Methodika
1988, Vol. II, Issue 1, pp. 53–78

Conditional Expectations: An Introduction to the Concept and its Applications in Empirical Sciences*)

Rolf Steyer

Abstract: An introduction is given to the concept of the conditional expectation that may be viewed as the generalized concept of regression. It is shown that the conditional expectation $E(Y|X)$ of a numerical random variable Y given a possibly multidimensional and not necessarily numerical random variable $X = (X_1, \dots, X_m)$ is an adequate concept to formulate very restrictive as well as very weak propositions on the dependency of Y on X. The focus is not on the dependency in the sample but on the "true" or theoretical dependency to be estimated and tested by means of samples. The properties of the residual $F := Y - E(Y|X)$ are studied in detail using some general rules of computation which are formulated and illustrated in the Appendices. These properties, such as the uncorrelatedness of F with each X_j, $j \in J = \{1, \dots, m\}$, are derived from the definition of F. No assumptions are required for the functional form of $E(Y|X)$, nor on the distributions of X, Y, or F. Two applications are treated in some detail. In the first one, the conditional expectation is used to unravel an old fallacy (going back to Simon, 1954) concerning uncorrelated errors and causality in path models. In the second application the conditional expectation is used to derive a new and simple procedure to determine a coefficient (proposed by Tack, 1980, 1986) for the "sensitivity" of a test-score variable with respect to a given set of situations.

Key words: Conditional expectations, path analysis, test theory, regression theory.

1. Introduction

About a hundred years ago, when the concepts of correlation and regression were introduced to the bio- and social sciences by Bravais (1846), Galton (1877, 1889), Pearson (1896, 1901), and Yule (1897), a new type of law was discovered. In contrast to the kind of laws that could be described by (deterministic) mathematical functions, the new kind of law was nondeterministic in nature. Galton (1877) proposed for instance, that the dependency of the weight of sweet peas on the weight of the mother plant can be described by a regression line. What was

---

*) The author thanks Prof. Dr. K. W. Schaie, B. Baltes-Götz, E. Erdfelder, J. Funke, M. Schmitt, and two referees for helpful comments on previous drafts of this paper.

Author's address: Dr. Rolf Steyer, Universität Trier, Fachbereich I – Psychologie, D-5500 Trier.

new about this regression (or "reversion" as Galton called it in that paper)?

A *first* important feature is that the dependence described by this regression line was lawful but *nondeterministic per se,* that is, the deviations from the regression line were not due to measurement error, but reflected an essential property of the phenomenon itself. Although regression analyses were known before Galton, they were applied mainly to deal with measurement error, for example, in geophysics (Adrain, 1818), astronomy (Littrow, 1818, 1833), and psychophysics (Fechner, 1860). These scientists assumed that the underlying laws could be formulated by deterministic mathematical functions, although it was not possible to observe them in their pure form, because the variables involved could not be measured without error.

While stochastic types of laws were already known earlier (e.g., Mendel, 1866), the new type of law introduced with the concepts of regression and correlation was concerned with *quantitative variables,* which is the *second* important feature. Thus, at the end of the last century, stochastic laws and concepts had emerged involving both qualitative (see also Yule, 1900, 1907, 1912) and quantitative variables.

Later on, the concept of regression was generalized to deal with *multiple independent variables* (multiple regression) that may also be *qualitative* in nature. This development is mainly ascribed to Fisher (1925) and became popular under the label "analysis of variance", first in the agricultural sciences and later on in sciences like psychology. The fact that analysis of variance (with fixed effects) is a special case of regression analysis was rediscovered for psychology mainly by Cohen (1968) in his well-known article "Multiple Regression as a General Data-Analytic System".

The title of Cohen's article reflects a tendency in modern methodology to treat correlation and regression as "data-analytic" methods and to neglect their importance as theoretical concepts which allow to formulate substantive scientific propositions much more precisely than was possible before these concepts were known. This tendency may also be responsible for ignoring an important development in probability theory due to Kolmogoroff (1933). This mathematician succeeded *first* to reformulate probability theory as a special branch of mathematical measure theory, thus establishing probability theory as a mathematical discipline. *Second,* Kolmogoroff developed the theory of *conditional expectations* that may also be viewed as a formalization and generalization of the concept of regression. In fact, these terms are now used synonymously by mathematicians (see, e.g., Bellach, Franken, and Warmuth, 1978).

The present paper is intended to popularize the concept of a conditional expectation, which may be applied almost always when lawful but nondeterministic dependencies are to be formulated. We begin with hinting at a problem in the usual formulation of the *linear* regression model and give a *general* characterization of the conditional expectation or *general* regression $E(Y \mid X)$. Next, we study the properties of the residual $F = Y - E(Y \mid X)$. This residual variable deserves special attention, because it is the crucial feature that discriminates between deterministic and stochastic dependencies. We then briefly review some well-known classes of models that can be formulated in terms of conditional expectations. Furthermore, an application is discussed in which the conditional expectation is used to derive a new and simple procedure to determine a coefficient for the "sensitivity" of a test-score variable with respect to a given set of situations. Finally, we outline some other applications that appeared in the methodological literature.

2. Simple Linear Regression

Let us look more closely at the concept of linear regression as a conceptual tool to describe the "true" dependence of a random variable, say Y, on another one, say X, both numerical. By "true" we mean that it is not the dependence in a sample but the theoretical dependence to be estimated by samples. As has already been discussed above, such a dependence is, in many cases, not deterministic. Therefore, an error variable, say F, has to be introduced. This variable might have a variance greater than zero because of at least four reasons. First, it might be due to measurement error of the regressand Y. Second, there might be other variables on which Y depends. Third, the true regressive dependence of Y and X might not be linear, and fourth, it might reflect the fact that the dependence of Y on X is nondeterministic per se. In all these situations, an equation may be formulated specifying how Y depends on X. In the cases of simple *linear* regression, the equation is

$$(1) \qquad Y = \alpha_{Y0} + \alpha_{YX} \cdot X + F,$$

where α_{Y0} and α_{YX} are real constants. In order to express the idea that the error variable F is, in some sense, random and not systematically related to X, the assumptions

$$(2) \qquad E(F) = 0 = C(F, X)$$

are added, that is, the expectation $E(F)$ of the error variable F is zero, as well as the covariance $C(F, X)$ of F and X. These are the usual assumptions of simple linear regression in that kind of literature which uses regression to formulate substantive scientific propositions and not only as a technique to analyse data (see e.g., Hummell and Ziegler, 1976; Kenny, 1979; or Jöreskog and Sörbom, 1981).

If we define the conditional expectation $E(Y \mid X = x)$ of Y given $X = x$ to be the best prediction of Y given $X = x$, then the Equations 1 and 2 are not sufficient to derive that the best prediction of Y given the value x of X is equal to $\alpha_{Y0} + \alpha_{YX} \cdot x$; a prediction, the model should imply, if it is supposed to make sense at all. This prediction can only be derived from the equation on Y if the additional assumption is made that the conditional expectation of the residual F given $X = x$ is zero, that is if

$$(3) \qquad E(F \mid X = x) = 0, \quad \text{for all values } x \text{ of } X.$$

This assumption on F is much stricter than assuming that the unconditional expectation $E(F)$ is zero. Using this last equation and some rules of computation for the conditional expectation $E(Y \mid X = x)$ of Y given $X = x$, Equation 1 now yields

$$(4) \qquad \begin{aligned} E(Y \mid X = x) &= E(\alpha_{Y0} + \alpha_{YX} \cdot X + F \mid X = x) \\ &= E(\alpha_{Y0} \mid X = x) + E(\alpha_{YX} \cdot X \mid X = x) + E(F \mid X = x) \\ &= \alpha_{Y0} + \alpha_{YX} \cdot x, \end{aligned}$$

the desired property that $\alpha_{Y0} + \alpha_{YX} \cdot x$ is the best prediction of Y given a value x of X.

Hence, the Equations 1 and 2 are not sufficient to define the "true" linear regression because they do not imply that the pairs of points $(x, E(Y \mid X = x))$ lie on a straight line. Therefore, the model defined by Equations 1 and 2 might be called a *linear quasi-regression* in order to distinguish it from the true linear regression.

3. The General Concept of a Conditional Expectation or Regression

The situation described above is a very special one in that there is only one regressand (Y) and one numerical regressor (X) and that the dependence of Y on X is characterized by the Equations 1 and 2. A more general task, however, is to formulate propositions on how the best prediction of Y given $X = x$, or in mathematical well-defined terms, on how the conditional expectations $E(Y \mid X = x)$ of a specified stochastic variable Y given $X = x$ vary along with the values x of a specified m-dimensional random variable $X = (X_1, \ldots, X_m)$, where $m \geq 1$ and where X is not necessarily numerical. An adequate concept for this purpose is the conditional expectation (synonymously, regression) $E(Y \mid X)$ of a random variable Y (with finite expectation) given a (usually other) random variable X.

The conditional expectation $E(Y \mid X)$ of a numerical random variable Y on a probability space $(\Omega, \mathscr{A}, P)$ (with finite expectation) given a random variable X (with values in an arbitrary set) is defined by the following two conditions:

(a) $E(Y \mid X)$ is measurable with respect to the sigma-algebra $\mathscr{A}(X)$ generated by X;
(b) $E(I_C \cdot E(Y \mid X)) = E(I_C \cdot Y)$, for all $C \in \mathscr{A}(X)$.

The definition is hard to understand with no background in basic concepts of measure theory. Condition (a) basically means that $E(Y \mid X)$ is a function of X. An alternative characterization may be useful. If Y can be assumed to have finite variance, $E(Y \mid X)$ may be conceptualized to be the optimal function $f(X)$ of X to predict Y. More formally, $E(Y \mid X)$ is that numerical measurable function $f(X)$ of X, which minimizes the following function of $f(X)$:

$$(5) \qquad E((Y - f(X))^2).$$

Note that no assumption has to be made concerning the type of $f(X)$ (see Appendix B for a proof). $E(Y \mid X)$ is a numerical random variable, the values of which are identical with the conditional expectations $E(Y \mid X = x)$ of Y given $X = x$. Hence, $E(Y \mid X)$ is a function of X. It may be regarded as a generalization of the best predictor of Y that often appears in the literature on applied statistics. Note, however, that $E(Y \mid X)$ is defined without reference and restriction to a certain equation. The linear quasi-regression, on the other hand, may also be defined to be that *linear* function $f(X) = \alpha_0 + \alpha_1 \cdot X$, $\alpha_0, \alpha_1 \in \mathbb{R}$, which minimizes 5. The conditional expectation is uniquely defined only up to probability 1. Therefore, equations on conditional expectations only hold "almost surely", in general, a mathematical detail that is not mentioned any more in this introductory presentation. Also note that we always assume a underlying probability space $(\Omega, \mathscr{A}, P)$ on which X, Y, and $E(Y \mid X)$ are random variables.

Applying the concept of a conditional expectation $E(Y \mid X)$ of Y given X, the (true) linear regression model can now be formulated as follows:

$$(6) \qquad E(Y \mid X) = \alpha_{Y0} + \alpha_{YX} \cdot X,$$

that is, $E(Y \mid X)$ is assumed to be a linear function of the numerical random variable X. As will be seen later on, this single assumption allows for the deduction of the Equations 1 to 4 and even many more important properties. If X is dichotomous with values zero and one, for example, there are only two different values of $E(Y \mid X)$, namely $E(Y \mid X = 0) = \alpha_{Y0}$ and $E(Y \mid X = 1) = \alpha_{Y0} + \alpha_{YX}$ (see Equation 6). In this case, $\alpha_{YX} = E(Y \mid X = 1) - E(Y \mid X = 0)$ is the difference between the two true group means of Y and Equation 6 will always hold. Hence, if X is dichotomous, there is no question whether or not the regression $E(Y \mid X)$ is linear. The only question is whether or not $\alpha_{YX} = 0$.

If, on the other hand, X is continuous, the values of the pair of variables $(X, E(Y \mid X))$ are the points on the regression line. If X has more than two values, Equation 6 is restrictive and may be wrong in a substantive application. However, propositions about conditional expectations do not have to be as restrictive as that contained in Equation 6. A less restrictive proposition would be, for example, to postulate that $E(Y \mid X)$ is an increasing function of X, which is a more precise formulation of "Y tends to increase with X". If the number of observations of X and Y is big enough, the values $E(Y \mid X = x)$ of $E(Y \mid X)$ can be estimated and it can be decided whether or not $E(Y \mid X)$ increases with X. Such a decision will usually be based on statistical decision procedures, of course.

Note that the proposition that $E(Y \mid X)$ increases with X is much weaker than propositions often found in psychological theories saying that Y increases with X. The latter proposition is deterministic and therefore must be rejected if there are two pairs (x_1, y_1), (x_2, y_2) of values of X and Y such that $x_1 < x_2$ and $y_1 > y_2$.

The arguments discussed above are intended to show that the formalization of psychological propositions does not necessarily lead to propositions that are unrealistically restrictive. Even weak propositions like the one discussed above may easily and *precisely* be formulated in terms of conditional expectations. What is advocated here, is the use of conditional expectations to formulate *theoretical scientific propositions* that should not be mixed up with descriptive propositions characterizing a given sample.

Using conditional expectations does not only allow for the precise formulation of *weak* propositions, but it also allows to introduce *any degreee of restrictiveness* up to deterministic propositions. The most restrictive models specify a function for $E(Y \mid X)$ and postulate $Y = E(Y \mid X)$, implying that there is no residual variable $F = Y - E(Y \mid X)$ with positive variance. Hence, the conditional expectation is an adequate concept for stochastic as well as for deterministic propositions.

Usually, however, propositions on conditional expectations allow for a residual F with positive variance. In this case it is important to know the properties of F studied in the following section.

4. The Residual and its Properties

In this section, we describe the most important properties of the residual $F := Y - E(Y \mid X)$, which hold regardless of the assumptions formulated about $E(Y \mid X)$; that is, whether $E(Y \mid X)$ is assumed to be a linear function of X as in the simple linear regression model, any other polynomial, or whether it is just a nondecreasing function of X, for instance. These properties even hold if X is not numerical at all such as in classical error (test) theory models (see Section 9 below).

In the following sections, the regressor X may be m-dimensional, or, synonymously, X may be a random vector consisting of m one-dimensional, not necessarily numerical random variables $X_1, \ldots, X_m$, that is, $X = (X_1, \ldots, X_m)$. In this case, the regression $E(Y \mid X)$ may also be written $E(Y \mid X_1, \ldots, X_m)$. An equivalent notation is $E(Y \mid X_j, j \in J)$, where $J = \{1, \ldots, m\}$ denotes a finite index set. However, the propositions to be made about the residual F even hold if J is a countably infinite index set, that is, if $J = \{1, 2, \ldots\}$ is the set of natural numbers. It will be assumed throughout that the regressand Y has finite expectation, say $E(Y)$, and that Y and X are random variables on the same probability space $(\Omega, \mathscr{A}, P)$.

Formulating substantive propositions by specifying assumptions on a conditional expectation $E(Y \mid X)$ of a random variable Y on a, possibly multidimensional, random variable $X = (X_1, \ldots, X_m)$, implicitly involves propositions on a residual variable F (see Section 2 for some interpretations of F) that is defined by

$$F := Y - E(Y \mid X). \tag{7}$$

We must ask then, what are these implicit propositions about F and what are its general properties? Note again that *no assumptions* are necessary to derive the equations in the following paragraphs (see the Appendix B).

First of all, and almost trivially, F and $E(Y \mid X)$ add together yielding Y. In formula,

$$Y = E(Y \mid X) + F, \tag{8}$$

which immediately follows from Equation 7. This additivity property of F is not self-evident. In psychophysics, for example, the power law is formulated with a multiplicative error variable (cf., e.g., Thomas, 1983), although the log-transformation yields an additive error again (see, e.g., Steyer and Erdfelder, 1984).

Second, and most importantly, the conditional expectation of the residual F given any measurable function $f(X)$ of X is zero, that is,

$$E(F \mid f(X)) = 0. \tag{9}$$

Equation 9 implies that the expectation of F given all combinations of values of the variables X_j, $j \in J$, is zero. Special cases of Equation 9 are $E(F \mid X) = 0$ (with $f(X) = X$) and $E(F \mid X_j) = 0$, $j \in J$ (with $f(X) = X_j$).

Equation 9 is a very fundamental property of F, because it allows to derive a number of other properties (see the Appendix B).

A *third* property of the residual F is that its (unconditional) expectation is equal to zero, that is

$$E(F) = 0. \tag{10}$$

This property is weaker than the second one, because it follows from the second (see the Appendix B), but not vice versa. Contrary to the latter, Equation 10 does not imply that the conditional expectations $E(Y \mid X = x)$ of F given $X = x$ are zero, for all values x of X. Considering Equations 9 and 10, we can say that F is *regressively independent* from any measurable function $f(X)$ of its regressor X: $E(F \mid f(X)) = E(F)$.

A *fourth* property of the residual $F = Y - E(Y \mid X_1, \ldots, X_m)$ is that its covariance with each X_j, $j \in J = \{1, \ldots, m\}$, is zero, provided that each X_j is numerical with finite expectation and finite variance (which implies that the covariance exists). In formula,

$$C(F, X_j) = 0, \quad \text{for all } j \in J = \{1, \ldots, m\}, \tag{11}$$

which implies that the coefficients of correlation of F and each X_j are zero. If $S(X)$ and $S(Y)$ denote the positive square roots of the corresponding variances, the correlation between two numerical random variables X and Y with finite expectation and variance is defined by

$$R(X, Y) := \begin{cases} \dfrac{C(X, Y)}{S(X) \cdot S(Y)}, & \text{if } S(X), S(Y) > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

A *fifth* property of the residual $F = Y - E(Y \mid X)$ is

$$E(F \mid E(Y \mid X)) = E(F), \tag{13}$$

that is, the residual F is regressively independent from the regression $E(Y \mid X)$ of Y given X. In fact, this is a special case of the second property with $f(X) := E(Y \mid X)$. This fifth property is mentioned separately only because it has a special well-known interpretation in the context of error (test) theory models (see Section 9).

Equation 13 also implies the *sixth* property

$$C(F, E(Y \mid X)) = 0, \tag{14}$$

which immediately follows from Equation 13 (see the Appendix B). According to this equation, the covariance of the residual F and the regression of Y given X (the conditional expectation) is zero.

Equation 14 may be used to derive the *seventh* property of F:

$$V(Y) = V(E(Y \mid X)) + V(F). \tag{15}$$

Hence, the variance of Y is a sum of the variance of the regression $E(Y \mid X)$ and the variance of the residual F. $V(E(Y \mid X))$ is that part of the variance of Y that is determined by X and $V(F)$ is the residual variance, that is, that part of $V(Y)$ that is not determined by X.

To summarize, the definition of the residual F (Equation 7) implies – *without any assumptions* – the seven properties discussed above. It should be noted however, that the definition of F, and especially F and each X_j being uncorrelated, do *not* imply that the absolute value variable $|F|$ or the conditional variance of F are independent from X. Hence, if such properties are desired to be part of a model, they have to be formulated as additional assumptions. It will be shown in Section 6 that the conditional expectation is also very useful to formulate propositions on how the variance of a random variable Y varies along with another random variable X. In Section 9 we show that the so-called "axioms" of classical test theory are special examples of the properties of the residual discussed above.

5. The Coefficient of Determination

A concept closely related to the residual $F = Y - E(Y \mid X)$ is the *coefficient of determination* $R(Y \mid X)^2$, which is defined for a numerical random variable Y with finite expectation $E(Y)$ and finite variance $V(Y)$ by

$$R(Y \mid X)^2 := \begin{cases} \dfrac{V(E(Y \mid X))}{V(Y)}, & \text{if } V(Y) > 0, \\ 0, & \text{if } V(Y) = 0, \end{cases} \tag{16}$$

that is, the coefficient of determination of the random variable Y with respect to the regression $E(Y \mid X)$ is defined to be the variance of $E(Y \mid X)$ devided by the variance of Y, provided that $V(Y)$ is greater than zero. Otherwise, $R(Y \mid X)^2$ is defined to be zero. The positive square root of $R(Y \mid X)^2$ is called the *coefficient of multiple correlation* and denoted by $R(Y \mid X)$.

Again, these definitions are not based on any assumptions about $E(Y \mid X)$ or X. Thus X may be nonnumerical such as in test theory models, for instance, where $R(Y \mid X)^2$ is also called the "reliability coefficient" (see Section 9).

It is easily seen that $R(Y \mid X)^2$ and $R(Y \mid X)$ may take on values between zero and one. They will be zero, if $E(Y \mid X) = E(Y)$, that is, if Y is regressively independent from X, because in this case, $V(E(Y \mid X)) = V(E(Y)) = 0$, since $E(Y)$ is a constant. They will be one, if $E(Y \mid X) = Y$ and $V(Y) > 0$, which implies $V(E(Y \mid X)) = V(Y)$ (see Equation 16). As the conditional expectation $E(Y \mid X)$ may be interpreted to be the best predictor of Y based on X, the coefficient $R(Y \mid X)^2$ may be interpreted to be that portion of the variance of Y that is determined by X.

If $R(Y \mid X)^2$ is the portion of the variance of Y that is determined by X, then $1 - R(Y \mid X)^2$ is that portion that is not determined by X. In fact,

$$1 - R(Y \mid X)^2 = V(F)/V(Y), \tag{17}$$

where $V(F)$ denotes the variance of the residual $F = Y - E(Y \mid X)$. This is easily seen, if we insert Equation 15 into 16:

$$R(Y \mid X)^2 = (V(Y) - V(F))/V(Y) = 1 - V(F)/V(Y). \tag{18}$$

Rearranging this equation then yields Equation 17. Combining 16 and 17 results in

$$\frac{V(E(Y \mid X))}{V(Y)} + \frac{V(F)}{V(Y)} = R(Y \mid X)^2 + \frac{V(F)}{V(Y)} = 1, \tag{19}$$

that is, the portion of the variance that is determined by X and the portion that is not determined by X add up to unity, another property which follows from the definition of the residual F by Equation 7.

6. The Conditional Variance and Covariance

Sometimes one is interested in formulating propositions on how the conditional variance $V(Y \mid X = x)$ of a random variable Y, given that another random variable X takes on a value x, varies along with the random variable X, which is possibly multidimensional and non-numerical. In psychophysical adjustment experiments, for instance, the conditional variance of the judgment increases with size of the given stimulus (see, e.g., Gescheider, 1976, p. 35; Steyer and Erdfelder, 1984). Another proposition often made is that the conditional variance of Y given $X = x$ does *not* depend on X. The conditional variance is an important concept, because it reflects the goodness of the prediction of Y based on X. Similarly, the *covariance* of two random variables Y_1 and Y_2 may be assumed to be dependent or independent on another one, X.

Analogously to the conditional expectation $E(Y \mid X)$ of Y given X, which is a random variable whose values are identical with the conditional expectations $E(Y \mid X = x)$ of Y given $X = x$, the conditional variance $V(Y \mid X)$ of Y given X is a random variable whose values are identical with the conditional variances $V(Y \mid X = x)$ of Y given $X = x$. The corresponding propositions hold, of course, for the conditional covariance of two random variables Y_1 and Y_2.

Let Y_1 and Y_2 be random variables with finite expectations $E(Y_1)$, $E(Y_2)$, and $E(Y_1 \cdot Y_2)$. The *conditional covariance* of Y_1 and Y_2 given $X = (X_1, \ldots, X_m)$ is then defined by

$$C(Y_1, Y_2 \mid X) := E((Y_1 - E(Y_1 \mid X)) \cdot (Y_2 - E(Y_2 \mid X)) \mid X), \tag{20}$$

that is, the conditional covariance $C(Y_1, Y_2 \mid X)$ of Y_1 and Y_2 given X is the conditional expectation of the random variable

$$Z := (Y_1 - E(Y_1 \mid X)) \cdot (Y_2 - E(Y_2 \mid X)) \tag{21}$$

given X. If $F_1 := Y_1 - E(Y_1 \mid X)$ and $F_2 := Y_2 - E(Y_2 \mid X)$, we may equivalently define $C(Y_1, Y_2 \mid X)$ to be the conditional expectation of the product of the residual variables F_1 and F_2. In formula,

$$C(Y_1, Y_2 \mid X) = E(F_1 \cdot F_2 \mid X). \tag{22}$$

Note that $E(F_1 \cdot F_2 \mid X) = C(F_1, F_2 \mid X)$, because $E(F_1 \mid X) = 0 = E(F_2 \mid X)$ (see Equation 9).

The *conditional variance* of a random variable Y given X is defined to be the conditional covariance of Y with itself, provided that Y has finite expectation $E(Y)$ and variance $V(F)$. In formula,

$$V(Y \mid X) := C(Y, Y \mid X), \quad \text{if } E(Y) \text{ and } V(Y) \text{ are both finite.} \tag{23}$$

The following equations can be derived from these definitions:

$$C(Y_1, Y_2 \mid X) = E(Y_1 \cdot Y_2 \mid X) - E(Y_1 \mid X) \cdot E(Y_2 \mid X), \tag{24}$$

$$V(Y \mid X) = E(Y^2 \mid X) - E(Y \mid X)^2. \tag{25}$$

Both equations closely correspond to those well-known properties of unconditional covariances and variances. If Y_1 and Y_2 are dichotomous (e.g., with values zero and one), conditional (local) stochastic independence, which plays an important role in probabilistic models of mental tests (see, e.g., Fischer, 1974), may now be formulated by

$$C(Y_1, Y_2 \mid X) = 0 \quad \text{or} \quad E(Y_1 \cdot Y_2 \mid X) = E(Y_1 \mid X) \cdot E(Y_2 \mid X). \tag{26}$$

7. Some Single Equation Models

The conditional expectation is a unifying concept, which makes it easy to see the communalities of different kinds of models resulting from different kinds of assumptions on $E(Y \mid X)$. The conditional expectation $E(Y \mid X)$ of Y given a numerical random vector $X = (X_1, \ldots, X_m)$ might be specified, for example, by

$$E(Y \mid X) = \alpha_{Y0} + \sum_{j \in J} \alpha_{Yj} \cdot X_j, \quad J = \{1, \ldots, m\}, \tag{27}$$

where the coefficients α_{Y0} and α_{Yj}, $j \in J$, are real constants. If the random variables X_j represent quantitative properties, this is often called a *regression model*, whereas it is also called an *analysis of variance model*, if the variables X_j represent qualitative properties or group membership (see, e.g., Bock, 1975; Moosbrugger and Steyer, 1983; Timm, 1975). In this case, the values of $E(Y \mid X)$ are the conditional expectations of Y in the smallest cells of the designs represented by the variables $X_1, \ldots, X_m$. In a 2×2 factorial design, for instance, $E(Y \mid X)$ does not have more than four different values, the four true cell means.

In *logit-linear models* (see, e.g., Cox, 1970; Haberman, 1978, 1979; Langeheine, 1980), where the dependent variable Y is dichotomous with $Y = 1$ or $Y = 0$, the conditional expectation $E(Y \mid X)$ of Y given $X = (X_1, \ldots, X_m)$ is specified by

$$(28) \qquad E(Y \mid X) = \frac{\exp(\beta_{Y0} + \sum_{j \in J} \beta_{Yj} \cdot X_j)}{1 + \exp(\beta_{Y0} + \sum_{j \in J} \beta_{Yj} \cdot X_j)}, \qquad J = \{1, \ldots, m\},$$

where again β_{Y0} and β_{Yj}, $j \in J$, are real constants. In these cases, in which Y is dichotomous with values zero or one, $E(Y \mid X)$ is also denoted by $P(Y = 1 \mid X)$ and called the *conditional probability* of $Y = 1$ given X. In this case, the values of $E(Y \mid X) = P(Y = 1 \mid X)$ are the conditional probabilities $P(Y = 1 \mid X = x)$ of $Y = 1$ given $X = x$.

8. A Recursive Path Model and an Old Fallacy

Sometimes models consist of *several* equations on conditional expectations such as in recursive *simultaneous equation systems* or *path models*. Consider, for example, the path diagram depicted in Figure 1. Models of this type can be described by specifying one equation for each dependent variable, for instance,

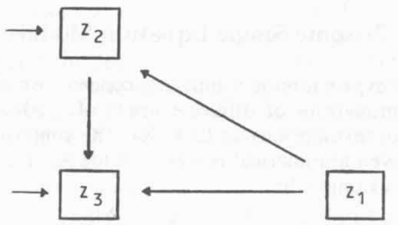

Figure 1: Path diagram for a recursive model with three variables.

$$(29) \qquad E(Z_2 \mid Z_1) = \alpha_{20} + \alpha_{21} \cdot Z_1$$

and

$$(30) \qquad E(Z_3 \mid Z_1, Z_2) = \alpha_{30} + \alpha_{31} \cdot Z_1 + \alpha_{32} \cdot Z_2,$$

where again the coefficients α_{ij} are real constants. All properties of the residuals

$$(31) \qquad F_2 := Z_2 - E(Z_2 \mid Z_1) \quad \text{and} \quad F_3 := Z_3 - E(Z_3 \mid Z_1, Z_2)$$

appearing as additional assumptions in the usual formulation of path analysis models, namely

$$(32) \qquad C(F_2, Z_1) = 0,$$

$$(33) \qquad C(F_3, Z_1) = C(F_3, Z_2) = 0,$$

$$(34) \qquad C(F_2, F_3) = 0,$$

$$(35) \qquad E(F_2) = E(F_3) = 0,$$

(see, e.g., Brandtstädter and Bernitzke, 1976; Kenny, 1979; or Hummell and Ziegler, 1976) and the additional properties

$$(36) \qquad E(F_3 \mid F_2) = 0,$$

$$(37) \qquad E(F_2 \mid Z_1) = E(F_3 \mid Z_1) = E(F_3 \mid Z_2) = E(F_3 \mid Z_1, Z_2) = 0,$$

can be *derived* from the definitions of F_2 and F_3 using the rules of computation for conditional expectations given in the Appendix A. Using the concept of the conditional expectation helps to distinguish between assumptions and logical consequences from definitions. Thus, Equations 32 to 37 being logical consequences of the assumptions formulated in Equations 29 to 31, makes immediately clear that the uncorrelatedness of residuals and the corresponding regressors can *not* be used to distinguish a usual regression model from a causal one, a fallacy still appearing in the literature (see, e.g., Hodapp, 1984, p. 26, or Kenny, 1979, p. 51) that goes back to a paper of Simon (1954, 1971). The fallacy is due to Simons misinterpretation of the uncorrelatedness of errors. Whereas his interpretation is that "all other" variables influencing Z_3 are uncorrelated with all other variables influencing Z_2 (see 1971, p. 10), the analysis above shows that the uncorrelatedness of the error variables F_2 and F_3 is a logical consequence of their definition and the equation 29 and 30. Hence, Simons conclusion that "correlation is proof of causation in the two-variable case if we are willing to make the assumption of time precedence and non-correlation of the error terms"

(1971, p. 10) is wrong. Non-correlation of errors does *not* imply that there is no omitted other variable, say W, influencing Z_1, Z_2, and Z_3.

The proof of this proposition is as follows: Since Equation 29 and 30 imply the uncorrelatedness of errors, it must only be shown that these two equations do *not* imply that the coefficients β_{iW}, $i = 1, 2, 3$ (representing the influence of the omitted variable W) in the following equations are zero:

$$E(Z_1 \mid W) = \beta_{10} + \beta_{1W} \cdot W, \tag{38}$$

$$E(Z_2 \mid Z_1, W) = \beta_{20} + \beta_{21} \cdot Z_1 + \beta_{2W} \cdot W, \tag{39}$$

$$E(Z_3 \mid Z_1, Z_2, W) = \beta_{30} + \beta_{31} \cdot Z_1 + \beta_{32} \cdot Z_2 + \beta_{3W} \cdot W. \tag{40}$$

In fact, using the rules of computation given in the Appendix A, it is easy to see that the Equations 38 to 40 with $\beta_{iW} \neq 0$, $i = 1, 2, 3$, even imply the Equation 29 and 30 (and therefore also the uncorrelatedness of the residuals F_2 and F_3), provided that one additionally assumes:

$$E(W \mid Z_1) = \alpha_{W0} + \alpha_{W1} \cdot Z_1, \tag{41}$$

$$E(W \mid Z_1, Z_2) = \beta_{W0} + \beta_{W1} \cdot Z_1 + \beta_{W2} \cdot Z_2. \tag{42}$$

A more adequate theory of causal regression models does not focus on the non-correlation of the residuals but on the invariance of the regression coefficients when passing from $E(Z_2 \mid Z_1)$ to $E(Z_2 \mid Z_1, W)$ and from $E(Z_3 \mid Z_1, Z_2)$ to $E(Z_3 \mid Z_1, Z_2, W)$ (see, e.g., Steyer, 1983, 1984, 1985 a, b, c, d).

9. Error Theory of Situation-Dependent Psychological Properties

Although there are a number of applications of the concept of conditional expectations in the psychological and methodological literature (see Section 10), a more detailed application might be in place here to further demonstrate the usefulness of the concept. In contrast to the application of the conditional expectation in path models, we now treat an example with nonnumerical regressors. We present and discuss assumptions which allow to compute a coefficient representing the "sensitivity" of a test-score variable with respect to a given set of situations. This coefficient has been proposed by Tack (1980, 1986).

The procedure and the assumptions refer to the following type of experiment: A unit is sampled and then randomly assigned to one of several given situations in which its values on two essentially equivalent variables (see Assumption (d) below) are assessed. After that another situation is drawn from the given set and the value of the unit on a third variable is observed that is also assumed to be essentially equivalent to the first two variables. The most remarkable result is that no assumption of uncorrelated errors is needed to determine the "sensitivity coefficient".

The classical error (test) theory in its modern version presented by Zimmerman (1975, 1976) considers the following type of experiments: A unit u (e.g., a person or person-situation combination) is sampled from a set U of observational units and the values of u with respect to m properties (which, in the simplest cases, might be dichotomous, such as solving or not solving a task) are observed. Hence, the set of possible outcomes of this kind of experiment is:

$$\Omega = U \times \Omega_1 \times \ldots \times \Omega_m, \quad m \in \mathbb{N}, \tag{43}$$

where each Ω_i, $i = 1, \ldots, m$, is the set of possible values with respect to the i-th property. In the usual simple cases, all the factor sets of Ω are finite. Hence we can chose the power set $\mathscr{P}(\Omega)$ to be the sigmaalgebra on which a (usually unknown) probability measure P is assumed to exist. We might just know that each unit $u \in U$ has the same chance to be sampled.

We consider the stochastic variables Y_i: $\Omega \rightarrow \mathbb{R}$, $i = 1, \ldots, m$ (the test-score variables) on this probability space that map the qualitative values of the m properties considered into the set of real numbers. Furthermore, we consider the projection $p_U: \Omega \rightarrow U$, the values of which are the units $u \in U$, that is,

$$p_U(\omega) = p_U((u, \omega_1, \ldots, \omega_m)) = u, \quad \text{for all } \omega \in \Omega. \tag{44}$$

We now define T_i to be the conditional expections or regressions

$$T_i := E(Y_i \mid p_U), \quad i = 1, \ldots, m. \tag{45}$$

If the units are persons, we may call the variables T_i person-regressions (or "true-score variables") of the variables Y_i. The values of such a person-regression T_i are the conditional expectations (or true scores) $E(Y_i \mid p_U = u)$ of Y_i given a person u. The general properties of the residual discussed in Section 3 may now be written: $Y_i = T_i + F_i$ (Eq. 8), $E(F_i \mid f(p_U)) = 0$ (Eq. 9), $E(F_i) = 0$ (Eq. 10), $E(F_i \mid T_i) = 0$ (Eq. 13), $C(F_i, T_i) = 0$ (Eq. 11), $V(Y_i) = V(T_i) + V(F_i)$ (Eq. 15). Note that none of these Equations is an axiom. Instead, each of it is an immediate consequence of Definition 45. Also note that the equation $E(F_i \mid f(p_U)) = 0$ also implies $E(F_i \mid T_j) = 0$ and $C(F_i, T_j) = 0$. Another interesting point is that the uncorrelatedness of "errors"

$$C(F_i, F_j) = 0, \quad i \neq j, \; i, j = 1, \ldots, m, \tag{46}$$

is *not* a consequence of 45 (see, e.g., Tack, 1980). Furthermore, the coefficient of determination (Eq. 12) may be written $V(T_i)/V(Y_i)$, which is called "reliability coefficient" in this context. Hence its definition does not depend on the "axiom" of uncorrelated errors.

Tack (1980) generalized this classical error (test) theory model introducing a set S of situations out of which one is observed. Hence, instead of 43, we now underly the set

$$\Omega = O \times S \times \Omega_1 \times \ldots \times \Omega_m, \tag{47}$$

of elementary events or possible outcomes of the experiment. We may interpret Ω in two different ways. *First*, we assume $U = O$. In this case Equation 47 characterizes another experiment than 43. Whereas in 43 the situation in which the properties of the person are observed is constant (and therefore not mentioned), it may now be different every time the experiment is run. *Second*, if on the other hand $U = O \times S$, then 43 and 47 may characterize the same type of experiment. The only difference is that in 47 we take account of the fact that another situation may be realized each time a person is sampled and his or her properties are observed. Both kinds of interpretation are possible. The first one leads to a generalization of the classical model, the second to its reinterpretation.

In any case we may now consider the projection $p_s : \Omega \to S$, the value $p_s(\omega)$ of which is the situation $s \in S$ drawn or observed. Furthermore, aside from the *person-regression* $E(Y_i \mid p_0)$ (the classical "true-score variable"), there is the *situation-regression* $E(Y_i \mid p_s)$ and the *person-situation-regression* $E(Y_i \mid p_0, p_s)$.

Tack (1980) has shown that the trivial equation

$$E(Y_i \mid p_0, p_s) = E(Y_i \mid p_0) + E(Y_i \mid p_0, p_s) - E(Y_i \mid p_0) \tag{48}$$

about the *person-situation-regression* $E(Y_i \mid p_0, p_s)$ leads – without additional assumptions – to the following decomposition of variances:

$$V(E(Y_i \mid p_0, p_s)) = V(E(Y_i \mid p_0)) + V(E(Y_i \mid p_0, p_s) - E(Y_i \mid p_0)). \tag{49}$$

The last variance may be interpreted as that part of the variance of Y_i that is not due to the persons but still systematic with respect to persons and situations. It contains the variance determined by the situations and the interaction of persons and situations (see Steyer, 1986, for details).

In the following sections we treat conditions under which the variance of $E(Y_i \mid p_0, p_s) - E(Y_i \mid p_0)$ (see Eq. 49) or the coefficient

$$V(E(Y_i \mid p_0, p_s) - E(Y_i \mid p_0))/V(Y_i) \tag{50}$$

can be determined from the variances and covariances of the variables Y_i.

Although the definition of the coefficient 50 does not depend on any assumptions, it cannot be determined from "observable" quantities (such as the variances and covariances of the Y_i) without any assumptions. This is the same state of affairs as with $V(E(Y_i \mid p_0))/V(Y_i)$, the classical coefficient of reliability, the determination of which rests on assumptions of parallel or essentially equivalent tests.

We now propose a procedure to determine the coefficient 50. For this purpose we have to enlarge the experiment discussed above: A person o is sampled from a set O and randomly assigned to a *first* situation s_1 which is drawn from a set S_1 of several given situations. In this situation his or her values on two (essentially equivalent) variables Y_1 and Y_2 are assessed. After that a *second* situation s_2 is randomly drawn from the set of situations and his or her value on a third variable Y_3 is observed (that is also assumed to be essentially equivalent to the first two variables with respect to the person-regression; see Assumption (d) below). Hence the set of possible outcomes or elementary events is now assumed to be

$$\Omega = O \times S_1 \times \Omega_1 \times \Omega_2 \times S_2 \times \Omega_3, \tag{51}$$

where $S_1 = S_2$ and $\Omega_1 = \Omega_2 = \Omega_3$. Aside from the above mentioned projections $p_0 : \Omega \to O$ and $p_1 : \Omega \to S_1$ (denoted p_s above), we may now additionally consider the projection $p_2 : \Omega \to S_2$.

Assuming the random assignment of persons to situations we have:

(a) the projections p_0, p_1 and p_2 are totally stochastically independent.

Possible dependencies in a sample do not affect the validity of this assumption. It can only be wrong if the person sampled is not randomly assigned to the two situations or if the two situations are not sampled independently.

Two additional assumptions are:

(b) $E(Y_i \mid p_0, p_1) = E(Y_i \mid p_0, p_1, p_2), \quad i = 1, 2,$

(c) $E(Y_3 \mid p_0, p_2) = E(Y_3 \mid p_0, p_1, p_2)$.

Whereas (b) will always be fulfilled in the type of random experiment described above, the Assumption (c) may be wrong in some applications. For instance differential learning effects caused by the first situations s_1 might influence the value of Y_3 assessed in situation s_2.

The next assumption is

(d) $E(Y_1 \mid p_0) = E(Y_2 \mid p_0) + \alpha = E(Y_3 \mid p_0) + \beta, \quad \alpha, \beta \in \mathbb{R},$

which means that Y_1, Y_2, and Y_3 are *essentially equivalent* with respect to their person-regressions. This assumption is well known from classical error (test) theory.

Furthermore, we assume

(e) $E(Y_1 \mid p_0, p_1) = E(Y_2 \mid p_0, p_1) + \gamma, \quad \gamma \in \mathbb{R},$

which means that Y_1 and Y_2 are also essentially equivalent with respect to their person-situation-regression. Whether or not assumptions d and e hold in a given application is not a logical but an empirical question.

Finally, we assume that the covariances of the residuals

$$\begin{aligned} G_i &:= Y_i - E(Y_i \mid p_0, p_1), \quad i = 1, 2, \\ G_3 &:= Y_3 - E(Y_3 \mid p_0, p_2) \end{aligned} \tag{52}$$

are equal:

(f) $C(G_1, G_2) = C(G_1, G_3) = C(G_2, G_3).$

In classical error theory the residuals are usually assumed to be uncorrelated. Hence, in this important respect (see the discussion of Tack, 1980) our assumptions are less restrictive than the classical ones.

The Assumptions (a) to (f) imply the following equations (see Steyer, 1986, for proofs):

$$\begin{aligned} C(Y_1, Y_2) = {} & C(E(Y_1 \mid p_0), E(Y_2 \mid p_0)) \\ & + C(E(Y_1 \mid p_0, p_1) - E(Y_1 \mid p_0), E(Y_2 \mid p_0, p_1) - E(Y_2 \mid p_0)) \\ & + C(G_1, G_2) \end{aligned} \tag{53}$$

$$C(Y_i, Y_3) = C(E(Y_i \mid p_0), E(Y_3 \mid p_0)) + C(G_i, G_3), \quad i = 1, 2, \tag{54}$$

$$\begin{aligned} & C(Y_1, Y_2) - C(Y_i, Y_3) \\ & = C(E(Y_1 \mid p_0, p_1) - E(Y_1 \mid p_0), E(Y_2 \mid p_0, p_1) - E(Y_2 \mid p_0)), \\ & \qquad i = 1, 2 \end{aligned} \tag{55}$$

$$C(Y_1, Y_2) - C(Y_i, Y_3) = V(E(Y_i \mid p_0, p_1) - E(Y_i \mid p_0)), \quad i = 1, 2. \tag{56}$$

Dividing this difference by the variance of Y_i yields the desired coefficient 50:

$$\frac{C(Y_1, Y_2) - C(Y_i, Y_3)}{V(Y_i)} = \frac{V(E(Y_i \mid p_0, p_1) - E(Y_i \mid p_0))}{V(Y_i)}, \quad i = 1, 2. \tag{57}$$

Adding the final assumption of equality of variances:

(g) $V(Y_1) = V(Y_2) = V(Y_3),$

leads to

$$R(Y_1, Y_2) - R(Y_i, Y_3) = \frac{V(E(Y_i \mid p_0, p_1) - E(Y_i \mid p_0))}{V(Y_i)}, \quad i = 1, 2. \tag{58}$$

Hence in this case the coefficient 50 can be determined by the difference of the correlations $R(Y_1, Y_2)$ and $R(Y_1, Y_3)$ or the difference of $R(Y_1, Y_2)$ and $R(Y_2, Y_3)$.

To *summarize*, contrary to the opinion of Pawlik (1976), we do not need a "theory of parallel situations" in order to determine the variance due to situations and the person-situation interaction of the psychological property considered. What we need are assumptions of essential equivalence of the person-regressions and the person-situation-regressions or true score variables (see Assumptions (d) and (e)), and the Assumption (a) that can be realized through the design of the experiment. Even the classical (and highly problematic) assumption of uncorrelated errors can be replaced by the much weaker assumption of equal error covariances.

10. Conclusion

Correlation and regression are often exclusively treated as data-analytic methods, neglecting the fact that these concepts are very useful to formulate theoretical propositions on nondeterministic dependencies between random variables. The theory of conditional expectations may be viewed as the general theory of regression. The conditional expectation (regression) is defined without reference to a certain regression equation. Propositions formulated in terms of a conditional expectation do not necessarily involve an equation at all, for instance, if we say that $E(Y \mid X)$ increases with X. Furthermore, as illustrated in the application above, the regressor X may also be nonnumerical.

In this paper, two applications of the conditional expectation have been treated in some detail. In the first one, it has been used to unravel an old fallacy concerning uncorrelated errors and the concept of causality in recursive path models. In the second, it has been applied to derive a procedure to determine a coefficient for the "sensitivity" of a test-score variable with respect to a given set of situations.

It has often been argued that formalization is inadequate in sciences like psychology. In fact, the assumptions about distributions and equal variances, for example, that are often made in analyzing sample data, are rather restrictive. Using the concept of a conditional expectation $E(Y \mid X)$, however, any proposition can be formulated on how the conditional expectations $E(Y \mid X = x)$ of a random variable Y given $X = x$ vary along with X without any further assumptions. The only presupposition is that the expectation $E(Y)$ is finite, which is (almost) no restriction for application in the empirical sciences.

Although no assumptions about $E(Y \mid X)$ or about the distributions of Y, X, or the residual F need be made, many properties of F (see

Equations 7 to 15) can be derived from its definition $F := Y - E(Y \mid X)$ using general rules of computation given in the Appendix A. These rules in turn are derived from Kolmogoroff's (1933) definition of the conditional expectation (see any book on mathematical probability theory such as Bauer, 1974; Breiman, 1968; Loève, 1977, 1978).

The concept of a conditional expectation has been applied in a number of papers, for example, on the theory of mental tests (Steyer, 1986; Tack, 1980, 1986; Zimmerman, 1975, 1976; Zimmerman and Williams, 1977), on causal models (Steyer, 1983, 1984, 1985 a, b, c, d), on the communalities of some stochastic models (Moosbrugger, 1983), to psychophysical models (Erdfelder and Steyer, 1984), and models for perceptual illusions (Steyer and Erdfelder, 1984). The papers by Zimmerman and Zimmerman and Williams were pioneering, showing among other things that the axioms of the classical theory of mental tests can in fact be deduced from the definition of the true scrore variable as a special conditional expectation. Some of the other papers mentioned above, which were inspired by Zimmerman's, showed further that such simplifications can also be obtained in other areas.

Appendix A:
Rules of Computation for Conditional Expectations

In this appendix some general rules of computation for conditional expectations are gathered. Their use is illustrated in Appendix B. We assume throughout that Y is a numerical random variable on an underlying probability space $(\Omega, \mathscr{A}, P)$ with finite expectation $E(Y)$ and that X is a random variable on $(\Omega, \mathscr{A}, P)$ with values in an arbitrary set. Hence, X may be multidimensional and/or nonnumerical. Proofs for Rules 1 to 5 may be found in Breiman (1968) or Bauer (1974), for example. A proof for Rule 6 is given in Steyer (1983, p. 144). Note that all equations on conditional expectations only hold with probability 1 (or P-almost surely).

Rule 1. The expectation of a conditional expectation (regression) $E(Y \mid X)$ is equal to the expectation of Y. In formula:

$$E(E(Y \mid X)) = E(Y).$$

Rule 2. (i) The conditional expectation of a weighted sum $\alpha \cdot Y + \beta \cdot Z$ of two random variables Y and Z with finite expectations (the weighted sum or linear combination is also a random variable with finite expectation) given X is equal to the weighted sum of the conditional expectations of Y and Z given X.

(ii) The conditional expectation of a real constant β given X is equal to the constant β itself. In formula:

(i) $E(\alpha \cdot Y + \beta \cdot Z \mid X) = \alpha \cdot E(Y \mid X) + \beta \cdot E(Z \mid X)$;

(ii) $E(\beta \mid X) = \beta, \quad \alpha, \beta \in \mathbb{R}$.

Rule 3. The conditional expectation of the regression $E(Y \mid X)$ given X is equal to the conditional expectation of Y given X. In formula:

$$E(E(Y \mid X) \mid X) = E(Y \mid X).$$

Rule 4. The conditional expectation of the regression $E(Y \mid X)$ given a family $(f_k(X),\ k \in K)$ of measurable functions of X is equal to the conditional expectation of Y given $(f_k(X),\ k \in K)$:

$$E(E(Y \mid X) \mid f_k(X), k \in K) = E(Y \mid f_k(X), k \in K).$$

Special cases of this rule are, for example:

(i) $E(E(Y \mid X) \mid X) = E(Y \mid X)$ (this is Rule 3),

(ii) $E(E(Y \mid X_1, X_2) \mid X_1) = E(Y \mid X_1)$,

(iii) $E(E(Y \mid X) \mid E(Y \mid X)) = E(Y \mid E(Y \mid X))$,

(iv) $E(E(Y \mid X_1, X_2, X_3) \mid X_1, X_2 + X_3) = E(Y \mid X_1, X_2 + X_3)$.

Rule 5. If $f(X)$ is a numerical measurable function of the m-dimensional random variable $X = (X_1, \ldots, X_m)$ such that $E(Y \cdot f(X))$ is finite, then the conditional expectation of the product of Y with $f(X)$ given X is equal to the product of $f(X)$ and $E(Y \mid X)$. In formula:

$$E(f(X) \cdot Y \mid X) = f(X) \cdot E(Y \mid X).$$

If we define $J = \{1, \ldots, m\}$, presuppose that the X_j, $j \in J$, are numerical and that Z is a random variable with finite expectation $E(Z)$, special cases of this general rule are:

(i) $E(X_j \mid X) = X_j$, if $j \in J$ and $E(X_j)$ is finite.

(ii) $E(Y \cdot X_j \mid X) = X_j \cdot E(Y \mid X)$,
if $j \in J$ and $E(Y \cdot X_j)$ is finite.

(iii) $E(X_j \cdot X_k \mid X) = X_j \cdot X_k$,
if $j, k \in J$ and $E(X_j \cdot X_k)$ is finite.

(iv) $E(Y \cdot E(Z \mid X) \mid X) = E(Z \mid X) \cdot E(Y \mid X)$.

(v) $E(E(Y \mid X) \mid E(Y \mid X)) = E(Y \mid X)$.

(vi) $E(E(Z \mid X_j, j \in J_0) \cdot Y \mid X) = E(Z \mid X_j, j \in J_0) \cdot E(Y \mid X)$,
if $J_0 \subset J$.

Rule 6. If Y is regressively independent from the numerical random vector $X=(X_1,\dots,X_m)$ then the covariance and the correlation of Y and each X_j, $j \in J=\{1,\dots,m\}$, is zero provided that the variances of Y and each X_j, $j \in J$ are finite. In formula:

If
$$V(Y)<\infty \quad \text{and} \quad V(X_j)<\infty, \quad j\in J,$$
and
$$E(Y \mid X_1,\dots,X_m)=E(Y),$$
then
$$C(Y,X_j)=0, \quad \text{for all } j\in J.$$

Appendix B: Proofs

The use of the rules of computation (referred to as R. 1 to R. 6) listed above is now illustrated to derive:

(a) the proposition that $E(Y \mid X)$ minimizes the following function of the measurable function $f(X)$: $E((Y-f(X))^2)$ if Y has finite variance;
(b) the properties of the residual discussed in Section 3;
(c) the formulae for the residuals of the recursive path model in Section 8.

We begin with point (a) which refers to *Section 3.* For simplicity define $X_0 := E(Y \mid X)$ and let $Z := f(X)$ be an arbitrary random variable measurable with respect to the sigma-algebra $\mathscr{A}(X)$ generated by X. First consider

$$\begin{aligned} E((Y-X_0)\cdot(X_0-Z) \mid X) &= (X_0-Z)\cdot E(Y-X_0 \mid X) && \text{(R. 5)}\\ &= 0 && \text{(Eq. 9)} \end{aligned}$$

Because of R. 1, this implies: $E((Y-X_0)\cdot(X_0-Z))=0$. Hence,

$$\begin{aligned} f(Z) &= E((Y-Z)^2) = E(((Y-X_0)+(X_0-Z))^2)\\ &= E((Y-X_0)^2)+E((X_0-Z)^2)+2\cdot E((Y-X_0)\cdot(X_0-Z))\\ &= E((Y-X_0)^2)+E((X_0-Z)^2). \end{aligned}$$

Because $E((Y-X_0)^2)$ is a constant, this proves that $f(Z)$ has its minimum for $Z=X_0=E(Y \mid X)$, which was to be shown.

Equation 9:

$$\begin{aligned} E(F \mid f(X)) &= E(Y-E(Y \mid X) \mid f(X)) && \text{(Eq. 7)}\\ &= E(Y \mid f(X)) - E(E(Y \mid X) \mid f(X)) && \text{(R. 2)}\\ &= E(Y \mid f(X)) - E(Y \mid f(X)). && \text{(R. 4)} \end{aligned}$$

Equation 10:

$$\begin{aligned} E(F) &= E(Y-E(Y \mid X)) && \text{(Eq. 7)}\\ &= E(Y)-E(E(Y \mid X))\\ &= E(Y)-E(Y). && \text{(R. 1)} \end{aligned}$$

Equation 11 follows from Equations 9, 10, and Rule 6. Equation 13, is a special case of 9, if one considers Equation 10. Equation 14 follows from Equation 13 and Rule 6.

Equation 15:

$$\begin{aligned} V(Y) &= C(Y,Y) = C(E(Y \mid X)+F, E(Y \mid X)+F)\\ &= V(E(Y \mid X))+V(F). && \text{(Eq. 14)} \end{aligned}$$

We now turn to *Section 8.* The properties of the residuals F_2 and F_3 defined by Equation 31 will now be derived beginning with Equation 37 and then going back to Equation 32.

$$\begin{aligned} E(F_3 \mid Z_1,Z_2) &= E(Z_3-E(Z_3 \mid Z_1,Z_2) \mid Z_1,Z_2) && \text{(Eq. 31)}\\ &= E(Z_3 \mid Z_1,Z_2) - E(E(Z_3 \mid Z_1,Z_2) \mid Z_1,Z_2) && \text{(R. 2)}\\ &= E(Z_3 \mid Z_1,Z_2) - E(Z_3 \mid Z_1,Z_2). && \text{(R. 3)} \end{aligned}$$

$$\begin{aligned} E(F_2 \mid Z_1) &= E(Z_2-E(Z_2 \mid Z_1) \mid Z_1) && \text{(Eq. 31)}\\ &= E(Z_2 \mid Z_1) - E(E(Z_2 \mid Z_1) \mid Z_1) && \text{(R. 2)}\\ &= E(Z_2 \mid Z_1) - E(Z_2 \mid Z_1). && \text{(R. 3)} \end{aligned}$$

Equation 36:

$$\begin{aligned} E(F_3 \mid F_2) &= E(F_3 \mid Z_2-E(Z_2 \mid Z_1)) && \text{(Eq. 31)}\\ &= 0, && \text{(Eq. 9)} \end{aligned}$$

because F_2 is a function of the regressors Z_1 and Z_2 and because F_3 is the residual with respect to the regressions $E(Z_3 \mid Z_1, Z_2)$. Except for Equation 34, the other equations are the general properties of the residual discussed in Section 3. Equation 34 immediately follows from Equation 36 and Rule 6. Hence it is shown that Equations 29 to 31 imply Equation 34.

We now show that Equations 38 to 42 imply Equations 29 and 30: This proves that F_2 and F_3 being uncorrelated does *not* imply that the coefficients β_{iW}, $i=1,2,3$, of the Equations 38 to 40 are zero. Although F_2 and F_3 are uncorrelated (see Eq. 34), an omitted variable W can exist "which influences" all variables (Z_1, Z_2, Z_3) in the model.

Equation 29:

$$\begin{aligned} E(Z_2 \mid Z_1) &= E(E(Z_2 \mid Z_1, W) \mid Z_1) && \text{(R. 4)} \\ &= E(\beta_{20} + \beta_{21} \cdot Z_1 + \beta_{2W} \cdot W \mid Z_1) && \text{(Eq. 39)} \\ &= \beta_{20} + \beta_{21} \cdot Z_1 + \beta_{2W} \cdot E(W \mid Z_1) && \text{(R. 2, R. 3)} \\ &= \beta_{20} + \beta_{21} \cdot Z_1 + \beta_{2W} \cdot (\alpha_{W0} + \alpha_{W1} \cdot Z_1) && \text{(Eq. 41)} \\ &= (\beta_{20} + \beta_{2W} \cdot \alpha_{W0}) + (\beta_{21} + \beta_{2W} \cdot \alpha_{W1}) \cdot Z_1, \end{aligned}$$

which is Equation 29 with

$$\alpha_{20} := \beta_{20} + \beta_{2W} \cdot \alpha_{W0} \quad \text{and} \quad \alpha_{21} := \beta_{21} + \beta_{2W} \cdot \alpha_{W1}.$$

Equation 30 can be derived analogously:

$$\begin{aligned} E(Z_3 \mid Z_1, Z_2) &= E(E(Z_3 \mid Z_1, Z_2, W) \mid Z_1, Z_2) && \text{(R. 4)} \\ &= E(\beta_{30} + \beta_{31} \cdot Z_1 + \beta_{32} \cdot Z_2 + \beta_{3W} \cdot W \mid Z_1, Z_2) && \text{(Eq. 40)} \\ &= \beta_{30} + \beta_{31} \cdot Z_1 + \beta_{32} \cdot Z_2 + \beta_{3W} \cdot E(W \mid Z_1, Z_2) && \text{(R. 2, R. 3)} \\ &= \beta_{30} + \beta_{31} \cdot Z_1 + \beta_{32} \cdot Z_2 \\ &\quad + \beta_{3W} \cdot (\beta_{W0} + \beta_{W1} \cdot Z_1 + \beta_{W2} \cdot Z_2) && \text{(Eq. 42)} \\ &= (\beta_{30} + \beta_{3W} \cdot \beta_{W0}) + (\beta_{31} + \beta_{3W} \cdot \beta_{W1}) \cdot Z_1 \\ &\quad + (\beta_{32} + \beta_{3W} \cdot \beta_{W2}) \cdot Z_2, \end{aligned}$$

which is Equation 30 with

$$\alpha_{30} := \beta_{30} + \beta_{3W} \cdot \beta_{W0}, \quad \alpha_{31} := \beta_{31} + \beta_{3W} \cdot \beta_{W1},$$

and

$$\alpha_{32} := \beta_{32} + \beta_{3W} \cdot \beta_{W2},$$

which was to be shown.

References

Adrain, R. (1818) Investigation of the figure of the earth, and of the gravity in different latitudes. *Transactions of the American Philosophical Society, 1,* 119–135.

Bauer, H. (1974) *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie.* Berlin: de Gruyter.

Bellach, J., Franken, P., Warmuth, E., Warmuth, W. (1978) *Maß, Integral und bedingter Erwartungswert.* Berlin: Akademie-Verlag.

Bock, R. D. (1975) *Multivariate statistical methods in behavioral research.* New York: McGraw-Hill.

Brandtstädter, J., Bernitzke, F. (1976) Zur Technik der Pfadanalyse. Ein Beitrag zum Problem der nichtexperimentellen Konstruktion von Kausalmodellen. *Psychologische Beiträge, 18,* 12–34.

Bravais, A. (1846) Analyse mathématique sur les probabilités des erreurs de situation d'un point. *Mémoires presentés par divers savants à l'Académie Royale des Sciences de l'Institut de France, 9,* 255–332.

Breiman, L. (1968) *Probability.* Reading, Mass.: Addison-Wesley.

Cohen, J. (1968) Multiple regression as a general data-analytic system. *Psychological Bulletin, 70,* 426–443.

Cox, D. R. (1970) The analysis of binary data. London: Methuen.

Erdfelder, E., Steyer, R. (1984) Zur Psychophysik einiger Größentäuschungen. *Psychologische Beiträge, 26,* 640–646.

Fechner, G. T. (1860) *Elemente der Psychophysik.* Leipzig: Breitkopf und Härtel.

Fischer, G. H. (1974) *Einführung in die Theorie psychologischer Tests.* Bern: Huber.

Fisher, R. A. (1925) *Statistical methods for research workers.* London: Oliver & Boyd.

Galton, F. (1877) Typical laws of heredity. *Nature, 15,* 492–495, 512–514, 532–533.

Galton, F. (1889) *Natural inheritance.* London: McMillan.

Gescheider, G. A. (1976) *Psychophysics. Method and Theory.* Hillsdale, N.J.: Erlbaum.

Haberman, S. J. (1978) *The analysis of qualitative data I.* New York: Academic Press.

Haberman, S. J. (1979) *The analysis of qualitative data II.* New York: Academic Press.

Hodapp, V. (1984) *Analyse linearer Kausalmodelle.* Bern: Huber.

Hummell, H. H., Ziegler, R. (1976) Zur Verwendung linearer Modelle bei der Kausalanalyse nicht-experimenteller Daten. In Hummell, H. J., Ziegler, R. (Eds.) *Korrelation und Kausalität.* Band 1, Stuttgart: Enke.

Jöreskog, K. G., Sörbom, D. (1981) *LISREL V. Analysis of linear structural equation models by maximum likelihood and least squares methods.* Uppsala: University of Uppsala, Department of Statistics.

Kenny, D. A. (1979) *Correlation and Causality.* New York: Wiley.

Kolmogoroff, A. N. (1933) *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Berlin: Springer.

Langeheine, R. (1980) *Log-lineare Modelle zur multivariaten Analyse qualitativer Daten. Eine Einführung.* München: Oldenbourg.

Littrow, J. J. (1818) Über die gerade Aufsteigung der vornehmsten Fixsterne. *Astronomie, 6,* 1–26.

Littrow, J. J. (1833) *Die Wahrscheinlichkeitsrechnung in ihrer Anwendung auf das wissenschaftliche und praktische Leben.* Wien: Becks Universitäts-Buchhandlung.

Loève, M. (1977) *Probability theory I.* Berlin: Springer.

Loève, M. (1978) *Probability theory II.* Berlin: Springer.

Mendel, G. (1866) Versuche über Pflanzen-Hybriden. *Verhandlungen des naturforschenden Vereins in Brünn, 4,* 3–47.

Moosbrugger, H. (1983) Modelle zur Beschreibung statistischer Zusammenhänge in der psychologischen Forschung. In Bredenkamp J., Feger, H. (Eds.) *Enzyklopädie der Psychologie.* Serie Forschungsmethoden der Psychologie, Band 4. Göttingen: Hogrefe.

Moosbrugger, H., Steyer, R. (1983) Uni- und multivariate Varianzanalyse mit festen Parametern. In Bredenkamp J., Feger, H. (Eds.) *Enzyklopädie der Psychologie.* Serie Forschungsmethoden der Psychologie, Band 4. Göttingen: Hogrefe.

Pawlik, K. (1976) Modell- und Praxisdimensionen psychologischer Diagnostik. In Pawlik, K. (Ed.) *Diagnose der Diagnostik: Beiträge zur Diskussion der psychologischen Diagnostik in der Verhaltensmodifikation.* Stuttgart: Klett.

Pearson, K. (1886) Mathematical contributions to the theory of evolution. – III. Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London,* Series A, *187,* 253–318.

Pearson, K. (1901) On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* (6th Series), *2,* 559–572.

Simon, H. A. (1954, 1971) Spurious correlation: A causal interpretation. *Journal of the American Statistical Association, 49,* 467–479. (Reprinted in Blalock, H. M., Jr. (1971). (Ed.), *Causal models in the social sciences.* London: Macmillan.)

Steyer, R. (1983) Modelle zur kausalen Erklärung statistischer Zusammenhänge. In Bredenkamp, J., Feger, H (Eds.) *Enzyklopädie der Psychologie.* Serie Forschungsmethoden der Psychologie, Band 4. Göttingen: Hogrefe.

Steyer, R. (1984) Causal linear stochastic dependencies: The formal theory. In Degreef, E., van Buggenhaut, J. (Eds.) *Trends in Mathematical Psychology.* Amsterdam: North-Holland.

Steyer, R. (1985 a) Causal regressive dependencies: An introduction. In Nesselroade, J. R., von Eye, A. (Eds.) *Individual development and social change: Explanatory analysis.* New York: Academic Press, 1985.

Steyer, R. (1985 b) The randomized experiment and the concept of weak causal regressive dependence. *Trierer Psychologische Berichte, 12,* Heft 5.

Steyer, R. (1985 c) The conditionally randomized experiment and the concept of conditional weak causal regressive dependence. *Trierer Psychologische Berichte, 12,* Heft 9.

Steyer, R. (1985 d) The theory of causal regressive dependencies: An overview. *Trierer Psychologische Berichte, 12,* Heft 11.

Steyer, R. (1986) Zur Fehlertheorie situationsabhängiger Merkmale: Ein Verfahren zur Bestimmung eines Kennwerts der Situationssensitivität. *Trierer Psychologische Berichte, 13,* Heft 9.

Steyer, R., Erdfelder, E (1984) Zur stochastischen Formulierung zweier psychophysikalischer Gesetze. *Psychologische Beiträge, 26,* 659–665.

Tack, W. H. (1980) Zur Theorie psychometrischer Verfahren: Formalisierung und Erfassung von Situationsabhängigkeit und Veränderung. *Zeitschrift für Differentielle und Diagnostische Psychologie, 1,* 87–106.

Tack, W. H. (1986) Reliabilitäts- und Effektfunktionen. – Ein Ansatz zur Zuverlässigkeit von Meßwertänderungen. *Diagnostica, 32,* 48–63.

Thomas, H. (1983) Parameter estimation in simple psychophysical models. *Psychological Bulletin, 93,* 396–403.

Timm, N. H. (1975) *Multivariate analysis with applications in education and psychology.* Monterey: Brooks/Cole.

Yule, G. U. (1897) On the theory of correlation. *Journal of the Royal Statistical Society, 60,* 812–854.

Yule, G. U. (1900) On the association of attributes in statistics: with illustrations from the material of the Childhood Society, & c. *Philosophical Transactions of the Royal Society of London* (Series A), *194,* 257–319.

Yule, G. U. (1907) On the theory of correlation or any number of variables treated by a new system of notation. *Proceedings of the Royal Statistical Society* (Series A), *79,* 182–193.

Yule, G. U. (1912) On the methods of measuring associations between two attributes. *Journal of the Royal Statistical Society, 75,* 579–652.

Zimmerman, D. W. (1975) Probability spaces, Hilbert spaces, and the axioms of test theory. *Psychometrika, 40,* 395–412.

Zimmerman, D. W. (1976) Test theory with minimal assumptions. *Educational and Psychological Measurement, 36,* 85–96.

Zimmerman, D. W., Williams, R. H. (1977) The theory of test validity and correlated errors of measurement. *Journal of Mathematical Psychology, 16,* 135–152.