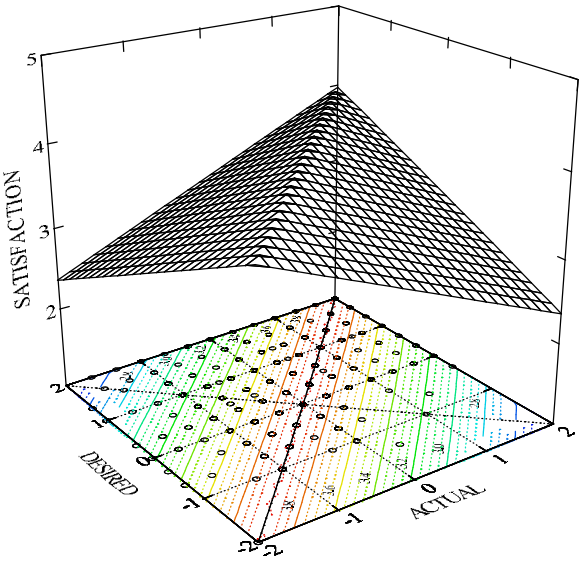
c. Received and referent other pay and pay satisfaction



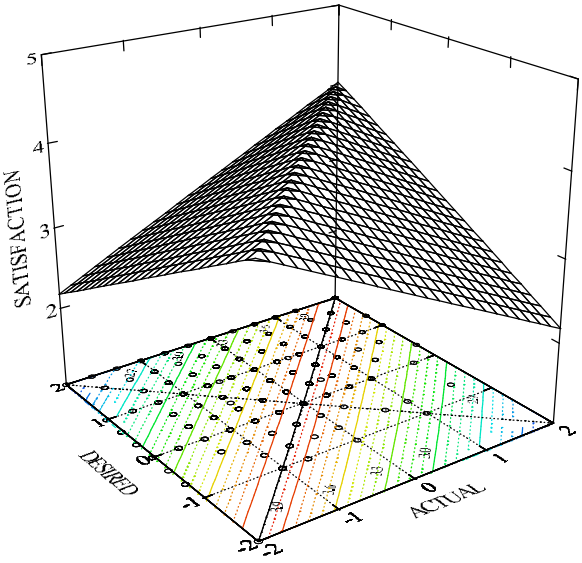
a. Authority



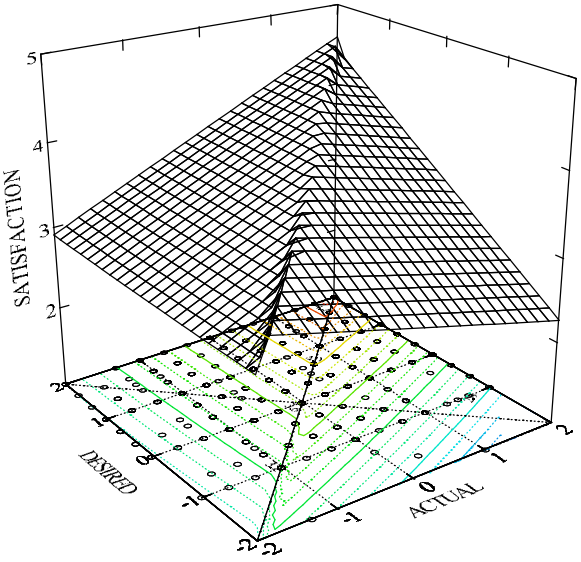
b. Relationships



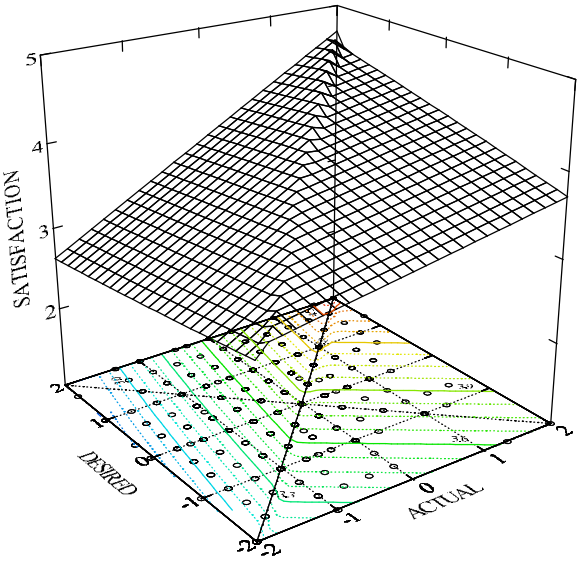
c. Variety



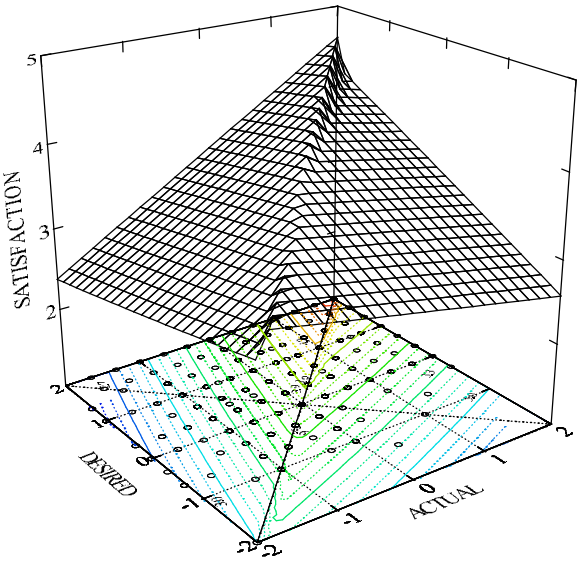
d. Autonomy



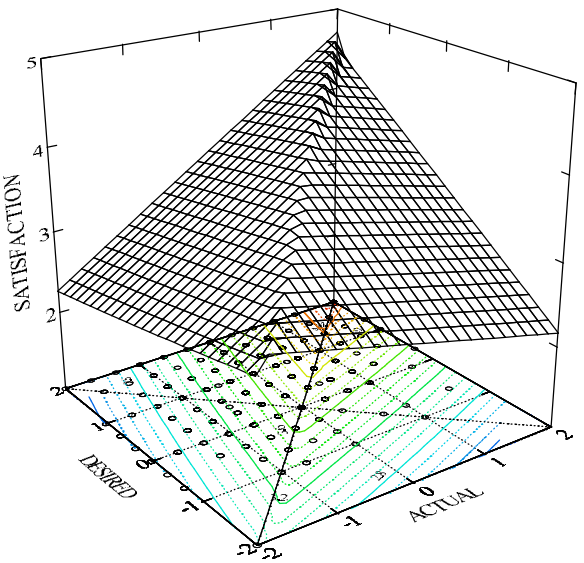
a. Authority



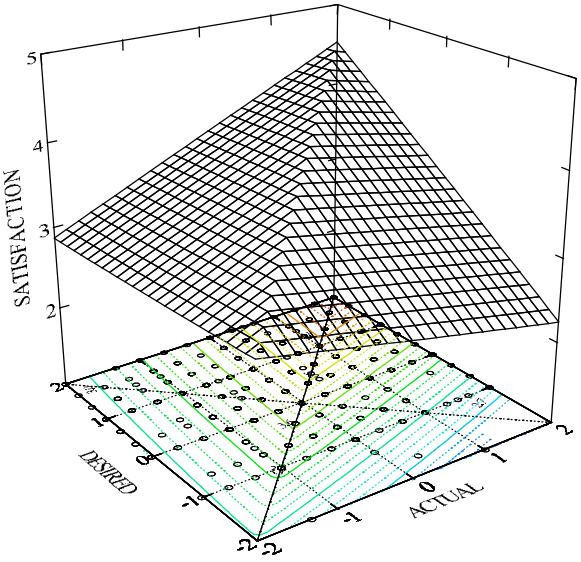
b. Relationships



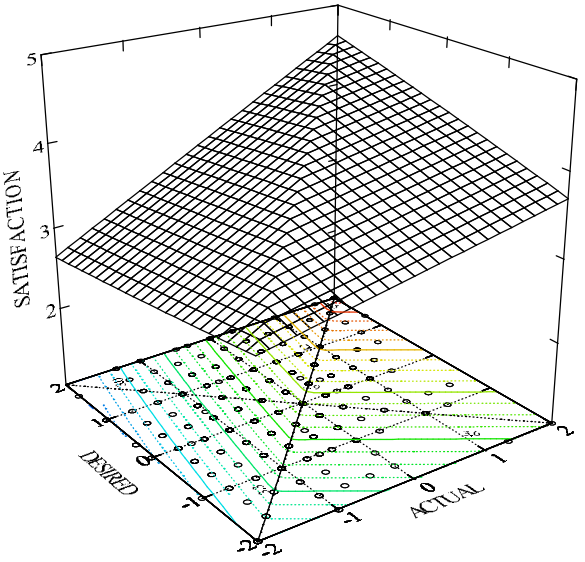
c. Variety



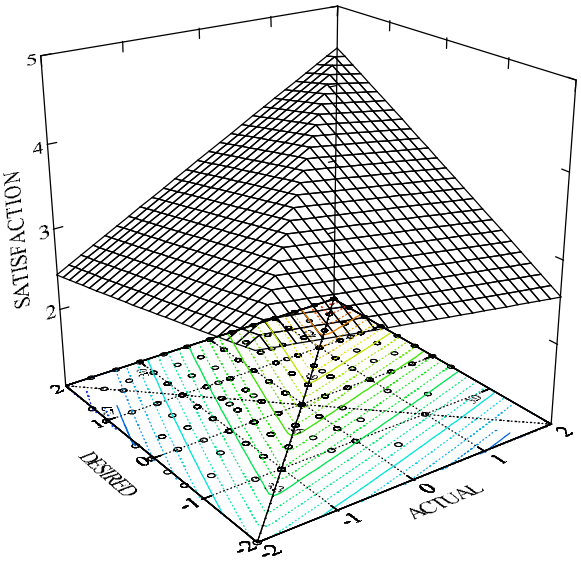
d. Autonomy



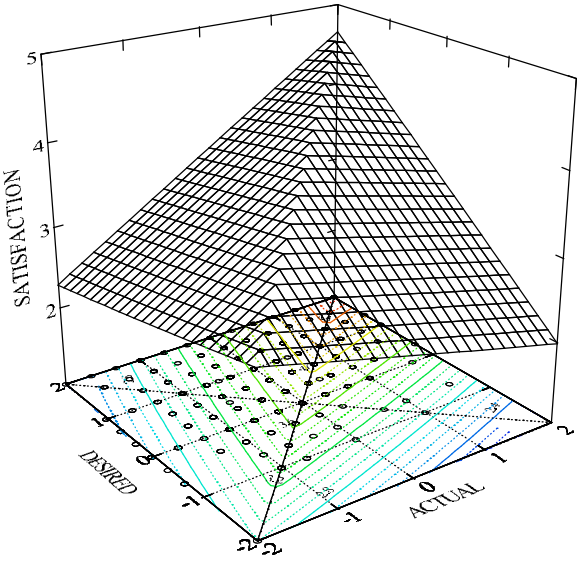
a. Authority



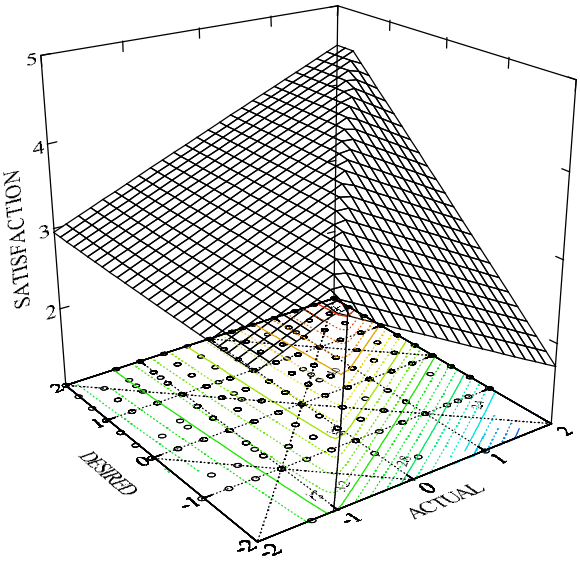
b. Relationships



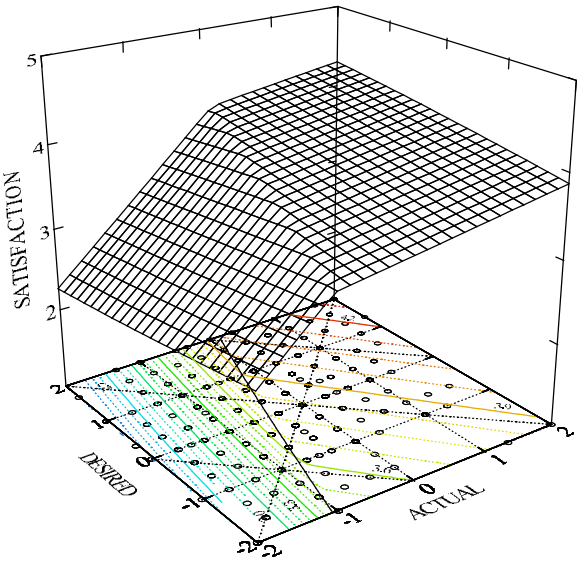
c. Variety



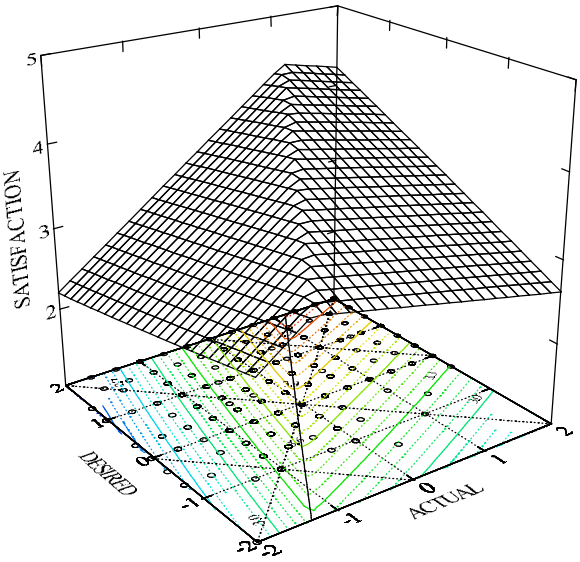
d. Autonomy



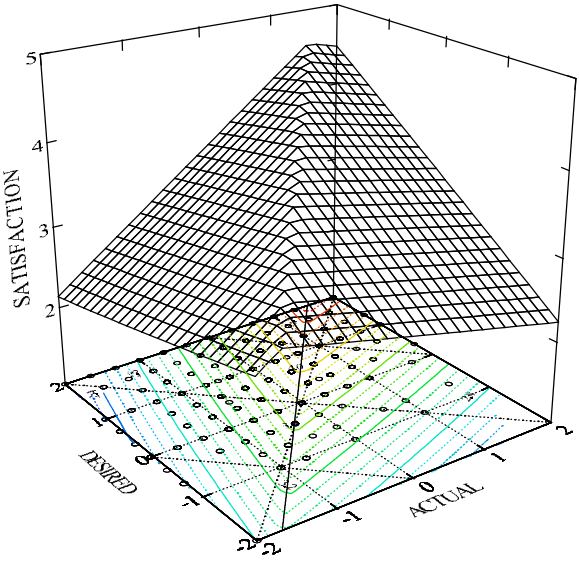
a. Authority



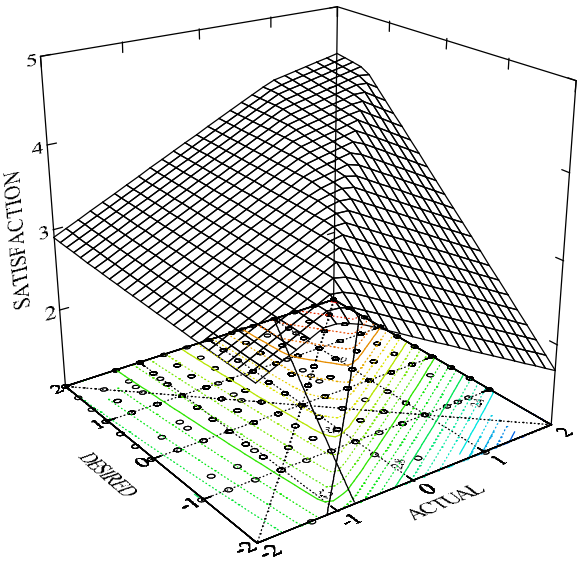
b. Relationships

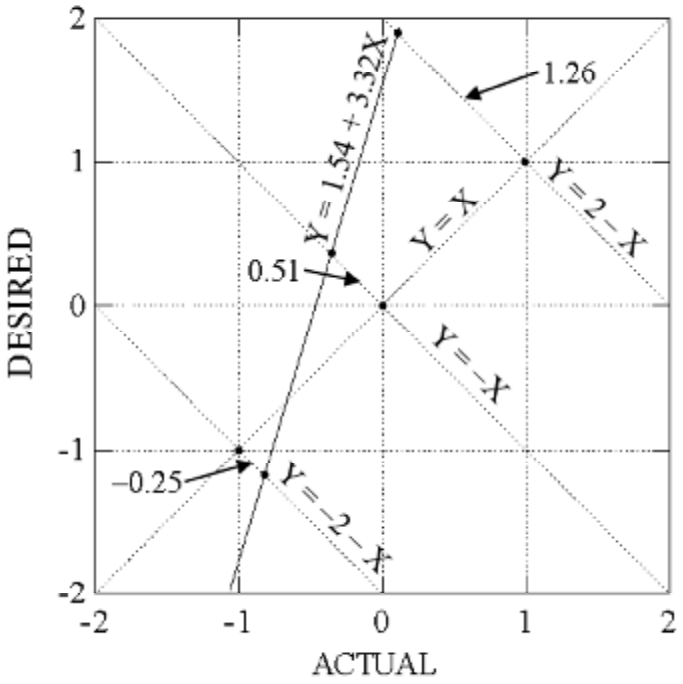


c. Variety



d. Autonomy


SATISFACTION
5
4
3
2
DESIRED
ACTUAL
-2
-1
0
1
2


2
1
0
-1
-2
-2
-1
0
1
2
DESIRED
ACTUAL
Y = 1.54 + 3.32X
1.26
Y = 2 − X
Y = X
0.51
Y = −X
−0.25
Y = −2 − X

Author Biographies

Jeffrey R. Edwards is the Belk Distinguished Professor of Organizational Behavior at the Kenan-Flagler Business School, University of North Carolina at Chapel Hill. He is past editor of *Organizational Behavior and Human Decision Processes*, past chair of the Research Methods Division of the Academy of Management, a fellow of the Academy of Management, the American Psychological Association, and the Society for Industrial and Organizational Psychology, and winner of the 2008 Distinguished Career Award from the Research Methods Division. His methodological research addresses difference scores, polynomial regression, response surface methodology, structural equation modeling, construct validation, and the development and evaluation of theory.

Mark E. Parry is the Ewing M. Kauffman/Missouri Endowed Chair in Entrepreneurial Leadership and Professor of Marketing at the Henry W. Bloch School of Management at the University of Missouri at Kansas City. He has served on the editorial board of the *Journal of Retailing* and has received various awards for his research, including the 2005 Excellence in Global Marketing Research Award from the American Marketing Association. His methodological work has addressed mathematical models of distribution channels, neural networks, cross-cultural comparisons, polynomial regression, and response surface methodology.