Testing fit patterns with polynomial regression models

Felix Schönbrodt
Ludwig-Maximilians-Universität München, Germany

Fit hypotheses, also labeled ‘congruence’, ‘discrepancy’, or ‘congruity’ hypotheses, contain the notion that an outcome is optimal when two predictor variables match well, while incongruent/ discrepant combinations of the predictors lead to a suboptimal outcome. Previous statistical frameworks for analyzing fit hypotheses emphasized the necessity of commensurable scales, which means that both predictors must be measured on the same content dimension and on the same numerical scale. In some research areas, however, it is impossible to achieve scale equivalence, because the predictors have to be measured with different methods, such as explicit attitudes (e.g., questionnaires) and implicit attitudes (e.g., reaction time task). In this paper, I differentiate *numerical congruence* from *fit patterns*, a concept that does not depend on the notion of commensurability, and hence can be applied to fit hypotheses with incommensurable scales. Polynomial regression can be used to test for the presence of a fit pattern in empirical data. I propose several new regression models for testing fit patterns which are statistically simpler and conceptually more meaningful than a full polynomial model. An R package is introduced which provides user-friendly functions for the computation, visualization, and model comparison of several fit patterns. An empirical example on implicit/explicit motive fit demonstrates the usage of the new methods.

Keywords: squared difference, congruence, fit, discrepancy, polynomial regression

DRAFT

Many psychological theories maintain the notion that the congruence (also called fit, match, correspondence, similarity, or discrepancy[1]) between two constructs has an effect on some outcome variable. One example is flow theory (Csikszentmihalyi, 1975), which maintains that one condition for flow experiences is an *optimal challenge*, which is defined as a situation where the perceived ability of the person fits the task difficulty. The optimal challenge can be operationalized with a discrepancy score between task difficulty and a person’s ability (e.g., Abuhamdeh & Csikszentmihalyi, 2009), where deviations into both directions should hinder flow experiences. Similar examples can be drawn from diverse fields of psychological research. For example, discrepancies between explicit and implicit attitudes predict dissonance reduction behavior (Briñol, Petty, & Wheeler, 2006), discrepancies between implicit and explicit self-concept predict scores in intelligence tests (Dislich et al., 2012), discrepancies between explicit and implicit self-esteem predict well-being (Schröder-Abé, Rudolph, & Schütz, 2007), and discrepancies between implicit and explicit motives predict well-being (e.g., Hofer & Chasiotis, 2003), psychosomatic symptoms (Baumann, Kaschel, & Kuhl, 2005), or relationship satisfaction and stability of couples (Hagemeyer, Neberich, Asendorpf, & Neyer, 2013). Furthermore, whole theories are built on the notion of optimal fit, like person-environment fit theory (Kristof-Brown & Billsberry, 2013) or regulatory fit theory (Higgins, 2000).

These examples document the widespread use of fit hypotheses. Several studies operationalize these hypotheses with an absolute or squared difference of the predictor variables. This approach, however, is challenged by theoretical considerations which state that for the operationalization of fit hypotheses it is necessary that both measures are commensurable (Edwards, 2002). The principle of commensurability, also known as ‘dimensional homogeneity’, states that if two quantities should be compared, added, or subtracted, they must be on the same content dimension (‘nominal equivalence’, Edwards & Shipp, 2007). For example, it does not make sense to subtract 2.1 meters from 0.6 kilograms. Furthermore, both measures must be assessed on the

---

Felix D. Schönbrodt, Department of Psychology, Ludwig-Maximilians-Universität München, Germany. Acknowledgements. I want to thank Jens Asendorpf, Giulio Costantini, Birk Hagemeyer, Gregor Kappler, Markus Maier, Marco Perugini, John Rauthmann, and Michael Zehetleitner for valuable comments on a previous version of the manuscript. Correspondence concerning this article should be addressed to Felix Schönbrodt, Leopoldstr. 13, 80802 München, Germany. Email: felix@nicebread.de. Phone: +49 89 2180 5217. Fax: +49 89 2180 99 5214.

[1]For clarification, the term ‘discrepancy’ here is used for undirected (aka. non-directional) difference measures (i.e., absolute differences $|X - Y|$ or squared differences $(X - Y)^2$). This paper is only concerned with squared differences as predictors, not with squared differences as outcomes (cf., Edwards, 1995).

same metric ('scale equivalence'). Two weights (which have the same content dimension) cannot be directly compared when they are assessed on different scales like pounds and kilograms. Nominal equivalence is a logical precondition for scale equivalence: When scales measure different constructs, scale equivalence cannot be meaningfully defined. These theoretical considerations led to the conclusion that '[c]ommensurate dimensions are required for the conceptualization and measurement of P-E [person-environment] fit [...]. Without commensurate dimensions, it is impossible to determine the proximity of the person and environment to one another, and the notion of P-E fit becomes meaningless.' (Edwards, Caplan, & Harrison, 1998, p. 31).

Commensurable scales typically are achieved by using the same response scale with different item stems, for example 'How much money do you actually earn?' and 'How much money would you like to earn?' (Edwards & Shipp, 2007). In practice, however, the actual degree of commensurability can be hard to determine as the different item stems might induce response biases, or differential item functioning might bias the meaning of the item when, for example, wifes and husbands are compared on the same item (see also Kristof, 1996). Moreover, and central to this paper, in some research areas it is impossible to achieve scale equivalence, because the predictors have to be measured with different methods. In fact, all of the studies described in the first paragraph compared incommensurable scales, such as reaction time tasks with Likert scales.

But if we cannot subtract kilograms from kilometers, how can we subtract milliseconds in a reaction time task from points on a Likert scale? These constructs are on inherently different measurement scales.[2] At first view, it seems that in the case of implicit/explicit attitude discrepancies and comparable situations the precondition of commensurability poses an insurmountable challenge. But is it really a compelling consequence that in these cases the notion of congruence 'becomes meaningless'?

In this paper, I propose an appropach that allows to test a certain type of fit hypotheses with incommensurable measures by differentiating *numerical congruence* from *fit patterns*. New statistical models based on polynomial regression provide a tool for describing and testing these fit patterns in empirical data. As will be shown, this approach does not rely on the commensurability of predictor variables. The relaxation of the commensurability precondition, however, comes at the cost that it is not possible to test several specific hypotheses, which could be tested with commensurable scales.

The paper is organized as follows. In the first part, numerical congruence and fit patterns are differentiated, two concepts which are sometimes mixed up. In the second part, conceptual and statistical problems of squared difference scores are discussed and related to polynomial regression, which provides an analytical framework within which both the problems and the solutions can be localized. The third part introduces new statistical models based on polynomial regression, which overcome the identified problems and allow to test fit patterns with incommensurable measures. Fourth, an empirical example demonstrates how to employ and interpret the new models. Finally, the limitations and implications of the proposed techniques are discussed.

Numerical Congruence vs. Fit Patterns

Science is communicated through verbal definitions, and problems arise when different scientific communities use the same label for different underlying constructs (e.g., Hmel & Pincus, 2002). Such a situation is present in the psychological literature concerning the term *congruence* (resp. fit, correspondence, discrepancy, congruity, or contingency), which refers to at least two different constructs. Henceforward, I will differentiate these two constructs as *numerical congruence* and *fit patterns*.

Numerical Congruence

It has been argued, probably most pronounced in the research tradition on person-environment fit, that the predictor variables must be commensurable in order to derive an meaningful quantification of the congruence between them (e.g. Bauer & Hussong, 2009; Caplan, 1987; Edwards, 2002; Edwards et al., 1998). For example, one could ask employees about the desired and the actual travel times in their job: 'How many days per months you want to travel?' (desired) and 'How many days per month do you actually travel?' (actual; cf. Edwards, 2002). A numerical comparison of these two quantities is directly possible: Employees with a perfect match (desired equals actual) are congruent, and increasing deviations mean increasing incongruence. As this notion of congruence is closely tied to the actual measurement scale on which the variables are located, henceforward I will call it *numerical congruence*. Numerical congruence requires both nominal and scale equivalence.

Fit Patterns

Other strands of psychology, however, conceive of congruence in a more abstract and conceptual way. For example, in client-centered therapy (Rogers, 2004) the term congruence is used for a good fit between the real, the perceived, and the ideal self. According to Rogers, during the process

[2] Of course, it could be debated whether, for example, implicit and explicit attitudes can be located on a common latent content dimension. In this case, we would have nominal equivalence, but still no scale equivalence (i.e., we would measure the same thing, but we only have a 'yard stick' for one predictor and a 'meter stick' for the other).

of therapy the ‘concept of self becomes increasingly congruent with his experience’ (Rogers, 1959, p. 205). In a comparable way, D. C. McClelland, Koestner, and Weinberger (1989) theorized about the congruence between implicit and explicit motives: ‘We are all familiar with individuals who express a desire to act in a certain way but seem unable to do so consistently. That is, their self-attributed and implicit motives are discordant. [...] We also believe [...] that through self-observation and analysis, greater congruence between the two types of motives can be achieved’ (p. 700). Another example is the congruence of verbal and non-verbal communication (e.g., Johnson, McCarty, & Allen, 1976; Mongrain & Vettese, 2003), where the verbal and the non-verbal channel can be more or less congruent. As a simplified example, if someone tells a serious or even sad fact (verbal channel), then only a low amount of smiling (non-verbal channel) is appropriate. If someone tells a somewhat funny thing, a medium amount of smiling is appropriate. If someone tells somethings really funny, bursting out in laughter is appropriate. A theory could predict that there is a certain intensity of smiling/laughing that ‘fits’ to the funniness of the verbal content. Both smiling more than the optimal combination and smiling less wouldn’t be perceived as appropriate.

These researchers have an conceptualization of congruence that does not presume commensurable scales. In fact, in none of the examples the components of congruence can be measured on the same dimension. Still a falsifiable hypothesis can be posed that relates combinations of the predictor variables to an outcome variable, such as ‘Too much and too less smiling is perceived as inappropriate’. Henceforward, I will call such patterns *fit patterns*.

A Formal Definition of Fit Patterns

The *fit pattern* is defined around two basic notions: First, a fit pattern is about an optimal match between the levels of two variables. For each level of X exists a matching level of Y which leads to an optimal response Z, and vice versa. Second, any deviation from these optimal combinations leads to suboptimal responses, where bigger deviations have a higher impact on the response variable than smaller deviations.

Described more formally, a weak and a strong version of the fit pattern can defined by four, respectively, five conditions[3]. For a *weak fit pattern*, the following four conditions must be satisfied:

First, for each value X_i exists a single value $Y_{opt.i}$ which maximizes a response variable Z. Second, for each Y_j exists a single value $X_{opt.j}$ which maximizes Z. Third, for all $Y \in [Y_{min}; Y_{opt.i}]$ and $X = X_i$, Z is a monotonically rising function of Y, for all $Y \in [Y_{opt.i}; Y_{max}]$ and $X = X_i$, Z is a monotonically falling function of Y (i.e., greater deviations of Y from $Y_{opt.i}$ lead to equal or greater reductions in Z). Fourth, in exactly the same way, for all Y_j is Z a monotonically rising (resp. falling) function with $X_{opt.j}$ as the vertex.

For a *strong fit pattern*, additionally the following condition must be satisfied:

Fifth, the relationship between optimal matching values must be symmetric: If X_i is the optimal value for Y_j, then Y_j must be the optimal value for X_i.

These definitions of weak and strong fit patterns do not entail the necessity of scale equivalence of X and Y. The predictor values are not directly compared, but rather their joint impact in the prediction of Z is investigated. Hence, in the example of implicit/explicit attitude discrepancy, it would not be valid to say ‘30 ms in an implicit reaction time task equal 1.4 points on an explicit Likert scale’ (which would assume scale equivalence). But one could say: ‘Increasing the implicit attitude by 30 ms has the same impact on dissonance reduction efforts as reducing the explicit attitude by 1.4 Likert points’ (for a similar argument about comparing the impact of incommensurable stimulus features, see Huang & Pashler, 2005).

Numerical congruence is logically independent from the notion of optimality or "optimal fit". The amount of numerical congruence of two variables is only determined by their standing on the same scale. Whether the amount of congruence furthermore is related to a maximization or minimization of an outcome variable is an independent question. Fit patterns, in contrast, have an inherent evaluative quality. They are only defined in relation to the optimization of an outcome variable. Hence, fit patterns can be seen as a subset of the broader concept of functional optimization. However, not every functional optimization would satisfy the conditions of a ‘fit pattern’. A model with two positive linear effects, for example, has an optimum at the highest level of both predictors, but it would not be a fit pattern according to the definition given above.

How can the presence of a fit pattern in a data set be detected? In the following sections it will be shown that discrepancy scores are *not* a generally valid approach to investigate fit patterns.

Squared Difference Scores Are Implicit Constraints of a Polynomial Regression

The shortcomings of difference and discrepancy scores have been repeatedly pointed out (for an overview and detailed discussion, see Edwards, 1994; Johns, 1981; Peter, Churchill, & Brown, 1993). In the following sections, I want to focus on two specific problems, dimensional reduction and scaling dependence. These problems can be framed as untested constraints. The next sections describe these constraints, and how a violation of them can result in a loss of information and statistical power.

[3]For simplicity of reading, all patterns are formulated as maximization patterns. The definitions can also be flipped for minimizing a response.

Expanding a Squared Difference

Several authors have pointed out that models with (directed) difference scores can be expressed as a constrained multiple regression (e.g., Cronbach & Furby, 1970; Griffin, Murray, & Gonzalez, 1999; Johns, 1981). In a similar manner, when squared difference scores $D = (X - Y)^2$ are used as a predictor in a regression model, the equation can be rewritten as (Edwards, 2002):

$$Z = c_0 + c_1(X - Y)^2 + e \tag{1}$$

Henceforward, this model will be called the *basic squared difference (SQD) model*. Expanding the equation yields:

$$Z = c_0 + c_1X^2 - 2c_1XY + c_1Y^2 + e \tag{2}$$

The regression weights of X^2 and Y^2 are constrained to have the same magnitude and the same sign, and the interaction term XY has two times the reversed weight. Removing these constraints and adding all lower order terms to the equation leads to the *general two-variable second-degree polynomial model*:

$$Z = b_0 + b_1X + b_2Y + b_3X^2 + b_4XY + b_5Y^2 + e \tag{3}$$

Note that throughout the paper the identifiers c_j (j = 0, ..., 5) are used for regression weights of constrained models and b_j for unconstrained models. In comparison to model (3), the following constraints are imposed by the SQD model (2): (1) $b_1 = 0$, (2) $b_2 = 0$, (3) $b_3 = b_5$, and (4) $b_4 = -2b_5$.

The basic SQD model is insensitive to the level on which numerical (in)congruence takes place (for example, $(2 - 1)^2$ equals $(10 - 9)^2$). Hence, any regression which only contains discrepancy scores (but not the main effects) implicitly assumes that the mean level effect is zero. (Algebraically, this is implied by the first two constraints of the SQD model, $b_1 = b_2 = 0$).

Fitting a full polynomial model to the data removes the implicit constraints of a squared difference score. However, two challenges are introduced by expanding the constrained squared difference equation to a full polynomial model.

Challenge 1: Polynomial Regression Results Can be Hard to Interpret

The first challenge of a polynomial regression is that results can be hard to interpret. Consider the example of desired (*DES*) and actual (*ACT*) amount of traveling (Edwards, 2002). Theories of person-environment fit predict that deviations from the ideal amount of traveling (too much and too less) lead to a reduced job satisfaction (*SAT*). Fitting a full polynomial model to the data of 366 MBA students led to the following result (Edwards, 2002, p. 373):

$$\begin{aligned} SAT = b_0 + 0.247 * ACT - 0.131 * DES - 0.130 * ACT^2 \\ + 0.231 * ACT * DES - 0.104 * DES^2 \end{aligned} \tag{4}$$

Does this result present evidence for a fit pattern or not? Interpreting these raw coefficients can be very difficult, and one might have problems to disentangle the unique and joint effects of the terms on the dependent variable in order to give a judgement about the fit hypothesis.

A practical solution for an easier interpretation is to plot the regression result as a response surface (e.g., Edwards, 2002; Myers, Montgomery, & Fitzgerald, 2009). In this plot, for each combination of predictor values on the X-Y-plane the predicted value of the response variable is plotted on the z axis, resulting in a three-dimensional response surface (see Figure 1). Alternatively to a three-dimensional plot, the values of the response variable can be displayed in different colors or as height lines in a contour plot (see Figure 4B).

From the plots of Figure 1 it is evident that numerical congruence is not a single point, but rather the line on the X-Y plane where X equals Y (line of numerical congruence, LOC[4]). A second important line is the line of numerical incongruence (LOIC), where X = -Y. This line is orthogonal to the LOC and intersects at the origin (X=0, Y=0).

Figure 1C provides the plot corresponding to regression equation (4). Now the interpretation is intuitive and straightforward: Congruent values of actual and desired travel time around the LOC have the highest satisfaction value, and increasing incongruence goes along with decreasing satisfaction. At this stage of analysis the pattern has not yet been tested for significance, but the plot clearly suggests a fit pattern.

Additionally to plotting, specific lines on the surface can be described numerically (Edwards & Parry, 1993; Edwards, 2002). Most importantly, the presence of a numerical congruence effect can be tested via the curvature along the LOIC, which is $b_3 - b_4 + b_5$. This quantity has been termed a_4 (Shanock, Baran, Gentry, Pattison, & Heggestad, 2010) and I will also use this label here. If predictors are commensurable, it has been argued that one can speak of a fit effect when a_4 is significant (Edwards & Cable, 2009).

Challenge 2: Increased Model Complexity Can Lead to Overfitting

The second challenge of the full polynomial model is the risk of overfitting the data – after all a model with one degree of freedom (basic squared differences) has been expanded to a model with five degrees of freedom. In fact, a

[4]For incommensurable scales, I use the term 'line of numerical congruence' to refer to the diagonal of the predictor space. At this line, predictors have the same numeric value, but this cannot be interpreted as 'semantic congruence'.

meta-analysis of 30 studies on person-environment fit (Yang, Levine, Smith, Ispas, & Rossi, 2008) investigated the incremental predictive validity of the non-linear terms, and showed that in many cases there was no gain beyond the simple linear terms X and Y. This suggests that we often might find overfitting in polynomial regression studies.

The Problem of Scale Dependence

Using polynomial regression as a framework, one problem of incommensurable scales can be analyzed more systematically. Many, if not most measures in psychology have arbitrary metrics without a natural zero point and spread (Blanton & Jaccard, 2006a). Simple and arbitrary operations, for example recoding a Likert scale from 1-5 to 0-4, or by standardizing at the population mean instead of the sample mean, can move the numerical zero point to virtually any position. If predictor variables are incommensurable, one of both can be rescaled without having a common anchor for both predictor scales. As will be shown, rescaling one predictor has only few consequences for unconstrained regression models, but a large impact on basic SQD models.

Rescaling predictors in an unconstrained polynomial regression. From an algebraic point of view, recentering or rescaling predictors in an *unconstrained* polynomial regression by additive shifts or multiplicative scaling factors is rather meaningless (J. Cohen, 1978; Dalal & Zickar, 2012). The essential statistics of the simultaneous model are invariant under linear transformations: R^2, the F and p value for the overall model, the estimated outcome values $\hat{Z}$, the proportion of variance in Z that is accounted for by X, Y, the interaction, or the squared terms, and the t and p values of the higher order terms. The regression weights for higher order terms in the polynomial model are not invariant, but their changes simply reflect the new scaling (i.e., the correlation with their untransformed counterparts is 1). Hence, the plotted response surface for rescaled predictors is exactly the same, only that the scales on the x and y axes are shifted and stretched. Another conclusion is that the choice of the center does not impact the statistical power to detect higher order effects.

The only parameters of a polynomial regression that are *not* invariant to rescaling are b_1 and b_2. Large changes in regression weights and their associated t and p values can occur. For the interpretation of these parameters it is recommended to center the predictors on a meaningful zero point before conducting an RSA (Aiken & West, 1991).

Rescaling predictors in the constrained squared difference model. What effect has a rescaled predictor on the response surface of a basic SQD model? As a thought experiment, imagine that explicit (E) and implicit (I) racial attitudes are located on the same latent scale. Theory predicts that dissonance reduction behavior is a function of implicit-explicit discrepancy – if both attitudes are congruent, $(E - I)^2 = 0$, no dissonance has to be reduced. The larger the discrepancy is, the higher are the dissonance reduction efforts (cf. Briñol et al., 2006). Now imagine that the manifest measure for the explicit attitude is biased towards the politically correct answer: all participants answer one scale point higher than their true explicit attitude E. The observed measure now is $E' = E+1$. In this case the outcome variable still is minimized, but on a discrepancy score of 1, $(E' - I)^2 = 1$.

Concerning the response surface, shifting the zero point of one predictor by an additive constant C corresponds to a lateral shift of the raw data away from the numerical LOC (see Figure 2B). The ridge of the basic squared difference model, however, is fixed to the numerical LOC. In such a case, the basic model still tries to find the best fit to the data, which will be a suboptimal fit (for an example, see Figure 2A). Hence, every analysis which employs discrepancy scores as predictors implicitly assumes that optimal fit is achieved at numerical congruence, and therefore confounds two independent concepts. Furthermore, statistical power is reduced to actually detect a shifted fit pattern.

When using arbitrary metrics also the spread of the predictor scales can be changed. If one of the predictors is rescaled by multiplication with a scaling constant S, geometrically the ridge is rotated away from the LOC (see Figure 2, panel C and D). Again, the basic SQD model provides a bad fit to the rotated and shifted data structure.

In the previous section, it has been stated that the choice of center and scale is rather irrelevant for the *unconstrained* polynomial model (except for interpretational reasons). Basic SQD models, in contrast, are highly dependent on the correct choice of center and scale. Simply shifting the zero point of one of the predictors can drive the statistical power to detect any existing discrepancy effect to zero (cf., Irwin & G. H. McClelland, 2001). The shift might be due to some biased self-insight, response tendencies induced by the framing of the question (‘ideal’ vs. ‘actual’), or response tendencies based on social norms. Given that numerous response biases can selectively affect the numerical zero point of one of the manifest predictor scales, hypothesis tests that are influenced by such simple transformations indeed are questionable (see also Blanton & Jaccard, 2006b).

To summarize, even if data follow a fit pattern on a latent dimension, a researcher will have a low statistical power to actually detect this pattern with a basic squared differences approach when one of the measured variables represents a rescaled version of the latent dimension.

The Proposed Solution: Turning Implicit Constraints Into Testable Hypotheses

Several problems of the basic squared difference model have been highlighted in the previous sections. First, the model imposes a dimensional reduction, which loses the information at which level of the predictors the squared difference was calculated. Second, optimal fit is constrained to happen at numerical congruence, which means that a shifted zero point or a rescaling of one of the predictors poses a serious problem to the basic SQD model. In the following sections, new statistical models based on polynomial regression are proposed that allow to test a fit pattern with incommensurable scales.

Modeling fit patterns with polynomial regression

Polynomial regression of the second degree can be used to test empirical data for the presence of a fit pattern. Some, but not all fit patterns can be described by a polynomial regression. Likewise, only some, but not all polynomial regressions satisfy the conditions for a fit pattern. Specifically, any second-degree polynomial regression that has non-zero quadratic terms with the same sign is an instantiation of the weak fit pattern:

$$b_3 b_5 > 0 \tag{5}$$

For a strong fit pattern, regression weights have to follow these constraints:

$$\begin{aligned} b_1 &= (b_2 b_4)/2b_5 \\ b_4 &= 2\sqrt{b_3 b_5} \end{aligned} \tag{6}$$

(For a derivation of these constraints, see Appendix A.)

In order to test for a fit pattern with polynomial regression, additional assumptions have to be made: (1) The misfit effect is symmetric relative to the ridge, (2) The misfit effect follows a quadratic form, and (3) The projection of the ridge on the X-Y plane has a linear form.

The last assumption implies that there exists a linear transformation of one predictor variable which is able to shift and/or rotate the ridge onto the numerical LOC. Additionally, the general assumptions for multiple regression have to be considered (e.g., Gelman & Hill, 2007).

Avoiding Overfitting and Testing Specific Assumptions: Five New Fit Models That Extend the Basic Squared Differences Model

Any polynomial regression result that satisfies the formal conditions of a fit pattern can be seen as evidence for a fit pattern. A potential criticism, however, could be that a second-degree polynomial regression is prone to overfitting (The weak fit pattern still has five degrees of freedom).

Therefore, in the following sections I propose five simpler fit models that are nested under the full polynomial model, but have fewer degrees of freedom. Furthermore, these models directly approach the problem of untested assumptions. Three implicit constraints of the basic SQD model have been identified, a) the assumption of no mean-level effect, b) the problem of shifted zero points, and c) the problem of different scale spread. The models developed in the following sections are based on a simple principle: Implicit constraints are successively turned into testable hypotheses by adding new parameters to the models. By adding the new parameters, the models have a higher statistical power to detect a fit pattern if the ridge is not located over the LOC. These new parameters can be seen as a reparametrization of the general polynomial model, which leads to a better interpretability of the results and targeted hypothesis tests.

One family of nested models contains *flat ridge models*. These models are derived by adding parameters to the basic squared difference model (2), which allow the ridge to be shifted (*shifted squared difference model, SSQD*) and additionally to be rotated (*shifted and rotated squared difference model, SRSQD*). Therefore, these models are able to describe flat fit patterns regardless of the choice of centering and scaling of the predictors.

A second family of nested models can be derived by additionally allowing the ridge to be tilted up- or downwards, which transforms the assumption of a flat ridge (i.e., the assumption of no mean-level effect) into a testable hypothesis. The basic squared difference model with a tilted ridge is called the *rising ridge model (RR)*. Adding parameters for shifting and rotating the surface leads to the *shifted rising ridge (SRR)* and *shifted and rotated rising ridge model (SRRR)*. Note that depending on the sign of the tilt and the curvature of the surface these models can also be 'rising troughs' or 'descending ridges', but for simplicity the label 'rising ridge' is used for all of these models.

Figure 3 shows how these models are nested within each other and how many free parameters (k) are estimated in relation to the intercept-only null model. All models are nested in the full polynomial model, and all constrained models satisfy the conditions for the weak fit pattern defined in (5). The tree of these models can be viewed from two perspectives. A bottom-up perspective would be to start with the basic squared difference model (which has $k = 1$ free parameter) and to increase model complexity by building additional parameters step-by-step into the model. A top-down perspective would be to start with the unconstrained second-degree polynomial model (which has $k = 5$ free parameters) and to impose an increasing number of constraints to arrive at each reduced model. Equations on the arrows in Figure 3 describe the testable constraint that is induced by each reduced model.

The following sections describe the parameters and constraints for these nested models.

The shifted squared difference model (SSQD model). If one of the predictors is centered to another zero point, the response surface is shifted laterally from the line of numerical congruence (see Figure 2, panels A and B, and Figure 1B). In such a case, the shifted squared difference model can be formulated as:

$$Z = c_0 + c_1[(X + C) - Y]^2 + e \tag{7}$$

where C is a shifting constant applied to the X predictor. Note that only a shift in one variable is necessary, as shifts in both variables can be algebraically reduced to one.

Expanding equation (7) yields the following:

$$Z = c_0 + 2c_1CX - 2c_1CY + c_1X^2 - 2c_1XY + c_1Y^2 + c_1C^2 + e \tag{8}$$

For a direct comparison, the equation for the full polynomial is reprinted:

$$Z = b_0 + b_1X + b_2Y + b_3X^2 + b_4XY + b_5Y^2 + e \tag{9}$$

Comparing the coefficients of (8) with the full polynomial model (9) reveals that following constraints are imposed: (1) $b_2 = -b_1$, (2) $b_3 = b_5$, (3) $b_4 = -2b_5$. Constraints 2 and 3 are the same as in the basic SQD model, and the shift of the ridge can be controlled by diametrically opposed weights for X and Y. After fitting the model, the amount of lateral shifting of the ridge can be expressed in the scale of the X variable[5]: $C := b_1/(2b_3)$. As in the SQD model, a_4 should be used for formally testing the fit effect: $a_4 := b_3 - b_4 + b_5 = 4b_3$.

It should be noted that in the absence of a specific hypothesis about the shift of the ridge C (and the other derived parameters) the results for these derived variables should be treated as exploratory, awaiting cross-validation.

By fitting this model, the optimal shift of the zero point of the X variable is performed. By definition, the SSQD model describes surfaces with a perfectly flat ridge.

The shifted and rotated squared difference model (SRSQD model). An implementation of arbitrary spreads into the shifted squared difference model introduces an additional free parameter which can rotate the surface (see Figure 2, panels C and D). The model can be formulated as:

$$Z = c_0 + c_1[(SX + C) - Y]^2 + e \tag{10}$$

where S is a scaling factor applied to the X predictor (note that again only a rescaling of one variable is necessary). Similar to the SSQD model, the SRSQD model describes surfaces with a perfectly flat ridge.

The SRSQD model imposes the following constraints on the parameters of equation (3): (1) $b_1 = (b_2b_4)/(2b_5)$, (2) $b_3 = b_4/4b_5$. It is actually hard to attach a specific meaning to these constraints. These constraints are algebraic conversions of equations (10) and (3), and it may help to point out that a polynomial regression with these specific constraints simply *is* the more meaningful equation (10).

After fitting the model, the amount of lateral shifting can be expressed in the scale of the X variable: $C := -\frac{1}{2}(b_2/b_5)$. The amount of rotation can be expressed as the scaling factor for X: $S := -b_4/(2b_5)$. After computing C and S, X can be rescaled using these values: $X' = SX + C$. This rescaling would rotate and shift the surface such that the ridge is exactly above the line of numerical congruence, and numerical congruence would predict an optimal outcome. Although this transformation does not mean that commensurable scales have been established, it can help to interpret the original variables in relation to the outcome.

For unrotated surfaces, a_4 is a direct measure for the misfit effect. With rotated surfaces, one can directly compute which value a_4 would have if X would be rescaled:

$$\begin{aligned} a_4' &:= a_4 \times 4/(S+1)^2 \\ &:= b_3/S^2 - b_4/S + b_5 \end{aligned} \tag{11}$$

In the case of rotated surfaces, a_4' is the relevant index that tests for the presence of a misfit effect, as a_4, which is above the line of numerical incongruence, represents an arbitrary cut that is not orthogonal to the ridge[6].

The rising ridge model (RR). The basic squared difference model constrains the ridge of the response surface to be flat. In order to make that contraint testable, the effect of the mean level, $M = (X + Y)/2$, can be incorporated into the model[7]:

$$Z = c_0 + b_MM + c_1(X - Y)^2 + e \tag{12}$$

Note that the parameter b_M denotes the effect of mean level in all following rising ridge models. If b_M is positive, the ridge is tilted in a way that high/high combinations of the predictors have a larger predicted response than low/low combinations. The RR model imposes the following constraints on the parameters of the full polynomial model (3): (1) $b_1 = b_2$, (2) $b_3 = b_5$, (3) $b_4 = -2b_5$. After fitting the model, the mean effect can be computed as $b_M := b_1 + b_2$. For formally testing the fit effect, a_4 should be used.

The shifted rising ridge model (SRR). Incorporating a shifted ridge into the RR model leads to the shifted rising ridge model, where variable X is shifted by value C and the mean level of shifted X and Y is $M = ((X + C) + Y)/2$:

$$Z = c_0 + b_MM + c_1((X + C) - Y)^2 + e \tag{13}$$

[5]The values of C, S, and b_M are deterministic functions of the regression weights b_1 to b_5 and are not free parameters which are estimated. To highlight this relationship, the operator ':=' is used for values that are derived from the regression weights.

[6]For unrotated surfaces, a_4 equals a_4'

[7]An alternative parametrization of the second-degree polynomial model, the Difference and Mean Model (DMM; A. Cohen, Nahum-Shani, & Doveh, 2010) also directly models the mean level effect and can be seen as the generalization of the more specific rising ridge models proposed here.

The following constraints are imposed on the full polynomial model (3): (1) $b_3 = b_5$, and (2) $b_4 = -2b_3$. After fitting the model, the mean effect can be computed as $b_M := b_1 + b_2$. The shift of the ridge can be computed as $C := (b_2 - b_1)/4b_3$. For formally testing the fit effect, again a_4 should be used.

The shifted and rotated rising ridge model (SRRR). Finally, allowing the SRR model to rotate leads to the shifted and rotated rising ridge model (SRRR), where the mean level is $M = ((SX + C) + Y)/2$:

$$Z = c_0 + b_M M + c_1((SX + C) - Y)^2 + e \tag{14}$$

The following constraint is imposed on the full polynomial model (3): (1) $b_4 = -2\sqrt{b_3 b_5}$. Due to the square root in the constraint, another constraint is implied: The equation is not defined when $b_3 b_5 \leq 0$, which is exactly the condition for a weak fit pattern. Hence, the SRRR model and all nested models satisfy the condition of a weak fit pattern.

After fitting the model, the shift of the ridge can be computed as as $C := -(2b_1 b_5 + b_2 b_4)/(4b_4 b_5)$, and the scaling factor for X as $S := -b_4/(2b_5)$. As in the SRSQD model, the misfit effect is represented by the rescaled a_4 parameter, $a'_4 := a_4 \times 4/(S + 1)^2 = b_3/S^2 - b_4/S + b_5$. The mean level effect along the ridge can be computed as $b_M := b_1/S + b_2$.

Strong and Weak Version of Fit Patterns

Rising ridge models only satisfy the conditions of the weak fit pattern, but not he strong version, as the optimal values are *not* symmetric (i.e., if X_i is the optimal corresponding value for Y_j, then Y_j will not be the optimal value for X_i). This is due to the linear effect of mean level, which is superimposed on the quadratic fit effect. Consider a rising ridge model where the ridge is aligned on the line of numerical congruence (RR model, see Figure 3). If you start on a point on the ridge and increase one predictor variable a bit, the response will increase due to the linear mean level effect, although one moves away from the ridge. At higher increments, however, the quadratic misfit effect will be stronger than the linear mean level effect and the response will decline again. To summarize, the combination of mean level effect and congruence effects leads to a violation of the fifth condition that defines the strong fit pattern. The flat ridge models, in contrast, satisfy the conditions for a strong fit pattern.

Hypothesis Tests With Incommensurable Predictor Scales

Commensurable scales allow tests that include comparisons of the predictor scales, such as 'satisfaction is highest when the actual travel time is one day more than the desired travel time'. In general, with commensurable scales the shift of the ridge C can be directly interpreted as 'the outcome is optimal when X is C units smaller than Y'. The rotation S can be interpreted as 'the outcome is optimal when Y equals S times X'. With incommensurable scales, however, numerical congruence has no semantic meaning and any statement involving comparisons like 'larger', 'smaller', or 'equal to' is meaningless. Hence, even when the presence of a fit pattern can be tested with any type of predictor scales, it has to be clear that many hypotheses cannot be tested when incommensurable scales are used.

Practical Considerations and an Empirical Example

In the following sections, I will consider practical questions on how to test fit patterns and demonstrate the usage with an empirical example.

An R Package to Test For Fit Patterns

Along with this paper a package called *RSA* (Schönbrodt, 2015b) has been released for the R Statistical Environment (R Core Team, 2014). This package automatically computes the squared and interaction terms of the predictor variables, checks raw data for outliers according to the Bollen and Jackman (1985) criteria, runs all models that have been described in this paper, compares them via several indices of model fit, and provides several functions for plotting the results. The *RSA* package computes the models using a path modeling approach with the package *lavaan* (Rosseel, 2012). This approach allows imposing the non-linear constraints on the parameters of the multiple regression. By default, regression weights are estimated using maximum likelihood. As squared and interaction terms necessarily introduce non-normality to the data, by default robust standard errors (SEs) are computed, which are robust against violations of the normality assumption. Beyond these default settings, many alternative estimation procedures and several methods for computing the standard errors can be selected in the *lavaan* package. See the *lavaan* documentation (Rosseel, 2012) for details. Furthermore, bootstrapped confidence intervals and p values can be computed for all regression coefficients, indices of the response surface (such as the surface parameter a_4 and a'_4), and the derived parameters of the constrained models (C, S, and b_M). Note that only the regression weights b_1 to b_5 are estimated (along with the intercept). All other indices are simply deterministic functions of the regression weights and therefore are not free parameters of the model. Furthermore, the constraints for the nested models effectively reduce the number of free parameters to be estimated. The number of free parameters that is actually estimated for each model can be seen in Figure 3.

Model selection

When several candidate models are available for a data set, one has to strike the right balance between underfitting (i.e., the model has too few parameters to fit the data) and overfitting (i.e., the model has too many parameters and starts to model the random noise). Both over- and underfitting reduce the replicability when the model is applied to a new data set. Model selection procedures can help to identify the most parsimonious model that neither has too few nor too many parameters.

Nested models often are compared via a χ^2 likelihood-ratio test (LRT). For multimodel selection purposes, however, this test is not recommended (Burnham & Anderson, 2002, Ch. 1.4.3, 6.9.3), as LRT does not provide a coherent framework for model selection. First, the appropriate α level is not clear. Recommendations range from .01 to .15, and others have argued that the α level should not be fixed but rather a function of the sample size (Good, 1982). Second, multiple testing is a problem if several, non-independent models are to be compared. In the tree of nested models (Figure 3) 25 potential LRTs can be computed, and there is no principled theory how to rank the models based on these tests. Third, non-nested models cannot be compared.

An additional index for model fit can be computed from the underlying path modeling approach: The comparative fit index (CFI). This index ranges from 0 to 1, and, as a rule of thumb, values >= .95 are considered as a relatively good fit (Hu & Bentler, 1999). However, this index only indicates underfitting (i.e., when the model has too few parameters), and therefore cannot serve as the only index for model selection.

Due to the shortcomings of these indices of model fit, I mainly focus on the corrected Akaike Information Criterion (AICc)[8] for model comparisons of the proposed fit patterns. This index is able to compare nested and non-nested models, and provides a coherent theory of model selection (Burnham & Anderson, 2002; Burnham, Anderson, & Huyvaert, 2011). AICc is a fit index that balances the complexity of a model with its predictive accuracy, and therefore is sensitive both for over- and underfitting. Generally, the model with the smallest AICc is considered the best model. The absolute size of AICc cannot be interpreted, as it depends on arbitrary constants in the calculation. The relevant measure is the difference in AICc, ΔAICc, between any two models. As a general rule of thumb, it has been suggested that a ΔAICc < 2 indicates that both models are essentially equally good, and that models with ΔAICc < 7 compared to the best model still have some support and should probably not be rejected yet. If ΔAICc is larger than 10, then the model is considered to be implausible compared to the best model (Symonds & Moussalli, 2011; Burnham et al., 2011).

For a better interpretation of the AICc differences, the ΔAICc can be transformed into model weights and evidence ratios. The model weight (also called ‘Akaike weight’; Burnham & Anderson, 2002) can be interpreted as the probability that this model is the best model of the set of candidate models, given the data (Wagenmakers & Farrell, 2004). Furthermore, one can compute the ratio of the weights of two models, which gives the evidence ratio. An evidence ratio of 2.5 between the best and the second-best model, for example, indicates that the better model is 2.5 times more likely than the other model.

When model selection is based on AICc, one should bear in mind that this only gives an index for the relative, but not absolute, plausibility of models. Even if all models are completely implausible, either theoretically or statistically, AICc will still select one ‘best’ model (Symonds & Moussalli, 2011). Hence, after selecting the best model via AICc, the absolute performance of the model should be evaluated. This can be done with R^2_{adj}, which is a useful descriptive statistic, but should not be used for model selection by itself (Burnham & Anderson, 2002; McQuarrie & Tsai, 1998).

An Empirical Example

The data analyzed in the following example are a subset from a study of motives in families (Schönbrodt, 2012). An increasing amount of publications investigates the effect of a misfit of explicit and implicit motives on stress symptoms and well-being (e.g., Baumann et al., 2005; Hagemeyer et al., 2013; Schüler, Job, Fröhlich, & Brandstätter, 2008; Brunstein, Schultheiss, & Grässmann, 1998; Hofer & Chasiotis, 2003; Hofer, Chasiotis, & Campos, 2006; Kazén & Kuhl, 2011; Kehr, 2004). The basic hypothesis is that a misfit between explicit motives (i.e., what you consciously strive for) and implicit motives (i.e., what gives you affective rewards) leads to impaired well-being (D. C. McClelland et al., 1989)[9].

This fit hypothesis is tested for the affiliation domain in the current data set. The implicit affiliation/intimacy motive (*IM*) of 362 persons was assessed by scoring the content of fantasy stories written to six ambiguous picture stimuli (Schultheiss & Pang, 2007; Winter, 1991). Raw motive scores ranged from 0 to 12 (M = 4.08, SD = 2.13). The number of coded motive scores partly depends on the length of the story, because in this particular coding system each sentence can be coded for the presence or absence of motive imagery. Hence, as longer stories tend to have higher scores, we computed density scores (i.e., motive scores per 1000 words) (Schultheiss & Pang, 2007) for subsequent analyses.

[8] In comparison to the classical AIC, AICc (Hurvich & Tsai, 1989) corrects for a bias when the sample size is small compared to the number of model parameters. When sample size increases, AICc converges to AIC.

[9] When reading the verbal models on implicit-explicit motive congruency literally it seems they only make explicit statements about the incongruent cases, without directly referring to the congruent counterparts. But from the overall reading of the literature it seems clear to me that the negative effects are seen in relation to the congruent alternative. Furthermore, nowhere a differentiation between low-, medium-, and high-level congruence is given, so without further qualification implicit-explicit congruence should be equally beneficial on any level of the trait.

The explicit intimacy motive (*EM*) was assessed using a Likert scale (Schönbrodt & Gerstenberg, 2012) ranging from 0 to 5 (M = 3.62, SD = 0.79). The explicit motive score was uncorrelated to the implicit motive score (r = .02). As an indicator of well-being, the Valence scale from the PANAVA inventory (Schallberger, 2005) has been assessed on a Likert scale ranging from 0 to 6 (M = 4.19, SD = 1.21). Higher values indicate more positive valence. (Actually, the data set has a multilevel structure with persons nested in families. For simplicity, this is ignored in the following demonstration.) Due to the different measurement methods, the scales for the implicit and the explicit motive do not have a common metric or a meaningful zero point. For an easier interpretation, all variables have been z-standardized to the sample mean and standard deviation.

In the following analyses, this data set will be analyzed for the presence of a fit pattern. In a confirmatory fashion we expected that a misfit of motives would lead to less positive valence. Additionally, a mean-level effect is explored, as fit on a high motive level could be associated with higher levels of well-being than fit on a low level (e.g., Hofer, Busch, & Kärtner, 2010). Due to the incommensurability of the predictor scales the numerical line of congruence and the numerical line of incongruence are meaningless. Therefore the ridge of a possible fit pattern could be shifted and/or rotated away from the line of numerical congruence. The employed models allow such a shift to find the optimal combinations of predictor variables, but the shift cannot meaningfully be interpreted.

By default, the main function of the `RSA` package calculates all models displayed in Figure 3, along with a model with two main effects only ($X + Y$) and an interaction model ($X + Y + XY$). No multivariate outliers have been detected using the criteria of Bollen and Jackman (1985) and the quantile-quantile plot of the residuals of the final model did not show strong violations of the normality assumption[10].

Model selection in the example data set. A detailed table of model indices is shown in Table 1. The best model according to AICc is the SRRR model with a model weight of .27, and the SRR and SRSQD model share the second place with a model weight of .17 each. The full polynomial also is in the range of equally plausible models with ΔAICc < 2, with a model weight of .15.

Inspecting the other fit indices also indicates that the SRRR model is a good choice. Its CFI is 1, and all models simpler than the SRRR have a CFI < .87. Furthermore, the χ^2-LR test indicates that it is not significantly worse than the full polynomial model ($\Delta\chi^2(1) = 0.90$, $p = .343$), and the adjusted R^2 is the highest of all models. Taken the overall evidence together, it seems reasonable to conclude that the SRRR model is the most parsimonious and best-fitting model for this data set.

Parameter estimates. Table 2 shows the estimated regression parameters for the unconstrained polynomial model (upper part) and the regression weights and the derived surface parameters of the SRRR model (lower part). Furthermore, robust standard errors, percentile bootstrapped CIs, and p values (10,000 replications) are reported. The combined impact of the raw regression coefficients b_1 to b_5 may be difficult to interpret, but a direct and meaningful interpretation of the surface parameters can be given. For rotated surfaces, the crucial test of the fit hypothesis is the CI and the p value of the curvature orthogonal to the ridge, expressed by a'_4 = -0.384, 95% bootstrapped CI [-0.658; -0.135], bootstrapped p < .001. As the surface is significantly bent down orthogonal to the ridge, the motive fit hypothesis is supported in the current example: Increasing misfit leads to increasingly impaired well-being.

Further conclusions can be drawn from the other parameters of the SRRR model. The parameter for the rotation of the ridge, S, does not significantly differ from an unrotated ridge where S would be 1 (bootstrapped p = .081; note that the p value for S tests $H_0 : S = 1$). Likewise, the parameter for the mean-level effect, b_M, is not significantly different from a flat ridge (bootstrapped p = .077). Still, according to the AICc criterion, the rotation and the slope of the ridge still add a little bit to the model's quality as the SRRR model is 1.57 times more likely than its unrotated counterpart (SRR) and 1.59 times more likely than its flat counterpart (SRSQD). The lateral shift of the ridge C, in contrast, passed the significance threshold (bootstrapped p = .031). But whenever predictor scales are incommensurable, any deviation of the ridge from the LOC, which is indicated by C and S, cannot be interpreted in terms of semantic congruence.

Figure 4 shows a visualization of the final SRRR model as a 3d plot (panel A) and as a contour plot (panel B). Generally, one should only interpret regions of the surface that are in the range of the original data. The surface plot tempts to focus on the corners, which often are very salient due to an pronounced upward or downward bend. These corners, however, usually are extrapolations of the surface into regions where no actual observations exist, and this extrapolation rests on very unlikely assumptions (Montgomery, Peck, & Vining, 2012). Hence, I strongly recommend to always show the raw data imposed on the surface plot or on the floor of the 3d cube, and to interpret the surface only in regions where actual data exists (see also Wilkinson & Task Force on Statistical Inference, 1999; Tufte, 2001). In Figure 4, additionally

[10] Traditionally, the necessity of the squared terms has been tested using LRTs. Removing the squared terms in the current data leads to a significantly worse fit ($\Delta\chi^2(2) = 7.74$, $p = .021$; CFI = .60; ΔAICc = 3.3). However, it is strongly recommend to use one paradigm (model selection via AICc) or the other (LRT), but not mixing them in the same analysis (Burnham et al., 2011). As described above, I focus on AICc here.

Table 1
Model Comparison for the Empirical Example, Ordered by ΔAICc

Model	k	AICc	ΔAICc	Model weight	Evidence ratio	CFI	R^2	p_{model}	R^2_{adj}
SRRR	4	7030.50	0.00	.27		1.00	0.048	.002	0.037
SRR	3	7031.41	0.90	.17	1.57	0.87	0.040	.002	0.032
SRSQD	3	7031.44	0.93	.17	1.59	0.86	0.040	.002	0.032
Full polynomial	5	7031.68	1.18	.15	1.80	1.00	0.050	.003	0.037
SSQD	2	7032.71	2.21	.09	3.02	0.69	0.031	.003	0.026
RR	2	7033.48	2.98	.06	4.43	0.64	0.029	.005	0.024
SQD	1	7034.32	3.82	.04	6.75	0.50	0.021	.005	0.019
$X + Y + XY$	3	7034.97	4.47	.03	9.34	0.60	0.031	.011	0.022
null	0	7040.12	9.62	.00	122.50	0.00	0.000		0.000
$X + Y$	2	7042.22	11.72	.00	350.73	0.00	0.005	.384	0.000

Note. k = Number of parameters; AICc = corrected Akaike Information Criterion; Evidence ratio = Ratio of model weights of the best model compared to each other model; CFI = Comparative fit index; R^2 = variance explained of the model; p_{model} = p value for explained variance of the model; R^2_{adj} = adjusted R^2.
Model abbreviations (see also Figure 3): SRRR = Shifted and rotated rising ridge model; SRR = Shifted rising ridge model; RR = Rising ridge model; SRSQD = Shifted and rotated squared difference model; SSQD = Shifted squared difference model; SQD = Squared difference model; $X + Y$ = Model with two linear main effects; $X + Y + XY$ = Moderated regression; null = Intercept-only model.

Table 2
Regression Coefficients b_1 to b_5 and Derived Model Parameters for the Full Polynomial and the Shifted and Rotated Rising Ridge (SRRR) Model.

Model	Estimate	robust *SE*	95% CI (lower)	95% CI (upper)	*p*
Full polynomial					
b_1	0.106	0.058	−0.010	0.221	.071
b_2	−0.036	0.058	−0.145	0.084	.554
b_3	−0.014	0.021	−0.055	0.042	.585
b_4	0.146	0.044	0.055	0.253	.003
b_5	−0.089	0.038	−0.161	−0.011	.023
SRRR Model					
b_1	0.126	0.056	0.019	0.236	.023
b_2	−0.041	0.057	−0.151	0.072	.481
b_3	−0.031	0.013	−0.064	−0.010	<.001
b_4	0.109	0.023	0.051	0.170	.004
b_5	−0.096	0.036	−0.164	−0.034	<.001
C	−0.684	0.242	−1.379	−0.085	.031
S	0.566	0.187	0.291	1.087	.081
b_M	0.181	0.098	−0.023	0.413	.077
a'_4	−0.384	0.145	−0.658	−0.135	<.001

Note. SRRR Model = Shifted and Rotated Rising Ridge Model. The p values test for the H_0 that the parameter is zero. The only exception is the S parameter, where the p value tests the $H_0 : S = 1$. Confidence intervals and p values are derived from a percentile bootstrap with 10,000 replications.

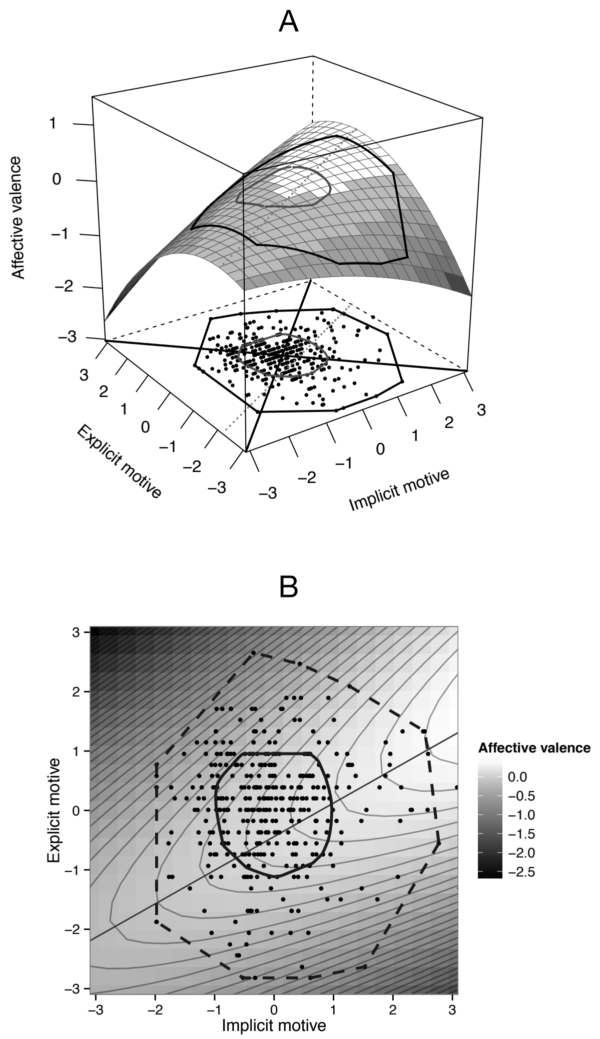

Figure 4. 3d and contour plot of the SRRR response surface of the empirical example.

a bagplot around the raw data points is projected onto the surface to give an additional visual aid for the 'interpretable region'. The bagplot (Rousseeuw, Ruts, & Tukey, 1999) is a bivariate extension of a boxplot, which describes the position of the inner 50% of points (the inner polygon, called *bag*) and separates outliers from inliers (the outer polygon, called *fence*).

To summarize, this example showed how the new models can be employed to test fit patterns with incommensurable scales. Even if one takes the considerable model uncertainty in the current example into account, there is quite strong evidence for a motive misfit pattern: All models in the candidate set (i.e., models with ΔAICc < 2) have a significant a_4, respectively a'_4 parameter. Concerning the rising ridge and the rotation, however, the evidence is not conclusive yet. The applied model selection procedure shows a slight preference for a tilted and rotated ridge. Based on the current data, however, models with a flat or unrotated ridge cannot be discarded yet, which is also reflected in the wide CIs of the corresponding parameters.

The data set is included in the `RSA` package, and a full script of the analyses is in Appendix B. Hence, readers can reproduce all results and plots reported here. For additional features of the RSA package, please consult Schönbrodt (2015a).

Discussion

Building upon the pioneering work of Edwards (1994, 2001, 2002), this paper extends existing polynomial regression techniques, and proposes a rationale and statistical models to test fit patterns with incommensurable scales.

Every research paper using absolute or squared difference scores relies on the usually unrealistic assumptions that both predictors are on the same numeric scale, that the optimal fit is exactly on the line of numerical congruence, and that there is no effect of the mean level. Whenever one of these assumptions is violated, the basic discrepancy models have a deficient fit to the data and therefore a low statistical power to detect existing fit patterns.

These constraints should be treated as hypotheses that can be tested empirically with polynomial regression. The proposed extensions of the basic SQD model give researchers the opportunity to appropriately test fit patterns without having to rely on unrealistic assumptions. In some cases, the more complex models will be reduced to the basic squared difference model. In many cases, however, the proposed models will improve the fit, and in some cases apparent null results will become significant. The new models can give further insights into the relationships of the variables, as they can directly test specific hypotheses and allow for more complex relationships like rising ridge. The reparametrizations of the general polynomial model are statistically simpler and conceptually more meaningful, as the parameters C, S, and b_M can be directly interpreted in terms of a shift, a rotation, and a tilt of the surfaces' ridge. Located at different levels of model complexity, the new models allow to balance complexity and parsimony.

Limitations

Polynomial regression and the proposed models are not without limitations. First, as pointed out in the introduction, the method is limited to situations where discrepancy scores serve as predictors. If these scores are used as outcome variables, other approaches have to be taken (Edwards, 1995). Second, as the equations and the `RSA` package work with manifest variables, it is assumed that the variables are assessed without measurement error. Cheung (2009) developed the Latent Congruence Model, which allows to test directed congruence hypotheses on a latent level, but only includes linear (but not squared or interaction) effects. So far, no consensus has been achieved on how to model latent interactions or latent quadratic terms (e.g., Klein & Moosbrugger, 2000; Mooijaart & Bentler, 2010; Harring, Weiss, & Hsu, 2012). Once these technical issues are resolved, however, there are no principal objections against latent modeling of polynomial surfaces. Third, a potential problem of polynomial regression is its susceptibility to outliers, as the squared terms exaggerate outliers even more. Therefore it is strongly recommended to inspect raw data for influential points and outliers (Bollen & Jackman, 1985; Wilcox, 2012). Fourth, polynomial regression (of second degree) is restricted to symmetric fit patterns. That means, 'too much' of something is constrained to have the same impact as 'too less'. Asymmetric fit patterns could be modeled by a polynomial of the third degree, but this implies fitting a large number of parameters (nine parameters compared to five in the second-degree polynomial), which bears the danger of overfitting even more. Hence, constrained third-degree models should be developed which test specific asymmetric hypotheses with fewer parameters. As another alternative, piecewise regression techniques (also known as *segmented regression*) can be used to model asymmetric fit patterns, which allows the regression line to take different slopes on each side of the breakpoint (e.g., Frost & Forrester, 2013). Readers interested in these models can refer to Keele (2008). Recent piecewise regression techniques can also search for the optimal breakpoint (Beem, 1995; Muggeo, 2003) and therefore are able to adjust for shifted zero points, but, to my knowledge, there is no solution yet for adjusting the spread of the scales. Finally, I want to point out once more that incommensurable predictors retrict the range of testable hypotheses compare to commensurable predictors. Data can be tested for the presence of fit patterns, but any hypothesis involving comparisons between the predictors principally cannot be tested.

Conclusions

The usage of discrepancy scores as predictor variables has been a topic of long debates. The introduction of response surface analysis into psychological science removed many of the problems surrounding difference scores. In this paper, the concept of fit patterns was introduced, which provides the theoretical base for testing fit hypotheses with incommensurable scales. New statistical models, namely the shifted (and rotated) squared difference models and their extensions with rising ridges, extend the statistical toolbox and enable researchers to test fit hypotheses without having to rely on unrealistic assumptions. These models have a higher statistical power to detect actual fit patterns and provide easily interpretable parameters. With these tools, new classes of hypotheses can be tested which would be hard or intpossible to test with conventional methods. Finally, a new open-source software provides user-friendly functions which hopefully make polynomial regression methodology more accessible to researchers from a wide range of scientific fields.

References

Abuhamdeh, S. & Csikszentmihalyi, M. (2009). Intrinsic and extrinsic motivational orientations in the competitive context: An examination of person situation interactions. *Journal of Personality*, *77*, 1615–1635. doi:10.1111/j.1467-6494.2009.00594.x

Aiken, L. S. & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA, US: Sage Publications, Inc.

Bauer, D. J. & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. *Psychological Methods*, *14*, 101–125. doi:10.1037/a0015583

Baumann, N., Kaschel, R., & Kuhl, J. (2005). Striving for unwanted goals: Stress-dependent discrepancies between explicit and implicit achievement motives reduce subjective well-being and increase psychosomatic symptoms. *Journal of Personality and Social Psychology*, *89*, 781–799. doi:10.1037/0022-3514.89.5.781

Beem, A. L. (1995). A program for fitting two-phase segmented-curve models with an unknown change point, with an application to the analyis of strategy shifts in a cognitive task. *Behavior Research Methods, Instruments & Computers*, *27*, 392–399.

Blanton, H. & Jaccard, J. (2006a). Arbitrary metrics in psychology. *American Psychologist*, *61*, 27–41. doi:10.1037/0003-066X.61.1.27

Blanton, H. & Jaccard, J. (2006b). Tests of multiplicative models in psychology: A case study using the unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *113*, 155–166. doi:10.1037/0033-295X.113.1.155

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A. & Jackman, R. W. (1985). Regression diagnostics: An expository treatment of outliers and influential cases. *Sociological Methods & Research*, *13*, 510–542. doi:10.1177/0049124185013004004

Briñol, P., Petty, R. E., & Wheeler, S. C. (2006). Discrepancies between explicit and implicit self-concepts: Consequences for information processing. *Journal of Personality and Social Psychology*, *91*, 154–170. doi:10.1037/0022-3514.91.1.154

Brunstein, J. C., Schultheiss, O. C., & Grässmann, R. (1998). Personal goals and emotional well-being: The moderating role of motive dispositions. *Journal of Personality and Social Psychology*, *75*, 494–508.

Burnham, K. P. & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. Springer Science & Business Media.

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, *65*, 23–35. doi:10.1007/s00265-010-1029-6

Caplan, R. D. (1987). Person-environment fit theory and organizations: Commensurate dimensions, time perspectives, and mechanisms. *Journal of Vocational Behavior*, *31*, 248–267. doi:10.1016/0001-8791(87)90042-X

Cheung, G. W. (2009). Introducing the latent congruence model for improving the assessment of similarity, agreement, and fit in organizational research. *Organizational Research Methods*, *12*, 6–33. doi:10.1177/1094428107308914

Cohen, A., Nahum-Shani, I., & Doveh, E. (2010). Further insight and additional inference methods for polynomial regression applied to the analysis of congruence. *Multivariate Behavioral Research*, *45*, 828–852. doi:10.1080/00273171.2010.519272

Cohen, J. (1978). Partialed products are interactions; partialed powers are curve components. *Psychological Bulletin*, *85*, 858–866. doi:10.1037/0033-2909.85.4.858

Cronbach, L. J. & Furby, L. (1970). How we should measure 'change': Or should we? *Psychological Bulletin*, *74*, 68–80. doi:10.1037/h0029382

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. San Francisco, CA: Jossey-Bass.

Dalal, D. K. & Zickar, M. J. (2012). Some common myths about centering predictor variables in moderated multiple regression and polynomial regression. *Organizational Research Methods*, *15*, 339–362. doi:10.1177/1094428111430540

Dislich, F. X. R., Imhoff, R., Banse, R., Altstötter-Gleich, C., Zinkernagel, A., & Schmitt, M. (2012). Discrepancies between implicit and explicit self-concepts of intelligence predict performance on tests of intelligence. *European Journal of Personality*, *26*, 212–220. doi:10.1002/per.827

Edwards, J. R. (1994). Regression analysis as an alternative to difference scores. *Journal of Management*, *20*, 683–689.

Edwards, J. R. (1995). Alternatives to difference scores as dependent variables in the study of congruence in organizational research. *Organizational Behavior and Human Decision Processes*, *64*, 307–324. doi:10.1006/obhd.1995.1108

Edwards, J. R. (2001). Ten difference score myths. *Organizational Research Methods*, *4*, 265–287. doi:10.1177/109442810143005

Edwards, J. R. (2002). Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In F. Drasgow & N. W. Schmitt (Eds.), *Advances in measurement and data analysis* (pp. 350–400). San Francisco: Jossey-Bass.

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. doi:10.2307/256822

Edwards, J. R. & Shipp, A. J. (2007). The relationship between person-environment fit and outcomes: An integrative theoretical framework. In C. Ostroff & T. A. Judge (Eds.), *Perspectives on organizational fit.* (pp. 209–258). The organizational frontiers series. Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Edwards, J. R. & Cable, D. M. (2009). The value of value congruence. *Journal of Applied Psychology*, *94*, 654–677. doi:10.1037/a0014891

Frost, D. M. & Forrester, C. (2013). Closeness discrepancies in romantic relationships: Implications for relational well-Being, stability, and mental health. *Personality and Social Psychology Bulletin*. doi:10.1177/0146167213476896

Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Good, I. (1982). C140. Standardized tail-area probabilities. *Journal of Statistical Computation and Simulation*, *16*, 65–66. doi:10.1080/00949658208810607

Griffin, D., Murray, S., & Gonzalez, R. (1999). Difference score correlations in relationship research: A conceptual primer. *Personal Relationships*, *6*, 505–518. doi:10.1111/j.1475-6811.1999.tb00206.x

Hagemeyer, B., Neberich, W., Asendorpf, J. B., & Neyer, F. J. (2013). (In)Congruence of implicit and explicit communal motives predicts the quality and stability of couple relationships. *Journal of Personality*, *81*, 390–402. doi:10.1111/jopy.12016

Harring, J. R., Weiss, B. A., & Hsu, J. C. (2012). A comparison of methods for estimating quadratic effects in non-linear structural equation models. *Psychological Methods*, *17*, 193–214. doi:10.1037/a0027539

Higgins, E. T. (2000). Making a good decision: Value from fit. *American Psychologist*, *55*, 1217–1230. doi:10.1037/0003-066X.55.11.1217

Hmel, B. A. & Pincus, A. L. (2002). The meaning of autonomy: On and beyond the interpersonal circumplex. *Journal of Personality*, *70*, 277–310.

Hofer, J. & Chasiotis, A. (2003). Congruence of life goals and implicit motives as predictors of life satisfaction: Cross-cultural implications of a study of Zambian male adolescents. *Motivation and Emotion*, *27*, 251–272.

Hofer, J., Chasiotis, A., & Campos, D. (2006). Congruence between social values and implicit motives: Effects on life satisfaction across three cultures. *European Journal of Personality*, *20*, 305–324. doi:10.1002/per.590

Hofer, J., Busch, H., & Kärtner, J. (2010). Self-regulation and well-being: The influence of identity and motives. *European Journal of Personality*, *25*, 211–224.

Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55. doi:10.1080/10705519909540118

Huang, L. & Pashler, H. (2005). Quantifying object salience by equating distractor effects. *Vision Research*, *45*, 1909–1920. doi:10.1016/j.visres.2005.01.013

Hurvich, C. M. & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*, 297–307.

Irwin, J. R. & McClelland, G. H. (2001). Misleading heuristics and moderated multiple regression models. *Journal of Marketing Research*, *38*, 100–109. doi:10.1509/jmkr.38.1.100.18835

Johns, G. (1981). Difference score measures of organizational behavior variables: A critique. *Organizational Behavior and Human Performance*, *27*, 443–463. doi:10.1016/0030-5073(81)90033-7

Johnson, D. W., McCarty, K., & Allen, T. (1976). Congruent and Contradictory Verbal and Nonverbal Communications of Cooperativeness and Competitiveness in Negotiations. *Communication Research*, *3*, 275–292. doi:10.1177/009365027600300303

Kazén, M. & Kuhl, J. (2011). Directional discrepancy between implicit and explicit power motives is related to well-being among managers. *Motivation and Emotion*, *35*, 317–327. doi:10.1007/s11031-011-9219-8

Keele, L. J. (2008). *Semiparametric regression for the social sciences*. New York, NY: Wiley.

Kehr, H. M. (2004). Implicit/explicit motive discrepancies and volitional depletion among managers. *Personality and Social Psychology Bulletin*, *30*, 315–327. doi:10.1177/0146167203256967

Klein, A. & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*, 457–474.

Kristof, A. L. (1996). Personorganization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology*, *49*, 1–49.

Kristof-Brown, A. L. & Billsberry, J. (Eds.). (2013). *Organizational fit: Key issues and new directions*. Wiley-Blackwell.

McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review*, *96*, 690–702.

McQuarrie, A. D. R. & Tsai, C.-L. (1998). *Regression and time series model selection*. Singapore: World Scientific.

Mongrain, M. & Vettese, L. C. (2003). Conflict over emotional expression: Implications for interpersonal communication. *Personality and Social Psychology Bulletin*, *29*, 545–555. doi:10.1177/0146167202250924

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis*. John Wiley & Sons.

Mooijaart, A. & Bentler, P. M. (2010). An alternative approach for nonlinear latent variable models. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*, 357–373. doi:10.1080/10705511.2010.488997

Muggeo, V. M. R. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine*, *22*, 3055–3071.

Myers, R. H., Montgomery, D. C., & Fitzgerald, C. M. (2009). *Response surface methodology: Process and product optimization using designed experiments*. Hoboken, New Jersey: John Wiley & Sons.

Peter, J. P., Churchill, G. A., & Brown, T. J. (1993). Caution in the use of difference scores in consumer research. *Journal of Consumer Research*, *19*, 655–662. doi:10.1086/209329

R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria. Retrieved from http://www.R-project.org/

Rogers, C. R. (1959). A theory of therapy, personality, and interpersonal relationships, as developed in the client-centered framework. In S. Koch (Ed.), *Psychology: A study of a science*. New York, NY US: McGraw-Hill.

Rogers, C. R. (2004). *On becoming a person: A therapist's view of psychotherapy*. London: Constable.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36.

Rousseeuw, P. J., Ruts, I., & Tukey, J. W. (1999). The bagplot: A bivariate boxplot. *The American Statistician*, *53*, 382–387.

Schallberger, U. (2005). *Kurzskala zur Erfassung der Positiven Aktivierung, Negativen Aktivierung und Valenz in Experience Sampling Studien (PANAVA-KS) [Short scales for the assessment of positive affect, negative affect, and valence in experience sampling studies]*. University of Zurich.

Schönbrodt, F. D. (2012). *The Munich 100 family study.* Munich, Germany.

Schönbrodt, F. D. (2015a). *Response Surface Analyses with the RSA package*. Munich, Germany.

Schönbrodt, F. D. & Gerstenberg, F. X. R. (2012). An IRT analysis of motive questionnaires: The Unified Motive Scales. *Journal of Research in Personality*, *46*, 725–742. doi:10.1016/j.jrp.2012.08.010

Schönbrodt, F. D. (2015b). *RSA: An R package for response surface analysis (version 0.9.7)*. Retrieved from http://cran.r-project.org/web/packages/RSA/index.html

Schröder-Abé, M., Rudolph, A., & Schütz, A. (2007). High implicit self-esteem is not necessarily advantagous: Discrepancies between explicit and implicit self-esteem and their relationship with anger expression and psychological health. *European Journal of Personality*, *21*, 319–339. doi:10.1002/per.626

Schüler, J., Job, V., Fröhlich, S. M., & Brandstätter, V. (2008). A high implicit affiliation motive does not always make you happy: A corresponding explicit motive and corresponding behavior are further needed. *Motivation and Emotion*, *32*, 231–242. doi:10.1007/s11031-008-9096-y

Schultheiss, O. C. & Pang, J. S. (2007). Measuring implicit motives. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 322–344). Guilford Press.

Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggestad, E. D. (2010). Polynomial regression with response surface analysis: A powerful approach for examining moderation and overcoming limitations of difference scores. *Journal of Business and Psychology*, *25*, 543–554. doi:10.1007/s10869-010-9183-4

Symonds, M. R. E. & Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, *65*, 13–21. doi:10.1007/s00265-010-1037-6

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd edition). Connecticut: Graphics Press.

Wagenmakers, E.-J. & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*, 192–196. doi:10.3758/BF03206482

Wilcox, R. R. (2012). *Introduction to robust estimation and hypothesis testing*. Amsterdam; Boston: Academic Press.

Wilkinson, L. & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*, 594–604. doi:10.1037/0003-066X.54.8.594

Winter, D. G. (1991). *Manual for scoring motive imagery in running text (4th edition)*. Ann Arbor: University of Michigan.

Yang, L. Q., Levine, E. L., Smith, M. A., Ispas, D., & Rossi, M. E. (2008). Person–environment fit or person plus environment: A meta-analysis of studies using polynomial regression analysis. *Human Resource Management Review*, *18*, 311–321. doi:10.1016/j.hrmr.2008.07.014

Appendix A
Conditions for Weak and Strong Fit Patterns

Constraints for the general polynomial model can be derived in order to identify models which satisfy the conditions for a fit pattern. The first and second condition of the *weak fit pattern* state that for each X_i there is a single optimal value Y_j which maximizes/minimizes the response variable Z, and vice versa. Every vertical cut through the surface of a two-variable second-degree polynomial has the functional form of a parabola. Hence, for each X and each Y there is a single maximum/minimum point, as long as the regression weights for the quadratic terms are not zero: $b_3 \neq 0$, and $b_5 \neq 0$. Furthermore, the quadratic terms must have the same sign, otherwise the optimal response for X_i would be the the worst response for Y_j. All these conditions can be summarized in one inequality:

$$b_3 b_5 > 0 \tag{15}$$

The third and fourth condition, which state that greater deviations from the optimal point have greater impact on the response than smaller deviations, is automatically satisfied by the general functional form of the parabola, as the response changes quadratically on both sides of the vertex.

To summarize, all polynomial regressions which follow inequality (15) satisfy the conditions for a weak fit pattern. This inequality condition is implicitly present in the constraint for the SRRR model, which is $b_4 = -2\sqrt{b_3 b_5}$. As the square root is only defined when (15) is satisfied, all SRRR models follow the weak fit pattern. Furthermore, as all constraints are passed down to nested models (Bollen, 1989), it can be concluded that all constrained models depicted in Figure 3 are instantiations of the weak fit pattern.

The condition for a *strong fit pattern* states that the optimal matches between X and Y values must be symmetric. The values $Y_{opt.i}$ where Z is maximal for a given X_i can be derived by setting the first derivative with respect to Y to zero:

$$\frac{dX}{dY} = b_2 + b_4X + 2b_5Y = 0 \tag{16}$$

Similarly, the values $X_{opt.j}$ where Z is maximal for a given Y_j can be derived by setting the first derivative with respect to X to zero:

$$\frac{dY}{dX} = b_1 + 2b_3X + b_4Y = 0 \tag{17}$$

Rearranging these equation leads to point-intercept equations describing the maximum lines on the X-Y-plane:

$$\begin{aligned} Y &= -\frac{b_2}{2b_5} - \frac{b_4}{2b_5}X \\ Y &= -\frac{b_1}{b_4} - \frac{2b_3}{b_4}X \end{aligned} \tag{18}$$

The condition for a strong fit pattern can be reformulated as that the maximum line for X is identical to the maximum line of Y. Equating the equations of (18) gives two constraints on polynomial regressions in order to satisfy the strong fit pattern: (1) $b_1 = (b_2b_4)/2b_5$, for equal intercepts, and (2) $b_4 = 2\sqrt{b_3b_5}$, for equal slopes.

The constraint for equal slopes is identical to the single constraint of the SRRR model. Hence, all constrained models shown in Figure 3 have equal slopes of the maximum lines. The constraint for identical intercepts is identical to one of the constraints of the SRSD model. Hence, all flat ridge models (i.e., SRSD, SSQD, and SQD) satisfy the conditions for a strong fit pattern.

Appendix B

Installing R and the RSA Package, Demo script

First, the R base system has to be installed. Installation files for R can be obtained from http://cran.r-project.org/ and are provided for all major operating systems (Windows, Mac OS, Linux). Detailed instructions for installation can be obtained from the R website (http://www.r-project.org). The R installation provides the base system and basic packages for standard statistical analyses such as multiple regression, ANOVAs, or factor analyses. There are, however, numerous additional packages with new functions, such as the RSA package. To install RSA, one has to launch the R console and to type `install.packages("RSA")`. R will automatically download the necessary files, and installs the package to your system. The RSA package relies on some other packages which will be automatically installed on the system as well. After installation, the RSA package can be loaded into the current R session by typing `library(RSA)`. Typing ?RSA opens the help file for the main function, in which also links to help files for other functions can be found.

Following script loads the built-in data set and performs all analyses shown in this paper. For the following script, it is assumed that the raw data is loaded into a data frame with the name `motcon2`. The variables of interest are called `IM` (implicit affiliaiton/intimacy motive), `EM` (explicit intimacy motive), and `VA` (affective valence). Response surface analysis can be run with a single line of code: `r1 <- RSA(VA ~ IM*EM, data=motcon2)`. This command evokes the main function `RSA` and defines the variables to be used from the data frame `motcon2`. The first parameter, `VA ~ IM*EM`, follows R's formula style and is read as '`VA` is a function of `IM` and `EM`' and defines the outcome and the predictor variables. The result of the `RSA` function is stored in a new object named `r1`.

By default, the `RSA` function estimates all models displayed in Figure 3 (absolute difference models and additional models can be computed on request). Summary information of the results can be obtained with `summary(r1)`. A detailed table of model comparisons can be printed with the command `aictab(r1)` (cf. Table 1). A detailed look on the parameters of a specific model (e.g., the SRRR model) can be obtained with: `getPar(r1, "coef", model="SRRR")`. Bootstrapped standard errors and CIs can be computed with the following commands: `c1 <- confint(r1, model="SRRR", method="boot", R=10000)` (Bootstrapped CIs are only computed for the final model as the procedure takes several minutes for each model). Finally, the model can be visualized using the `plot` command: `plot(r1, model="SRRR", xlab="Implicit motive", ylab="Explicit motive", zlab="Affective valence", type="3d")` (see Figure 4).

The full analysis presented in this paper can be done with the following script:

```
## lines beginning with # are comments

# if not already done:
# install the RSA package
# (only has to be done once)
# install.packages("RSA")

# load the RSA package for
# the active session
library(RSA)

# open the help page
?RSA

## Motive congruency example
# load the built-in data set
data(motcon2)

# Compute the RSA and store the result into
# the new variable r1
r1 <- RSA(VA ~ IM*EM, data=motcon2)

# Show summary of the RSA
summary(r1)
```

```
# Confidence intervals with robust standard errors
CI.full <- confint(r1, model="full")
CI.SRRR <- confint(r1, model="SRRR")

# Percentile bootstrap CIs and p values
# This takes several minutes
CI.boot.full <- confint(r1, model="full",
        method="boot", R=10000,
        parallel="multicore", ncpus=4)

CI.boot.SRRR <- confint(r1, model="SRRR",
        method="boot", R=10000,
        parallel="multicore", ncpus=4)

# Table with model comparisons
a1 <- aictab(r1, plot=TRUE)
print(a1)


# Show all RSA parameters of the final model
# with robust SEs, p values, and CIs
getPar(r1, "coef", model="SRRR")


## Plot the final model
# for information on the several options,
# see ?plotRSA

# 3d plot
plot(r1, axes="PA1", model="SRRR",
project=c("points", "PA1", "LOC", "LOIC", "hull"),
points=FALSE, xlim=c(-3.1, 3.1), ylim=c(-3.1, 3.1),
param=FALSE, legend=FALSE, bw=TRUE,
pal.range="surface", pal="flip",
xlab="Implicit motive", ylab="Explicit motive",
zlab="Affective valence")

# contour plot
plot(r1, points=TRUE, model="SRRR", axes="PA1",
xlim=c(-3.1, 3.1), ylim=c(-3.1, 3.1), showSP=FALSE,
legend=TRUE, bw=TRUE, pal.range="surface", pal="flip",
xlab="Implicit motive", ylab="Explicit motive",
zlab="Affective valence", type="contour")


## Additional functions

# interactive, rotatable 3d plot
plot(r1, model="SRRR", type="interactive")

# open an interactive widget with control
# sliders for regression weights
demoRSA()
```

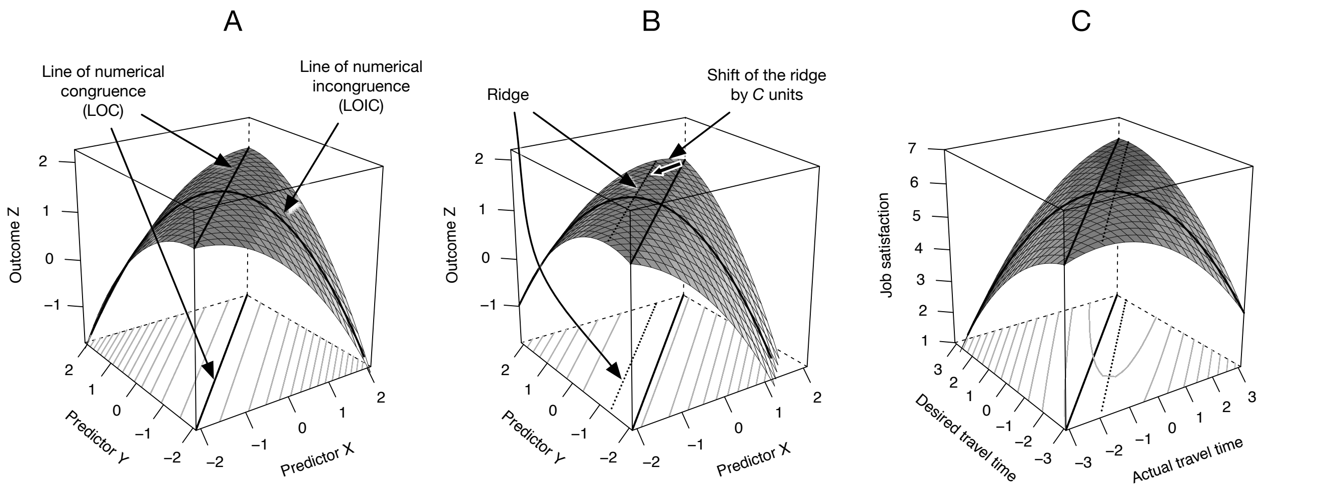

Figure 1. Three examples of response surface plots. (A) Basic squared differences model, (B) Shifted squared differences model, shifted by $C = 0.75$ units of the X axis, (C) Travel time example from Edwards (2002).

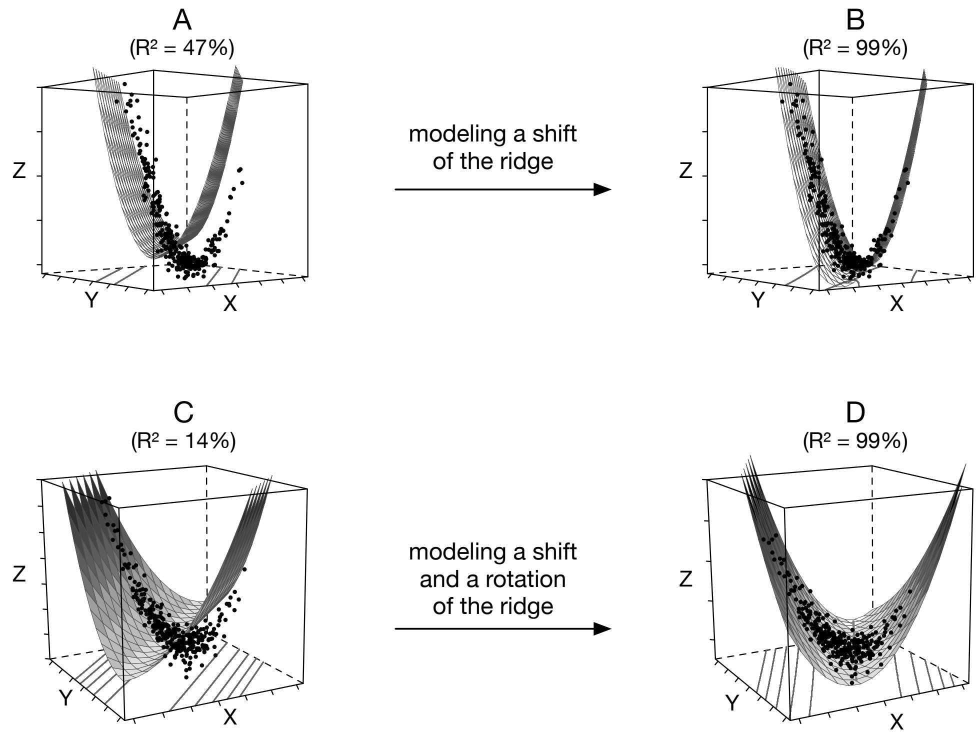

Figure 2. The effect of choosing an inappropriate center and scale. The basic squared difference model (panels A and C) is fixed on the X=Y line, and has an R^2 of 47% (panel A), respectively 14% (panel C). The shifted squared difference model (SSQD; panel B) and the shifted and rotated squared difference model (SRSQD; panel D) have an R^2 of 99%.

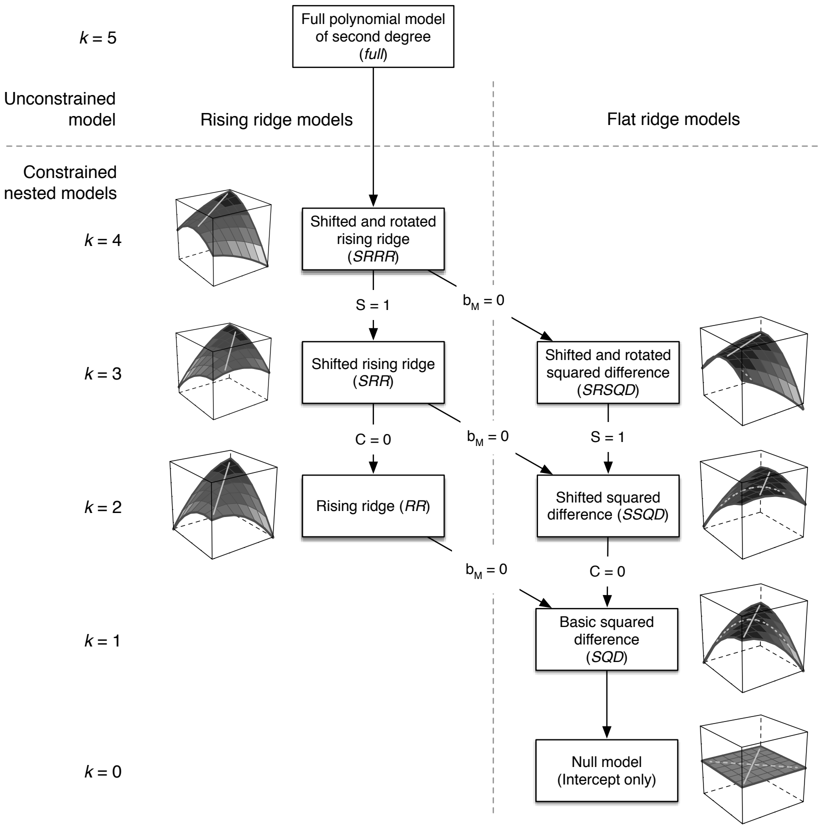

Figure 3. A tree of nested models. *k* is the number of free parameters in relation to the intercept-only null model. Equations on the arrows describe the testable constraint that is induced by each reduced model.