UNIVERSITY OF APPLIED SCIENCES

THM
TECHNISCHE HOCHSCHULE MITTELHESSEN

Kai Bruchlos

Two Variable Linear Regression Model with Errors in Variables

THM-Hochschulschriften Band 1E

Kai Bruchlos

Two Variable Linear Regression Model with Errors in Variables

THM-Hochschulschriften Band 1E

THM-Hochschulschriften Band 1E

Technische Hochschule Mittelhessen
Fachbereich Mathematik, Naturwissenschaften, Datenverarbeitung

Herausgeber der THM-Hochschulschriften:
Der Präsident der Technischen Hochschule Mittelhessen

Die Hochschulschriften sind online abrufbar:
www.thm.de/bibliothek/thm-hochschulschriften

ISSN (Print) 2568-0846
ISSN (Online) 2568-3020

Preface

The aim of this paper is to specify the assumptions and statements of various models of linear regression analysis, show the relationships between the models and illustrate the possible applications.

The actual task of regression analysis is an optimisation task: From a set of functions, find the one that „best“ matches a given set of points P_i. — We will limit ourselves here to the simplest case: Determine the straight line „closest“ to the points $(x_1, y_1), \ldots, (x_n, y_n)$ lies. — It is astonishing that I am not aware of any source that mentions this optimisation task in connection with the regression analysis. The formulation of the task as an optimisation problem is so important because it is the starting point for the mathematical development of the different model variants. So we start with the optimisation problem in Chapters 1 and 2 and, in contrast to inductive statistics, we deliberately refer to this section as regression calculation as a subfield of descriptive statistics.

It is equally astonishing that reference is very rarely made to a central prerequisite for the models of simple regression analysis: The observed values $x_1, \ldots, x_n$ of the independent variable are not only fixed, but must be the same for each repetition of the sampling.[1] As far as I know, only Fuller 1987 and Gujarati 1988 write this so clearly in textbooks.[2] This requirement is usually fulfilled for time series, but not for other features, especially economic ones. The observed values $y_1, \ldots, y_n$ of the dependent variable, on the other hand, can change when the sample is repeated and are subject to random fluctuations.

It is also noteworthy that the basic mathematical model, which enables the transition from descriptive to inductive statistics, is only presented in one recent source known to me, namely in Fisz 1989, p. 91 et sq.[3] Other authors at least implicitly use the basic mathematical model, but speak at best of conditional expected values.[4] In most cases, only the properties of the

[1] The violation of this requirement has fundamental consequences: Compare Gujarati 1988, p. 417.

[2] Fuller 1987, p. 1.Gujarati 1988 chooses the following formulation on p. 19: *The explanatory variables, on the other hand, are assumed to have fixed values (in repeated sampling), …*

[3] An older source is Cramér 1946, p. 270 et sq.

[4] Cp. Miller 1986, p. 222.

distributions that sum up to the model are listed.[5] Although the different models and model variants can be represented very well with the help of the basic model.

Furthermore, it is puzzling that during the development phase of the measurement error model in the 1940s to 1960s, the work of Kummell 1879 was not taken into account accordingly, although Deming 1948 cites his result.[6] The problems with the functional model, which were only solved in 1969, would not have happened in this way.

The measurement error model (MEM) treats the two features (variables) X and Y equally: The feature values of both features are measured with random deviations. This gives the MEM an advantage: It is possible to estimate the regression line symmetrically, i.e. it does not matter whether x is the independent variable and y the dependent variable or vice versa.[7] Here, too, it is astonishing that this property is very rarely mentioned in the literature. This also applies to the independence with regard to scaling (multiples of units).

If the quotient of the error variances σ_δ^2 and σ_ε^2 is known, then the functional model is by far to be favoured (Lemma 2.2.3, Theorem 2.2.20, Theorem 3.0.4):

- The model is the generalisation of simple regression.
- The model is symmetrical.
- The estimators are consistent.
- The model is independent in terms of scaling.

The restriction here to the two variable, i.e. the observation of only two features, is of no particular significance. As econometrics shows, the results can be generalised relatively easily into multidimensionality. However, the formalism becomes more complex.

As a rule, the observed values $x_1, \ldots, x_n$ are all different, just like the observed values $y_1, \ldots, y_n$. But this does not have to be the case. Whether all observed values $x_1, \ldots, x_n$ or $y_1, \ldots, y_n$ may be the same depends on the model. In simple regression analysis, at least two of the $x_1, \ldots, x_n$ must be different. Otherwise the Theorem 1.1.1 does not apply.[8]

[5] Cp. Fuller 1987, p. 30 et sqq.; Kendall and Stuart 1979, p. 403 et sqq.; Schach and Schäfer 1978, p. 155 et sqq.

[6] Deming 1948, p. 184. – On p. 181, Madansky 1959 only refers to Deming 1948 in order to prove the different names of a procedure. The result of Kummel mention Madansky 1959, p.202 and Lindley 1947, p. 241. However, the references to the results of Kummell 1879 do not seem to have had any consequences.

[7] Cp. Schach and Schäfer 1978, p. 159 et sq., footnote +).

[8] Note Georgii 2004, p. 317, example 12.1.

I would like to thank my colleague Prof Dr Manfred Börgens for his many tips and comments, especially on independence with regard to scaling.

Brietlingen, Germany September 2017 Kai Bruchlos

Contents

1 Simple Regression Analysis

In this chapter, the basic properties of simple regression are provided, which are referred to in the extension to the measurement error model . The starting point of the regression analysis, the optimisation problem and a possible solution are formulated in section 1.1. In section 1.2, we switch to inductive statistics with the presentation of the basic model of probability theory in order to then obtain statements on the quality of LS estimators in section 1.3.

1.1 Regression Calculation

The regression calculation is a subfield of descriptive statistics and assumes that observed values $x_1, \dots, x_n$ and $y_1, \dots, y_n$ of two continuously scaled features X and Y exist, whereby there should be a functional relation of the type $f(x) = y$. Probability theory is not used.

Here we only consider the linear relation between two features X and Y, i.e. the straight line[1]

$$y = \alpha + \beta \cdot x \ .$$

If $n > 2$, i.e. if more than two points (x_i, y_i) are given, then the linear system of equations

$$\begin{aligned} y_1 &= \alpha + \beta \cdot x_1 \\ &\vdots \\ y_n &= \alpha + \beta \cdot x_n \end{aligned}$$

is overdetermined. Normally, this system of equations cannot be solved exactly. A linear equation

$$y = a + b \cdot x \ ,$$

the so-called **regression function**, here **regression line**, must therefore be determined, to which the points (x_i, y_i) have the smallest possible „distance“. This is an optimisation task. The **method of least squares**[2] is usually

[1] With regard to the notation of the linear regression model mentioned below, this should actually read $\eta = \alpha + \beta \cdot x$. However, this notation is not common in this context.

[2] Cp. Pestman 1998, p. 183 et sq.

used to solve this, which minimises the sum of the squares of the individual deviations

$$e_i := y_i - (a + b \cdot x_i) \ .$$

We are therefore looking for the minimum of the function

$$S : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, \ (a, b) \mapsto \sum_{i=1}^{n} (y_i - a - b \cdot x_i)^2 \ .$$

Differential calculus provides the local minimum with the coordinates

$$b = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad \text{and} \qquad a = \bar{y} - b \cdot \bar{x} \ .$$

a and b are called **least squares estimators**, in short **LS estimators**. In total we have:

Theorem 1.1.1:[3] *Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be observed values of the features X and Y in statistical units of a population. If the relation $y = \alpha + \beta \cdot x, \ \alpha \in \mathbb{R}, \ \beta \in \mathbb{R} \backslash \{0\}$ is assumed, then*

$$b := \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad \textit{and} \qquad a := \bar{y} - b \cdot \bar{x}$$

are the LS estimators for β and α.

Corollary 1.1.2: *The point $(\bar{x}, \bar{y})$ lies on the regression function $y = a + bx$.*

Remark 1.1.3: (i) The method of least squares is based on two assumptions due to the minimisation of deviations

$$e_i := y_i - (a + b \cdot x_i) \iff y_i = a + b \cdot x_i + e_i$$

- $x_1, \ldots, x_n$ are fixed values and would not change if the sampling was repeated.

- $y_1, \ldots, y_n$ can deviate from the values $a + b \cdot x_1, \ldots, a + b \cdot x_n$, where $(x_1, y_1), \ldots, (x_n, y_n)$ belong to the same sample.[4]

[3] Pestman 1998, p. 185, Proposition IV.1.1.

[4] From a different sample $(x_1, \tilde{y}_1), \ldots, (x_n, \tilde{y}_n)$ other LS estimators than a and b are usually calculated, such as $\tilde{a}$ and $\tilde{b}$. And then, of course, $\tilde{y}_1, \ldots, \tilde{y}_n$ deviate from the values $\tilde{a} + \tilde{b} \cdot x_1, \ldots, \tilde{a} + \tilde{b} \cdot x_n$.

This can be interpreted to mean that x_i is measured without error, y_i with error.

(ii) The notation $b = \hat{\beta}$ and $a = \hat{\alpha}$ is often used at this point. This notation is deliberately avoided here, as we are not yet in the inductive statistics.

(iii) The aim of the regression calculation is to adjust the parameters of a given function type $f(x) = y$ „as well as possible“ to the points $(x_1, y_1), \ldots, (x_n, y_1)$. „As well as possible“ is usually understood to mean the determination of a minimum with the help of differential calculus. Especially in the two variable linear regression calculation, only the minimum of S is determined when using the least squares method. How well this minimum solves the optimisation task is not yet known. Or in the language of inductive statistics: The properties of the LS estimators and their quality are still open. For this we need a probabilistic model for regression.

So far, we have assumed that x is the independent variable[5] and y is the dependent variable[6]. However, the reverse can also be the case. The following notations are common:

Definition 1.1.4: Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be observed values of the features X and Y in statistical units of a population.

(i) If the functional relation $f(x) = y$ is assumed, then there is a **regression of** y **on** x.

(ii) If the functional relation $f(y) = x$ is assumed, then there is a **regression of** x **on** y.

There are cases in which the dependency of the features X and Y is factually established, i.e. it is factually justified whether there is a regression of y on x or a regression of x on y. But there are also cases in which the dependency cannot be objectively justified or even makes no sense. For these cases, it would be advantageous if the regression function $x = a' + b'y$ (in the y-x coordinate system) corresponded to the regression function $y = a + bx$ (in the x-y coordinate system). The following now applies:

$$y = a + bx \iff x = -\frac{a}{b} + \frac{1}{b}y$$

In particular, $b' = 1/b$ must therefore be true. Looking at the two estimators

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \text{and} \qquad b' = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2},$$

it is clear that $b' = 1/b$ does not apply in general.[7] To summarise, we have seen:

[5] Other names: Explanatory variable, regressor, predictor, exogenous variable

[6] Other names: Explained variable, regressand, predictand, endogenous variable

[7] Cp. Schach and Schäfer 1978, p. 159, footnote +.

Remark 1.1.5: (i) In simple linear regression (with LS estimators), a regression of y on x generally leads to a different regression line than a regression of x on y. Simple linear regression is **asymmetrical**.[8]

(ii) The asymmetry of simple linear regression has its origin in the fact that the KQ method uses independent and dependent variables differently: The independent variable is measured without error, the dependent variable is not – see remark 1.1.3.

Description of the model of the linear regression calculation:

$x_1, \ldots, x_n$	fixed values of the feature X, do not change when the sampling is repeated
$y_1, \ldots, y_n$	values of the feature Y
$y_i = a + b \cdot x_i + e_i$	regression equations
$y = a + b \cdot x$	regression function (regression line)
$\eta = \alpha + \beta \cdot x$	„true“ relation

1.2 Basic Model of Probability Theory

In order to be able to say something about the quality of the LS estimators from Theorem 1.1.1, a probabilistic model is to be used. The only probabilistic model of regression analysis known to me is by Fisz 1989 and is presented on pages 91 et sqq:

Let (X, Y) be a two-dimensional random variable with continuous density $f(x, y)$ and continuous marginal densities f_X and f_Y.

Theorem 1.2.1:[9] *The following applies to the conditional expected values:*

$$\mathbb{E}(X|Y=y) = \int_{-\infty}^{\infty} x \cdot \frac{f(x,y)}{f_Y(y)}\, dx \quad \textit{as well as} \quad \mathbb{E}(Y|X=x) = \int_{-\infty}^{\infty} y \cdot \frac{f(x,y)}{f_X(x)}\, dy$$

If X and Y are stochastically independent, then $\mathbb{E}(X|Y=y) = \mathbb{E}(X)$ and $\mathbb{E}(Y|X=x) = \mathbb{E}(Y)$.

Definition 1.2.2:[10] The set of points in the x-y plane with the coordinates

$$\big(\mathbb{E}(X|Y=y), y\big)$$

is called **regression curve of the random variable X with respect to Y**, the set of points with the coordinates

$$\big(x, \mathbb{E}(Y|X=x)\big)$$

[8] Cp. Miller 1986, p. 169; Schach and Schäfer 1978, p. 159, footnote +; Madansky 1959, p. 175.

[9] Fisz 1989, p. 91, (3.7.1'), p. 92.

[10] Fisz 1989, p. 92, Definition 3.7.1.

is called **regression curve of the random variable Y with respect to X**.

Remark 1.2.3: In general, the two regression curves do not overlap.[11] The regression is therefore fundamentally asymmetrical, see Remark 1.1.5.

The regression curve of the random variable X with respect to Y or the regression curve of the random variable Y with respect to X corresponds to the regression of x on y or the regression of y on x in the regression calculation. We will establish the connection between the two terms in the next section.

1.3 Properties of the Estimators

In the following, we only consider the regression curve of the random variable Y with respect to X and establish the connection to the regression of y on x with the

Assumption 1 $$\mathbb{E}(Y|X=x) = \alpha + \beta \cdot x,\ \alpha \in \mathbb{R},\ \beta \in \mathbb{R}\backslash\{0\}\ .$$

The approach for the regression curve of the random variable X with respect to Y is analogue and the results are the same.

Let $x_1, \ldots, x_n$ be fixed realisations of X, which do not change even if the sampling is repeated, and let $Y_{x_1}, \ldots, Y_{x_n}$ be random variables whose distribution is the conditional distribution of Y under $X = x_i$. The overdetermined linear system of equations

$$\begin{aligned} \mathbb{E}(Y_{x_1}) &= \alpha + \beta \cdot x_1 \\ &\vdots \\ \mathbb{E}(Y_{x_n}) &= \alpha + \beta \cdot x_n \end{aligned}$$

thus exists.

Theorem 1.3.1:[12] *Let $y_1, \ldots, y_n$ be realisations of $Y_{x_1}, \ldots, Y_{x_n}$. The LS estimators a and b for α and β are unbiased and a linear combinations of $Y_{x_1}, \ldots, Y_{x_n}$, so-called* ***linear*** *estimators in $Y_{x_1}, \ldots, Y_{x_n}$.*

Remark 1.3.2: (i) An estimator is a random variable and a as well as b are not. However, a and b are realisations of the corresponding estimators. Since the formalism for defining the corresponding estimators is of secondary importance here, we take the realisation of an estimator as the estimator.

(ii) If the realisations $x_1, \ldots, x_n$ can change when the sampling is repeated, for example if they cannot be measured without error, then it is difficult to make a statement about the unbiasedness of the LS estimators a and b, since

[11] Cp. Fisz 1989, p. 92.
[12] Pestman 1998, p. 187, Proposition IV.1.2.

$\mathbb{E}(b)$ cannot be calculated as in Theorem 1.3.1.[13] With the consistency of the LS estimators, it is the other way round: The denominator of b is n times the sample variance. This converges in probability against the variance[14], which here has the value 0, since $x_1, \ldots, x_n$ are fixed values and not realisations of sample variables $X_1, \ldots, X_n$.

For further statements on the LS estimators, we need the following assumptions:

<u>Assumption 2</u> $Y_{x_1}, \ldots, Y_{x_n}$ are stochastically independent.

<u>Assumption 3</u> It applies $\mathbb{V}(Y_{x_1}) = \cdots = \mathbb{V}(Y_{x_n}) =: \sigma^2$.

Remark 1.3.3: (i) From assumption 2, the modelling could also be carried out using the errors e_i: The assumptions are formulated for the random variables $E_{x_i} := Y_{x_i} - (\alpha + \beta \cdot x_i)$.

(ii) Assumption 3 means that the standard deviation in the data measurement is always the same. This property is called **homoscedasticity** in econometrics.[15]

Theorem 1.3.4:[16] *The following applies under assumptions 1 to 3:*

$$\mathbb{V}(b) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$\mathbb{V}(a) = \frac{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2 + n\cdot(\bar{x})^2\right)\cdot\sigma^2}{n\cdot\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$Cov(a,b) = -\frac{\bar{x}\cdot\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

Theorem 1.3.5:[17] **by Gauß-Markov** *With assumptions 1 to 3, the LS estimators a and b have the lowest variance among the unbiased linear estimators and are the only ones with this property.*

Remark 1.3.6: The LS estimators a and b are also called **BLUE estimators** (best linear unbiased estimator).[18]

<u>Assumption 4</u> $Y_{x_i} \sim N(\alpha + \beta \cdot x_i, \sigma^2)$ applies for all x_i .

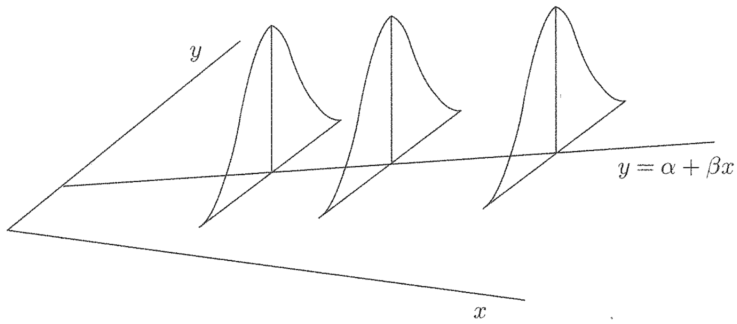
[13] Gujarati 1988 claims on p. 417 that the KQ estimators are not unbiased.
[14] Fisz 1989, p. 367.
[15] Cp. Gujarati 1988, p. 316 et sqq.
[16] Pestman 1998, p. 187, Proposition IV.1.2.
[17] Pestman 1998, p. 189, Proposition IV.1.3. Cp. Gujarati 1988, p. 63.
[18] Cp. Gujarati 1988, p. 63.

Remark 1.3.7: (i) With assumption 4, assumptions 1 and 3 are irrelevant and can therefore be omitted.

(ii) If assumption 4 applies,we are talking about **normal regression analysis**.[19]

Theorem 1.3.8:[20] *With assumptions 1 to 4, the LS estimators a and b are also the ML estimators for α and β.*

The following diagram[21] is intended to give an impression of what normal regression analysis means. The distribution of Y_{x_i} essentially only differs in the size of the expected value. And the graph shows how much the realisations y_i can deviate from the expected value $\mathbb{E}(Y_{x_i})$, vary around it:[22]



Before we can specify the description of the model, we need to define the deviations, the errors for $i = 1, \ldots, n$:

$$E_{x_i} := Y_{x_i} - (\alpha + \beta \cdot x_i)$$

It is $E_{x_i} \sim N(0, \sigma^2)$ and $E_{x_1}, \ldots, E_{x_n}$ are stochastically independent.

Description of the normal linear regression model:

(X, Y)	two-dimensional random variable
$x_1, \ldots, x_n$	$\star$)
$y_1, \ldots, y_n$	$\star\star$)
$Y_{x_i} = \alpha + \beta \cdot x_i + E_{x_i}$	regression equations
$y = a + b \cdot x$	regression function
$\eta = \mathbb{E}(Y_x) = \mathbb{E}(Y \mid X = x) = \alpha + \beta \cdot x$	„true“ relation
$Y_{x_1}, \ldots, Y_{x_n}$	are stochastically independent
$Y_{x_i} \sim N(\alpha + \beta \cdot x_i, \sigma^2)$	
$E_{x_i} \sim N(0, \sigma^2)$	

[19] Cp. Pestman 1998, p. 194.

[20] Pestman 1998, p. 198, Proposition IV.3.4.

[21] The graphic is taken from Fahrmeir [u.a.] 2011, p. 478. Cp. Gujarati 1988, p. 55.

[22] Instead of $y = \alpha + \beta \cdot x$ it must be $\eta = \alpha + \beta \cdot x$.

$\star$): Fixed realisations of the random variable X do not change when sampling is repeated.
$\star\star$): Realisations of the random variables $Y_{x_1}, \ldots, Y_{x_n}$

2 Measurement Error Model

The starting point of the simple regression analysis are observed values $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ of two features X and Y, for which there should be a functional relationship of the type $f(x) = y$. If we assume the relationship $y = \alpha + \beta \cdot x$, then the resulting optimisation problem can be solved using the least squares method. The method of least squares assumes that $x_1, \ldots, x_n$ are very specific fixed values, in particular measured without error, whereas $y_1, \ldots, y_n$ can change when the sampling is repeated, and are therefore subject to error – see Remark 1.1.3 (i).

If X is an economic feature, then $x_1, \ldots, x_n$ are usually measured with errors in the same way as $y_1, \ldots, y_n$. And this applies not only to economic features, but to many others — most? — too.[1] The model of the simple regression analysis must therefore be changed with regard to the non-existent errors of $x_1, \ldots, x_n$ so that the optimisation problem can also be solved with the least squares method for $x_1, \ldots, x_n$ with errors. The model in which both $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ can change when sampling is repeated is called **measurement error model**, or **MEM** for short.

The aim of this chapter is to show the consequences of changing the model assumption with regard to $x_1, \ldots, x_n$. In section 2.1, we start again with the regression calculation.

Ideally, the measurement error model is a generalisation of the simple regression model or, in other words, the simple regression model is a special case of the measurement error model. It will be shown in section 2.2 that a variant of the measurement error model, the so-called functional model, is the generalisation of the simple regression model (Lemma 2.2.3).

2.1 Regression Calculation

We start with observed values $\xi_1, \ldots, \xi_n$ and $y_1, \ldots, y_n$ of two continuously scaled features X and Y as in the usual regression calculation. All observations can be measured with errors, i.e. there are „random errors“[2] $\delta_1, \ldots, \delta_n$,

[1] Cp. Miller 1986, p. 221.

[2] A distinction is often made between improper errors, systematic errors and random errors as error types. A improper error occurs when there is a lack of care or concentration when measuring. The systematic error includes environmental influences and device

$\varepsilon_1, \ldots, \varepsilon_n$ and the error-free values $x_1, \ldots, x_n, \eta_1, \ldots, \eta_n$, which cannot be observed, so that

$$\xi_1 = x_1 + \delta_1, \ldots, \xi_n = x_n + \delta_n \quad \text{as well as} \quad y_1 = \eta_1 + \varepsilon_1, \ldots, y_n = \eta_n + \varepsilon_n$$

applies.[3]

Again, we consider the linear relationship

$$y = \alpha + \beta \cdot \xi \quad \text{respectively} \quad \eta = \alpha + \beta \cdot x \ .$$

If $n > 2$, then the system of linear equations

$$\begin{aligned} y_1 &= \alpha + \beta \cdot \xi_1 \\ &\vdots \\ y_n &= \alpha + \beta \cdot \xi_n \end{aligned}$$

is overdetermined. Normally, this system of equations cannot be solved exactly. The regression line

$$y = a + b \cdot \xi \ ,$$

must therefore be determined from which the points (ξ_i, y_i) have the smallest possible „distance“. As before, the method of least squares minimizes the sum of the squares of the individual deviations

$$e_i := y_i - (a + b \cdot \xi_i) \ .$$

At this point, there is a decisive change compared to the simple regression. If a and b are determined with the error-free values x_i and η_i, then $\eta_i = \alpha + \beta \cdot x_i = a + b \cdot x_i$, and therefore[4]

$$e_i = \eta_i + \varepsilon_i - a - b \cdot (x_i + \delta_i) = \eta_i - (a + b \cdot x_i) + (\varepsilon_i - b \cdot \delta_i) = \varepsilon_i - b \cdot \delta_i \ .$$

The least squares method is related to the slope b of the regression line that is to be calculated using the method. And this has unpleasant consequences for the properties of the LS estimators a and b. – For comparison, this is not the case with simple regression:

$$e_i = \eta_i + \varepsilon_i - a - b \cdot x_i = \eta_i - (a + b \cdot x_i) + \varepsilon_i = \varepsilon_i$$

Formally, we can proceed as with simple regression and obtain the calculation formulas for a and b according to theorem 1.1.1. With the LS estimator, the MEM is of course also asymmetric.

errors. The random error is considered unavoidable and arises from the imperfection of the measuring instruments and human perception.

[3] $x_1, \ldots, x_n$ are still the error-free values. This significantly increases the readability of the following notations.

[4] Cp. Schach and Schäfer 1978, p. 153.

Description of the model of the linear regression calculation:

$x_1, \ldots, x_n$	error-free values of the feature X
$\eta_1, \ldots, \eta_n$	error-free values of the feature Y
$\xi_i := x_i + \delta_i,\ i = 1, \ldots, n$	observed values of the feature X
$y_i := \eta_i + \varepsilon_i,\ i = 1, \ldots, n$	observed values of the feature Y
$y_i = a + b \cdot \xi_i + e_i$	regression equations
$y = a + b \cdot \xi$	regression function
$\eta = \alpha + \beta \cdot x$	„true“ relation

2.2 Probabilistic models

The change in the model assumption that x_i can no longer be measured without error has a significant impact on the estimation of the parameters a and b of the regression line. Initially, the LS estimator proves to be not consistent. Then, until 1969, there was only one very specific model — the structural model with additional information — in which a quality estimate, namely a consistent one, was possible at all.[5] This probably also explains why most textbooks still only present the structural model. Since the mid-1970s, there has also been a consistent estimator in the functional model with additional information.

We will now introduce the functional and the structural model and then first show in section 2.2.1 that the LS estimators are not consistent for both models. Then, in section 2.2.2, we look at the ML estimation in both models. Functional and structural model are based on the basic model from section 1.2 and combine the regression curve with the regression calculation.

Starting from the two-dimensional random variable (X, Y) of the basic model, let the conditional distribution of Y under $X = x$ be a normal distribution, more precisely

$$Y_{|X=x} \sim N(\alpha + \beta \cdot x, \sigma_\varepsilon^2)$$

with $\alpha \in \mathbb{R},\ \beta \in \mathbb{R} \backslash \{0\},\ \sigma_\varepsilon^2 > 0$. In particular,[6]

$$\mathbb{E}(Y_{|X=x}) = \mathbb{E}(Y|X = x) = \alpha + \beta \cdot x$$

and[7]

$$\mathbb{V}(Y_{|X=x}) = (1 - \varrho_{X,Y}^2) \cdot \sigma_Y^2 \ .$$

Here $\varrho_{X,Y}^2$ is the correlation coefficient of X and Y and σ_X^2 is the variance of X.

[5] Cp. Schach and Schäfer 1978, p. 165. See sentence 2.2.14.

[6] Pestman 1998, p. 30, Theorem I.6.2.

[7] Fisz 1989, p. 159, (5.11.5). Cp. Pestman 1998, p. 479, Exercise 17.

Let $x_1, \ldots, x_n$ be fixed realisations of X, which do not change even with repeated sampling, but cannot be observed and $Y_{x_1}, \ldots, Y_{x_n}$ are stochastically independent random variables whose distribution is the conditional distribution of Y under $X = x_i$, i.e.

$$Y_{x_i} \sim N(\alpha + \beta \cdot x_i, \sigma_\varepsilon^2) \ .$$

The errors of the dependent variable are

$$E_{x_i} := Y_{x_i} - (\alpha + \beta \cdot x_i)$$

for $i = 1, \ldots, n$. It applies $E_{x_i} \sim N(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon^2 = (1 - \varrho_{X,Y}^2) \cdot \sigma_Y^2$ and $E_{x_1}, \ldots, E_{x_n}$ are stochastically independent.[8] The errors of the independent variable are still missing: Let $D_{x_1}, \ldots, D_{x_n}$ be stochastically independent random variables defined on the measurable spaces of X with the following properties:

1. $D_{x_1}, \ldots, D_{x_n}, E_{x_1}, \ldots, E_{x_n}$ are stochastically independent.
2. $D_{x_i} \sim N(0, \sigma_\delta^2)$ with $\sigma_\delta^2 := (1 - \varrho_{X,Y}^2) \cdot \sigma_X^2$ for $i = 1, \ldots, n$.

The special choice of the variance of the distribution, σ_δ^2 guarantees the symmetry of the functional model.

The values of the independent variable observed with errors are now missing. Let $V_{x_i} := x_i + D_{x_i}$ for $i = 1, \ldots, n$. The following applies: $V_{x_i} \sim N(x_i, \sigma_\delta^2)$.[9] This provides the functional model:

Definition 2.2.1:[10] The model

(X, Y)	two-dimensional random variable
$x_1, \ldots, x_n$	error-free values of the random variable X
$y_1, \ldots, y_n$	observed realisations of $Y_{x_1}, \ldots, Y_{x_n}$
$V_{x_i} := x_i + D_{x_i}$	Random variable of observed values of X
$v_1, \ldots, v_n$	observed realisations of $V_{x_1}, \ldots, V_{x_n}$
$Y_{x_i} = \alpha + \beta \cdot x_i + E_{x_i}$	regression equations
$y = \hat{\alpha} + \hat{\beta} \cdot v$	regression function
$\eta = \mathbb{E}(Y \mid X = x) = \alpha + \beta \cdot x$	„true“ relation
$Y_{x_i} \sim N(\alpha + \beta \cdot x_i, \sigma_\varepsilon^2)$	
$E_{x_i} \sim N(0, \sigma_\varepsilon^2)$	
$D_{x_i} \sim N(0, \sigma_\delta^2)$	
$D_{x_1}, \ldots, D_{x_n}, E_{x_1}, \ldots, E_{x_n}$	stochastically independent

[8] Cp. Bruchlos 2015, p. 90, Lemma 1.
[9] Pestman 1998, p. 30, Theorem I.6.1 (i).
[10] Bruchlos 2015, p. 90 et sq., Lemma 2 (ii). Cp. Kendall and Stuart 1979, p. 407 et sq.; Schach and Schäfer 1978, p. 152 et sq.

is called **functional model**.

Remark 2.2.2: $v_1, \ldots, v_n$ corresponds to $\xi_1, \ldots, \xi_n$ in the linear regression model.

The normal regression analysis is a special case of the functional model:

Lemma 2.2.3:[11] *If we set $D_{x_i} \equiv 0$ in the functional model, then we have the normal regression model.*

For the structural model, let the random variables $X_1, \ldots, X_n$, defined on the measurable spaces of X, be normally distributed, more precisely[12]

$$X_i \sim N(x_i, \sigma_\delta^2/2),\ i = 1, \ldots, n\ .$$

Let the stochastic independent random variables $\tilde{D}_{x_1}, \ldots, \tilde{D}_{x_n}$ be the errors of the independent variable, defined on the measurable spaces of X, with the following properties:

1. $\tilde{D}_{x_1}, \ldots, \tilde{D}_{x_n}, E_{x_1}, \ldots, E_{x_n}$ are stochastically independent.
2. $\tilde{D}_{x_i}, X_i$ are for $i = 1, \ldots, n$ stochastically independent.[13]
3. $\tilde{D}_{x_i} \sim N(0, \sigma_\delta^2/2)$ for $i = 1, \ldots, n$.[14]

Now let $W_{x_i} := X_i + \tilde{D}_{x_i}$. Then $W_{x_i} \sim N(x_i, \sigma_\delta^2)$.[15]

[11] Bruchlos 2015, p. 90 et sq., Lemma 2 (iii).

[12] The special choice of the variance of the distribution, $\sigma_\delta^2/2$ guarantees the symmetry of the structural model.

[13] This is required so that W_{x_i} is correspondingly normally distributed. The requirement of stochastic independence corresponds to the non-systematic, random error. Cp. Fuller 1987, p. 3, (1.1.3).

[14] The special choice of the variance of the distribution, $\sigma_\delta^2/2$ guarantees the symmetry of the structural model.

[15] Pestman 1998, p. 34, Theorem I.6.6.

Definition 2.2.4:[16] The model

(X, Y)	two-dimensional random variable
$x_1, \ldots, x_n$	error-free values of the random variable X
$y_1, \ldots, y_n$	observed realisations of $Y_{x_1}, \ldots, Y_{x_n}$
$W_{x_i} := X_i + \tilde{D}_{x_i}$	Random variable of observed values of X
$w_1, \ldots, w_n$	observed realisations of $W_{x_1}, \ldots, W_{x_n}$
$Y_{x_i} = \alpha + \beta \cdot x_i + E_{x_i}$	regression equations
$y = \hat{\alpha} + \hat{\beta} \cdot w$	regression function
$\eta = \mathbb{E}(Y\|X = x) = \alpha + \beta \cdot x$	„true“ relation
$Y_{x_i} \sim N(\alpha + \beta \cdot x_i, \sigma_\varepsilon^2)$	
$X_i \sim N(x_i, \sigma_\delta^2/2)$	
$E_{x_i} \sim N(0, \sigma_\varepsilon^2)$	
$\tilde{D}_{x_i} \sim N(0, \sigma_\delta^2/2)$	
$\tilde{D}_{x_1}, \ldots, \tilde{D}_{x_n}, E_{x_1}, \ldots, E_{x_n}$	stochastically independent
$\tilde{D}_{x_1}, \ldots, \tilde{D}_{x_n}, X_1, \ldots, X_n$	stochastically independent

is called **ultrastructural model**.[17] If $X_i \sim N(\mu, \sigma_\delta^2/2),\ \mu \in \mathbb{R}$ applies for $i = 1, \ldots, n$, then we have the **structural model**.[18]

Remark 2.2.5: (i) The following mnemonic is useful for the names of the models: Structural like stochastic, functional like fixed.[19]

(ii) $w_1, \ldots, w_n$ corresponds to $\xi_1, \ldots, \xi_n$ in the linear regression calculation.

(iii) The description of the structural model changes in three places compared to the ultrastructural model:

$Y_{x_i} = \alpha + \beta \cdot \mu + E_{x_i}$	regression equations
$Y_{x_i} \sim N(\alpha + \beta \cdot \mu, \sigma_\varepsilon^2)$	
$X_i \sim N(\mu, \sigma_\delta^2/2)$	

(iv) The regression function in the functional and ultrastructural model, i.e. $y = \hat{\alpha} + \hat{\beta} \cdot v$ and $y = \hat{\alpha} + \hat{\beta} \cdot w$ differ from the regression function $y = a + b \cdot x$ in the simple case in three aspects: We do not yet have any estimators, i.e. the procedure is exactly the opposite: There are no estimators that have proven themselves in practice and are analysed with regard to their probabilistic

[16] Cp. Kendall and Stuart 1979, p. 400 et sq., p. 403; Schach and Schäfer 1978, p. 152 et sq., p. 155.

[17] Dolby 1976, p. 39 et sq. Cp. Kendall and Stuart 1979, p. 407.

[18] Cp. Kendall and Stuart 1979, p. 400 et sq., p. 403; Schach and Schäfer 1978, p. 152 et sq., p. 155; Fuller 1987, p. 1 et sqq., especially (1.1.1), (1.1.2) and (1.1.3).

[19] Cp. Fuller 1987, p. 2.

statements. The regression function now differs in both variables from the „true“ relation $\eta = \alpha + \beta \cdot x$ and the regression function has a different independent variable than the regression equations $Y_{x_i} = \alpha + \beta \cdot x_i + E_{x_i}$. This indicates the difficulty of the estimation.

(v) A common presentation of the MEM, here for the structural model:[20]

$$\begin{aligned} Y_{x_i} &= \alpha + \beta \cdot x_i + E_{x_i} & W_{x_i} &= X_i + \tilde{D}_{x_i} \\ X_i &\sim N(\mu, \sigma_x^2) & & \\ E_{x_i} &\sim N(0, \sigma_\varepsilon^2) & \tilde{D}_{x_i} &\sim N(0, \sigma_\delta^2) \\ \tilde{D}_{x_i}, E_{x_i}, X_i & & &\text{stochastically independent} \end{aligned}$$

The basic model is not mentioned, the random variables of the errors are not deduced and the regression function is not specified. In particular, this means that the correlations between the variances of the errors and the different variables of the regression function and the regression equations are not stand out. The description of the functional model is even more spartan.[21]

A major advantage of the two measurement error models is that the regression line can be estimated symmetrically, i.e. the regression line (regression function) for the regression of y on v or w is the same as for the regression of v or w on y.[22] This is due to the fact that both variables are not observed without error. With corresponding estimators, it is therefore no longer important whether v or w is the independent variable and y the dependent variable or vice versa. What property must the estimators fulfil for this?

Analogous to the linear relationship $\eta = \alpha + \beta \cdot x$ in the regression calculation, the following applies in the MEM

$$\eta = \alpha + \beta \cdot x$$

in the x-η coordinate system. For the symmetry,

$$x = -\frac{\alpha}{\beta} + \frac{1}{\beta} \cdot \eta$$

must be in the η-x coordinate system. If $\tilde{a}, \tilde{b}$ are the estimators for α, β and $\check{a}, \check{b}$ are the estimators for $-\frac{\alpha}{\beta}, \frac{1}{\beta}$, then the MEM is **symmetric** if and only if[23]

$$\check{b} = \frac{1}{\tilde{b}} \quad \text{and} \quad \check{a} = -\frac{\tilde{a}}{\tilde{b}}\ .$$

This property must be checked for the estimators.

[20] Fuller 1987, p. 13 or p. 30. Cp. Schach and Schäfer 1978, p. 152, p. 155; Kendall and Stuart 1979, p. 400, p.403.

[21] Cp. Schach and Schäfer 1978, p. 152, p. 163; Kendall and Stuart 1979, p. 407 et sq.

[22] Cp. Fuller 1987, p. 30.

[23] Cp. Schach and Schäfer 1978, p. 159 et sq., footnote +).

If $\tilde{a} := \bar{y} - \tilde{b} \cdot \bar{w}$, then because of

$$\check{a} = -\frac{\tilde{a}}{\tilde{b}} = -\frac{\bar{y} - \tilde{b} \cdot \bar{w}}{\tilde{b}} = \bar{w} - \frac{1}{\tilde{b}} \cdot \bar{y}$$

the following applies in particular:

Lemma 2.2.6: *If $\tilde{a} := \bar{y} - \tilde{b} \cdot \bar{w}$ respectively $\tilde{a} := \bar{y} - \tilde{b} \cdot \bar{v}$, then the MEM is symmetrical if and only if $\check{b} = \frac{1}{\tilde{b}}$.*

In contrast to the procedure described here, the functional and structural models are often initially defined without a distribution assumption in order to show that the LS estimators are not consistent without any distribution assumption.[24] For all further statements, however, the normal distribution assumption is necessary.

The aim of the next two sections is to determine the regression function

$$y = \hat{\alpha} + \hat{\beta} \cdot v \quad \text{respectively} \quad y = \hat{\alpha} + \hat{\beta} \cdot w \ ,$$

i.e. the estimation of α and β using the observed values $(v_1, y_1), \ldots, (v_n, y_n)$ in the functional model respectively $(w_1, y_1), \ldots, (w_n, y_n)$ in the ultrastructural model.

2.2.1 LS Estimator

For the estimation of α and β, the LS estimator of the simple regression calculation is initially suitable, see section 1.1. With regard to Remark 1.3.2, it is not unbiasedness that is considered, but the consistency:

Theorem 2.2.7:[25] *In the structural model, the LS estimator b converges for $n \to \infty$ P-almost surely to*

$$\beta \cdot \frac{\sigma_\delta^2/2}{\sigma_\delta^2/2 + \sigma_\delta^2/2} = \frac{\beta}{2} \ .$$

The LS estimators b and a are therefore not consistent estimators for β and α.

Remark 2.2.8: The description of convergence in Theorem 2.2.7 is more general in the literature:

$$\beta \cdot \frac{\sigma_x^2}{\sigma_x^2 + \sigma_d^2}$$

Here, $\sigma_x^2 := \mathbb{V}(X_i)$ and $\sigma_d^2 := \mathbb{V}(\tilde{D}_{x_i})$. The description in Theorem 2.2.7 is a consequence of the construction of W_{x_i}.

[24] Cp. Schach and Schäfer 1978, p. 155; Kendall and Stuart 1979, p. 402 et sq.;Miller 1986, p. 223 et sq.

[25] Madansky 1959, p. 177; Gujarati 1988, p. 418, (13.6.10); Schach and Schäfer 1978, p. 155; Miller 1986, p. 223; Fuller 1987, p. 3, (1.1.6). For the statement regarding the LS estimator a see Georgii 2004, p. 210, Theorem 7.19. a and b are calculated with (w_i, y_i).

Corollary 2.2.9:[26] *In the functional model, the LS estimators b and a are not consistent estimators for β and α.*

Remark 2.2.10: Of course, the estimators a and b still do not provide a symmetrical MEM – see the end of section 1.1.

2.2.2 ML Estimator

Theorem 1.3.8 suggests that the next estimation method to be considered is maximum likelihood. We need a distribution type for the ML method. Also with regard to theorem 1.3.8, the normal distribution is a suitable distribution type. The structural and functional models are defined accordingly. As a quality characteristic we try to show consistency and not unbiasedness – compare Remark 1.3.2.

The following statement clearly restricts the choice of models:

Theorem 2.2.11:[27] *In the ultrastructural model, the ML estimators for α and β are not consistent. This does not change even if σ_δ^2 is known.*

A consistent estimator is currently only known for the special case, the structural model. The condition $X_i \sim N(\mu, \sigma_\delta)$, $\mu \in \mathbb{R}$ for $i = 1, \ldots, n$ means that there is only one „true“ value, μ and the measured values scatter around it. This only applies to very specific examples.[28] Nevertheless, only the structural model is usually considered in the literature.

In the functional and structural model, the ML estimator for α or β generally does not exist. In both models, further information is required to calculate the ML estimator, such as the information that the quotient of the error variances λ or $\tilde{\lambda}$ is known.

2.2.2.1 Structural Model

When calculating the ML estimators for α and β, the problem arises that five equations with six unknowns are to be solved uniquely. Therefore applies

Theorem 2.2.12:[29] *In the structural model, the ML estimators for α and β generally do not exist.*

If we assume that

$$\tilde{\lambda} := \frac{2 \cdot \sigma_\varepsilon^2}{\sigma_\delta^2}$$

is known, there is initially a restriction (overdetermined system of equations):

[26] a and b are calculated with (v_i, y_i).

[27] Dolby 1976, p. 43.

[28] Cp. Fuller 1987, p. 34; Madansky 1959, p. 198. Note Schach and Schäfer 1978, p. 156 et sq.

[29] Miller 1986, p. 224 et sq.; Schach and Schäfer 1978, p. 157. Cp. Kendall and Stuart 1979, p. 404.

Theorem 2.2.13:[30] *If $\tilde{\lambda}$ is known in the structural model and $\alpha = 0$, the ML estimator for β does not exist.*

However, if there is no straight line through the origin, the following applies:

Theorem 2.2.14:[31] *If $\tilde{\lambda}$ is known in the structural model and $\alpha \neq 0$, then*

$$\tilde{b} := \theta + \sqrt{\theta^2 + \tilde{\lambda}}$$

with

$$\theta := \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2 - \tilde{\lambda} \cdot \sum_{i=1}^{n}(w_i - \bar{w})^2}{2 \cdot \sum_{i=1}^{n}(y_i - \bar{y}) \cdot (w_i - \bar{w})}$$

and

$$\tilde{a} := \bar{y} - \tilde{b} \cdot \bar{w}$$

are the consistent ML estimators for β and α if the estimator of the covariance of W and Y, i.e. the empirical covariance $s_{W,Y}$ is not equal to zero. Here $\bar{y}$ respectively $\bar{w}$ is the arithmetic mean of $y_1, \ldots, y_n$ respectively $w_1, \ldots, w_n$. If $s_{W,Y} = 0$, then $\tilde{b} = 0$. $\tilde{a}$ and $\tilde{b}$ are also the method of moments estimators for α and β.[32] *With these two estimators for α and β, the structural model is symmetrical.*[33]

Remark 2.2.15: (i) There are various representations of the ML estimator for β from Theorem 2.2.14. However, there are also incorrect formulas, with an age-old error reappearing in Miller 1986, p. 226 et sq.[34]

(ii) The regression line, which is calculated with the estimators from Theorem 2.2.14 and is also referred to as **structural straight line**, lies between the regression lines from w to y and from y to w determined with KQ estimators. The intercept of all three lines is $(\bar{w}, \bar{y})$:[35]

[30] Creasy 1956, p. 69.

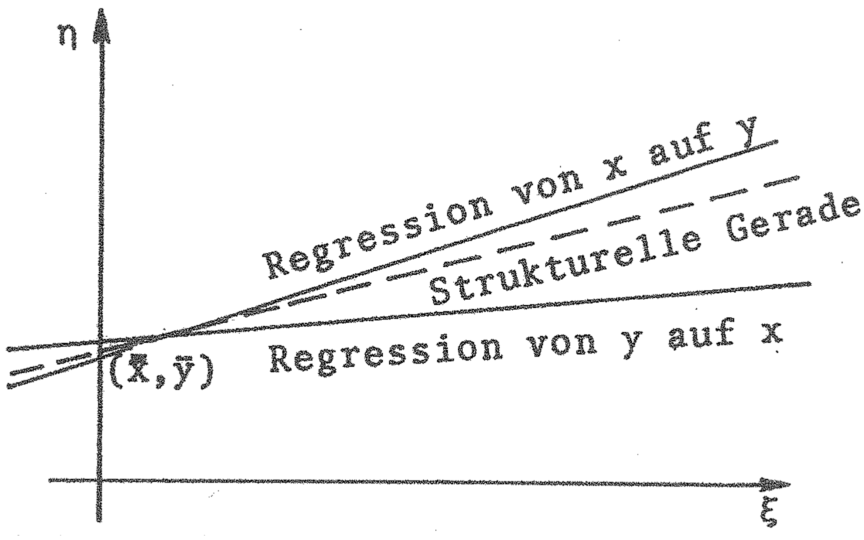
[31] Kendall and Stuart 1979, p. 405, p. 410 et sq.; Schach and Schäfer 1978, p. 158 et sq. Cp. Fuller 1987, p. 31 et sq. For a description, see Miller 1986, p. 226.

[32] Fuller 1987, p. 31; Miller 1986, p. 227.

[33] Cp. Schach and Schäfer 1978, p. 160, footnote +). Note Lemma 2.2.6.

[34] Cp. Kendall and Stuart 1979, p. 405; Schach and Schäfer 1978, p. 158, footnote +).

[35] Schach and Schäfer 1978, p. 159 et sqq. In the notation of Schach, Schäfer, x here corresponds to w and ξ here to x. Cp. Motulsky and Christopoulos 2004, p. 50. See also Casella and Berger 2002, p. 583, taking into account Remark 2.3.4.



The regression lines of w on y („Regression von x auf y“) and of y on w („Regression von y auf x“) are upper and lower bounds for the structural line. Schach and Schäfer 1978 on page 160 gives an estimate of how far apart the straight lines are.

The following applies to straight line through the origin:

Theorem 2.2.16:[36] *If $\alpha = 0$ in the structural model, the ML estimator for β is*

$$\tilde{b} := \frac{\bar{y}}{\bar{w}} \ .$$

With this estimator, the structural model is symmetrical.

Theorem 2.2.17:[37] *If σ_ε^2 is known in the structural model and $\alpha \neq 0$, then*

$$\tilde{b} := \frac{\sum\limits_{i=1}^{n} y_i^2 - n \cdot \bar{y}^2 - (n-1) \cdot \sigma_\varepsilon^2}{\sum\limits_{i=1}^{n} w_i \cdot y_i - n \cdot \bar{w} \cdot \bar{y}}$$

and

$$\tilde{a} := \bar{y} - \tilde{b} \cdot \bar{w}$$

the consistent ML estimators for α and β.

Theorem 2.2.18:[38] *If σ_δ^2 is known in the structural model and $\alpha \neq 0$, then*

$$\tilde{b} := \frac{\sum\limits_{i=1}^{n} w_i \cdot y_i - n \cdot \bar{w} \cdot \bar{y}}{\sum\limits_{i=1}^{n} w_i^2 - n \cdot \bar{w}^2 - (n-1) \cdot \sigma_\delta^2/2}$$

[36] Miller 1986, p. 229. Note Kendall and Stuart 1979, p. 403, (29.19).
[37] Kendall and Stuart 1979, p. 405, p. 410 et sq.; Fuller 1987, p. 14 et sq.
[38] Kendall and Stuart 1979, p. 405, p. 410 et sq.

and

$$\tilde{a} := \bar{y} - \tilde{b} \cdot \bar{w}$$

the consistent ML estimators for α and β.

Remark 2.2.19: (i) If either σ_ε^2 or σ_δ^2 is known in the structural model, the question arises as to whether it makes sense to consider the symmetry at all, since the error variance of only one variable is known. If we nevertheless ask about the symmetry, the following example for Theorem 2.2.17 shows that the ML estimators for α and β are asymmetrical:
We consider the „ideal body weight index“[39]

$$m = 0.9 \cdot l - 100 \ ,$$

where the ideal body weight m is given in kilograms and the body height l in centimetres. Without measurement error, this means („the truth“):

body height x_i	195	163	178
ideal body weight η_i	75.5	46.7	60.2

The estimators are of course $\tilde{a} = -100$ and $\tilde{b} = 0.9$ as well as $\breve{a} = 111.\bar{1}$ and $\breve{b} = 1.\bar{1}$, which means that there is symmetry. – For the values

body height w_i	196	162	180
ideal body weight y_i	76.3	46.6	59.6

measured with errors, where $\sigma_\varepsilon^2 = 0.71$ is rounded, the rounded estimators $\tilde{b} = 0.877$ and $\breve{b} = 1.146$ result. Since $\breve{b}^{-1} = 0.872$, the regression lines are not symmetrical.
A corresponding example shows that the ML estimators for α and β of the Theorem 2.2.18 are also asymmetric.

(ii) If σ_ε^2 and σ_δ^2 are known, ML estimators for α and β can be available under certain conditions. Compare Kendall and Stuart 1979, p. 405 et sq.

2.2.2.2 Functional Model

In the structural model, the problem of the underdetermined system of equations arises when calculating the ML estimators for α and β. This problem could be solved by making additional assumptions. A similar problem now occurs when calculating the ML estimators in the functional model: They cannot be calculated without an additional assumption, as the likelihood

[39]This index is based on the Broca index, taking into account the lower limit of the normal weight range of 18.5 of the body mass index.

function does not have a local maximum, but a saddle point.[40] In the literature, only the additional assumption that

$$\lambda := \frac{\sigma_\varepsilon^2}{\sigma_\delta^2}$$

is known is now cited. Is this the only way to calculate the ML estimators in the functional model?

Theorem 2.2.20:[41] *If λ is known in the functional model, then*

$$\tilde{b} := \theta + \sqrt{\theta^2 + \lambda}$$

with

$$\theta := \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2 - \lambda \cdot \sum_{i=1}^{n}(v_i - \bar{v})^2}{2 \cdot \sum_{i=1}^{n}(y_i - \bar{y}) \cdot (v_i - \bar{v})}$$

and

$$\tilde{a} := \bar{y} - \tilde{b} \cdot \bar{v}$$

are the consistent ML estimators for β and α if the estimator of the covariance of V and Y, i.e. the empirical covariance $s_{V,Y}$ is not equal to zero. Here $\bar{y}$ respectively $\bar{v}$ is the arithmetic mean of $y_1, \ldots, y_n$ respectively $v_1, \ldots, v_n$. With these two estimators for α and β, the functional model is symmetrical.[42]

Remark 2.2.21: (i) The functional model is rarely discussed in the literature. In Fuller's book, the theorem 2.2.20 is not mentioned, it is only stated that the ML estimation does not lead to the goal in multidimensional models, p. 103 et sq.

(ii) Since the estimators of Theorem 2.2.14 and Theorem 2.2.20 are identical for a given sample, the **functional straight line** naturally also lies between the regression lines of v on y and of y on v determined with KQ estimators.

2.2.3 Evaluation of the Models

As can be seen from Lemma 2.2.3, the functional model is the generalisation of the simple regression model and for this reason alone is preferable to the structural model. The structural model cannot be a generalisation of the simple regression model, even in terms of the original optimisation task, as the domain of definition of the regression function, i.e. the values for X, only contains one number, namely μ. In the case of the simple regression model and the functional model, the domain of definition consists of a real interval.

[40] Kendall and Stuart 1979, p. 407 et sq.;Dolby 1976, p. 43; Schach and Schäfer 1978, p. 165.

[41] Kendall and Stuart 1979, p. 410 et sq.; Schach and Schäfer 1978, p. 167 et sq.

[42] Cp. Schach and Schäfer 1978, p. 160, footnote +). Note Lemma 2.2.6.

What do applications to the structural model look like? There may only be one „true“ feature expression for X. Madansky cites the example of the yield strength of artillery shells,[43] Fuller uses the example of the hen pheasant population in Iowa,[44] thus very specific applications. The number of possible applications of the structural model is very small. The ultrastructural model would be preferable, but no estimators are known for it, see Theorem 2.2.11.

Estimators are only available in the functional and structural model if another property of the model is known. Estimators exist in the structural model for the following additional information:

1. The quotient of the error variances $\tilde{\lambda}$ is known, compare Theorem 2.2.14.
2. It is a straight line through the origin, i.e. it is $\alpha = 0$, compare Theorem 2.2.16.
3. The error variance σ_ε^2 is known, compare Theorem 2.2.17.
4. The error variance σ_δ^2 is known, compare Theorem 2.2.18.

In the functional model, there are estimators if the following additional information is available:

1. The quotient of the error variances λ is known, compare Theorem 2.2.20.

As a rule, it will be known whether a specific application is a straight line through the origin. The quotient of the error variances λ or $\tilde{\lambda}$ is often also available with the value 1.[45] This case always occurs when both features are measured in the same way. But how should the error variance σ_δ^2 or σ_ε^2 of the independent or dependent variable be known? I do not have an example of this. Fuller proposes the estimation of the error variance — this is of course not the knowledge of the same — from a large number of independently repeated measurements.[46]

In principle, the MEM enables symmetrical estimation of the regression line. Why is this an advantage? It often happens that a linear relation exists or is suspected between two features X and Y, but there is no effect of one feature on the other, i.e. it is not possible to speak of an explanatory (independent) and an explained (dependent) variable. There is a clear mode of action for the two features „fertiliser quantity“ and „wheat yield“ The amount of fertiliser influences the amount of wheat harvested. There is no mode of action for the features „body weight“ and „body height“ of humans.

[43] Madansky 1959, p. 198.

[44] Fuller 1987, p. 34.

[45] Kendall and Stuart 1979 write about this on page 405: *This is the classical method of resolving the identifiability problem.*

[46] Fuller 1987, p. 13 et sq.

If there is no mode of action between the features X and Y and there is no possibility of symmetrical estimation, then the result of the regression analysis depends on which feature is used as the independent variable. The regression analysis therefore does not provide a satisfactory result.

It is difficult to judge how often the case where there is no mode of action occurs in relation to the case where there is a mode of action. There are various examples in which there is no mode of action, which means that the symmetrical estimate is of interest and should not be neglected.

2.3 Regression by Kummell

The estimators for α and β of the Theorems 2.2.14 and 2.2.20 have already been proposed by Kummell 1879.[47] However, the derivation is done in a completely different way, namely with the help of the Taylor polynomial.[48] Significantly, this approach also results in an underdetermined system of equations.[49]

In some natural sciences, Kummell's regression is referred to as **Deming regression**,[50] although this is not very appropriate. This is because Deming 1948 deals with the adjustment calculation mostly using the method of least squares[51] and only mentions Kummell's regression in passing with reference to him.[52]

Why the research of the 1940s did not take Kummell's results into account, to which Deming refers, is incomprehensible to me. Results would have been achieved more quickly, especially with the functional model.

Orthogonal regression is a special case[53] of Kummell regression:

Definition 2.3.1:[54] Let the model of the regression calculation be given with the regression function $y = \hat{\alpha} + \hat{\beta} \cdot \xi$. If we use the squared Euclidean distance to determine the regression function, i.e.[55]

$$\sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} \cdot x_i)^2 + (\xi_i - x_i)^2 \to \min! \ ,$$

then we have the **method of orthogonal regression**, also known as **total least squares**.

[47] Kummell 1879, p. 101, equation (f) is the estimator for α, equation (h) and the following remarks the estimator for β.

[48] Kummell 1879, p. 98. Note Deming 1948, p. 38, equation (4).

[49] Kummell 1879, p. 100: „Unless we make a certain assumption with regard to the weights of observed quantities, we cannot solve, by a direct process, even these eq'ns.“

[50] Cp. Motulsky and Christopoulos 2004, p. 50; Linnet 1998,

[51] Deming 1948, p. iii.

[52] Deming 1948, p. 184.

[53] Note Remark 2.3.4.

[54] Cramér 1946, p. 275.

[55] Madansky 1959, p. 37, (1.3.14); Casella and Berger 2002, p. 582.

Remark 2.3.2: (i) The Euclidean distance of a point to a straight line is the shortest distance between this point and the straight line. The segment defined by the shortest distance is perpendicular (orthogonal) to the regression line.[56]

(ii) What is the difference between the KQ method and the method of orthogonal regression? The KQ method minimises the distance only with respect to the y-axis, the method of orthogonal regression with respect to both axes.

Theorem 2.3.3:[57] *The method of orthogonal regression leads to the estimators*

$$\hat{b} := \theta + \sqrt{\theta^2 + 1}$$

with

$$\theta := \frac{\sum\limits_{i=1}^{n}(y_i - \bar{y})^2 - \sum\limits_{i=1}^{n}(\xi_i - \bar{\xi})^2}{2 \cdot \sum\limits_{i=1}^{n}(y_i - \bar{y}) \cdot (\xi_i - \bar{\xi})}$$

and

$$\hat{a} := \bar{y} - \tilde{b} \cdot \bar{v} \ .$$

Remark 2.3.4: The estimators of the method of orthogonal regression correspond to those in the Theorem 2.2.14 (structural model) with $\tilde{\lambda} = 1$ and those in the Theorem 2.2.20 (functional model) with $\lambda = 1$. Kummell already mentions this special case with reference to Adcock.[58]

The method of orthogonal regression is the starting point to derive the estimators of the Theorems 2.2.14 and 2.2.20 (for any $\tilde{\lambda}, \lambda$) in a third way. Glaister 2001 shows this on pages 105 et sq.[59]

[56] Casella and Berger 2002, p. 581.

[57] Fuller 1987, p. 38, (1.3.21), (1.3.27). There is an error in the specification of the conditions for equation (1.3.27). It must be $\sigma_{ee} = \sigma_{uu}$. Casella and Berger 2002, p. 582.

[58] Kummell 1879, p. 101 et sq.

[59] Note that for Glaister 2001 $\mu = 1/\lambda$.

3 Independence of scaling

If we consider the feature X body height and the feature Y body weight in humans and assume that there is the linear relationship

$$\eta = f(x) = \alpha + \beta \cdot x$$

between these two features. Then it would be fatal for the estimation of the regression line if the estimation depends on whether height is given in centimetres or metres or weight in kilograms or grams.

What effect does a change in scaling have on a linear relation? Let's assume that body height is no longer given in metres but in centimetres. This leads to the line

$$\check{f}(x) = \alpha + \check{\beta} \cdot x \ ,$$

which has the same ordinate intercept as $f(x)$, but is 100 times flatter. So that the linear relation remains unchanged,

$$\check{\beta} = \frac{1}{100} \cdot \beta$$

must apply. – If, on the other hand, the body weight is not given in kilograms but in grams, this leads to

$$1000 \cdot f(x) = 1000 \cdot (\alpha + \beta \cdot x) \ .$$

This means that the straight line $f(x)$ is only adapted to the change in the scaling of the ordinate axis, is shifted accordingly, but remains unchanged.

The change in scaling must therefore be taken into account in the following way so that the same linear relation exists: For

$$f(x) = \alpha + \beta \cdot x \ ,$$

let x be specified in the unit Λ and $f(x)$ in the unit Θ. If $\check{x}$ is now specified in the unit $p \cdot \Lambda$ and $\check{f}(\check{x})$ in the unit $q \cdot \Theta$, where $p, q \in \mathbb{R}\backslash\{0\}$, then

$$\check{f}(\check{x}) = q \cdot \alpha + \frac{q}{p} \cdot \beta \cdot \check{x} \ .$$

If the estimators change accordingly, the scaling of a unit does not play a role in the estimation:

Definition 3.0.1: Let the regression function $y = f(\xi) = \hat{\alpha} + \hat{\beta} \cdot \xi$ be given[1], where ξ is in the unit Λ and $f(\xi)$ is in the unit Θ.

(i) Let $f_p(\check{\xi}) = \hat{\alpha}_p + \hat{\beta}_p \cdot \check{\xi}$ be the regression function for which $\check{\xi}$ is specified in the unit $p \cdot \Lambda$, $p \in \mathbb{R}\backslash\{0\}$. The regression function $f(\xi)$ is called **independent of the scaling with respect to** ξ if $f_p(\xi) = \hat{\alpha} + \frac{1}{p} \cdot \hat{\beta} \cdot \check{\xi}$ applies.

(ii) $f_q(\xi) = \hat{\alpha}_q + \hat{\beta}_q \cdot \xi$ is the regression function where $f_q(\xi)$ is given in the unit $q \cdot \Theta$, $q \in \mathbb{R}\backslash\{0\}$. The regression function $f(\xi)$ is called **independent of the scaling with respect to** $f(\xi)$ if $q \cdot f(\xi) = f_q(\xi)$ applies.

Remark 3.0.2: (i) The estimators $\hat{\alpha}, \hat{\beta}, \hat{\alpha}_p, \hat{\beta}_p, \hat{\alpha}_q, \hat{\beta}_q$ in Definition 3.0.1 are calculated in the same way.

(ii) The „unit“ is the dimension of the scale in which the feature is measured. These can be physical units such as metres or milliamperes, but also economic units such as the number of unemployed in thousands or euro cents. The feature can usually be measured in different units, which can be converted into each other. For example, if we consider the feature „time period“, then this can be measured in the units year or second.

(iii) What does a change in the scaling of the unit mean? If ξ has the unit $p \cdot \Lambda$, then $\frac{1}{p} \cdot \xi$ has the unit Λ. For example, if ξ has the unit centimetre ($= 100 \cdot$ metre), then $\frac{1}{100} \cdot \xi$ has the unit metre. The same applies to $f(\xi)$.

Lemma 3.0.3: *The regression function $y = a + b \cdot x$ respectively $y = a + b \cdot \xi$ of the regression calculation with LS estimators is independent of the scaling with regard to x respectively ξ and y.*

Proof: Let $\check{x} := p \cdot x, \check{y} := q \cdot y$ and $\check{y} = \check{a} + \check{b} \cdot \check{x}$ with $p, q \in \mathbb{R}\backslash\{0\}$.
To show: $\check{y} = q \cdot a + \frac{q}{p} \cdot b \cdot \check{x}$
We have

$$\check{b} = \frac{\sum_{i=1}^{n} (p \cdot x_i - p \cdot \bar{x})(q \cdot y_i - q \cdot \bar{y})}{\sum_{i=1}^{n} (p \cdot x_i - p \cdot \bar{x})^2} = \frac{p \cdot q \cdot \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{p^2 \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{q}{p} \cdot b$$

and

$$\check{a} = q \cdot \bar{y} - \check{b} \cdot \overline{(p \cdot x)} = q \cdot \bar{y} - \frac{q}{p} \cdot b \cdot p \cdot \bar{x} = q \cdot \bar{y} - q \cdot b \cdot \bar{x} = q \cdot a \ .$$

□

Theorem 3.0.4: *The regression function $y = \tilde{a} + \tilde{b} \cdot w$ of the structural model with known $\tilde{\lambda}$ with the estimators from Theorem 2.2.14 as well as the regression function $y = \tilde{a} + \tilde{b} \cdot v$ of the functional model with known λ with the estimators from Theorem 2.2.20 is independent of the scaling with respect to w respectively v and y.*

[1] This regression function is representative of all regression functions listed in this paper.

Proof: Let $\breve{w} := p \cdot w, \breve{y} := q \cdot y$ and $\breve{y} = \breve{a} + \breve{b} \cdot \breve{w}$ with $p, q \in \mathbb{R}\backslash\{0\}$.
To show: $\breve{y} = q \cdot \tilde{a} + \frac{q}{p} \cdot \tilde{b} \cdot \breve{w}$
Firstly, we have[2]

$$\breve{\lambda} = \frac{q^2}{p^2} \cdot \tilde{\lambda} \ .$$

From this follows

$$\breve{\theta} = \frac{q^2 \cdot \sum_{i=1}^{n} (y_i - \bar{y})^2 - \frac{q^2}{p^2} \cdot \tilde{\lambda} \cdot p^2 \cdot \sum_{i=1}^{n} (w_i - \bar{w})^2}{2 \cdot p \cdot q \cdot \sum_{i=1}^{n} (y_i - \bar{y}) \cdot (w_i - \bar{w})} = \frac{q}{p} \cdot \theta$$

and therefore

$$\breve{b} = \breve{\theta} + \sqrt{\breve{\theta}^2 + \breve{\lambda}} = \frac{q}{p} \cdot \theta + \sqrt{\left(\frac{q}{p} \cdot \theta\right)^2 + \frac{q^2}{p^2} \cdot \tilde{\lambda}} = \frac{q}{p} \cdot \tilde{b} \ .$$

The point is now that

$$\breve{a} = q \cdot \bar{y} - \breve{b} \cdot \overline{(p \cdot w)} = q \cdot \bar{y} - \frac{q}{p} \cdot \tilde{b} \cdot p \cdot \bar{w} = q \cdot \tilde{a} \ .$$

□

Remark 3.0.5: The difference between the estimators of the structural model respectively the functional model on the one hand and those of the orthogonal regression on the other hand is that $\tilde{\lambda}$ respectively λ are 1 in the orthogonal regression – compare Remark 2.3.4. For the proof of Theorem 3.0.4, however, it is essential that $\tilde{\lambda}$ respectively λ is the quotient of the error variances. Therefore, the orthogonal regression is not independent of the scaling with regard to ξ or y. This has already been pointed out by Wald 1940 on page 284.

[2] Georgii 2004, p. 107, Theorem 4.23, a).

4 Technical Questions

In the following, questions are listed for which I do not know the answer:

1. Are there other estimators in the structural or functional model besides the estimator from Theorem 2.2.14 using the method of moments?

2. The estimators in Theorem 2.2.14 (structural model) are also method of moments estimators. Does this also apply to the corresponding Theorem 2.2.20 (functional model)?

3. In the functional model, the ML estimator is available for a given λ. Are there ML estimators under other conditions, for example with the knowledge of σ_ε^2 or σ_δ^2? Can other estimators be constructed? What properties do they have?

4. Does the property that the structural straight line lies between the two LS regression lines — compare Remark 2.2.15 — also apply under other conditions? Which statements regarding the functional straight line are possible in the functional model — compare Remark 2.2.21 — if λ is not known?

5. Theorem 2.2.14 gives an estimator for the case $s_{W,Y} = 0$. Is a corresponding statement possible in the parallel case of the Theorem 2.2.20 for $s_{V,Y} = 0$?

6. Is it possible to construct an estimator in the ultrastructural model?

7. In the functional model — also in the (ultra-)structural model — the regression equations $Y_{x_i} = \alpha + \beta \cdot x_i + E_{x_i}$ have a different independent variable than the regression function $y = \hat{\alpha} + \hat{\beta} \cdot v$. Can the functional model be constructed in such a way that $Y_{x_i} = \alpha + \beta \cdot v_i + E_{x_i}$? Note Remark 2.2.5, (ii).

References

Bruchlos, Kai. „A test for the slope in the functional measurement error model.“ *Acta et Commentationes Universitatis Tartuensis de Mathematica.* 19.2 (2015): 87 - 95.

Casella, George and Roger L. Berger. *Statistical Inference.* 2nd ed. Pacific Grove: Duxbury / Thomson Learning, 2002.

Cramér, Harald.*Mathematical methods of statistics.* 1st ed. Princeton: Princeton University Press, 1946.

Creasy, Monica A. „Confidence Limits for the Gradient in the Linear Functional Relationship.“ *Journal of the Royal Statistical Society, Series B.* 18 (1956): 65 - 69.

Deming, W. Edwards. *Statistical Adjustment of Data.* 4th ed. New York: John Wiley& Sons, 1948.

Dolby, G. R. „The ultrastructural relation: A synthesis of the functional and structural relations.“ *Biometrika.* 63.1 (1976): 39 - 50.

Fahrmeir, Ludwig, Rita Künstler, Iris Pigeot, Gerhard Tutz. *Statistik. Der Weg zur Datenanalyse.* 7th ed. corr. reprint. Heidelberg [et al.]: Springer, 2011.

Fisz, Marek. *Probability Theory and Mathematical Statistics.* 3rd ed. New York [et al.]: John Wiley& Sons, 1967.

Fuller, Wayne A. *Measurement Error Models.* 1st ed. New York: John Wiley& Sons, 1987.

Georgii, Hans-Otto. *Stochastik.* 2. bearb. Aufl. Berlin, New York: Walter de Gruyter, 2004.

Glaister, P. „Least squares revisited.“ *The Mathematical Gazette.* 85.502 (2001): 104 - 107.

Gujarati, Damodar N. *Basic Econometrics.* 2nd ed. New York [et al.]: McGraw-Hill, 1988.

Kendall, Maurice and Alan Stuart, *The Advanced Theory of Statistics. Volume 2: Inference and Relationship.* 4th ed. London: Griffin, 1979.

Kummell, Chas. H. „Reduction of Observation Equations Which Contain More Than One Observed Quantity.“ *The Analyst.* 6.4 (1879): 97 - 105.

Lindley, Dennis V. „Regression lines and the linear functional relationship.“ *Journal of the Royal Statistical Society, Supplement.* 9 (1947): 218 - 244.

Linnet, Kristian. „Performance of Deming regression analysis in case of misspecified analytical error ratio in method comparison studies.“ *Clinical Chemistry.* 44.5 (1998): 1024 - 1031.

Madansky, Albert. „The Fitting of Straight Lines When Both Variables Are Subject To Error.“ *Journal of the American Statistical Association.* 54 (1959): 173 - 205.

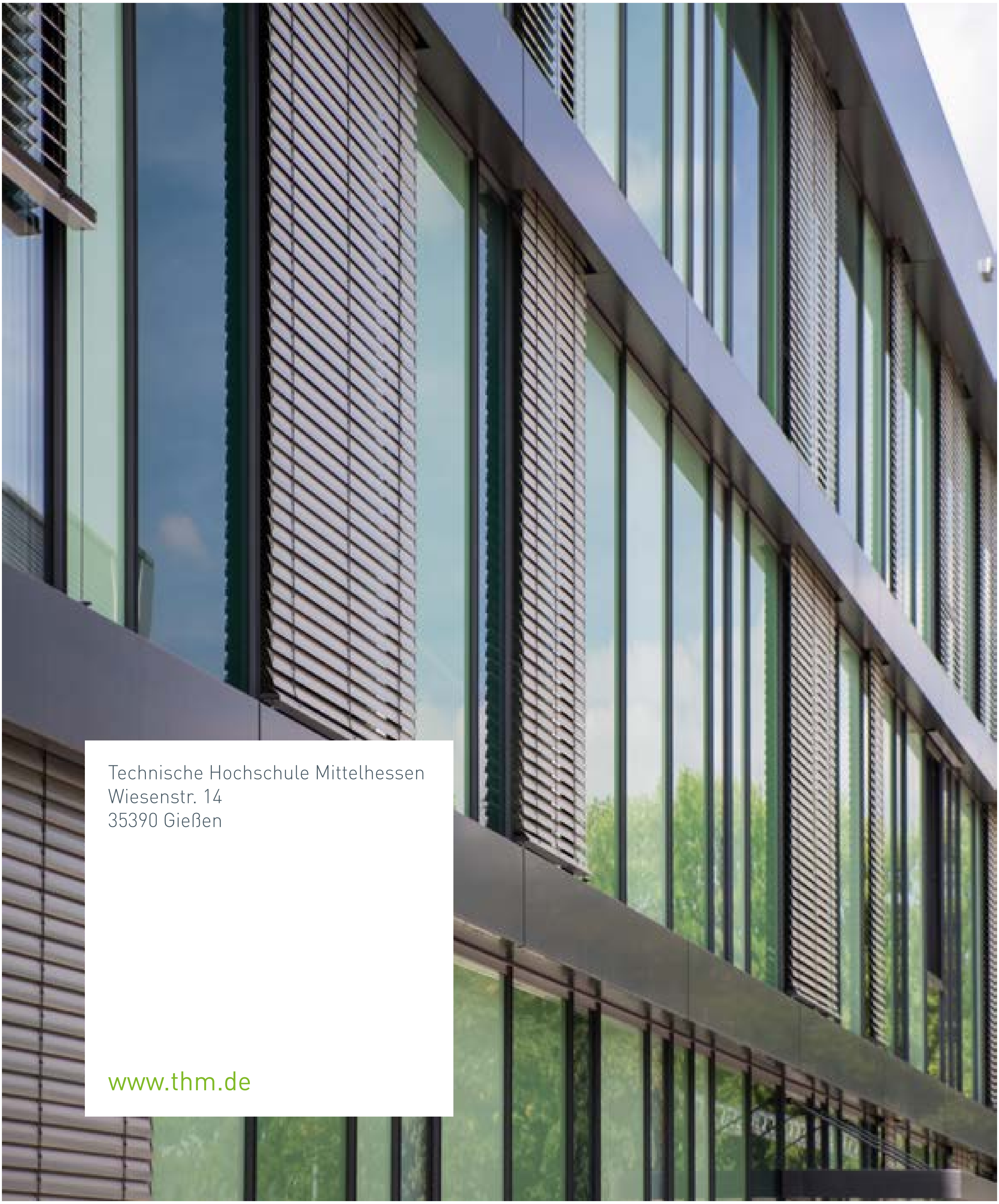
Miller, Rupert G. *Beyond ANOVA, Basics of Applied Statistics.* 1st ed. New York [et al.]: John Wiley& Sons, 1986.

Motulsky, Harvey and Arthur Christopoulos. *Fitting Models to Biological Data Using Linear and Nonlinear Regression.* 1st ed. Oxford: Oxford University Press, 2004.

Pestman, Wiebe R. *Mathematical Statistics.* 1st ed. Berlin, New York: de Gruyter, 1998.

Schach, Siegfried and Thomas Schäfer. *Regressions- und Varianzanalyse.* 1st ed. Berlin [et al.]: Springer, 1978.

Wald, Abraham. „The fitting of straight lines if both variables are subject to error.“ *The Annals of Mathematical Statistics.* 11.3 (1940): 284 - 300.



Technische Hochschule Mittelhessen
Wiesenstr. 14
35390 Gießen

www.thm.de