Modelling and testing interactive relationships within regression analysis

Thome, Helmut

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Diese Version ist zitierbar unter / This version is citable under:
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-33129

Modelling and Testing Interactive Relationships within Regression Analysis

*Helmut Thome**

Abstract: Regression analysis is one of the major re-
search tools in the social sciences, but this technique is
not often used to its full capacity. In most cases appli-
cations are restricted to linear additive models even
though theoretical considerations may point to non-
linear and/or interactive models. In this article the sub-
stantive interpretation of interactive models is clarified
and the often heard objections that interactive models
are not suitable for interval level data and are vitiated
by multicollinearity problems are shown to be unwar-
ranted.

1. Introduction

Regression analysis is one of the major research tools used in the social
sciences including the study of social history. It is a model-oriented ap-
proach, i. e., it presupposes a theoretical hypothesis which specifies a re-
lationship among variables that can be formally expressed in one or more
equations. In most social science applications, regression analysis is con-
fined to single equation models of a linear-additive form, such as

$$(1\text{-}1) \qquad Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_K + u$$

Variables $\{X_1, \ldots, X_K\}$ are assumed to be the »independent« or »predictor«
variables which »influence« or »determine« (to a certain degree) the »de-
pendent« variable Y, each X_k (k= 1,2,...,K) »adding« something to Y in a
linear fashion. The X-Variables are also called »regressor variables,« the
Y-variable is sometimes referred to as the »criterium« or »response variab-
les The »u« represents »errors« or »disturbances« which may result from
measurement flaws or from other variables (»implicit« variables) which
may influence Y, but have been (for good or bad reasons) left out of the

* Address all communications to Helmut Thome, Zentrum für Histori-
sche Sozialforschung, Zentralarchiv für Empirische Sozialforschung,
Universität zu Köln, Bachemerstr. 40, D-5000 Köln 41.

equation. Whereas Y and u are always treated as »random« variables (usually assumed to follow a normal distribution), the X's may be either »fixed« (e. g. by experimental design) or random as well. Y and, by implication, u have to be metrically scaled, while the X-variables may be metric or categorical. The constant »α« (the »intercept«) and the coefficients $\{\beta_1,...,\beta_K\}$ are called »parameters« of the (theoretical) model and usually have to be estimated from the data. The intercept states a level (often fictitious) which Y reaches when all the regressor variables take on a value of zero. In most cases theoretical interest centers upon the β- coefficients, which are also called »structural coefficients« or »slopes« or »slope-coefficients.« According to model (1-1) a unit increase in X_k (k= 1,2,...,K) leads to an increase in the level of Y in the amount of β_k under the condition that all the other regressor variables in the equation are being held constant. Mathematically speaking, the slope β_k is the 1st partial derivative with respect to X_k. Thinking in causal terms, the slope β_k may be interpreted as an »effect« parameter indicating the **specific** influence which X_k exerts upon Y when all the other regressors have been »partialed out,« In the linear-additive model any effect parameter β_k is hypothesized to be constant over the whole range of values of X_k and all the other regressor variables.

The parameters of model (1-1) are estimated from the observed values of the Y- and X-variables, in most cases by the method of Least Squares (OLS: Ordinary Least Squares). To indicate the transition from the purely theoretical statement in (1-1) to the task of empirical estimation on the basis of sample data one usually changes notation:

(1-1') $Y= a + b_j X^{\wedge} \; b_2 X_2 + \; ... + \; b_K X_2 + \; e$

The regression coefficients {a, $b_1,...,b_K$ } are considered to be »estimators« of the corresponding parameters in model (1-1): $a = \hat{\alpha}, b_k = \hat{\beta}_k$ where the hat indicates the quality of being an estimator. They are mathematically determined in such a way as to minimize the sum of squared errors over all n cases: Σe_i^2 = min., i=1,2,...,n (»n« being the number of observations). The estimators are said to be BLUE (Best Linear Unbiased Estimators) if certain assumptions have been met.[1] The most important prerequisite is that the model has been correctly specified, i. e., that its functional form (e.g. linearity, additivity) has been stated adequately and that all »relevant« variables have been included in the model. Specifically, there should be no variable left out of the equation which is related to the dependent variable and to at least one of the regressors. Violation of this assumption leads to inconsistent parameter estimation, i.e. to flaws in estimators that cannot be overcome by increasing sample sizes even when the sample approaches or becomes identical with an empirical population beyond which one does not want to generalize.

Researchers have often lamented that the data needed to include all relevant variables in the estimation equation are not available. At the same time, however, they have been much less concerned about the appropriateness of the functional form in which the model is stated. Yet, it is often in this respect that the model is unnecessarily deficient even though sufficient data is available. It is not the lack of data which prevents a researcher to correctly specify functional forms, rather it is his lack of imagination or thoughtfulness which might cause trouble.

The researcher, for example, must carefully consider the possibility that the relationship between variables is **non**-linear, turning the slope coefficient from a constant to a variate. Figure 1, for example, depicts a relationship which is quadratic in form:

$$(1\text{-}2) \quad Y = \alpha + \text{ß}X + \delta X^2 + u \; ; \text{ß} > 0, \; \delta < 0, \; |\delta| < |\text{ß}|$$

Figure 1

An example of a non-linear relationship
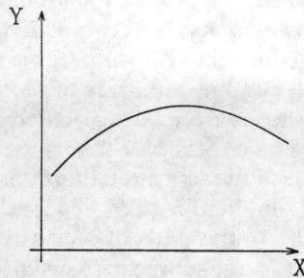
With low values in X a unit increase in X leads to relatively high positive increments in Y. But the slope decreases with growing X-values and eventually turns negative. The slope at any value of X can again be determined by caculating the first derivative of equation (1-2) with respect to X:

$$(1\text{-}3) \quad dY/dX = \text{ß} + 2\delta X$$

Setting (1-3) to zero and solving for X results in

$$(1\text{-}3') \quad 0 = \text{ß} + 2\delta X$$
$$X = \text{ß}/2\delta$$

Thus the value of X where the slope becomes zero depends on the relative size of the **ß**- and **δ**-Parameter, which need to be estimated.

In certain countries the relationship between personal income and age has been found to follow such a pattern. The history of social movements might reveal similar patterns between the intensity of protest actions and

the degree of suppression administered by government. If 6 and 8 change signs, the relationship between Y and X turns from an inverted U to a non-inverted (but somewhat stylized) U-shape.

Equation (1-2) causes no additional estimation problems when compared with equation (1-1), because both models are »linear in their parameter and can be estimated by OLS. There are numerous possibilities to express varying functional forms in such a way as to keep the equation »linear in the parameters« (thereby estimable by means of OLS) and yet posit a non-linear relationship between the variables involved.^

In model (1-2) the impact of X on Y (measured in terms of the 1st derivative) is made dependent on the level reached by X itself. In the social sciences however it is often quite conceiveable that the impact a variable X_1 exerts upon a variable Y may depend on the value of another variable X_2 (or vice versa) or on the values of a whole set of other variables. For example, the influence of socio-economic status (SES) on one's political orientation (PO) may depend on membership in one or the other religious denomination (RD). In the literature different terminologies have been used to label this type of relationship. It is said, for example, variable RD »moderates« or »specifies« the relationship between SES and PO, or variable SES »interacts« with variable RD in its impact on variable PO.

In the social sciences generally and in the study of social history specifically, interactive relationships have rarely been specified in models set up for regression analysis. One may wonder, why this is the case, given the fact that in theoretical discussions interactive relationships have been proposed (explicitly or, more often, implicitly) quite frequently. Critics of the so-called »quantitative approach« to the study of social history have often argued that quantitative analysis is not appropriate or unfeasible in many instances, because the impact a certain »factor« may have upon another factor is thought to be dependent upon the historical »context.« What is meant by context dependency, however, could often be more clearly stated in terms of interactive relationships[2a]. Therefore, the argument about context dependency might be turned around to make the case not against but for »quantitative« analysis (provided the theoretical constructs can be translated into measureable variables and sufficient data are available).

Apart from principal considerations about the »logic« of historical investigation, in the past there have been three major arguments raised against the idea of incorporating interactive relationships in regression models: (1) The analytical meaning of the regression coefficients is hard to specify or is even impossible to be clarified at all. (2) Models incorporating interactive relationships run, when estimated, into severe problems of multicollinearity, thereby producing highly unstable estimates. (3) If the estimated parameters are to be interpretable at all, the variables must be measured on a ratio-scale level hardly obtainable in the social sciences.

Based primarily upon an article by Friedrich (1982), which has appa-
rently received little attention among social science practitioners, this pa-
per will show that there is little validity to the above mentioned argu-
ments. Thereby students of sociology or social history may be encouraged
to consider interactive relationships in their theoretical thinking and to
incorporate them in regression models.

The following section will explicate the meaning of interactive relation-
ships within the context of tabular analysis, which may be more familiar to
some of the readers. The third and fourth section demonstrate, how such
relationships can be modeled and tested in regression analysis. The next
two sections turn, first to the problem of multicollinearity and then to the
level of measurement argument. The seventh section considers interaction
between qualitative and metric variables and introduces the notion of co-
variance analysis.

2. Interaction in tabular analysis

The following example is taken from data on members of the German
national assembly that was convened in Frankfurt during 1848/49. The
original research was carried out by Heinrich Best, University of Colo-
gne.[3]   Factor analysing documented roll-call behavior, he constructed,
among other things, an index showing the »left-right« orientation of the
individual assembly members. In the examples following below this index
has been broken down to a dichotomy which serves as the dependent va-
riable (Y_1). A second variable (X_1) indicates the constitutional tradition of
each representative's constituency. Again, for reason of simplicity, only
two categories have been defined: (a) territories which were ruled by an
absolute monarchy at least until 1848, (b) territories which were governed
under some form of constitutional law (»constitutional regime«) before
1848. (But keep in mind that the units of analysis are not the territories but
the individuals who represent territories classified in this way.) The biva-
riate relationship between variables Y_1 and X_1 is shown in Table 1.

As can be seen by comparing column frequencies, representatives from
absolute monarchies are more likely to lean towards the »right« than
representatives of constitutional regimes. The percentage difference is
64.7 % - 44.8 % =   19.9 %, which is usually taken as indicating a relation-
ship of »moderate strength.«

Now let us look at a third variable (Z), the religious denomination of the
individual member, which again has been dichotomized into »catholic« or
»non-catholic« confession. Table 2 represents the bivariate distribution of
variables Y_1 and Z.

Table 1

Bivariate relationship between
constitutional tradition and ideology

Ideology	Abs. Monarchy	Const. Regime	Row Total
Right	299 (64.7 %)	126 (44.8 %)	425 (57.2 %)
Left	163 (35.3 %)	155 (55.2 %)	318 (42.8 %)
Column Total	462 (100.0 %)	281 (100.0 %)	743 (100.0 %)

Table 2

Bivariate relationship between
religious denomination and ideology

Ideology	Catholics	Non-Catholics	Row Total
Right	181 (57.3 %)	244 (57.1 %)	425 (57.2 %)
Left	135 (42.7 %)	183 (42.9 %)	318 (42.8 %)
Column Total	316 (100.0 %)	427 (100.0 %)	743 (100. %)

Table 3

Three-dimensional relationship between
personal confession, type of constitution, and ideology

Ideology	Catholics		Non-Catholics	
	Abs. Mon. (1)	Const.Reg. (2)	Abs. Mon. (3)	Const.Reg. (4)
Right	132 (56.7 %)	49 (59.0 %)	167 (72.9 %)	77 (38.9 %)
Left	101 (43.3 %)	34 (41.0 %)	62 (27.1 %)	121 (61.1 %)
Column Total	233 (100%)	83 (100%)	229 (100%)	198 (100%)

Surprisingly perhaps, no relationship appears: catholics are just as much inclined towards the right or the left as non-catholics. But religious denomination is not an irrelevant variable at all. This becomes apparent in the three-dimensional distribution of Table 3. In this table variable Z can be read as a »control variable« whose categories define conditions under which »partial« relationships between Y_i and X_i can be re-examined. Several results emerge:

1. The relationship between constitutional tradition (or »context«) and left-right orientation virtually disappears among catholics, though it shows increased strength among non-catholics (with d% = 72.9 % - 38.9 % = 34%). Religious confession apparently »modifies« (and »specifies«) the relationship between constitutional context and left-right orientation. If one assumes that constitutional rule is a better seedbed than absolute monarchies for developing leftist orientations, one might conclude that non-catholics are more amenable than catholics to political tendencies prevalent in their immediate environment. (But substantive interpretation is beyond the purpose of this paper.)

2. By some eyeballing we can exchange the positions of variables X_i and Z without altering table 3. Comparing columns (1) and (3) we recognize that non-catholics from absolute monarchies are more likely to be right-oriented (72.9 %) than catholics from absolute monarchies (56.7 %). This relationship is reversed when we look at the representatives from constitutional regimes (comparing columns 4 and 2): Non-catholics are considerably less likely to lean towards the right (38.9 %) than catholics (59.0 %). So, when holding the constitutional context constant a relationship between religious denomination and left/right-orientation appears in the subtables, a relationship which differs depending on the constitutional context given.

3. Such a pattern of causal connectedness is called an »inter action«: Variables X_i and Z »interact« in their influence upon variable Y_i. The relationship between X_i and Y_i depends upon which category of the third variable, Z, is given at the same time. Equally, the relationship between Y and Z depends upon which category of the (then) third variable, X_1, is also given. The interaction-component in a three or more variable relationship is symmetric. By looking at the difference between the percentage differences (»second-order« difference)obtained in the partial tables we get a rough indication as to the strength of the interaction. The percentage difference in the subtable made up of columns 3 and 4 in Table 3 is d% = 34. In the sub-table made up of columns 1 and 2 it is d% = -2.3. So the second-order difference indicating interaction is 34 % - (-2.3 %) = 36.3 %. Looking at the sub-tables when variables Z and X change places, i. e. looking at the partial tables made up by columns 3 and 1 and columns 4 and 2, respectively, we get the second order difference of (72.9 % - 56.7 %) -

Table 4

Bivariate relationship between territory
and constitutional voting

Voting	Austrian	Mixed	Prussian	Row Total
"großd."	170 (74.2 %)	109 (54.0 %)	102 (27.0 %)	381 (47.1 %)
"kleind."	59 (25.8 %)	93 (46.0 %)	276 (73.0 %)	428 (52.9 %)
Column Total	229 (100 %)	202 (100 %)	378 (100 %)	809 (100 %)

Table 5

Three-dimensional relationship between religious
denomination, political territory, and constitutional voting

| Const. Voting	Catholics			Non-Catholics		
	Austri-an (1)	Other (2)	Pruss-ian (3)	Austri-an (4)	Other (5)	Pruss-ian (6)
großd.	156 (76.1%)	56 (61.5%)	15 (36.6%)	11 (57.9%)	51 (47.7%)	74 (23.2%
kleind	49 (23.9%)	35 (38.5%)	26 (63.4%)	8 (42.1%)	56 (52.3%)	245 (76.8%
Column Total	205 (100%)	91 (100%)	41 (100%)	19 (100%)	107 (100%)	319 (100%)

(38.9 % - 59.0 %) = 36.4 %. Apart from rounding errors the two second-order differences are the same.

The concept of interaction may be further clarified by looking at a three-variable relationship in which interactive components (in the sense defined above)[4] are absent. Two new variables have been constructed from the same set of data. The dependent variable (Y_2), again a dichotomy, depicts the representatives' stand in the constitutional conflict that the Frankfurt Assembly was not able to resolve: Judging from his roll-call behavior, did he favor a »larger« German Empire including Austria (»großdeutsche« Lösung) or did he opt for the »smaller« Germany that would exclude Austria and transfer the leading role solely to Prussia (»kleindeutsche« Lösung). The independent variable (X_2) assigns each representative's constituency to one of three larger political territories: the first comprises Austria and states close to it (like Bohemia); the second includes Rhine-Prussia, Franconia, and the smaller southern states; the third groups together Prussia, Schleswig, Saxony and the smaller states in middle and northern Germany. In Tables 4 and 5 these territories have been given the shorthand labels: »Austrian«, »Mixed«, »Prussian« (region).

Table 4 indicates a fairly strong (monotone) relationship between territorial affiliation and the representative's stand in the constitutional debate. We introduce again religious denomination (Z) as a control variable leading to the three-dimensional distribution in Table 5.

In both subtables (for catholics as well as for non-catholics) the relationship between »territory« and »constitutional vote« is of about the same strength (the two coefficients for Cramer's V are almost equal), i.e., it is **not** specified by personal religious affiliation. Representatives from Prussian oriented territories are more likely to opt for the »smaller« Germany than their colleagues from the Austrian-oriented areas, and this holds for catholics (d%= 63.4-23.9= 39.5) and non-catholics (d% = 76.8 - 42.1 = 34.7) nearly alike. On the other hand, it is also true, that given the same territorial affiliation (within each category) catholics are more likely to opt for the »großdeutsche Lösung« than their non-catholic counterparts. And again, the percentage differences within each territorial category are roughly the same. So, we must conclude that there is no (at least no relevant) interaction between territorial affiliation and religious denomination with respect to their impact upon constitutional voting. Each of these two determining variables, X_2 and Z, has some impact on the dependent variable, but the impact of one is not influenced or shaped by the other. Such a pattern of relationship is often called »additive« in opposition to the interactive relationship exemplified in Table 3. Both models may be graphically prensented as in Figure 2.

29

Figure 2: **Additive (a) and interactive (b) relationships**
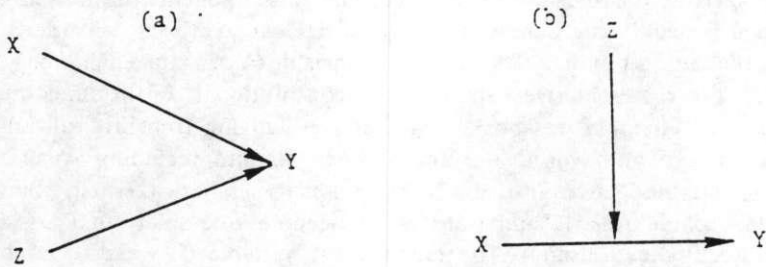
When interaction occurs, a third variable Z influences the relationship between two other variables X and Y, giving it a different shape or strength depending which value or category of Z is realized at the same time. In a purely additive 3-variable-relationhip there are only lines of influence running between variables; the relationship between each pair of variables is the same across all values or categories of the third variable. (An altogether different question is whether or not the conditional relationships in the subtables of a three-dimensional distribution differ from the respective unconditional (bivariate) relationship in the two-dimensional distribution.)

Both, the interactive and the purely additive model can be easily generalized to higher-dimensional tabels. Then second-order interaction might occur in which a fourth variable specifies the (first-order) interaction between two other variables. Such complexities are, however, hard to disentangle by means of conventional tabular analysis, which does not specify a parametric model, but relies solely on percentage differences and/or simple coefficients of association (like Cramer's V).

3. Introducing and Interpreting Multiplicative Terms in Regression Equations

The example worked out for illustration in this section is taken from a research project conducted by W. H. Schröder at the Center for Historical Social Research.[5] The dependent variable (Y) contains the percentage proportion of votes which the SPD was able to attract in each of 395 districts in the 1912 election to the German »Reichstag« (two cases had missing data). Two independent variables will be considered. The degree of urbanization (X_1) and the percentage proportion of protestants living in

the election district (X_1). The degree of urbanization is measured as the percentage proportion of the population in each election district living in communities of more than 2000 inhabitants. We start with the simple additive regression model given in equation (3-1) making the usual assumptions.^ Applying least square estimation and weighing cases by size of the electorate we arrive at the following results (standard errors, first, and t-values, second, are given in parenthesis):

$$(3-1) \qquad SPD = -11.79 + .39(URBAN) + .28(PROT) + e$$

$$\begin{array}{ccc} (1.42) & (.017) & (.0146) \\ (-8.3) & (22.7) & (19.0) \end{array}$$

$$R^2 = .71 \qquad\qquad F = 480.76$$

The outcome is hardly surprising. The proportion of SPD votes is positively related to urbanization and the proportion of protestants; together they explain 71 % of the variance.

In a next step one might hyothesize that the impact of religious denomination upon party preferences will be weaker in an urban surrounding as compared to rural settings. Simplifying matters one may assume that urban settings create melting pots in which traditional group loyalties (including religious denomination) tend to erode. In a new model we consequently expect the slope of $E(Y)$ on PROT to decrease when urbanization increases. This would be a case of interaction as explicated in section 2: We expect urbanization to influence (specify) the **relationship** between religious denomination and party preference (as measured by election results).

This interaction hypothesis can be formaly expressed in the following equation:

$$(3-2) \qquad \beta_2 = c + dX_1 , \quad d < 0$$

where β_2 symbolizes the slope of PROT. We assume that this slope (i. e. the impact of PROT upon the proportion of SPD votes) be linearly dependent on the level of urbanization (X_1). Other forms of relationships might be considered, e.g., a quadratic one. But for reasons of simplicity (and conforming to common practice), we accept the linear sub-model. Substituting equation (3-2) into equation (3-1) gives

$$(3\text{-}3) \qquad Y = \alpha + \beta_1 X_1 + (c + dX_1)X_2 + u$$
$$= \alpha + \beta_1 X_1 + cX_2 + d(X_1X_2) + u$$

or, in more common notation

$$= \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u \;, \; X_3 = X_1 X_2, \; \beta_3 < 0$$

Comparing equation (3-3) with equation (3-1), we note that a multiplicative term $(X_1 X_2)$ has been added. By actually performing the multiplication we construct a third regressor variable (X_3) as a composite of the two original regressors. The equation remains linear in the parameters and can again be estimated by the method of least squares. However, three slope coefficients $\{\beta_1, \beta_2, \beta_3\}$ must be interpreted with respect to only two substantive Variables, X_1 and X_2. How is this to be done?

As noted before, in the purely additive model of type (1-1) the coefficients $(\beta_1, \beta_2,...,\beta_K)$ attached to the regressor variables $(X_1, X_2, ... , X_K)$ are »slopes« telling us how much the expected value of Y increases in response to a unit increase in X_k (k= 1,2,...,K), »holding constant« all the other regressor variables. Formally, the slope of a function is obtained by the first (partial) derivative. In model (3-1) the 1st derivative with respect to X_2 is

$$(3\text{-}4) \qquad dY/dX_2 = \beta_2$$

In the linear-additive model the slopes are constants, in nonlinear models or models including multiplicative terms the slope varies with the value of X. A parabolic function has already been given in equation (1-2) so that we can immediately turn to the multiplicative model in equation (3-3).

Since we are only interested in expected values we drop the error term and write

$$(3\text{-}5) \qquad E(Y \mid X = x) = Y^* = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 X_2)$$

The slopes, i. e., the partial derivatives, of X_1 and X_2 are given by

$$(3\text{-}6) \qquad dY^*/dX_1 = \beta_1 + \beta_3 X_2$$
$$dY^*/dX_2 = \beta_2 + \beta_3 X_1$$

Thus, the coefficient B_3 of the multiplicative term has a clearcut meaning: it represents the amount of change in the slope of E(Y) on X_1 which follows from a unit increase in X_2. In the same way it also represents the amount of change in the slope coefficient of X_2 that follows from a unit

increase in X_1 Although sub-model (3-2), which cannot be estimated, expresses a one-way relationship, in the final model (3-3) the interaction between X_1 and X_2 is symmetric. With respect to our substantive example: if it were true that with increased urbanization religious denomination made less of a difference in voting behavior, it should also hold that with increasing proportions of protestants the impact of urbanization on SPD-voting diminished. This is less disturbing than it might look at first glance. If (1) urbanization increases, and, consequently, (2a) the proportion of SPD votes increases, and (2b) at the same time the influence of religious confession decreases, then (3) the SPD-gains in the cities (over against rural areas) must be larger among catholics than among protestants (otherwise the difference between catholics and protestants should not have become smaller)[7]. This implies, on the other hand, that the overall SPD-gains brought about by increased urbanization are lower when the proportion of protestants in the electorate is larger. Thus, there must be a numerical symmetry even though the causal chain might be asymmetric. (Causality is, at any case, a theoretical imputation.)[8]

Having clarified the meaning of the coefficient $ß_3$ (attached to the multiplicative term), we now turn to coefficients $ß_1$ and $ß_2$ in equation (3-5). As is obvious from equation (3-6), the coefficient $ß_1$ represents the slope of the relationship between $E(Y)$ and X_1 only under the very specific condition that $X_2 = 0$. In the same manner $ß_2$ represents the slope of $E(Y)$ on X_2 only under the condition that $X_1 = 0$.[9] This constitutes the crucial difference in the meaning of the regression coefficients as we move from purely additive to multiplicative (interaction) models. In the former case (see model (3-1)) the slopes are constants; in the latter model (3-5) the slope of $E(Y)$ on one of the regressors is **conditioned** by the given value of the other regressor and must be specifically calculated for this value.

Returning to our substantive example we obtain from least squares estimation the following paramaters of the multiplicative model (standard errors and t-values are again given in parenthesis):

```
(3-7) SPD = .71 + .15(URBAN) + .06(PROT) + .004(URBANxPROT) + e
            (2.04) (.034)        (.031)        (.0005)
            (.349) (4.47)        (1.99)        (7.99)
            R² = .75                         F = 393.19
```

The coefficient estimated for the multiplicative term seems to be unduly small (for significance testing see next section), but we must keep in mind that we have multiplied two percentage values, thereby arriving at a new scale unit. Contrary to our theoretical assumption that coefficient is posi-

tive: with increasing urbanization the impact of the confessional variable on voting SPD does not decrease, but rather increases. Post festum one might speculate that urbanization in an early stage threatens personal and cultural identities in various ways thereby exciting defensive reactions that, at least temporarily, lead to the strengthening of traditional group loyalties« and group conflicts. This »hypothesis« is of course ad hoc and would need to be elaborated and then tested by new sets of data. Situational factors such as acute power conflicts between political parties or between the state and the churches would also need to be taken into account.

According to model (3-5) how much does the slope of E(Y) on PROT vary given the actually observed range of urbanization values? The lowest URBAN-value is 7 %, the highest is 100 %. Accordingly the lowest and highest slopes are

$$(3\text{-}8) \quad b_{min}(PROT) = b_2 + b_3X_1$$
$$= .06 + 0.004 \cdot 7 = 0.088$$
$$b_{max}(PROT) = .06 + 0.004 \cdot 100 = 0.46$$

The average degree of urbanization is 61 %. Under this condition the slope of E(Y) on PROT becomes

$$(3\text{-}9) \quad b_m = 0.06 + 0.004 \cdot 61 = 0.30$$

which is nearly identical to the unconditional slope of E(Y) on PROT in the purely additive model. In the literature this coefficient is sometimes called the »main effect« interpreted as the average impact of the independent variable (here PROT) upon the dependent variable (SPD-votes) across all levels of the »control« or »moderator« variable (Jaccard/Turrisi/Wan 1990: 14 f.)

4. Standard Errors and Significance Testing

The research hypothesis that there is an interaction between two regressors can be tested against the nullhypothesis that there is no interaction. This can be done by way of the F-change-test, which is regularly applied when the statistical significance of a new regressor (or several new regressors) added to an already estimated model is to be evaluated. The test relates the additional proportion of variance explained by the additional variable (in this case, the product-term of the interacting variables) to the proportion of variance left unexplained by the extended model. The F-change value can be calculated from the coefficients of determination: $R^2_{(1)}$ for the purely additive model, and $R^2_{(2)}$ for the extended model which includes the multiplicative term; both coefficients of determination need to be divided by the appropriate degrees of freedom:[10]

(4-1)
$$F_{change} = \frac{(R^2_{(2)} - R^2_{(1)})/q}{(1 - R^2_{(2)})/(n-(K+1)-q)} = \frac{R^2_{change}(n-(K+1)-q)}{(1 - R^2_{(2)})q}$$

$$= \frac{(.751 - .710)(395-2-1-1)}{.249} = 64.4$$

(The coefficients of determination are given above with equations (3-1) and (3-7). K+1= 3 represents the number of parameters (including the intercept term) in the reduced model; q= 1 gives the number of variables added to the first model, in this case a single product term).

With one degree of freedom in the numerator and 391 degrees of freedom in the denominator, an F-value of 64 indicates a significance level well below the .001 level. Thus, the multiplicative term is »highly significant in the statistical sense, even though the gain in explained variance ($R^2_{change} = .041$) does not look very impressive.

The same result can be obtained from a t-test applied to the coefficient, b_3, attached to the multiplicative term. In general the squared t-value applied to a single regressor is equal to the F-change value with one degree of freedom in the numerator and n-1-K degrees in the denominator. In our example from equation (3-7):

(4-2) $t^2_{391} = F_{1;391}$
       $(7.99)^2 = 63.8$

which is, within rounding errors, identical to the value given in equation (4-1). In principle, the statistical significance of each of the terms in the regression equation can be tested by its respective t-value.[11]

In our example we note a marked decline in the t-values for b_1 and b_2 as we move from model (3-1) to model (3-7), though both still indicate a significance level of $\alpha < 0.05$. In order to understand this reduction of t-values, we must remember that b_1 (b_2) now represents the slope of E(Y) on X_1 (X_2) under the condition that X_2 (X_1) equals zero. If the slopes are conditioned, so are the standard errors and t-values, too. That is, one must calculate standard errors not just for the estimates b_1 and b_2 but, following equation (3-6), for the sums ($b_1 + b_3x_2$) and ($b_2 + b_3x_1$) treating $X_1 = x_1$ and $X_2 = x_2$ as given constants.

For two random variables, U and V, and a constant c it generally holds that[12]

(4-3)   (a) $\text{var}(cU) = c^2\text{var}(U)$
         (b) $\text{var}(U + V) = \text{var}(U) + \text{var}(V) + 2\text{cov}(U,V)$
         (c) $\text{cov}(U,cV) = c\,\text{cov}(U,V)$

Treating the estimates b_1, b_2, and b_3 in the multiplicative model as random variables and the specific conditions $X_1 = x_1$ and $X_2 = x_2$ as constants, we can, based on the theorems in (4-3), derive estimates for the conditional standard errors (s) of the slopes defined in (3-6) according to

(4-4) $$s(b_1+b_3x_2) = \sqrt{\text{var}(b_1) + x_2^2\text{var}(b_3) + 2x_2\text{cov}(b_1,b_3)}$$

for the slope of $E(Y)$ on X_1 under the condition of $X_2=x_2$

and

$$s(b_2+b_3x_1) = \sqrt{\text{var}(b_2) + x_1^2\text{var}(b_3) + 2x_1\text{cov}(b_1,b_3)}$$

for the slope of $E(Y)$ on X_2 under the condition of $X_1=x_1$.

A variance-covariance matrix of (estimated) regression coefficients is provided by any standard computer package for statistical analysis. In our example we obtain the following information:

	b_1	b_2	b_3
b_1	.00114		
b_2	7.913E-04	9.237E-04	
b_3	-1.493E-05	-1.362E-05	2.508E-07

The variances are given in the diagonal, the covariances in the lower triangle. For illustration, let us determine the value of $s(b_2+b_3x_1)$ for the slope coefficient of PROT ($=X_2$) under the condition that URBAN $= x_1 = 61$:

(4-5) $$s(b_2+b_3x_1) = \sqrt{.0009237 + (61)^2(.0000002508) + 2\cdot61(-.00001362)}$$

$$= \sqrt{.00019231} = .0139$$

which is again very close to the standard error given for b_2 in the purely additive model (3-1). This however does not equal the minimal standard error for the conditional slope of $E(Y)$ on PROT. The minimal standard

errors are obtained by differentiating equations (4-4) with respect to X_1 and X_2, setting the first derivatives equal to zero and solving the resulting equations for X_1 and X_2, respectively (Friedrich 1982: 190):

$$(4\text{-}6) \qquad X_1 = -\text{cov}(b_2, b_3)/\text{var}(b_3)$$

$$\text{(the minimizing condition for } s(b_2 + b_3 x_1))$$

and

$$X_2 = -\text{cov}(b_1, b_3)/\text{var}(b_3)$$

$$\text{(the minimizing condition for } s(b_1 + b_3 x_2))$$

Consequently the minimal standard error for the relationship of $E(Y)$ on X_2 (= PROT) will be given under the condition of

$$(4\text{-}7) \qquad X_1 (= \text{URBAN}) = -(-.00001362)/.0000002508) = 54.3$$

Under this condition the slope coefficient for $E(Y)$ on PROT takes on the value of

$$(4\text{-}8) \qquad b(\text{PROT} \mid \text{URBAN} = 54.3) = b_2 + b_3(54.3)$$
$$= 0.06 + 0.004(54.3) = .278$$

using equations (3-6) and (3-7). Its standard error, according to equation (4-4), is given by

$$(4\text{-}9) \quad s(b_2 + b_3 \cdot 54.3) = \sqrt{.0009237 + (54.3)^2(.0000002508) + 2 \cdot 54.3(-0,00001362)}$$
$$= \sqrt{.0001848} = .01359$$

This standard error is even lower than that of the corresponding slope b_2 in the purely additive model given in equation (3-1). This holds generally: if there is interaction, the minimal standard error for the conditional slope in an interactive model is always lower than the corresponding (unconditional) slope in the additive model (Friedrich 1982: 812 f.).

As a final step we now calculate the t-value for the conditional slope with minimal standard error determined by equation (4-8):

(4-10)
$$t = \frac{b_2 + b_3 x_1}{s(b_2 + b_3 x_1)} = \frac{.28}{.01359} = 20.6$$

This value is considerably higher than the t-value (1.99) given for b_2 in equation (3-7), where the condition was URBAN= 0. As Friedrich notes: »Not until conditional slopes and t tests are calculated within the observed range of experience of the variables can valid conclusions be drawn. Statistically insignificant b_1's, b_2's, and b_3's may nevertheless combine to produce statistically significant conditional effects.« (Friedrich 1982: 821).

5.   Multicollinearity

The inclusion of a multiplicative term into a regression model regularly introduces multicollinearity, and equally regularly this implication has been turned into an argument against this type of modelling approach. Before assessing the validity of the argument we first look at the consequences that ensue from »multicollinearity« in purely additive models.

The term denotes the degree of linear interdependence between the regressor variables. In the extreme case, a regressor X_i may be completly dependent upon the other explanatory variables included in the model (in other words: regressing X_i upon the other X's produces $R^2 = 1$). In this case the regression algorithm breaks down, and (partial) regression coefficients cannot be computed. Such a situation is unlikely to occur in practice. Hence, interest centers on situations where X_i can be predicted from the other X's to a rather »high« extent such that the coefficient of determination falls into an interval of, say, $0.5 < R^2 < 1$.

In such a case the estimates of the slope coefficients remain unbiased, i. e., on the (very) long run with repeated sampling the sample coefficients will »average out« on the »true« (population) value. However, the efficiency of estimates decreases as multicollinearity increases. Due to its increased standard error,[13] in any specific sample the estimate **may** contain a larger error component. This means, among other things, that the estimate may react very sensitively to small changes in samples. The problem is further aggravated by the covariances between the estimates.

The stronger the correlation between two regressors X_i and X_j, the stronger, with reversed sign, the covariance between the slope estimators. If the

regressor variables are positively correlated, the covariance between the sample coefficients is negative. Hence, if b_k overestimates the population coefficient, b_j tends to underestimate it (or vice versa). If the regressor variables are negatively correlated, the sample coefficients will jointly tend to over- or underestimate the population coefficient. To sum up, a high degree of multicollinearity may jeopardize significance testing and greatly distort the information on the (relative) strength of the explanatory power of regressor variables. All these negative consequences of high multicolli-nearity are ultimately due to an increase in the standard error of the esti-mates. Thus, »we expect the standard errors to give adequate warning of collinearity« (Johnston 1972, p. 163).

We now turn to multicollinearity in interactive models such as equation (3-7). In section 4 it has been shown that

(a) substantive interest does not center on the coefficients b_1, b_2, b_3 singly, but on the conditional slopes ($b_1 + b_3 X_2$) and ($b_2 + b_3 X_1$)

(b) although the standard errors $s(b_1)$, $s(b_2)$ and $s(b_3)$ are generally (not necessarily: Friedrich 1982, p. 813) larger in an interactive model than in a purely additive model, the conditional standard errors $s(b_1 + b_3 X_2)$ and $s(b_2 + b_3 X_1)$ will for at least some values of X_1 and X_2 be smaller than $s(b_1)$ and $s(b_2)$ in the purely additive model.

Let us again look at formula (44) for calculating the conditional stan-dard error in an interactive model. The third term under the root includes the covariance between the estimates b_1 (resp. b_2) and b_3. In many instan-ces this term will be negative, thereby tending to offset the increase in $var(b_1)$ or $var(b_2)$ which usually results from introducing a multiplicative term $X_1 X_2$ into the equation. Whether or not the covariance between b_1 (or b_2) and b_3 is negative depends on the correlation between the multiplica-tive term and its constituent variables.[14] In our example the correlation of URBEV with URBAN and with EV12 is r_{31}= .678 and r_{32}= .715, and the covariances between b_1 resp. b_2 and b_3 are negative. Even in cases where these covariances are positive, the X_2- or X_1-value of interest might still be negative.[15]

Thus, in general, multicollineartiy between the multiplicative term and its constituent variables, causes fewer problems than multicollinearity bet-ween explanatory variables that are additively combined to account for the variance in the dependent variable.

As a rough check on the severity of the problem one may calculate not only the conditional standard errors where they are minimal, but also those for the lowest and highest observed values of X_1 resp. X_2. For exam-ple, the minimal standard error (s= 0.01359) for the conditional slope of PROT was given above at an urbanization value of $x_1 = 54.3$. The lowest observed value of URBAN is $x_1 = 7.0$, and the conditional standard error of

b(PROT) at that point is s= 0.0273. The highest observed value of URBAN is $x_1=$ 100, and the conditional standard error there is 0.0266. Thus, the conditional slopes increase as one moves up and down the scale away from the point of minimum.[16] The respective conditional slope estimates have already been given in equations (3-8) and (4-8). Even the smallest of them, b(PROT$|x_1=7$) = 0.088, is more than three times as large as its standard error.

Friedrich (1982: 818) presents further evidence from Monte Carlo studies which support his conclusion that »interactive models,..., do yield sample estimates of population parameters that are accurate and reasonably and specifiably stable«.[17]

6. The Level of Measurement Argument

If a variable X_1 is measured on an interval scale (as opposed to ratio scales), its zero point is arbitrary, and adding or subtracting a constant c produces a new scale $X_2' = X_1 + c, c \# 0$, which is equivalent to the old one. (This is not true for ratio scales.) When regressing a variable Y, in a first run, on X_1 and X_2, then on X_1' and X_2 in a purely additive model, the slope coefficients remain the same in both runs, only the intercept changes. This does not hold for the interactive model. If two variables constitute a multiplicative term, adding a constant to one of the original variables causes the coefficients of the other to change. Therefore, it has often been argued that interactive models are useless and their coefficients are meaningless if one or more of the variables in the product term have been measured on interval scales (see, e.g., Allison 1977). As Friedrich (1982) has shown, the critics' objections are unwarranted.

In order to illuminate the problem, we slightly modify our regression model in equation (3-7) by subtracting the amount of 12 from each value of the confession variable PROT (thereby »downgrading« it to an interval scale):

(6-1)   PROZSPD$=a+b_1$(URBAN)$+b_2$(PR0T-12)$+b_3$ [URBAN*(PROT-12)]$+e$

Table 6 presents the results of this new analysis in addition to those given via equation (3-7). Differences are starred.

Table 6:

	Model		
	(3-7)	(6-1)	
R^2	.751	.751	
interc.	.714	1.440	
b_1	.151	.199	(*)
b_2	.060	.060	
b_3	.004	.004	
β_1	.241	.317	(*)
β_2	.113	.113	
β_3	.613	.564	(*)
$s(b_1)$	.034	.029	(*)
$s(b_2)$	.030	.030	
$s(b_3)$	.0005	.0005	
$t(b_1)$	4.474	6.959	(*)
$t(b_2)$	1.989	1.989	
$t(b_3)$	7.991	7.991	

(The ß's represent standardized slopes)

We note that the t-value of the multiplicative term has not changed; significance tests on interaction components are not affected by additive transformations of scales, even though the standardized regression coefficients of multiplicative terms do change (Southwood 1978:1166,1168)/[18]) There is indeed a change in the standardized and unstandardized coefficients of URBAN (X_1). To understand this, we must remember that b_1 in an interactive model represents the slope of X_1 only under the very specific condition that $X_2 = 0$. But $X_2 = 0$ in model (3-7) has become $X_2^* = -12$ in model (6-1). So, a change in b_1 is to be expected. Under the condition of $X_2^* = -12$ the slope of X_1 should be exactly the same as under the condition of $X_2 = 0$. We can easily check for this by calculating

(6-2) $b(X_1 | X_2^* = -12) = .199 + .004(-12) = .151$

which indeed equals the value of b_1 in model (3-7). The coefficients in model (6-1) convey exactly the same substantive information as the coefficients in model (3-7). The conditional slope b_2 of PROT has not changed since the conditioning variable X_1 (URBAN) has not been transformed, its zero point has not been moved along the scale. The same logic of reasoning extends to the conditional beta-coefficients and standard errors.[19]

Thus, we may conclude: »Rescaling then has no effect whatsoever on the description of the conditional relationship at any particular point or on the outcome of a test of significance ... Though the numerical values of the coefficients obtained may change, the results of interest and the substantive answers will not.« (Friedrich 1982, S. 823)

7. Interaction between metric and categorical variables

Let us assume that $X_1 :=$ URBAN is not available as a (nearly) continuous variable (percentages), but has been coded into three ranks or categories denoting »low«, »medium«, and »high« urbanization. We can easily construct such a variable by assigning to its first category those election districts ($n_1 = 78$ cases) which have up to one third of their population living in communities of more than 2,000 inhabitants; the second category is comprised of election districts ($n_2 = 140$) which have between one third and two thirds of their population living in communities of more than 2,000 inhabitants; and the last category is comprised of election districts ($n_3 = 177$) which have more than two thirds of their population living in communities of more than 2,000 inhabitants.

In regression analysis ordinal variables are usually treated just like categorical variables. They are introduced into regression models by first transforming them into so-called »dummy variables.« If the number of categories of variable X equals g, then (g-1) artificial variables D_1, D_2,..., $Dg-1$ need to be constructed according to the following rules:[20] Arbitrarily select one of the g categories of X as a so-called reference or base category. In our example we choose the »medium«-level of urbanization as our base. If a case, i.e., an election district, belongs to this category, it has the value of zero assigned to it on all of the (g-1)= 2 dummy variables to be created. If a district displays a »low« level of urbanization, it is given the value »1« on the first »dummy,« D_1, and the value »0« on the second »dummy,« D_2. If a district has reached a »high« level of urbanization, it obtains a »0« on D_1 and a »1« on D_2. In regression analysis these dummy variables are technically treated in the same manner as any metric regressor variable. Let us examine two models (in sample notation):

(7-1) $PROZSPD = a' + b_1'D_1 + b_2'D_2 + b_3'(PROT) + e$

(7-2) $PROZSPD = a + b_1D_1 + b_2D_2 + b_3(PROT) + b_4(D_1*PROT) + b_5(D_2*PROT) + e$

The estimation results for both models are given in Table 7.

Table 7:   **Estimation results**

Model

	(7-1)	(7-2)
a	8.060	7.534
bi	-11.094	-2.075
b_2	15.349	8.728
b_3	.265	.273
b_4		-.176
b_5		.101
t(a)	5.858	4.203
t(bj)	-7.538	-.799
t(b_2)	13.231	3.418
t(b_3)	16.615	11.459
t(b_4)		-4.575
t(b_5)		2.900
R^2	.667	.705

Model (7-1) formalizes the hypothesis that the proportion of votes for the SPD is additively related to the proportion of protestants in the electorate and to the level of urbanization. It is further assumed that the slopes of PROT are the same for any category of urbanization. It is only the level, i. e., the intercept of the SPD-votes, that changes as one moves from one category of urbanization to the next. For »medium« districts the intercept is a'= 8.06, for »low« districts it is a'+b₁'= -3.03, for »high« districts it is a'+ b_2'= 23.41. Thus, the dummy coefficients b_1' and b_2,' in model (7-1) state differences in the level of the dependent variable which occur when the proportion of protestants is »held constant« at any value. According to model (7-1) the predicted value of PROZSPD in low urbanization districts is a'+b₁'+b_3,'(PROT); for high urbanization districts it is a'+ b_2,'+ b_3,'(PROT); and for medium urbanization districts it is a'+ b_3,'(PROT).

We already know, however, from the analyses in previous sections that model (7-1) is not correct. There is interaction between urbanization and the proportion of protestants (resp. catholics) in the electorate: the slope of PROT varies depending on the level of urbanization. Model (7-2) forma- lizes this assumption by introducing two product terms which result from multiplying (case by case) the two dummy variables with PROT. Techni- cally these multiplications do not differ from generating the product term of two metric variables. The slope of PROT is again a **conditioned** para- meter depending on the values of D_1 and D_2:

43

$$(7-3) \qquad \frac{d(\text{PROZSPD})}{d(\text{PROT})} = b_3 + b_4 D_1 + b_5 D_2$$

If the case (district) belongs to the »medium« category of urbanization, i.e., if $D_1 = D_2 = 0$, the slope $b(\text{PROT})$ is equal to $b_3 = .273$ (see table 7.1). If it is classified into the »low« urbanization category, we have $D_1 = 1$ and $D_2 = 0$, and therefore $b(\text{PROT}) = b_3 + b_4 = .097$. Finally, if the district has a »high« degree of urbanization, $b(\text{PROT}) = b_3 + b_5 = 0.374$. Thus, the coefficients of multiplicative terms involving dummy variables represent **differences** of slopes that appear as one moves from the reference category (»medium« urbanization«) to the other categories represented by dummy variables.

If interaction is present, the coefficients of D_1 and D_2 are also conditioned (by PROT), and consequently they do not directly translate into level differences in the dependent variable. For instance, let us assume that PROT= 10 in model (7-2). The expected value of SPD votes in a district with »low« urbanization then is

$$(7-4) \qquad \begin{aligned} E(\text{PROZSPD}) &= a + b_1 D_1 + b_3(\text{PROT}) + b_4 (D_1 \cdot \text{PROT}) \\ &= 7.534 - 2.075 + .273(10) - .176(10) \\ &= 6.429 \end{aligned}$$

For a district with »medium« urbanization and PROT= 10 we would expect

$$(7-5) \qquad \begin{aligned} E(\text{PROZSPD}) &= a + b_3(\text{PROT}) \\ &= 7.534 + .273(10) = 10.264 \end{aligned}$$

Thus the level-difference between »low« and »medium« districts is not $b_1 = -2.075$, but $b_1 + b_4(\text{PROT}) = -3.835$, given that PROT=10. The coefficients b_1 and b_2 represent differences in the level of SPD voting »caused« by differences in the level of urbanization only under the very specific condition that PROT= 0. In our example these level differences increase as PROT increases. In the same vain, in model (7-2) the standard errors and t-values of PROT and the urbanization dummies are **conditional.** So, the very low t-value (-.799) of b_1 is not a sufficient reason to drop this term from the model. It only tells us that there is no statistically significant difference between »low« and »medium« urbanization in its impact upon SPD voting as long as the proportion of protestants is zero or very low.

The statistical significance of the overall slope differences (i.e., interaction between the categorical and the metric variable) can be evaluated by

way of the F-change test described in section 4 (see equation (4-1)).[21] The extended model which includes the interaction components is now given by equation (7-2), the reduced model which excludes the interaction components is given by equation (7-1). Using the notation of equation (4-1), we have

$$(7-6) \qquad F_{change} = \frac{(R^2_{(2)} - R^2_{(1)})/q}{(1 - R^2_{(2)})/(n-(K+1)-q)} = \frac{R^2_{change}(n-(K+1)-q)}{(1 - R^2_{(2)})q}$$

$$= \frac{(.7049 - .6674)(395-(3+1)-2)}{(1-.7049)2} = 24.7$$

The value of $F_{2,389} = 24.7$ is significant at $\alpha < .00005$, thereby confirming the results of section 4. If the null-hypothesis of no interaction had not been rejected, one could test if the intercept in the relationship between PROZSPD and PROT differs depending on the level of urbanization. The extended model would then be equation (7-1), the reduced model would be PROZSPD = a + b(PROT) + e. The test for differential intercepts always presupposes homogeneous slopes; it is not valid when the categorical and the metric variable interact (see Johnston 1972, S. 193 ff.). This way of testing the statistical significance of differential slopes and intercepts in models which involve one or more categorical and one or more metric regressor variables is commonly called »covariance analysis« (Johnston 1972: 194; Wonnacott/Wonnacott 1970: 77).

Instead of estimating and evaluating models such as given with equation (7-2), sociologists and historians often unnecessarily split their sample into the classes defined by the categorical variable.[22] In our example they would estimate model

$$(7-7) \qquad PROZSPD = \alpha + \beta(PROT) + u$$

within three sub-samples, the first of which containing those election districts with »low« urbanization only, the second those with »medium« and the third those with »high« urbanization. For the sake of argument, we will also follow this procedure and compare its results (given in Table 8) with those given for model (7-2) in Table 7.

Table 8: Coefficients and t-values for model (7-4) as estimated within 3 sub-samples

	Level of urbanization		
	low	medium	high
Intercept a	5.459	7.534	16.262
b(PROT)	.096	.273	.373
t_a	3.575	3.845	8.967
t_b	3.909	10.483	14.786
R^2	.167	.444	.555

The slope and intercept estimates in Table 8 can exactly be reproduced (save rounding errors) by those in Table 7. Starring the coefficients of Table 8 (with the specific sub-sample indicated in parenthesis) and retaining the notation of Table 7 for model (7-2), we have

$$
\begin{aligned}
(7\text{-}8) \quad a^*(low) &= (a + b_1) = (7.534 - 2.075) = 5.459 \\
a^*(medium) &= a = \cdot 7.534 \\
a^*(high) &= (a + b_2) = (7.534 + 8.728) = 16.262 \\
b^*(low) &= (b_3 + b_4) = (.273 - .176) = .097 \\
b^*(medium) &= b_3 = .273 \\
b^*(high) &= (b_3 + b_5) = (.273 + .101) = .374
\end{aligned}
$$

By splitting the sample we lower the degrees of freedom (i.e., the power of statistical tests), we cannot calculate partial and multiple correlation coefficients for urbanization plus confession, and, even worse, we forsake the opportunity of testing the significance of differential slopes. However, we must keep in mind that model (7-2), just as any other regression model estimated by ordinary least squares, assumes homogeneous error variance. If we have reason to believe[23] that this assumption is untenable, then we must estimate the slope coefficients for each category separately in order to obtain group-specific error estimates (Pindyck/Rubinfeld 1981: 114).

8. Concluding Remarks

There are several issues related to interactive models in regression analysis which have not been dealt with in this article. A fuller treatment would include, for example, a more detailed discussion of various **substantive** models of interaction and the problem of translating them into regression equations (see Southwood 1978). Another point would concern the difference between non-linear and interactive relationships (see again Southwood 1978: 1169 -1174). One would also need to clarify how »interaction« formulated within logistic regression models differs from »interaction«

evaluated in linear regression analysis (see Hanushek & Jackson 1977, Chpt. 7; Jagodzinski & Kühnel & Terwey 1989). Even with these topics left uncovered, it is hoped that the present article helps to clarify the most basic concepts and some of the pre-suppositions related to the analysis of interaction within the framework of linear regression models.

Bibliography

Allison, D., Testing for interaction in multiple regression, in: American Journal of Sociology 83 (1977), p. 144 - 153.

Belsley, David A../Kuh, Edwin/Welsch, Roy E., Regression diagnostics: Identifying influential data and sources of collinearity, New York etc.: Wiley, 1980.

Berry, William D./Feldman, Stanley, Multiple regression in practice, Beverly Hills, London, New Delhi 1985 (Sage University Paper Series on Quantitative Applications in the Social Sciences, No. 50).

Best, Heinrich, Die Männer von Bildung und Besitz. Struktur und Handeln parlamentarischer Führungsgruppen in Deutschland und Frankreich 1848/49, Düsseldorf: Droste, 1990.

Draper, N. R./Smith, H., Applied regression analysis, New York etc.: Wiley, 1981 (2nd edition).

Friedrich, Robert J., In defense of multiplicative terms in multiple regression equations, in: American Journal of Political Science, Vol. 26 (1982), p. 797 - 833.

Jaccard, James/Turrisi, Robert/Wan, Choi K., Interaction effects in multiple regression, Beverly Hills, London, New Delhi 1990 (Sage University Paper Series on Quantiatative Applications in the Social Sciences, No. 72).

Johnston, J., Econometric methods, New York etc.: McGraw-Hill, 1972 (2nd edition).

Kenny, David A., Correlation and causality, New York etc. 1979.

Kerlinger, Fred N./Pedhazur, E. J., Multiple regression in behavioral research, New York etc.: Holt, Rinehart & Wilson, 1973.

Kmenta, Jan, Elements of econometrics, New York: Macmillan, 1986.

Linder, Arthur/Berchtold, Willi, Statistische Auswertung von Prozentzahlen. Probit- und Logitanalyse mit EDV, Basel and Stuttgart: Birkhäuser, 1976.

Miller, Michael K./Farmer, Frank L., Substantive nonadditivity in social science research, in: Quality & Quantity, Vol. 22 (1988), p. 221 - 237.

Pindyck, Robert S./Rubinfeld, Daniel L., Econometric models and econometric forecasts, New York etc.: McGraw-Hill, 1981 (2nd edition).

Ragin, Charles C, The comparative method. Moving beyond qualitative and quantitative strategies. Berkeley etc.: Univ. of Calif. Press, 1987.

Rao, Potluri/Miller, Roger LeRoy, Applied Econometrics, Belmont: Wadsworth, 1971.

Rochel, Hubertus, Planung und Auswertung von Untersuchungen im Rahmen des allgemeinen linearen Modells, Berlin etc.: Springer, 1983.

Schröder, Wilhelm Heinz, Sozialdemokratische Reichstagsabgeordnete und Reich stags wählen von 1898 bis 1912, Beschreibung des Datensatzes (mimeo), Köln 1988.

Southwood, Kenneth E., Substantive theory and statistical interaction: Five models, in: American Journal of Sociology, Vol. 83 (1978), p. 1154 - 1203.

Thome, Helmut, Grundkurs Statistik für Historiker, Teil II: Induktive Statistik und Regressionsanalyse, Historical Social Research/Historische Sozialforschung, Supplement No. 3, Köln 1990.

Tufte, Edward R., Data analysis for politics and policy, Englewood Cliffs, N. J.: Prentice-Hall, 1974.

Wonnacott, Ronald J./Wonnacott, Thomas H., Econometrics, New York etc.: Wiley, 1970.

Notes

(1) For a review of these assumptions see any textbook in statistics or econometrics, e. g., Kmenta (1986).

(2) For an introduction see, e.g. Berry/ Feldman (1985), Tbfte (1974), Thome (1990).

(2a) An alternative approach based on Boolean methods has been proposed by Ch. C. Ragin (1987).

(3) I am grateful to Prof. Best who made these data available to me. For further reference see Best (1990).

(4) Within the context of log-linear models 1st order interactions are already defined for bivariate relationships.

(5) I am grateful to my colleague for making the data available. For documentation of the data set, see Schröder (1988).

(6) The validity of the assumption of linearity and constant error variances is generally doubtful when dealing with proportions or percentage values. Further problems arise from using aggregate data. We will gloss over these difficulties here, but they have been discussed at some length in Thome (1990).

(7) In this argument we make use of the assumption that the individual-level relationship has the same direction as the aggregate-level relationship; an assumption which is not generally true, but can be justified in this case.

(8) An element of asymmetry, however, may be introduced by reducing equation (3-3) to something like:

(a) $Y = a + \beta_1 X_1 + \beta_2(X_1 X_2) + u$

This model assumes that X_2 has no effect upon Y if $X_1 = 0$, even though X_1 has an effect on Y even in the case where $X_2 = 0$. But it holds for both regressors that the effect of one of them, its magnitude, depends upon the value of the other one. Obviously this model is interpretable only when the regressors have been measured on a ratio scale, for which a natural zero point is defined. Model (a) may be further reduced to

(b) $Y = a + \beta(X_1 X_2) + u$

where none of the regressor variables is assumed to have an impact on Y as long as the other regressor takes on a value of zero.

(9) If zero-values in the regressor variables have not been observed, i.e., if these conditions have not actually been realized, this interpretation may cause problems, since we would generalize »beyond the observed range of experience« (Friedrich 1982, p. 806).

(10) Note that the F-change value is not identical with the difference of the F-values calculated separately for the purely additive (the reduced) model and the extendend model including the multiplicative term.

(11) One must, however, recognize that multiple t-tests invoke the problem of inflated type I error rates. To overcome this difficulty some researchers divide the alpha-level (error risk) aimed at, say a = 0.05, by dividing this value by the number of tests performed, thereby making rejection of the null-hypothesis more difficult (for references see Jaccard/Tbrrisi/Wan 1990: 28). The problem is of no concern in the present context.

(12) See, e. g., Kenny (1979: 17 ff.).

(13) For formulae see any textbook on econometrics, e. g., Pindyck/Rubinfeld 1981, p. 78, 99 ff.

(14) The factors determining the sign and the magnitude of this correlation are quite complex. For a brief discussion and further reference see Friedrich (1982: 811, fn. 7).

(15) This points to the possibility of rescaling the independent variables as a technique of dealing with multicollineartiy. See section 6 below.

(16) As Friedrich notes (and further elaborates) this corresponds to the increase in prediction error also in purely additive models as one moves away from the center of the regression survace. »In either case, that is, descriptions of the relationship between the dependent variable and one independent variable at a particular level of another independent variable become less certain with increasing distance from

the center of the regression surface. The only difference is that in the additive model this uncertainty is 'hidden' in the often-ignored standard errors for the intercept and the slope of the other independent variable, while in the interactive model it emerges explicitly in the standard error for the slope of the independent variable of interest as well« (Friedrich 1982: 817).

(17) Further support is given by a more recent article by Miller & Farmer (1988).

(18) Adding a constant to a variable X_1 does not change the variance of X_1, but it does change the variance of the product term (X_1, X_2), and consequently its beta-coefficient and its correlation (Pearson's r) with other variables.

(19) In passing it should be noted that the standardized (or »beta«-) coefficients estimate the relationship between the dependent variable and the independent variables at different places than the unstandardized coefficients, namely at the means of X_1 and X_2, not at their respective zero points. When the researcher wants to use standardized coefficients in interactive models, he or she should first transform Y, X_1 and X_2 into z-scores and then compute the product term $Zx(1)^\wedge X(2)^*$ rather than transforming the product term X_1X_2 into z-scores (an explanation for this is given by Friedrich 1982: 824).

(20) Actually, there are different possibilities for coding categorical variables, only one will be considered here (see Kerlinger/Pedhazur (1973); Rochel (1983); Pindyck/Rubinfeld (1981: 135 ff.).

(21) For further details of the testing procedure, see Johnston (1972: 192 ff.).

(22) Occasionally sample splitting has been (mistakingly, I think) recommended as a way of overcomming multicollinearity problems in interactive models.

(23) For formal testing procedures, see Pindyck/Rubinfeld (1981: 114, 144 ff.) As Rao and Miller (1972: 90 f.) warn: »The researcher should keep in mind that the computed standard errors do not necessarily reflect the true efficiency, or the appropriateness, of assumptions regarding the variance of the error terms...He should also keep in mind that he cannot use the same data for testing a nullhypothesis on equivalence of the variances in the two categories [in the example which they use there are only two categories, H. T.] while also estimating the regressions under the assumption that the null hypothesis is true.«