Institut für
Ökonomische Bildung

Westfälische
Wilhelms-Universität
Münster

Fachbereich 04: Wirtschaftswissenschaften

IÖB-Diskussionspapier

3/08

Predicting Birth-Rates Through German Micro-Census Data

A Comparison of Probit and Boolean Regression

Rainer Hufnagel

IÖB-Diskussionspapier 3/08

März 2008

ISSN 1860-8159

Predicting Birth-Rates Through German Micro-Census Data A Comparison of Probit and Boolean Regression

Rainer Hufnagel

Abstract

This paper investigates the complex interrelationships of qualitative socio-economic variables in the context of Boolean Regression. The data forming the basis for this investigation are from the German Micro-census waves of 1996 – 2002 and comprise about 400 000 observations. Boolean Regression is used to predict how birth events depend on the socio-economic characteristics of women and their male partners.

Boolean Regression is compared to Probit. The data set is split into two halves in order to determine which method yields more accurate predictions. It turns out that Probit is superior, if a given socio-economic type is substantiated by less than about 30 observations, whereas Boolean Regression is superior to Probit, if a given socio-economic type is verified by more than about 30 observations. Therefore a "hybrid" estimation method, combining Probit and Boolean Regression, is proposed and used in the remainder of the paper. Different methods of interpreting the results of the estimations are introduced, relying mainly on simulation techniques.

With respect to the reasons for the prevailing low German fertility rates, it is evident that these could be decisively higher if people had higher incomes and earned more with relative ease.

From a methodological perspective, the paper demonstrates that Scientific Use Files of socio-economic data comprising hundred thousands or even millions of observations, and which have been made available recently, are the natural field of application for Boolean Regression. Possible consequences for future social and economic research are discussed.

JEL classification: C11, C15, J11.

Im Internet unter:
http://www.wiwi.uni-muenster.de/ioeb/downloads/forschen/paper/IOEB_DP_03_2008.pdf

Prof. Dr. Rainer Hufnagel; WWU Münster;
Abteilung Haushaltswissenschaft und ihre Didaktik; Philippistraße 2; D 48149 Münster;
(0049)(0)251/8329234; Hufnagel@hw.uni-muenster.de

1. Introduction

“What concerns the economy, is entangled”, wrote the German satirist Kurt Tucholsky in his Summary of Economics.[1] This sentence, written in 1931, and intended partly seriously, partly humorously, is still frequently used in academic debates in Germany. It is an all-purpose argument that social or economic life is too complex to be captured effectively by a handful or even dozens of quantitative equations, with which socio-economic models normally work. The argument is almost always true, but has, nevertheless, been rejected as too simplistic. What else but quantitative models was at hand? To be sure, there were empathic case studies and descriptive reasoning which focussed on the sense and logic of human behaviour. However, even a large collection of case studies does not constitute a general theory, and descriptively deduced theorems are often too complex or fuzzy to be falsificable. In short, the argument was accepted grudgingly , because there was no serious alternative.

In the 1990s, however, a new academic discipline was established: “Complexity Theory”. Its development is closely associated with problems arising mainly from the fields of biology and artificial intelligence.[2] Given the needs of economics described in the above paragraph, it seemed a natural development that economics would turn to this new theory of complexity to overcome its methodological problems. The interdisciplinary bridge between complexity theory and economics was established by the so-called evolutionary economics, which is concerned mainly with the development of markets, firms and products,[3] and not so much with socioeconomic behaviour.[4] An adaptation of complexity theory for socioeconomic purposes, especially household theory, was proposed by Hufnagel (2000). This theoretical approach has remained uncontested so far. However, the theory still needs an empirical foundation. This will be set out in more detail in the present paper.

A study of the so-called Boolean grids belongs to the paradigmatic core of complexity theory.[5] Boolean grids can be described as follows. Assume n Boolean variables (BV) x_1, x_2, …, x_n. Boolean variables can only assume the value 0 or 1 (or “false” or “true”, “off” or “on”, and so on). A function $y=f(x_1,\dots,x_n)$ mapping a vector of BVs to another BV y is referred to as a Boolean function. Each Boolean function f can be represented by a Boolean normal form[6]:

(1-1) $$f(x_1,\dots,x_n) = \bigoplus_{m=1}^{2^n} a_m \cdot T_m(x_1,\dots,x_n)$$

with

(1-2) $$T_m(x_1,\dots,x_n) = \prod_{i=1}^{n} (x_i)^{j_i^{<m>}} \cdot (1-x_i)^{(1-j_i^{<m>})}$$
$$(j_1^{<m>},\dots,j_n^{<m>}) \in \{0,1\}^n$$

[1] „Was die Weltwirtschaft angeht, so ist sie verflochten“, Tucholsky (2002, p. 195).
[2] See the well-known books of Kauffman (1995) and Holland (1995).
[3] See Arthur, B., Durlauf, St., Lane, D. (1997) and, as another example, Frenken (2006).
[4] Regarding the field of *political analysis*, see the reasoning in Braumoeller (2003, pp. 210-215) on “causal complexity”.
[5] See again Kauffman (1995) and Hufnagel (2000).
[6] Quine (1962,1974), Rudenau (1974).

$\oplus$ means the Sum Modulo 2, which is the same as a logical "either ..or..". The coefficients a_m are Boolean as well.

The state that a BV assumes at a discrete point or interval of time t shall be denoted by $x_i^{<t>}$. Assume that n Boolean functions $f_1,\ldots,f_n$ are given . A Boolean grid is an iterative dynamic defined by

$$
\begin{aligned}
& x_1^{<t>} = f_1\left(x_1^{<t-1>},\ldots.,x_n^{<t-1>}\right) \\
& x_2^{<t>} = f_2\left(x_1^{<t-1>},\ldots.,x_n^{<t-1>}\right) \\
& \ldots \\
& \ldots \\
& x_n^{<t>} = f_n\left(x_1^{<t-1>},\ldots.,x_n^{<t-1>}\right)
\end{aligned}
\tag{1-3}
$$

The dynamic can evolve in various different ways. The vector $\left(x_1^{<t>},\ldots,x_n^{<t>}\right)$ may remain static or proceed in a simple periodicity. On the other hand, it can yield a highly complex periodicity, what Kauffman called quasi-chaotic behaviour. At the frontier of quasi-chaotic and static behaviour we find sub-critical behaviour of the iteration. It is sufficiently stable to ensure an identity over time, and yet, it is flexible enough to react to and balance out disturbances, and finally to evolve. These are properties, that, from a biological perspective, a living being needs, or, in socio-economic terms, an institution needs.

Using Boolean grids in socio-economics seems to be a logical approach. Among the variables frequently used are many that are naturally binary: male/female, employed/unemployed, child/adult, and so on. All qualitative or ordinal variables, such as professions, activities, degrees, nationalities, attitudes, norms etc., are easily encoded by binaries. The few remaining truly metric variables, such as age and income, can be reduced to binaries by classification. An institution is defined as a set of rules. Rules are easily transposed into logical sentences and these, in turn, into Boolean expressions.[7] Hence, institutions can be modelled as a Boolean grid. These considerations promise an interesting and useful new field of research, because it is a fundamental issue to determine whether a given institution is too static, is sub-critical enough to evolve, or too chaotic to be sustainable.

Whether a Boolean grid behaves statically, sub-critically or quasi chaotically, depends on the Boolean functions by which it is constituted. Kauffman's finding is that the behaviour of a grid tends more towards flexibility, the more variables the Boolean functions f_i in (1-3) depend on and the more $\wedge$s and $\vee$s are needed to express the Boolean function. Accordingly, if we are interested in the properties of a Boolean grid, we must determine the shape of the functions f_i in (1-3), which is equivalent to estimating the coefficients a_m in Equation (1-1). This exercise is referred to as Boolean Regression (BR).[8]

Before we can consider the new domain of a social science based on the use of Boolean grids, we must, therefore, be able to master BR. At first glance, this seems fairly straightforward. It is logical to develop BR through a simple analogy to a common multivariate linear regression.

---

[7] E.g.: "All adults who are not prisoners and not subject to legal control are entitled to vote". Assume y=entitled to vote, x_1=adult, x_2=prisoner, x_3=under legal control. The above rule can then formally be written as $y \Leftrightarrow x_1\wedge(\neg(x_2\vee x_3))$. For further examples and a detailed analysis of the general considerations relating to the applicability of Boolean grids in socio-economics, see Hufnagel (2000).

[8] See Boros et al. (1995), Ruczinski et al. (2002).

We need:

a) a measure of how well the estimated function fits the data.
b) an algorithm to find the best fitting function
c) a measure of the significance of the estimated coefficients a_m.

A final test of any method fulfilling a)-c) should be to split the data into two halves. One half is used to estimate a best fitting function and the other half to control subsequently, how well this function predicts in a new data set. The author made several attempts with BR, using data sets drawn from the German socio-economic panel (GSOEP) and from the Time Use Survey (TUS) of the "Statistisches Bundesamt Deutschland", Germany's national office of statistics. These data sets typically involve hundreds or thousands of people. The author used methods suggested by Boros et al. (1995) and his own considerations. Using the data-splitting test, the results turned out to be disappointing. The specific problems involved with Boolean regression can be outlined as follows.

Let us suppose one makes a preliminary selection of 75 regressors x_i for the purposes of a socio-economic investigation. Naturally, not all will be significant, so that, in a publication, we may ultimately present a table with 10 – 20 significant regressors. These are typical magnitudes, considering, for example, investigations on such topics as the estimation of Mincer's wage equation etc. If we started a Boolean regression with 75 variables, however, we would have to estimate 2^{75} coefficients a_m. Now, $2^{75} \approx 38 \cdot 10^{21}$, i.e. 38 pentillions, a number whose magnitude can be regarded as astronomic. This causes problems, concerning b) and c) mentioned above.

Ad b): The optimizing algorithm must be programmed carefully, so that the computer concludes its calculations within a reasonable period of time.

Ad c): Assume that we define a measure of significance as follows. The coefficients a_m can only take the values 0 or 1. Let be a_m the true value, $\hat{a}_m$ the estimated value. Define a measure of significance α by $\alpha = \Pr ob(a_m \neq \hat{a}_m)$. If we set, for example, α=0.0001 and n=25, we would expect $0.0001 \cdot 2^{25} = 3355$ wrongly estimated coefficients, which would severely reduce the chances of obtaining a useful prognosis.

A further point is that there is still another method for dealing with one binary variable which depends on a set of binary regressors: Probit.[9] This method can conveniently be used, because it is nowadays included in standard statistical software packages, yields significant results when using hundreds of observations and takes at the most a few minutes to perform. So, it has not only to be shown that BR can work in a reliable manner, but also that it is superior to Probit, at least in some aspects. Else it would not be worthwhile to exert oneself with this method, as a more comfortable and faster method is available.

As consequence of these difficulties that the author experienced, he came to the conclusion that only the use of large data sets was really appropriate for BR. A potential opportunity for testing this conjecture arose over the past few years, because a scientific use file (SUF) from the German micro-census was made available to researchers. The German micro-census is a random drawing of 0.5 % of the population and hence comprises hundreds of thousands of observations. It seems sensible to use these recently available data to try out BR again and to compare it to Probit.

[9] Or Logit. For details, see Paragraph 2.1.

Fertility was chosen as the focus of the regression. Given the low birth-rates in Germany, the topic is fundamentally relevant in the first instance. Furthermore, there is a large body of literature, both theoretical and empirical, on this classic field of socio-economic research, so that we are unlikely to be misled by invalid results caused by the method used, because comparisons with the enormous body of existing knowledge are possible.

A comparison of Probit to BR is also of interest from an epistemological point of view. BR relies on classical logic, descriptively formulated by Greek philosophers (therefore also called Aristotelian logic), handed down by scholars and algebraically formalized in the 19th century by Boole and Frege. Logic reasoning can be implemented in technical equipment by using such switches as relays, valves or transistors. Conventional science must argue coherently, that is, it must follow the rules of logic.[10] However, logic is not a natural gift that people possess intrinsically, it has to be taught and learned. Medieval students had first to pass the Trivium, consisting of grammar, rhetoric and dialectics, the first and last of these containing formal logic. Nowadays, students of mathematics or philosophy generally need to pass a course on logic in order to obtain their degrees. Think of the Star-Trek TV-series. Such characters as Mr. Spock or Mr. Data are often contrasted in an amusing manner to real humans. People challenge each other to think or act logically. Obviously, they could behave in quite different manner as well.

But what is this illogical thinking? Over the past decades, researchers have tried to formalize our everyday half-logic, half-intuitive thinking by means of so-called fuzzy logic.[11] In classical logic, a sentence or a composition of sentences can only have a truth value of 0 or 1. In fuzzy logic, a truth value can assume any value in the interval [0,1]. Any function f, mapping a vector $(x_1,\ldots,x_n)\in[0,1]^n$ to a number $y\in[0,1]$ can be seen as a fuzzy logic operator. In this sense, the Probit-prediction $\Phi(a_1 \cdot x_1 + \ldots + a_n \cdot x_n)$, with real coefficients a_i and the standard normal distribution Φ can be interpreted as a kind of a fuzzy-logic Or.

The human brain does not consist of determined switches in the form of transistors. It consists of neurons, which are cells that collect input information by means of synapses and convey an output through their axon. The input potentials x_i are weighted by weights a_i, the output activity y is an S-shaped function of the sum of the weighted inputs.[12] Accordingly, a neuron processes information in a similar manner to a Probit–prediction. One neuron alone cannot simulate all logic operations as *and, or, either-or, not, implies* and *equivalent to*. However, if one arranges them into a neuronal network consisting of several layers of neurons, this network is able to emulate the operations of classical logic and to form types, i.e. to define internal entities which fulfil given properties.[13] In large areas of the human brain[14], 6 layers of neurons can be found. Having more than two layers of neurons is a property that occurred during the evolution of animals. Simple animals, such as jelly fish, are restricted to simple 1 or 2 layer reception-action systems.

Therefore, in terms both of biological evolution and in human (scientific) history, we find a path leading from an intuitive Probit-like logic to the possibility of information processing through classical logic and classification. We may suppose that this path led to a higher level

[10] "The epistemology of science involves applying a series of tests:
The test for coherence consists of determining whether the conclusions adduced follow logically from the assumptions which have been made…." Eichner (1986, p.5).
[11] Zimmermann et al. (1984).
[12] Spitzer (2000, p. 22).
[13] Spitzer (2000, p. 125 onwards).
[14] As well as in the brains of other mammals.

of mental capability, at least concerning some aspects of life and reasoning. On the other hand, intuitive thinking and simple animals still survive. Therefore, simplicity must have its advantages as well.

Given these considerations, there is more than mere technical interest in comparing Probit and BR. We should also investigate the question what form of social perception offers what kind of advantages or disadvantages. Further implications of this investigation for social science theory are in the concluding Section 6.

The remainder of the paper is structured as follows. Section 2 reviews how Probit works and contains a discussion of BR. Section 3 briefly outlines fertility theory, provides an overview of the micro-census data used and the variables selected for this investigation. Section 4 compares the predictive power of Probit and BR by splitting the data set into two halves. Based on the insights gained from Section 4, the entire data set is used for a BR and the results are reported in Section 5. The final Section 6 firstly discusses the implications for social policy of the results presented in Section 5. Secondly, the possible consequences for methodology and the future social science are considered.

2. Probit and Boolean Regression

This section sets out the stochastic models underlying Probit and BR. There exist fine explanations of Probit in textbooks[15], however, because some details of interpretation are subsequently discussed in this paper, the model is summarised in Section 2.1 of this paper for convenience. Concerning BR, Boros et al. (1995) should be mentioned. However, in this paper, we do not follow their approach to BR, which opposes Logit (and hence Probit) through the notion that Logit predicts probabilities, whereas BR directly predicts outcomes of 0 or 1 (Boros et al. 1995, p. 203). Furthermore, their concept of significance was not found to be particularly helpful when dealing with socio-economic data. Accordingly, a different concept is developed in some detail in Section 2.2.

2.1 Probit

Let be given a BV $y \in \{0,1\}$ and n variables x_i (i=1,..,n). Let be N the number of observations of these variables, denoted by $\left(y^{[\nu]}, x_1^{[\nu]},..., x_n^{[\nu]}\right) \quad \nu = 1,..., N$. We wish to predict the value y given a certain observation $\left(x_1^{[\nu]},..., x_n^{[\nu]}\right)$. In order to do so, we attempt to determine n+1 real coefficients a_i (i=0,..,n) so that

(2-1)

$$y^{[\nu]} = 1 \iff a_0 + a_1 \cdot x_1^{[\nu]} + ... + a_n \cdot x_n^{[\nu]} \geq 0 \qquad \forall \nu \in \{1,.., N\} .$$

(2-1) consists of N inequalities in n+1 unknowns. If N>>n+1, there will normally be no solution to this system. Therefore, a further random or unobserved variable ε is introduced, combined with the assumption that ε assumes for each observation a value $\varepsilon^{[\nu]}$ so that

(2-2)

$$y^{[\nu]} = 1 \iff a_0 + a_1 \cdot x_1^{[\nu]} + ... + a_n \cdot x_n^{[\nu]} + \varepsilon^{[\nu]} \geq 0 \qquad \forall \nu \in \{1,.., N\}$$

holds. ε is assumed to be standard normal distributed. The density function of the standard normal distribution is denoted by $\phi(.)$ and its cumulative distribution function by $\Phi(.)$. From (2-2), we obtain:

(2-3) $$\Pr ob(y = 1) = \Pr ob(\varepsilon^{[\nu]} \geq -a_0 - a_1 \cdot x_1^{[\nu]} - ... - a_n \cdot x_n^{[\nu]}) = \Phi(a_0 + a_1 \cdot x_1^{[\nu]} + ... + a_n \cdot x_n^{[\nu]}) .$$

Accordingly, given a set of N observations $\left(y^{[\nu]}, x_1^{[\nu]},..., x_n^{[\nu]}\right) \quad \nu = 1,..., N$, we can determine the coefficients a_i by maximising the log-likelihood-function

(2-4)

$$LLH = \sum_{\{\nu : y^{[\nu]}=1\}} \log(\Phi(a_0 + a_1 \cdot x_1^{[\nu]} + ... + a_n \cdot x_n^{[\nu]})) + \sum_{\{\nu : y^{[\nu]}=0\}} \log\left(1 - \Phi(a_0 + a_1 \cdot x_1^{[\nu]} + ... + a_n \cdot x_n^{[\nu]})\right) .$$

[15] E.g. Judge et al. (1988, pp. 785 ff.), Baltagi (2002, pp. 331 ff.), Woolridge (2006, pp. 582 ff.).

Now, let be $(a_0,..,a_n)$ the vector of coefficients that maximises the LLH. We can interpret this vector, the result of the Probit, according to Equations (2-2) and (2-3), in two ways.

1. Direct Prediction

For a given observation of the independent variables $\left(x_1^{[\nu]},...,x_n^{[\nu]}\right)$, we predict $y^{[\nu]}=1$ if and only if:

(2-5) $a_0 + a_1 \cdot x_1^{[\nu]} + ... + a_n \cdot x_n^{[\nu]} \geq 0$ *i.e.* $\Phi(a_0 + a_1 \cdot x_1^{[\nu]} + ... + a_n \cdot x_n^{[\nu]}) \geq 0.5$.

2. Probabilistic interpretation

For a given observation of the independent variables $\left(x_1^{[\nu]},...,x_n^{[\nu]}\right)$, the probability that y=1 is

$\Phi(a_0 + a_1 \cdot x_1^{[\nu]} + ... + a_n \cdot x_n^{[\nu]})$.

This latter interpretation makes sense especially if all the x_i are Boolean. Then, the set of all N observations $\left(x_1^{[\nu]},...,x_n^{[\nu]}\right)$ can be split into 2^n subsets τ_m, characterised by

(2-6) $\left(x_1^{[\nu]},...,x_n^{[\nu]}\right) \in \tau_m \quad \Leftrightarrow \quad T_m(x_1^{[\nu]},...,x_n^{[\nu]}) = 1$,

where T_m is given by (1-2). Let N_m be the cardinality of τ_m. T_m indicates the properties (x_i) that must be fulfilled or that must not be fulfilled, so that an observation belongs to a certain subset or micro-type. In the set of N observations, micro-type m will have N_m members. For K_m of them shall hold y=1. By

(2-7) $k_m = \dfrac{K_m}{N_m}$

the frequency of y=1 in micro-type m is defined (for $N_m>0$). By

(2-8) $p_m = \Phi(a_0 + a_1 \cdot x_1^{[\nu]} + ... + a_n \cdot x_n^{[\nu]})$ for $\left(x_1^{[\nu]},...,x_n^{[\nu]}\right) \in \tau_m$

we hence obtain a predictor for the actual frequency k_m of y=1 for micro-type m based on the Probit-procedure.

A measure α for the significance of a coefficient a_i is given by defining α as the probability that the increment of the LLH is random when x_i is added to the list of regressors.[16]

There are many ways to define a measure for the goodness of fit of Probit.[17]

Regarding the Maximum-Likelihood-approach inherent in Probit, one can define, for example, McFadden's Pseudo measure of determination as:

[16] For a more detailed explication, see Woolridge (2006, pp. 587-588).
[17] Again, a detailed discussion can be found in Woolridge (2006, pp. 589 ff.).

(2-9) $$\psi - R^2 = 1 - \frac{LLH_{full}}{LLH_0} \quad .$$

LLH_{full} is the Log-Likelihood of a model with all regressors $x_1,\ldots,x_n$ having been estimated. LLH_0 is the Log-Likelihood of a model with only intercept a_0 and no regressors. It holds that $0 \le \psi\text{-}R^2 < 1$ and the fit is the better, the greater $\psi\text{-}R^2$.

With respect to the direct prediction interpretation of Probit, simple hit-rate criteria can be formulated. In terms of the probabilistic interpretation, one can compute the correlation between p_m and k_m.

These concepts are used in Sections 4 and 5 of this paper, in which Probit is compared to BR. Next, however, we expose BR, also using Maximum-Likelihood-Estimation and also discerning a direct prediction and a probabilistic interpretation.

2.2 Boolean Regression

Assume N observations of n+1 BVs $\left(y^{[\nu]}, x_1^{[\nu]},\ldots, x_n^{[\nu]}\right)$ $\nu = 1,\ldots,N$. We look for a Boolean function f that is able to describe or approximate the observations:

(2-10) $$y = f(x_1,\ldots,x_n).$$

In general, there will be no Boolean function f, so that (2-10) holds for all observations. Therefore, we must formulate a stochastic model

(2-12) $$y^{[\nu]} = f(x_1^{[\nu]},\ldots, x_n^{[\nu]}) \;\oplus\; \varepsilon^{[\nu]} \qquad \nu = 1,\ldots., N$$

with a Boolean random variable ε. There is only one parameter for determining the distribution of ε, namely p=Prob(ε=1). $\oplus$ means the sum Modulo 2 or a logical either-or, i.e. “anti-equivalence”.

Each Boolean function can be represented as

(1-1) $$f(x_1,\ldots., x_n) = \bigoplus_{m=1}^{2^n} a_m \cdot T_m(x_1,\ldots., x_n)$$

with

(1-2)
$$T_m(x_1,\ldots, x_n) = \prod_{i=1}^{n} (x_i)^{j_i^{<m>}} \cdot (1 - x_i)^{(1 - j_i^{<m>})}$$
$$(j_1^{<m>},\ldots., j_n^{<m>}) \in \{0,1\}^n$$

and Boolean coefficients a_m.

The T_m are disjoint in the following sense.

(2-11) For a given $\nu \in \{1,.., N\}$ there is one and only one $\mu \in \{1,\ldots, 2^n\}$ with $T_\mu(x_1^{[\nu]},...,x_n^{[\nu]}) = 1$.

All other T_m are 0 for this observation ν.

For a given $m \in \{1,\ldots, 2^n\}$, the set of observations is split into 4 subsets whose cardinalities are given by :

(2-13)

	T_m=0	T_m=1	
y=0	$N-K-N_m+K_m$	N_m-K_m	N-K
y=1	$K-K_m$	K_m	K
	$N-N_m$	N_m	N

There are N observations, for K of them, y=1 holds and for N-K, y=0 holds. For a given m, there are N_m observations with T_m=1 and hence $N-N_m$ observations with T_m=0. For K_m observations, T_m=1 and y=1 hold and hence, for N_m-K_m observations, T_m=1 and y=0 hold. Finally, we have $N-K-N_m+K_m$ cases with y=0 and T_m=0 and $K-K_m$ cases with y=1 and T_m=1. Observe:

$$\sum_m N_m = N \qquad \sum_m K_m = K$$

If we substitute (1-1) into (2-12) we obtain:

(2-14) $$y^{[\nu]} = \bigoplus_{m=1}^{2^n} a_m \cdot T_m(x_1^{[\nu]},...,x_n^{[\nu]}) \oplus \varepsilon^{[\nu]} \qquad \nu = 1,...,N \quad .$$

The task of determining the coefficients a_m in (2-14) is referred to as Boolean regression (BR), as mentioned briefly in the introduction. If, for a certain ν, equation (2-14) holds with $\varepsilon^{[\nu]}$=0, we talk of a correct prediction or a hit. If $\varepsilon^{[\nu]}$=1, we talk of an error.

There are (at least) 3 criteria by which the coefficients a_m in (2-14) can be determined:

a) Maximise the likelihood of the observations

b) Maximise the number of hits in the set of observations

c) Maximise the expected number of hits in another sample than the one observed, but drawn from the same basic sample.

a) and b) lead to the same algorithm. To proof this equivalence, put:

$$a_m = 1 \qquad if \quad \frac{K_m}{N_m} > \frac{1}{2}$$

(2-26) $$a_m = 0 \qquad if \quad \frac{K_m}{N_m} \le \frac{1}{2}$$

$$a_m = 0 \;\; or \;\; a_m = 1 \qquad if \quad N_m = 0$$

.

Proof of the equivalence of maximising likelihood and maximising hit-rates:

Each observation ν can be mapped to a certain T_μ by (2-11). Depending on a_μ , $\varepsilon^{[v]}$ in (2-14) is determined by the following scheme:

$\varepsilon^{[v]}$	a_μ=0	A_μ=1
$y^{[v]}$=0	0	1
$y^{[v]}$=1	1	0

.

As we had set Prob(ε=1)=p, the probability of observation ν depends on a_μ as follows:

(2-15)

	a_μ=0	A_μ=1
$y^{[v]}$=0	1-p	p
$y^{[v]}$=1	p	1-p

.

The product over the probabilities of all observations ν is the product of the probabilities of all observations that are mapped[18] to m=1, multiplied by the product of all observations that are mapped to m=2, …and so on, until … times the product of the probabilities of all observations that are mapped to the final m=2^n. Hence, the approach (2-14), using (2-13) and (2-15), leads to the likelihood-function

(2-16) $$LH = \prod_{m=1}^{2^n} \left\{ p^{N_m - K_m} \cdot (1-p)^{K_m} \right\}^{a_m} \cdot \left\{ (1-p)^{N_m - K_m} \cdot p^{K_m} \right\}^{1-a_m} .$$

By means of a simple transformation, we obtain

$$LH = \prod_{m=1}^{2^n} (1-p)^{N_m - K_m} \cdot p^{K_m} \cdot \left\{ p^{N_m - 2K_m} \cdot (1-p)^{-N_m + 2K_m} \right\}^{a_m} .$$

[18] Compare (2-11).

By using (2-13) $\sum_m N_m = N \qquad \sum_m K_m = K$, it finally follows:

$$\text{(2-17)} \qquad LH = (1-p)^{N-K} \cdot p^K \cdot \prod_{m=1}^{2^n} \left\{ \left(\frac{p}{1-p} \right)^{N_m - 2K_m} \right\}^{a_m} \qquad .$$

We must now differentiate between two cases.

<u>I. p<½</u>

In this case, we have $\frac{p}{1-p} < 1$ and hence

$$\left(\frac{p}{1-p} \right)^{N_m - 2K_m} > 1 \qquad if \ \frac{K_m}{N_m} > \frac{1}{2}$$

$$\left(\frac{p}{1-p} \right)^{N_m - 2K_m} \leq 1 \qquad if \ \frac{K_m}{N_m} \leq \frac{1}{2}$$

$$\left(\frac{p}{1-p} \right)^{N_m - 2K_m} = 1 \qquad if \ K_m = N_m = 0$$

.

Looking at equation (2-17), we see that the likelihood is increased, or at least not decreased, if we put

$$a_m = 1 \qquad if \ \frac{K_m}{N_m} > \frac{1}{2}$$

$$\text{(2-18)} \qquad a_m = 0 \qquad if \ \frac{K_m}{N_m} \leq \frac{1}{2}$$

$$a_m = 0 \ or \ a_m = 1 \qquad if \ N_m = 0$$

.

Applying (2-18), we obtain the following Likelihood-function

$$\text{(2-19)} \qquad LH_I = (1-p)^{N-K} \cdot p^K \cdot \prod_{m: N_m > 0 \wedge \frac{K_m}{N_m} > \frac{1}{2}} p^{N_m - 2K_m} \cdot (1-p)^{-N_m + 2K_m} \qquad .$$

II. p≥½

In this case, we have $\frac{p}{1-p} \geq 1$ and hence

$$\left(\frac{p}{1-p}\right)^{N_m - 2K_m} \geq 1 \qquad if \;\; \frac{K_m}{N_m} \leq \frac{1}{2}$$

$$\left(\frac{p}{1-p}\right)^{N_m - 2K_m} \leq 1 \qquad if \;\; \frac{K_m}{N_m} > \frac{1}{2}$$

$$\left(\frac{p}{1-p}\right)^{N_m - 2K_m} = 1 \qquad if \;\; 0 \leq K_m = N_m = 0$$

.

Considering equation (2-17), we see that the likelihood is increased, or at least not decreased if we put

$$a_m = 1 \qquad if \;\; \frac{K_m}{N_m} \leq \frac{1}{2}$$

(2-20)
$$a_m = 0 \qquad if \;\; \frac{K_m}{N_m} > \frac{1}{2} \qquad .$$

$$a_m = 0 \; or \; a_m = 1 \qquad if \;\; K_m = N_m = 0$$

Applying (2-20), we obtain the following Likelihood-function

(2-21)
$$LH_{II} = (1-p)^{N-K} \cdot p^K \cdot \prod_{m: N_m > 0 \wedge \frac{K_m}{N_m} \leq \frac{1}{2}} p^{N_m - 2K_m} \cdot (1-p)^{-N_m + 2K_m} \qquad .$$

We define:

$$K1 = \sum_{m: N_m > 0 \wedge \frac{K_m}{N_m} > \frac{1}{2}} K_m \qquad N1 = \sum_{m: N_m > 0 \wedge \frac{K_m}{N_m} > \frac{1}{2}} N_m$$

(2-22)

$$K2 = \sum_{m: N_m > 0 \wedge \frac{K_m}{N_m} \leq \frac{1}{2}} K_m \qquad N2 = \sum_{m: N_m > 0 \wedge \frac{K_m}{N_m} \leq \frac{1}{2}} N_m$$

.

It holds

(2-23) N1 + N2 = N K1 + K2 = K.

By the definitions (2-22), we obtain from (2-21) and (2-19):

$$
(2\text{-}24)\qquad
\begin{aligned}
LH_I &= (1-p)^{N-K-N1+2K1} \cdot p^{K+N1-2K1} \\
LH_{II} &= (1-p)^{N-K-N2+2K2} \cdot p^{K+N2-2K2}
\end{aligned}
.
$$

These likelihood-functions are maximised by

$$
(2\text{-}25)\qquad
\begin{aligned}
p_I &= \frac{K+N1-2K1}{N} \\
p_{II} &= \frac{K+N2-2K2}{N}
\end{aligned}
.
$$

From (2-23) and (2-25) follows:

$$
(2\text{-}27)\qquad p_I + p_{II} = 1 \qquad .
$$

From (2-22) and (2-25) follows:

$$
(2\text{-}28)\qquad
\begin{aligned}
p_I &= \frac{K}{N} + \frac{\overbrace{N1-2K1}^{<0}}{N} < \frac{K}{N} \\
p_{II} &= \frac{K}{N} + \frac{\overbrace{N2-2K2}^{\geq 0}}{N} \geq \frac{K}{N}
\end{aligned}
.
$$

K/N is now given empirically by the set of observations. If K/N < ½, we obtain p_I<½ from the first line of (2-28), which is consistent with assumption I. Moreover, from (2-27), in this case, follows p_{II}≥½, also consistent with assumption II. The converse applies, if K/N≥½, we obtain from the second line of (2-28) p_{II}≥½, which is consistent with assumption II, and then, from (2-27) p_I<½, consistent with assumption I. To resume both solutions (2-18) and (2-20) of the task of maximising the likelihood of (2-14) are feasible.

From (2-24) and (2-25), we obtain the maximum likelihoods:

$$LH_I^{\max} = (1-p_I)^{(1-p_I)\cdot N} \cdot p_I^{\,p_I \cdot N}$$

(2-29)

$$LH_{II}^{\max} = (1-p_{II})^{(1-p_{II})\cdot N} \cdot p_I^{\,p_{II} \cdot N}$$

.

By substituting (2-27) into (2-29), we obtain

(2-30) $$LH_I^{\max} = LH_{II}^{\max}$$.

Therefore, for task a), both solutions, (2-18) and (2-20) seem to be equal. Because we had set p=Prob(ε=1), the expected value of correct predictions in (2-14) will be (1-p)·N. Hence, (2-18), with p<½, maximises the number of hits in (2-14), whereas (2-20) maximises the number of errors. From a practical point of view, whether we use (2-18) and (2-20) makes no difference, because, if we are able to predict as incorrectly as possible, we also are able to predict as correctly as possible – relying on the "tertium non datur". To express it differently:

(2-14) $$y^{[\nu]} = \bigoplus_m a_m \cdot T_m(x_1^{[\nu]},...,x_n^{[\nu]}) \oplus \varepsilon^{[\nu]} \qquad \nu = 1,...,N$$

and

(2-31) $$\neg y^{[\nu]} = \bigoplus_m a_m \cdot T_m(x_1^{[\nu]},...,x_n^{[\nu]}) \oplus \varepsilon^{[\nu]} \qquad \nu = 1,...,N \quad .$$

are dual approaches to the same problem with regard to the contents. (2-18) and (2-20) are dual solutions. Because (2-18) directly maximises the hit-rate, we confine our analysis to this approach.

Next, we provide a further proof of why (2-18) maximises the number of hits. Suppose we set all a_m=0 in (2-14). The number of hits H is then N-K. Looking at (2-13), we can see that the number of hits changes by

(2-32) $$\Delta H_m = K_m - (N_m - K_m) \; .$$

We maximise H, if we set all a_m=1 which fulfil $\Delta H_m>0$, i.e. $2K_m>N_m$, which leads us again to (2-18) or identically (2-26) **q.e.d.**

Now let us return to task c). We draw another sample from the same basic one as the sample underlying (2-14). How can we determine the coefficients a_m in (2-14), so that the expected value of hits in a new random sample is maximised. In other words, how can we find a Boolean function that is appropriate for an ex-ante prognosis? We cannot simply use the solution of a) and b) for c), because, with (2-26), all coefficients a_m with N_m=0 remain undetermined. Indeed, subject to the circumstances of typical empirical socio-economic research, there are inevitably many undetermined coefficients. If we work, for example, with 40 regressors, there would be 2^{40} coefficients a_m to be determined, about one trillion. It would not matter if N were 1000, or 100 000 or even one million, there will always be about one trillion undetermined coefficients.

If N_m has been 0 in a given sample, it is not very likely that it will be greater 0 in a new sample, but it might happen. Because there are many cases with N_m=0, in another sample, we can expect observations that are mapped to a T_μ with an undetermined a_μ. Accordingly, for these observations, no or only arbitrary prognosis is possible, whether y=1 or y=0. Therefore, to complete BR, we have to find ways to fill this gap. There are many ways of doing so and we will construct some in the remainder of this section and test them in Section 4.

Let be π_m the probability or frequency in the basic sample that y=1, given T_m=1.

(2-33) $$\pi_m = \Pr ob(y=1 \mid T_m = 1).$$

We approach the task c) as follows. The sample to be drawn, i.e. the sample for which we provide an ex-ante-prognosis, will have N' observations. The number of observations with T_m=1 shall be N_m'. The expected value of the number of observations with y=1 will be K', with

(2-34) $$K' = \sum_{m=1}^{2^n} \pi_n \cdot N_m^{'} \ .$$

If we set all a_m=0, the expected number of hits H' in the new sample will be:

(2-35) $$H' = N' - K' \quad .$$

If we change a certain a_m from 0 to 1, H' will change by

(2-36) $$\Delta H' = \left(\pi_m - (1-\pi_m)\right) \cdot N_m^{'} = \left(2\pi_m - 1\right) \cdot N_m^{'} .$$

Accordingly, the strategy for maximising H' is simple. Put:

(2-37) $$\begin{aligned} a_m &= 0 \quad if \quad \pi_m \le \frac{1}{2} \\ a_m &= 1 \quad if \quad \pi_m > \frac{1}{2} \end{aligned}$$

.

Thus, to complete c), we need estimates for the π_m, which have to be based on the known observations.

An initial possibility is to use Probit-estimators, as described in Section 2.1.

A second possibility is to use the Laplacian estimators $\hat{k}_m = \frac{K_m}{N_m}$. In this case, (2-37) would coincide with (2-26), the maximum-likelihood-estimators. However, the problem remains of what to do if N_m=0.

A third possibility is to use Bayesian estimators. We consider this in more detail below.

If we do not have any information on the basic sample, we can only make suppositions based on π_m.[19] We formalize this supposition by a probability density function $g(\pi_m)$. This density should be sufficiently flexible, but also reasonably tractable in algebraic calculations. It is usual[20], for a purpose such as ours, to assume that $g(\pi_m)$ follows a Beta-distribution:

(2-38) $$g(\pi_m) = \frac{1}{B(\alpha,\beta)} \cdot (1-\pi_m)^{\beta-1} \cdot \pi_m^{\alpha-1} \quad .$$

$\alpha \geq 0$ and $\beta \geq 0$ are real parameters, $B(\alpha,\beta)$ is given by $B(\alpha,\beta) = \frac{\Gamma(\alpha)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, where $\Gamma(.)$ denotes Euler's Gamma-function.

For a given π_m, the probability that, among N_m drawings, there will be K_m with y=1 (compare to this (2-13) and (2-33)), is:

(2-39) $$\Pr ob(K_m \mid \pi_m) = \binom{N_m}{K_m} \cdot \pi_m^{K_m} \cdot (1-\pi_m)^{N_m-K_m} \quad .$$

Given the density function $g(\pi_m)$, the probability that the number of drawings with y=1 given T_m=1 assumes the value K_m is:

(2-40) $$\Pr ob(K_m) = \int_0^1 \Pr ob(K_m \mid \pi_m) \cdot g(\pi_m) \cdot d\pi_m \; .$$

We substitute (2-39) and (2-38) into (2-40) and evaluate the integral, thus obtaining:

(2-41) $$\Pr ob(K_m) = \frac{\binom{N_m}{K_m} \cdot B(K_m+\alpha, N_m-K_m+\beta)}{B(\alpha,\beta)} \quad .$$

By $\text{Prob}(\pi_m|K_m)$, we denote the probability density function that we assume π_m to follow, if we have had the information, that among N_m drawings with T_m=1, there was K_m times found y=1. $g(\pi_m)$ is the a-priori distribution, a supposition made before the drawings. $\text{Prob}(\pi_m|K_m)$ is the a-posteriori distribution, the density we assume after obtaining the information from the drawings. How the information from our empirical observation changes the a-priori density to the a-posteriori density, is given by Bayes' theorem, which can be deduced from the axioms and definitions of probability theory:

(2-42) $$\Pr ob(\pi_m \mid K_m) = \frac{\Pr ob(K_m \mid \pi_m) \cdot g(\pi_m)}{\Pr ob(K_m)} \quad .$$

[19] The basic idea behind Bayesian inference and a discussion on its epistemological status are provided, for example, by Chalmers (2006, pp. 141).
[20] Compare to this point Evans et al. (2000, p. 34).

Substituting (2-38), (2-39) and (2-41) into (2-42) shows that $\text{Prob}(\pi_m|K_m)$ is just another Beta-distribution, which combines a-priori assumptions (α,β) and a-posteriori information (K_m,N_m) in a specific manner:

(2-43) $$\Pr ob(\pi_m \mid K_m) = \frac{(1-\pi_m)^{N_m-K_m+\beta-1} \cdot \pi_m^{K_m+\alpha-1}}{B(K_m+\alpha, N_m-K_m+\beta)} \quad .$$

The mode $\hat{\pi}_m$, the expected value $\bar{\pi}_m$, and the standard deviation σ_m are given by the following formula:

(2-44) $$\hat{\pi}_m = \frac{K_m+\alpha-1}{N_m+\alpha+\beta-2}$$

(2-45) $$\bar{\pi}_m = \frac{K_m+\alpha}{N_m+\alpha+\beta}$$

(2-46) $$\sigma_m^2 = \frac{(K_m+\alpha)\cdot(N_m-K_m+\beta)}{(N_m+\alpha+\beta)^2\cdot(N_m+\alpha+\beta+1)} \quad .$$

If we set $\alpha=1$ and $\beta=1$ the a-priori distribution $g(\pi_m)$ assumes the special form of a rectangular distribution, i.e. each π_m is assigned the same probability a-priori. This is known as the Laplacian principle. The mode then becomes $\frac{K_m}{N_m}$, and if we estimate π_m using its mode, we obtain the estimators

(2-47) $$\hat{k}_m = \frac{K_m}{N_m},$$

which can therefore be characterised as "Laplacian".

If we leave the determination of α and β open, and if we use the expected value for (hence unbiased) estimation, we obtain the "Bayesian" estimators

(2-48) $$\overset{\frown}{k}_m = \frac{K_m+\alpha}{N_m+\alpha+\beta}.$$

For great K_m and N_m, i.e. for many observations, the parts of α and β become neglectible. That is, it does not matter which a-priori assumption had been made. The Bayesian estimator approaches the Laplacian estimator. On the other hand, even if $N_m=0$, the Bayesian exists, unlike the Laplacian estimator. However, an a-priori assumption for α and β has to be made.

In this paper, α and β are chosen as follows. For each $m\in\{1,\ldots,2^n\}$ with $N_m \geq 30$, we evaluate $k_m = \frac{K_m}{N_m}$. We represent all the values yielded for k_m in a histogram. We approximate this histogram by a Beta-distribution density and hence determine an α and β. Section 4 contains an explanation of how this works in practice.

So far, we have developed three methods of determining estimators for the π_m and hence to solve task c) (prognosis with new data) of BR:

1. Probit-estimation of π_m , given by (2-8)

2. Laplacian or Maximum-Likelihood estimation of π_m, given by (2-47).

3. Bayesian estimation of π_m, given by (2-48).

Because the estimation of the π_m is a central part of BR, as shown, we can interpret the results in two different ways, as demonstrated in Section 2.1 yet for the case of Probit.

<u>1. Direct prediction</u>

We predict y=1 if $\pi_m>½$ and y=0 if $\pi_m\leq½$. The criterion for goodness of fit is the hit-rate.

<u>2. Probabilistic interpretation</u>

Through the request $T_m=1$, the basic sample is separated into 2^n disjoint classes. For each of these classes, π_m yields the probability that y=1.

By means of (2-43), we can construct estimators for π_m and confidence intervals for these estimators. For greater K_m and N_m, (2-42) peaks as its mode and can be approximated by a normal distribution. This makes the (approximate) calculation of confidence intervals more convenient. For example, we could construct central confidence intervals for π_m by adding and subtracting $\lambda\cdot\sigma_m$ to $\overline{\pi}_m$ with the usual values for λ as, for example, $\lambda=1.65$ or $\lambda=2.57$.

By means of (2-43), it is also possible to construct a concept of significance for direct prediction. Looking at (2-37), we decide for $a_m=1$, when the estimated value of π_m is greater than ½ . The probability of having committed an error when deciding for $a_m=1$, is, therefore,

(2-49) $$\alpha = \int_0^{1/2} \Pr ob(\pi_m \mid K_m) d\pi_m \qquad .$$

The following two tables give some idea of the requirements for the magnitude of N_m if one wants to attain reasonable levels of significance in practice. The values for α and σ are evaluated according to (2-49) and (2-46) for some values of $\overline{\pi}_m$ and N_m.

(2-50)

		$\overline{\pi}_m$		
	α	0.55	0.63	0.74
	25	0.279	0.084	0.005
N_m	100	0.16	0.005	$2\cdot10^{-7}$
	1000	$8\cdot10^{-4}$	$7\cdot10^{-17}$	$2\cdot10^{-17}$

		$\overline{\pi}_m$		
	σ	0.55	0.63	0.74
	25	0.094	0.091	0.082
N_m	100	0.049	0.048	0.043
	1000	0.016	0.015	0.014

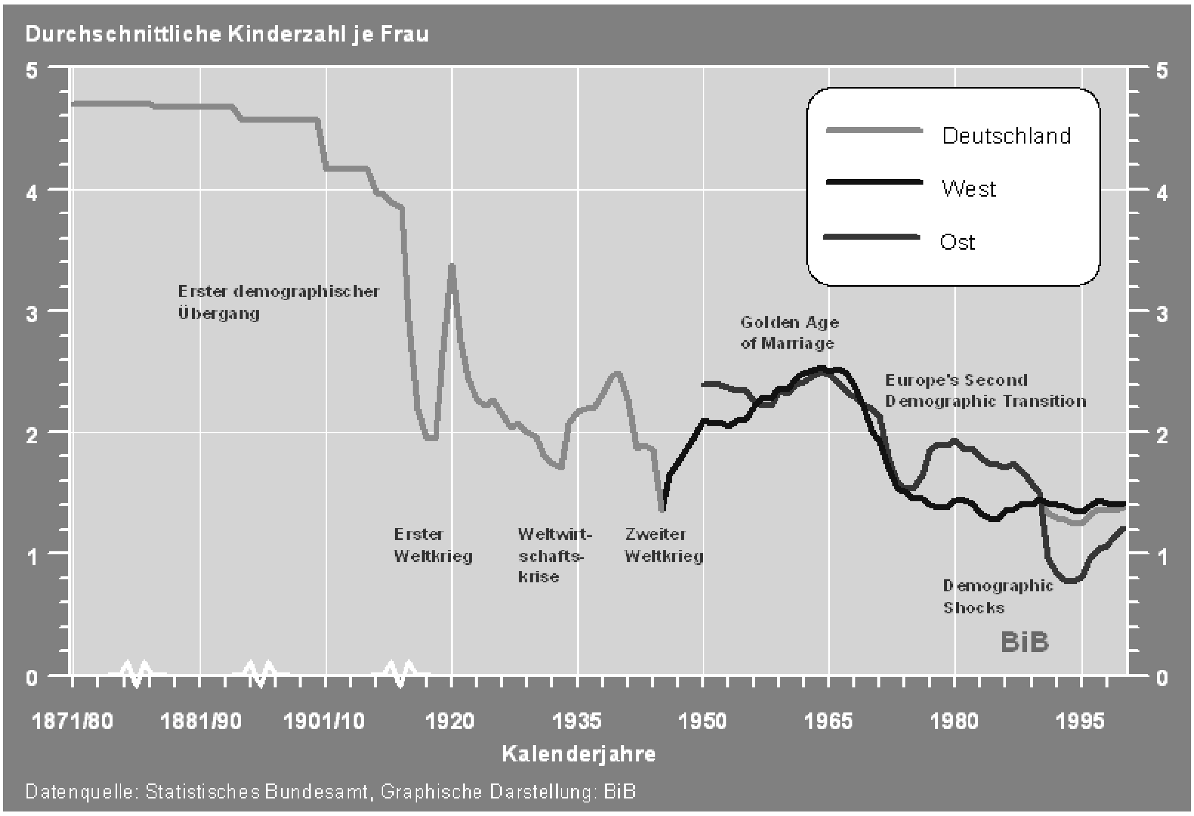
The table shows that, depending on $\bar{\pi}_m$, at least N_m>25 should be required, but hundreds or even thousands of observations per micro-type τ_m might be necessary, or at least desirable. Because the number of micro-types is 2^n, when n is the number of regressors, BR is only likely to be successful in practice, if large data sets are at hand. Therefore, German micro-census data with hundred thousands of observations were chosen to try out BR.

To conclude, in this section, we have developed a method of BR with three sub-methods, depending on how the estimators for π_m are chosen. Below, we test and compare these sub-methods using German micro-census data and a classical theme of socio-economics: fertility. The results of the tests and comparisons are shown in Section 4. In the following Section 3, we give a brief overview of fertility theory and the data used for this investigation.

3. Investigating Birth Rates in Germany by Micro-census Data

Since the end of the post-war Baby-boom in the mid-seventies, fertility rates in Germany have been far below the level 2 that would be necessary for an autonomous reproduction of the population (through own births), as shown in Fig. 1. This development has been widely discussed in terms of social science and social policy, but the diagnoses and recommendations differ. More on this theme can be found in Hufnagel (2007). This present paper will be restricted to a brief overview of three classical theories that provide an anchor to these debates and to the interpretation of empirical results. Further on, we briefly introduce the data set used for this investigation, namely the German micro-census scientific use file (SUF), which has recently been made available by the Statistisches Bundesamt Deutschland.

Fig. 1 Fertility rates in Germany from 1871-2000



thus correspond to the average number of children a woman conceives during her lifetime ("Durchschnittliche Kinderzahl je Frau"). The most current number provided by the Statistisches Bundesamt is 672 700 children born in Germany in 2006 what means a fertility rate of 1.33.

3.1 Fertility theory

Children are not directly comparable with other goods such as coal or wine (or countless other items), especially as one cannot simply buy or sell them on a market. An analogy is nevertheless appropriate, because they yield utility, but also entail costs. In terms of the usual commodities, we can differentiate between an inferior and a superior income reaction, depending on whether demand falls or rises with increasing income. Using this terminology, the essential core of Malthus' (1803) fertility theory suggests the thesis that children are superior goods. In times of economic prosperity and in more prosperous classes of society, more children are born and bred than in economically difficult times and than among the lower-income group of the population. As the main instruments for reducing the number of children, Malthus recommended delaying or avoiding marriage and sexual abstinence in wedlock. If the "poorer classes" did not rely on the above instruments, child mortality caused by poverty would automatically lead to a decline in the number of adult offspring. That 20th century has made available less joyless methods of birth control und very substantially reduced child mortality in the developed countries, does not lower the significance of Malthus' thesis. In Fig. 1, looking at the three great collapses in birth statistics after the two World Wars and during the Great Depression, we find an impressive confirmation of Malthus' ideas.

On the other hand, there has been a secular trend of decreasing fertility rates in the developed countries during the past century, and in the poorer countries fertility rates are higher than in the wealthy countries, both facts in obvious contradiction to Malthus' theory. In order to fill this theoretical gap, some supplements are clearly needed. Among these, the most important stem from Becker (1981) and Easterlin (1973).

According to Becker, decreasing birth-rates in the industrialised countries during the 20th century can be explained mainly by better job opportunities for women and rising female wage-rates. He agrees with the assumption that children are superior "commodities", he even justifies this by indicating that there are no close substitutes for them. However, he then points to the familiar economic insight that not only income, but also the price of a good determines demand: the higher the price, the lower the demand. The price of children can be derived from the costs of children. The costs of children comprise direct costs (food, clothes, school fees etc.) and indirect or opportunity costs. Opportunity costs occur, because children have to be cared for. The time that a parent devotes to child care cannot then be devoted to paid employment. The cost per hour of child care is thus related directly to the prevailing or potential market wage rate of the parent. According to Mincer's (1974) wage equation, the wage rate depends on human capital and especially on an individual's level of formal and vocational education. Because the attending parent is generally the mother, Becker claims that the price of children is higher, the higher a women's level of education and work force participation. Accordingly, declining fertility rates in the industrialised countries during the 20th century are only due to improvements in female education and increased participation rates. However, this argument only holds when two assumptions are made. Firstly, society or women regarded themselves as mainly responsible for child rearing. Secondly, there are no or not enough cheap or state-subsidised public institutions which provide child care. Looking again at Fig. 1, many demographers believe that the fertility decline in the last quarter of the past century in Germany is convincingly explained by Becker's framework.

Yet, another prominent extension of Malthus' theory is provided by Richard Easterlin. He extends Malthus' approach, saying that it is not the absolute, but the relative economic position that counts. "Relative" may mean in relation to others or in relation to the past. Accordingly, if one generation has experienced growth, even starting from a low level, they will feel (subjectively) prosperous, and if the following generation experiences economic stagnation, it will feel subjectively poor, even at an objectively higher economic level as the preceding generation. If one's neighbours are economically better off, one will feel poorer, even if the own economic conditions per se are not bad. Therefore, Easterlin's thesis is that fertility will decline if economic development stagnates or is not as good as expected. Alternatively expressed, if aspirations grow faster than the economic base, birth-rates will decrease. Looking at Fig. 1, Easterlin's theory seems to provide a plausible explanation of the collapse of birth rates in Eastern Germany during the 1990's.[21]

[21] Hufnagel (2004).

In addition, the work of the ecologist Colinvaux (1980) should be mentioned. His theory is broadly similar to those of Malthus and Easterlin, but he again stresses that population is able to regulate its offspring, depending on the perceived prospects for their future.[22]

A declining population will cause problems, at least in the long run, for the infrastructure, labour market and social systems. Given the low starting level, it seems a reasonable goal to increase German birth rates to a level of about 1.7 again.[23] There is less unanimity on how to achieve this goal. If the diagnosis is conducted in terms of Malthus and Easterlin, the remedy would be to foster rising incomes in Germany and provide more direct subsides to young families. If the diagnosis is more along the lines of Becker, more subsides to public child care and appeals to fathers to commit more to looking after their children[24] are necessary. Those agreeing with Easterlin's diagnosis on aspiration levels, complain about "softening of young people in Germany".

The theories of Malthus, Becker an Easterlin all are falsificable and all have empirical content. Statistical investigations abound, the result is that the causes may coexist parallel with one another. Which one dominates, may depend on the particular time and country.[25] We now add a further statistical investigation, which is new in approach and worth doing in so far as a new data set is used and a new method tested - and a recent time period is covered.

3.2 Variables from the German micro-census used for this investigation

The SUF to the German micro-census comprises several hundred thousands people for whom several hundred variables are available. It covers the years 1995-2002. Because variable definitions have partly changed, we restrict the analysis to the years 1996-2002. Given that we wish to consider the data for the mother and the father of a child, we restrict the analysis to (heterosexual) couples (married or unmarried). Because the objective is to investigate fertility, the analysis is restricted to couples in which the woman is aged 15-46. Thus we use[26] 357 502 married couples and 63 256 unmarried couples.

The dependent variable for our investigation is BIRTH, that is, whether the wife had born a child in the past year. However, this item is not available in the micro-census. We therefore use a proxy, the number of children aged 0-2 living in the household and belonging to the family. As independent variables, all variables that describe economic conditions, socio-economic status, income, human capital, attitudes and child-care availability are of interest, given the theoretical basics outlined in Paragraph 3.1. There are about one hundred variables relating to these criteria.[27]

[22] A description and discussion of his ideas can be found in Hufnagel (2007).
[23] See *Bundesministerium für Familie, Senioren, Frauen und Jugend* (2005).
[24] See also Hufnagel (2002).
[25] See also the surveys of Schultz (1973) and Macunovich (1998).
[26] After some of the usual data cleansing for missing or implausible values.
[27] A shortlist, a description and significances (Probit) of them can be found in Hufnagel (2007).

Tab. 1 Descriptive Statistics of the Boolean Variables Used for this Investigation

420 758 Couples (married or living together, heterosexual), drawn from the German micro-census SUF, waves 1996-2002. The variable assumes the value 1, if the property given below under “Description” is true, otherwise, it assumes the value 0.

Variable	Description	Mean
BIRTH	A child in the age 0-2 lives in the household and family	0.166
ALTERF1	Woman’s age < 25	0.068
ALTERF2	Woman’s age 25-29	0.145
ALTERF3	Woman’s age 30-34	0.226
ALTERF4	Woman’s age 35-40	0.289
EHE	Couple is married	0.850
ALTERM1	Man’s age <31	0.169
ALTERM2	Man’s age 31-40	0.435
MITHM	Man is helping member of a farmer’s family	0.001
UNTM	Man is an entrepreneur (large or middle sized firm)	0.061
ICHAGM	Man is an entrepreneur (small firm)	0.045
ARBM	Man is a worker	0.384
NETM	Man does not pursue paid work	0.091
ANGM	Man is a clerk	0.346
BAUER	Man is a farmer	0.010
BEAM	Man is a public servant	0.061
CLASSM1	Man’s net income < 1000 DM p.m.	0.039
CLASSM2	Man’s net income 1000 – 2500 DM p.m.	0.283
CLASSM3	Man’s net income 2500 – 4500 DM p.m.	0.474
CLASSM4	Man’s net income 4500 – 7500 DM p.m.	0.152
CLASSM5	Man’s net income > 7500 DM p.m.	0.034
AZUBIM	Man is in vocational training	0.004
SOLDATM	Man is a soldier	0.007
ASUM	Man is looking for paid work	0.059
TETM	Man is working part time	0.010
BEFRM	Fixed-term employment (Man)	0.049
WALCM	Man works more than 45 hors per week	0.170
POLYTECM	Man has polytechnic school leaving degree[2)]	0.138
LEHREM	Man has passed vocational training	0.529
FSM	Man has passed technical college	0.135
UNIM	Man has university degree	0.105
DEUTSCHM	Man is German	0.898
LEHREF	Woman has passed vocational training	0.552
FSF	Woman has passed technical college	0.099
FHSF	Woman has degree of university of applied sciences	0.032
ABIF	Woman has Abitur[1)]	0.175
UNIF	Woman has university degree	0.076
DEUTSCHF	Woman is German	0.899
NBL	Observation is from territory of former GDR	0.212
BOOM	Observation is from years 2000-2002	0.414

1) School leaving exam qualifying for university.
2) This is a school leaving degree that was common in the former GDR.

Given the results of preliminary Probits and the consideration of plausible causalities in the light of what is known from population economics[28], their number was reduced to 39. This reduction was necessary in order to create a reasonable relationship between the number of possible micro-types ($2^{39}=5{,}5\cdot10^{11}$) and the number of observations (420 758). Accordingly, the point of departure is the 40 variables listed and described in Table. 1. The field of incomes and human capital is covered very well, the usual demographic control for age etc. is given. What is lacking is a variable on public child-care facilities and variables relating to attitudes.[29] Therefore, with the data used in this investigation, we are able to prove the theories of Malthus and Becker, but that of Easterlin only insofar as it concerns the objective base of relative income.

The main interest of this paper, however, is to test BR as a method. For this reason, we can accept that, in terms of content, not all of what might be desirable, is covered. In the next section, we compare BR versus Probit by splitting the data set, which still leaves about 200 000 observations for estimation and 200 000 observations for control.

---

[28] For example, women's net incomes and household net incomes were not included in the list of regressors, because it is not clear whether the woman has no baby because she works or whether she does not work because there is a baby in the family. Another example is number of children living in the household. If we state in a regression that it has a negative influence on actual birth-events, we could explain this by saying that people who already have many children are less likely to have more. If we state that there is a positive relationship, we could explain this by saying that the number of children currently living in the household is an indicator of the preference for children. Therefore, in ambiguous cases like these, possible regressors were omitted, even if they were available in the micro-census.

[29] Which would be given in GSOEP. However, micro-census has the advantage of comprising far more observations than GSOEP and other data sets covering attitudes.

4. Comparing the Predictive Power of Probit and Boolean Regression

In this section, we proceed as follows. The data set described in Section 3.2 is split into two halves:

PAIR – number of observation is even
IMPAIR – number of observation is odd.

Since we can alter the sequence of the observations by a random- number-based process before they are split into two halves, it is possible to secure a random drawing when PAIR and IMPAIR are formulated. In Tables 2 and 3, the results are shown firstly for the sequence of observations which was just given by the SUF, and secondly for two more random drawings generated by different seeds of the random number generator.

Next, we must list some further denotations, partly already known from Section 2.2.

BIRTHp stands for the value the variable that BIRTH assumes for the members of data-set PAIR,

BIRTHi stands for the value the variable that BIRTH assumes for the members of data-set IMPAIR.

(4-1) $T_m = \prod_{i=1}^{39} x_i^{j_i^m} \cdot (1-x_i)^{1-j_i^m} \qquad (j_1^m,...,j_{39}^m) \in \{0,1\}^{39} \qquad m \in \{1,..,2^{39}\}$

is a given $\wedge$-form of the 39 Boolean variables x_i mentioned in Table 1 (besides BIRTH).

(4-2) Np_m is the number observations in data set PAIR with T_m=1
Ni_m is the number observations in data set PAIR with T_m=1

(4-3) Kp_m is the number of observations in data set PAIR with T_m=1 and $BIRTHp_m$=1
Ki_m is the number of observations in data set PAIR with T_m=1 and $BIRTHi_m$=1 .

We now set two goals according to task c) in Section 2.2 (ex-ante prognosis), for both a direct prognosis interpretation and a probabilistic interpretation. Based on the information in the data set PAIR, we wish to predict:

I. the frequencies $ki_m = \dfrac{Ki_m}{Ni_m}$ in data set IMPAIR

II. whether BIRTHi is 0 or 1 in data set IMPAIR.

According to Section 2, there are at least three possibilities to estimate ki_m:

(4-4) $\hat{ki}_m = \frac{Kp_m}{Np_m}$ *Laplacian estimator*

(4-5) $\widehat{ki}_m = \frac{Kp_m + \alpha}{Np_m + \alpha + \beta}$ *Bayesian estimator*

(4-6) pi_m *Probit estimator*

In (4-5), α and β have to be assumed as parameters of a Beta-distribution. For (4-6), we estimate the coefficients of the linear form by means of the data set PAIR and evaluate the linear form and probabilities associated with these coefficients by means of the values that the regressors assume in the data set IMPAIR. For reasons that will become clear below, we define a 4th estimator:

(4-7) $hi_m = \begin{cases} \widehat{ki}_m & if \quad Np_m \geq 30 \\ pi_m & if \quad Np_m < 30 \end{cases}.$

If Np_m=0 the Laplacian estimator is not defined.

Next, we have to choose the parameters α and β in (4-5). In order to do so, the author took all 41 152 T_m from data set PAIR (SUF order) with $N_m \geq 30$ and produced a plot of the frequencies of the k=Kp/Np. This plot is shown in Figure 2. Looking at the shape of the histogram, the author decided to set α=1. Using the SAS-Procedure UNIVARIATE, the other parameter of the Beta-distribution was estimated at β=5.08. For the following analysis, these values are used in (4-6).

Finally, we calculate the actual frequencies

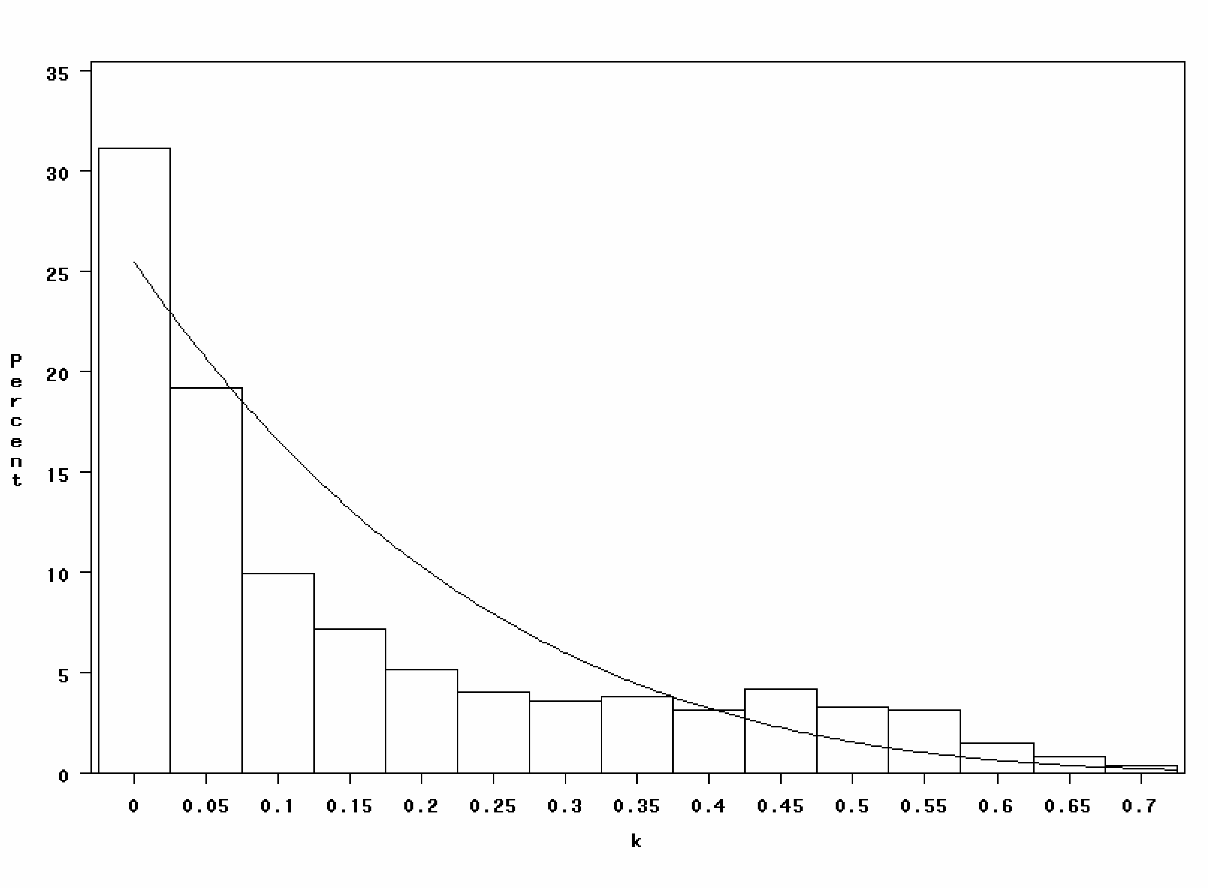
(4-8) $ki_m = \frac{Ki_m}{Ni_m}$

of BIRTHi=1 given T_m=1 in the control data set IMPAIR. We control the goodness of the estimators $\hat{ki}_m, \widehat{ki}_m, pi_m, hi_m \geq$ by correlating them with ki_m for every m with

(4-9) Np_m >0 and Ni_m >0.

The correlation coefficients that emerge are shown in Table 2, in which we vary the requests for the Nps and use 3 different ways to divide the data into the halves PAIR and IMPAIR. The first stems from the order given by the SUF itself and the other two are generated by a random drawing with different seeds. As can be seen from Table 2 and Figure 3, Probit estimators are better than Laplacian and Bayesian estimators, provided the number of observations on which they are based is less than about 30. Within this range, Bayesian estimators are again better than Laplacian estimators. Within the range Np>30, we find that the order reverses. The difference between Laplacian and Bayesian estimators becomes meaningless, but both are visibly superior to Probit.

Fig. 2 Histogramm of the k=Kp/Np in data set PAIR with Np≥30 (SUF order)



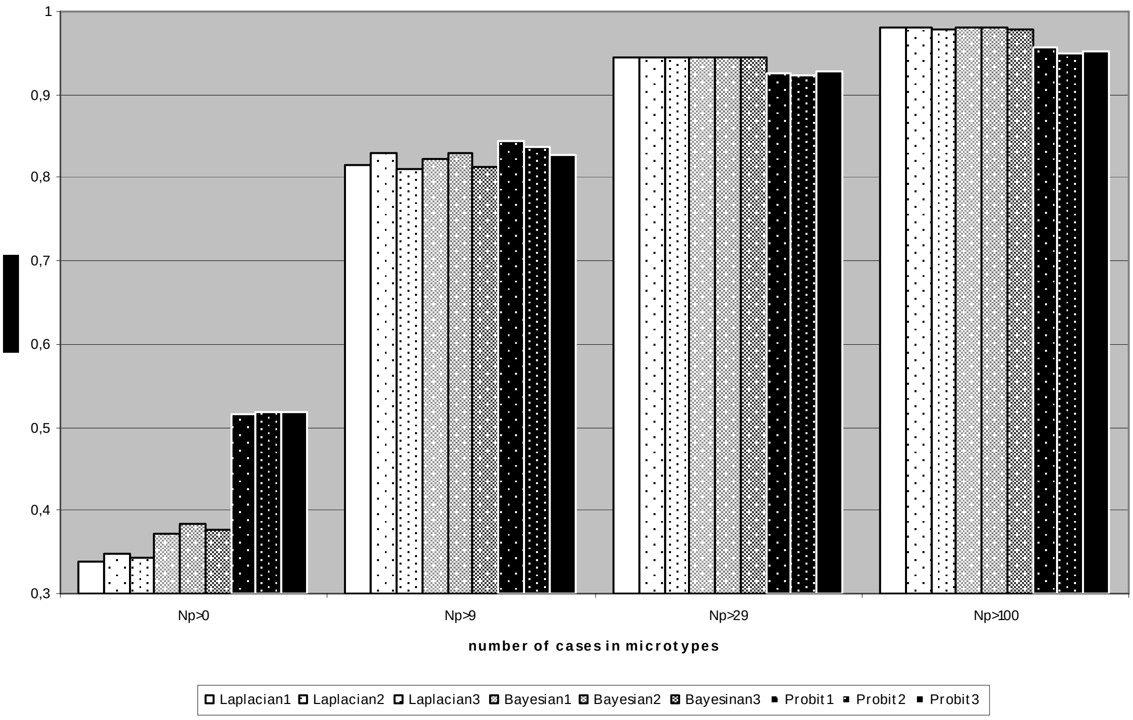
The frequencies of k=Kp/Np are calculated on the basis of 41152 observations with $Np_m \geq 30$ in the data set PAIR. The black line shows the density of a Beta-distribution with α=1 and β=5.08. β was estimated by means of the SAS-Procedure UNIVARIATE.

Tab. 2 Correlations of 4 different estimators of birth rates with actual rates in control data set

Estimator	Np>0		Np≥10		Np≥30		Np≥100	
	R	#	r	#	R	#	r	#
$\hat{k}i_m$	**0.33844**	**22535**	**0.81574**	**2965**	**0.94480**	**897**	**0.98092**	**231**
	0.34808	22530	0.82954	2978	0.94522	906	0.98061	227
	0.34442	22612	0.80961	2966	0.94491	881	0.97868	230
$\widehat{ki}_m$	**0.37240**	**22535**	**0.82291**	**2965**	**0.94566**	**897**	**0.98073**	**231**
	0.38471	22530	0.82989	2978	0.94423	906	0.98082	227
	0.37567	22612	0.81406	2966	0.94564	881	0.97842	230
pi_m	**0.51491**	**22535**	**0.84305**	**2965**	**0.92588**	**897**	**0.95592**	**231**
	0.51704	22530	0.83811	2978	0.92390	906	0.94853	227
	0.51728	22612	0.82663	2966	0.92748	881	0.95261	230
hi_m	**0.51715**	**22535**	**0.84893**	**2965**	**0.94566**	**897**	**0.98073**	**231**
	0.51926	22530	0.84403	2978	0.94423	906	0.98082	227
	0.51785	22612	0.83325	2966	0.94564	881	0.97842	230

r = correlation coefficient of estimator with fractions (4-8).
\# = number of $\wedge$-forms T_m with Np_m>n and Ni_m>0.
Significance of r is better than 0.0001 in all cases.
Bold numbers refer to a data division according to the SUF order of observations, the other numbers refer to a random drawing based on two different random number seeds.

Fig. 3 Correlations of 3 different estimators of birth rates with actual rates in control data set



The figure shows the dependency of the correlations of three kinds of estimators on the number of observations. Exact values are given in Tab. 2. The values of the Laplacian estimators are white, the values of the Bayesian estimators grey and the values of the Probit estimators are black.

The explanation of this phenomenon is probably as follows. The standard deviations of the Laplacian estimators and the Baysian estimators are large for small Nps. That Probit estimators are better under these conditions shows that there is a certain structure in the dependency of the birth rates in terms of the variables mentioned in Table 1. Obviously, there must be at least some variables that generally increase or decrease birth rates, mostly without regard to the cofactors with which they are combined in the T_ms. Therefore, for T_ms with a small Np_m, it seems to be better to interpolate the additive effect of the variables involved in T_m, as Probit does, rather than to rely on estimated frequencies as in the the Laplacian approach. For a similar reason, Bayesian estimators seem more accurate than Laplacian, because in this case birth rates are gradually interpolated by their overall mean $\alpha/(\alpha+\beta)$. For large Np_ms, however, the combined effect of variables in the T_m can be measured with a sufficient precision and, as Tab. 2 shows, play a decisive part in yielding better predictions than simple interpolation techniques. The turning point seems to be at observation numbers of about 30.

For those familiar with complexity theory and its metaphor of fitness landscapes[30], the entire matter can be expressed easily. Globally, the landscape is smooth, but if more precise measurement is possible, we can see a roughness that merits our attraction if we wish to increase our predicative capacity.
Given these considerations, we must assume that the threshold value of about 30 depends on the structure of the data on hand. One might imagine constructing a data set, for which the values of the dependent variable are chosen arbitrarily. In such a case, the N-threshold might

[30] See Kauffman (1995, pp 163), Frenken (2006, p. 11)

be less. On the other hand, if there were a nearly totally linear dependency on the variables, N might be even greater. Consequently, we can state that the threshold N must be estimated from the data on hand. For other data sets, it might assume different values. Accordingly, further empirical investigation of this point is necessary. We will come back on this in Section 6.2.

The observations made in Table 2 and Figure 3 and the above comments suggest that it would be advisable to use a hybrid composition of Probit and Bayesian (or Laplacian) estimators. For values of N less than 30, the Probit predicators are used, otherwise Bayesian predictors are used. hi_m (3-7) is just such an estimator and its overall superiority with respect to the data on hand is clear from its construction. Other suggestions for constructing such a hybrid estimator are easily formulated.[31] However, we will not pursue this point further in the present paper.

Instead, we turn to the question of how well our regression procedures work, if the objective is the correct prediction of a Birth event. With respect to the considerations in Section 2.2, we should predict for the data set IMPAIR as follows:

(4-10)
$$\begin{aligned} &\hat{B}IRTH = 1 \qquad if \ \Pr ob(BIRTH = 1) > 0.5 \\ &\hat{B}IRTH = 0 \qquad else \end{aligned}$$

.

For Prob(BIRTH=1) we might substitute any of the estimators (4-4) – (4-7). The goodness of this prediction rule can be tested by comparing the predicted values $\hat{B}IRTH$ with the actual values BIRTHi in the data set IMPAIR. The prediction rule commits an error if:

(4-11)
$$\hat{B}IRTH \neq BIRTHi \ ,$$

that is, if the predicted value is not the same as the real value of BIRTH in the data set IMPAIR. Let the number of errors be F. We define an “error-rate” ER and a “hit-rate” HR by

(4-12)
$$\begin{aligned} &ER = \frac{F}{Ni} \\ &HR = 1 - ER \end{aligned}$$

with Ni=210379, the number of observations in data set IMPAIR. Table 3 shows error-rates and hit-rates for Bayesian[32] and Probit estimators. At first glance, hit-rates of more than 80 %, as shown in Tab. 3, look quite well. However, one has to consider that the mean of the variable BIRTH in the data set IMPAIR is 0.1655679. Therefore, if we always simply predicted BIRTH=0, the error-rate would also be 0.1655679. The error-rates that we obtain from Bayesian or Probit based prediction are not even one percent better than this trivial natural error-rate.

---

[31] In this context, Braumoeller (2003, 2004) should be mentioned. He also proposes a combination of Probit and Boolean calculus although his approach is very different from what is proposed in this paper. Our approach estimates the Boolean relationships, whereas Braumoeller confirms a Boolean relationship, which must be formulated as an ex-ante hypothesis by the analyst.

[32] If a certain Type T_m occurred in data set IMPAIR, but not in data set PAIR, the Bayesian estimator was fixed to be 0.

Tab. 3 Hit-rates obtained from directly predicting BIRTH-events with Bayesian and Probit estimation

	Error rate	**Hit rate**
(1) Bayesian	0.1622358	0.8377642
(2) Probit	0.1646885	0.8353115
(3) Mean of BIRTH in IMPAIR	0.1655679	
(3) – (1)	0.0033321	
(3) – (2)	0.0008794	

Therefore, for the data on hand, it seems to be of little use to pursue direct Boolean forecasting any further. Predicting probabilities, however, proves useful, especially as we are dealing with a large number of observations, so that probabilities yield actual frequencies. Hence, in the next section, we apply the methods, tested in this section by splitting the data set, to the whole sample of 420 758 observations.

5. Boolean Estimation of the Complete Data Set - Results

Based on the results of Section 4, we will investigate the complete data set of 420 758 observations in this section by means of hybrid estimators. Descriptive statistics of the data set were provided in Table 1. First, in Table 4, we show the result of a pure Probit estimation.

Tab. 4 Probit estimates for BIRTH (N=420 758 observations)

Variable	Coeff.	α	Variable	Coeff.	α
ALTERF1	1.5727	0.0001	AZUBIM	0.4245	0.0042
ALTERF2	1.5603	0.0001	SOLDATM	0.3899	0.0073
ALTERF3	1.3339	0.0001	ASUM	0.1145	0.0001
ALTERF4	0.7479	0.0001	TETM	0.1681	0.0001
EHE	1.0076	0.0001	BEFRM	0.0654	0.0001
ALTERM1	0.5897	0.0001	WALCM	-0.0095	0.2489
ALTERM2	0.4414	0.0001	POYTECM	-0.0928	0.0001
MITHM	0.7813	0.0001	LEHREM	-0.0314	0.0001
UNTM	0.5718	0.0001	FSM	-0.0125	0.2025
ICHAGM	0.6449	0.0001	UNIM	0.0674	0.0001
ARBM	0.5662	0.0001	DEUTSCHM	-0.0943	0.0001
NETM	0.7177	0.0001	LEHREF	0.0320	0.0001
ANGM	0.5387	0.0002	FSF	0.0294	0.0078
BAUER	0.6537	0.0001	FHSF	0.0888	0.0001
BEAM	0.4834	0.0008	ABIF	0.0631	0.0001
CLASSM1	0.2115	0.0001	UNIF	0.1218	0.0001
CLASSM2	0.1364	0.0001	DEUTSCHF	-0.0283	0.0095
CLASSM3	0.4205	0.0001	NBL	-.1028	0.0001
CLASSM4	0.6234	0.0001	BOOM	0.0690	0.0001
CLASSM5	0.6538	0.0001	Ψ-R^2	20.9 %	

α is a measure of the significance of the regressors. ψ-R^2 denotes the McFaddens pseudo coefficient of determination, a measure of the goodness of fit. For details, see the end of Section 2.1.

The Probit results show some general trends in our data set. Fertility is higher among women younger than thirty, than among those older than thirty. The same holds for men. Fertility is higher for married couples than for unmarried couples. With respect to men's professions little difference can be found between blue and white collar workers, but civil servants and soldiers seem to have a slightly lower level of fertility, and farmers and small-firm entrepreneurs and unemployed people, a slightly higher fertility. The dependency on men's income is U-shaped. The lowest income cohort has a higher value than the second lowest. For the other income classes birth rates rise with increasing income. Men looking for work or who are employed part-time or temporarily, evidence a lower fertility rate. On the other hand, those working more than forty hours per week also have relatively few young children. Male levels of vocational training seem to have little influence, university degrees, however, increase fertility rates marginally. The same holds, more or less, for women (sic!). This is an obvious contradiction to Becker's theory of fertility and a relatively new phenomenon for Germany. Among current young German parents or might-be parents, the income effect seems to dominate the price effect of children.[33] In summary, couples living in Eastern Germany and those with German nationality have fewer children. In the relatively prosperous years or 2000-2002, the number of births was higher than in the precedent years. Therefore, from our Probit results, we acquire a picture that very reminiscent of Malthus' and Easterlin's

[33] There is a special paper on this subject: Hufnagel (2008a).

explanations. Low incomes, poor training, insecure employment or joblessness and poor general economic conditions all contribute to low fertility rates in Germany.

Having obtained an overview on the data structure in this manner, we turn to a more detailed analysis by BR. Based on the insights from Section 4, we determine the Bayesian estimators $\widehat{k}_m$ for the complete data set given by (4-5), for the probability of a birth event. In this respect, we restrict the analysis to those derived from $\wedge$-products T_m with $N_m \geq 30$. We find 1849 such products, covering 232 489 couples, i.e. 55 % of all observations. The values of N_m range between 30 and 4182, their mean is 125 and the median 56. The mean of the $\widehat{k}_m$ is 15.8 %, their lowest value is 0, their highest value 71 %. Figure 4 contains the histogram of the values of $\widehat{k}_m$.

Fig 4 Histogram of the density distribution of the Bayesian estimators $\widehat{k}_m$

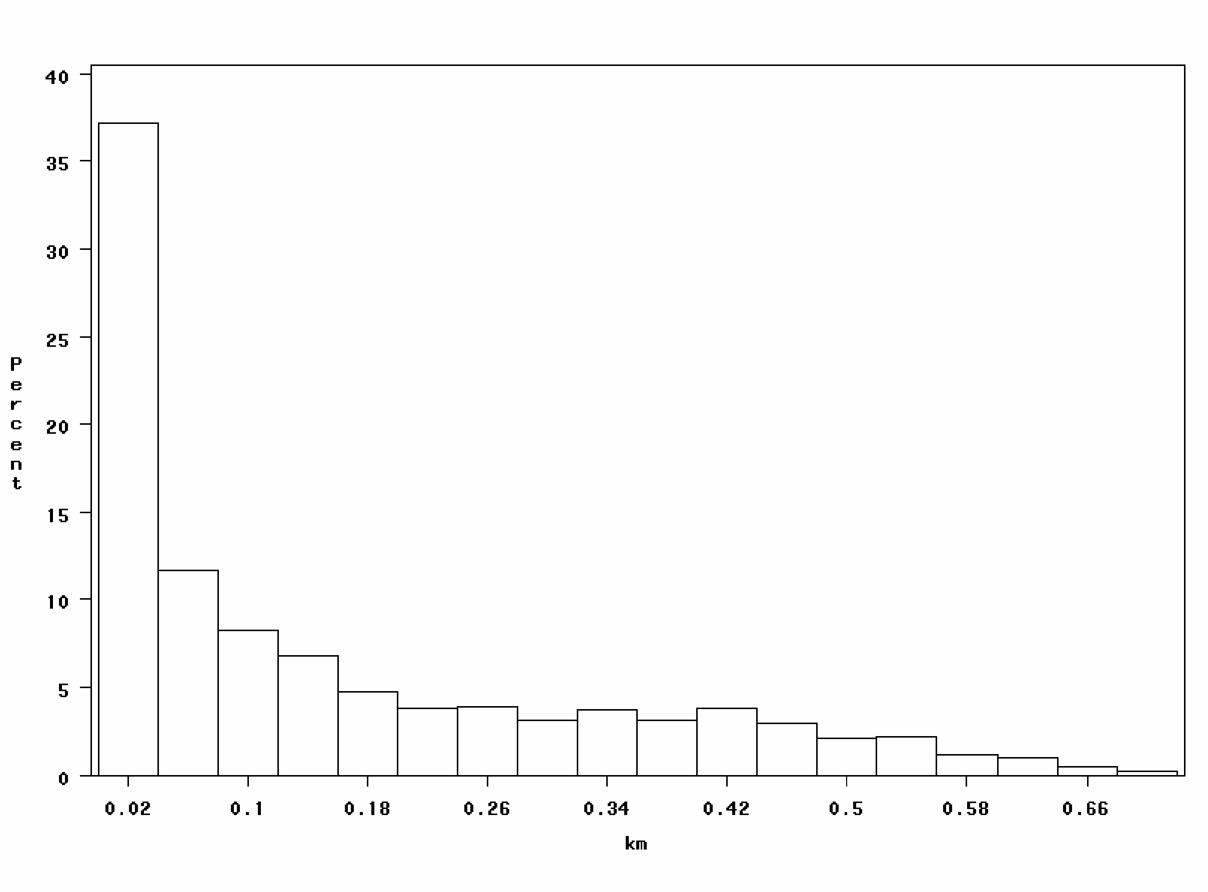


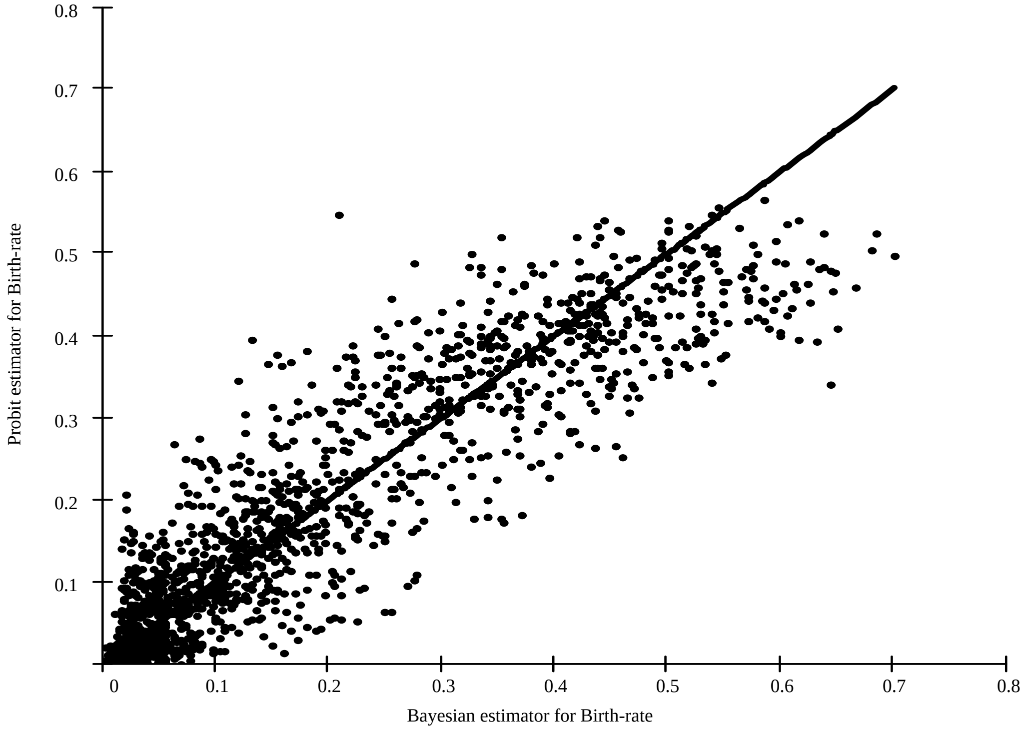
Table 5 gives some examples of the T_m-products and their characteristics. Many of the micro-types with very low birth-rate values reveal that the wife is older than 40, such as No. 4. However, there are also micro-types, where the woman is younger than 40, for example, No. 274 and No. 448, with very low birth-rates. In the latter two cases, the couple is not married. Couple No. 274 are qualified workers, couple No. 448 have university degrees, both with average earnings.

Tab. 5 Micro-Types T_m with very low, average and very high birth-rates

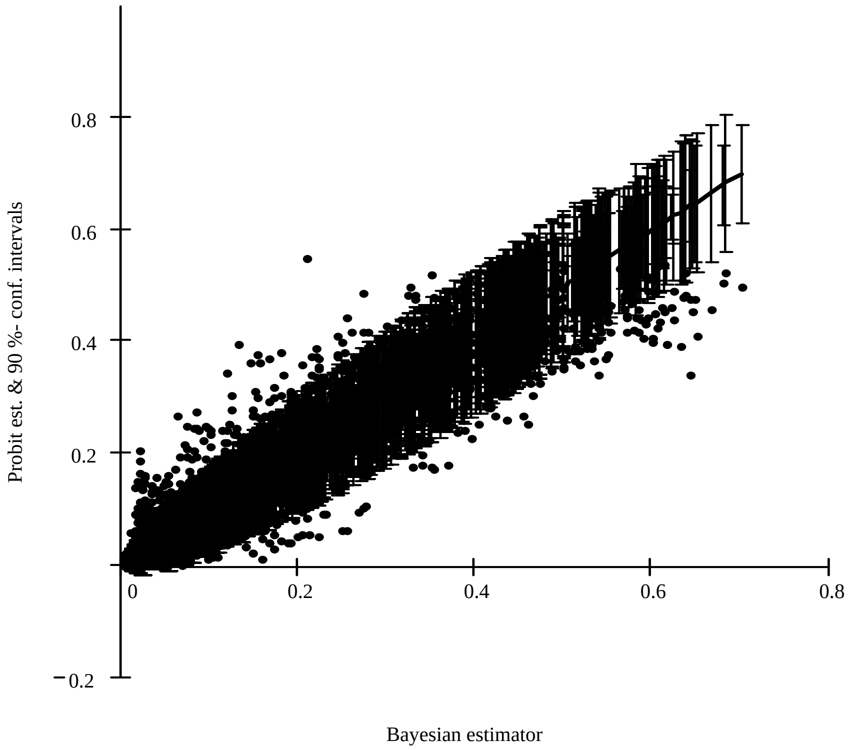
	$m^{1)}$	**4**	**274**	**448**	**1123**	**1158**	**1203**	**1841**	**1847**	**1849**
	N_m	1776	41	66	336	120	106	151	110	71
	Birth-rate	0.0028	0.021	0.028	0.152	0.161	0.176	0.643	0.681	0.701
	$\sigma^{2)}$	0.0013	0.021	0.019	0.019	0.032	0.059	0.038	0.043	0.052
1	ALTERF3	0	0	1	0	1	0	0	0	0
2	ALTERF2	0	1	0	0	0	0	0	0	1
3	ALTERF1	0	0	0	0	0	0	1	1	0
4	ALTERF4	0	0	0	1	0	1	0	0	0
5	EHE	1	0	0	1	1	1	1	1	1
6	ALTERM2	0	1	1	1	1	1	0	0	1
7	ALTERM1	0	0	0	0	0	0	1	1	0
8	MITHM	0	0	0	0	0	0	0	0	0
9	UNTM	0	0	0	0	0	0	0	0	0
10	ICHAGM	0	0	0	0	0	1	0	0	0
11	ARBM	1	0	0	1	0	0	1	0	0
12	NETM	0	0	0	0	1	0	0	1	0
13	ANGM	0	1	1	0	0	0	0	0	1
14	BAUER	0	0	0	0	0	0	0	0	0
15	BEAM	0	0	0	0	0	0	0	0	0
16	CLASSM1	0	0	0	0	0	0	0	0	0
17	CLASSM2	1	0	0	0	1	0	0	1	0
18	CLASSM3	0	1	1	1	0	0	1	0	0
19	CLASSM4	0	0	0	0	0	1	0	0	1
20	CLASSM5	0	0	0	0	0	0	0	0	0
21	AZUBIM	0	0	0	0	0	0	0	0	0
22	UNIF	0	0	1	0	0	0	0	0	0
23	SOLDATM	0	0	0	0	0	0	0	0	0
24	ASUM	0	0	0	0	1	0	0	1	0
25	FHSF	0	0	0	0	0	0	0	0	0
26	TETM	0	0	0	0	0	0	0	0	0
27	FSF	0	0	0	0	0	0	0	0	0
28	LEHREF	1	1	0	0	1	1	0	0	1
29	NBL	1	0	0	0	1		0	0	0
30	DEUTSCHF	1	1	1	0	1	1	1	0	1
31	BEFRM	0	0	0	0	0	0	0	0	0
32	ABIF	0	0	1	0	0	0	0	0	0
33	UNIM	0	0	1	0	0	0	0	0	0
34	WALCM	0	0	0	0	0	1	0	0	0
35	FSM	0	1	0	0	0	0	0	0	0
36	LEHREM	1	0	0	0	0	1	1	0	1
37	POLYTECM	1	0	0	0	1	0	0	0	0
38	DEUTSCHM	1	1	1	0	1	1	1	0	1
39	BOOM	0	1	0	0	1	1	1	0	1

1) The Micro-Types T_m were ranked by ascending birth-rate
2) Standard-deviations of the Bayesian estimators “birth-rate” in the line above.

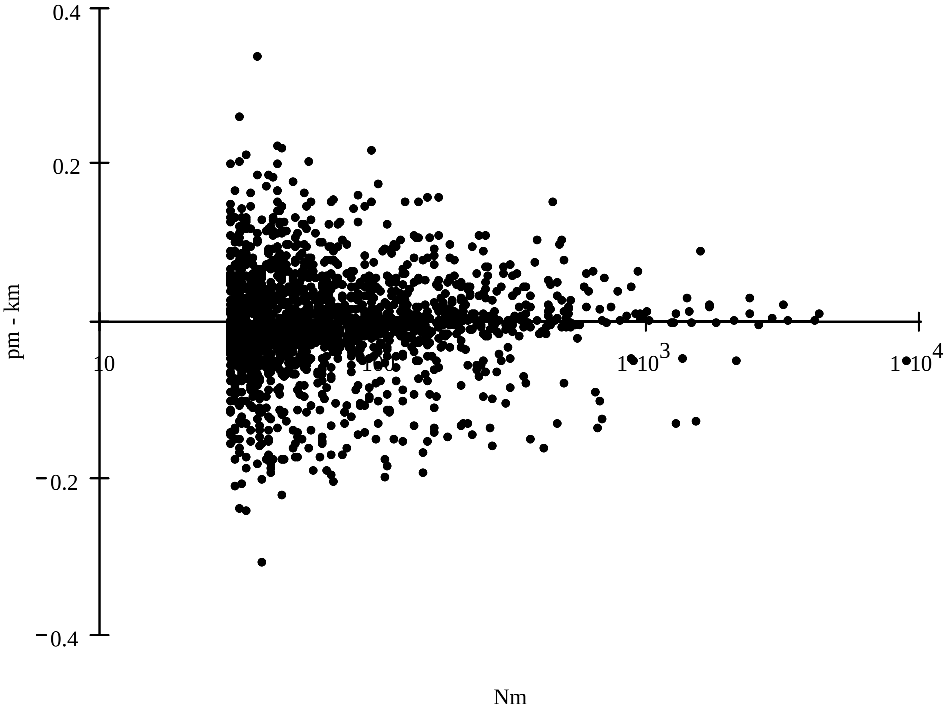
Fig. 5 A comparison of Bayesian and Probit estimators of Birth-rate



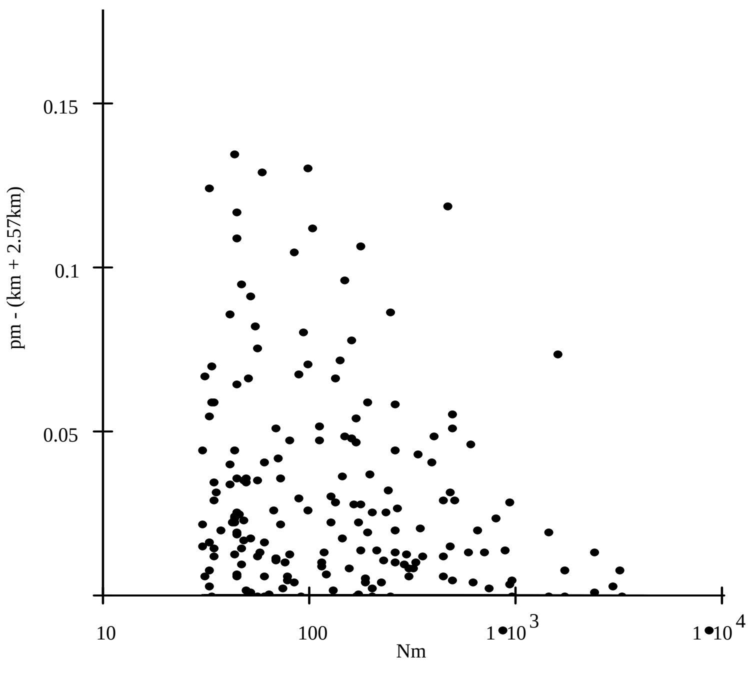
a) The points represent combinations of Bayesian and Probit estimators for the birth-rate for all 1849 Micro-types with $N_m \geq 30$. Comparing with the 45°-identity-line, we see remarkable differences between the two kinds of estimators.



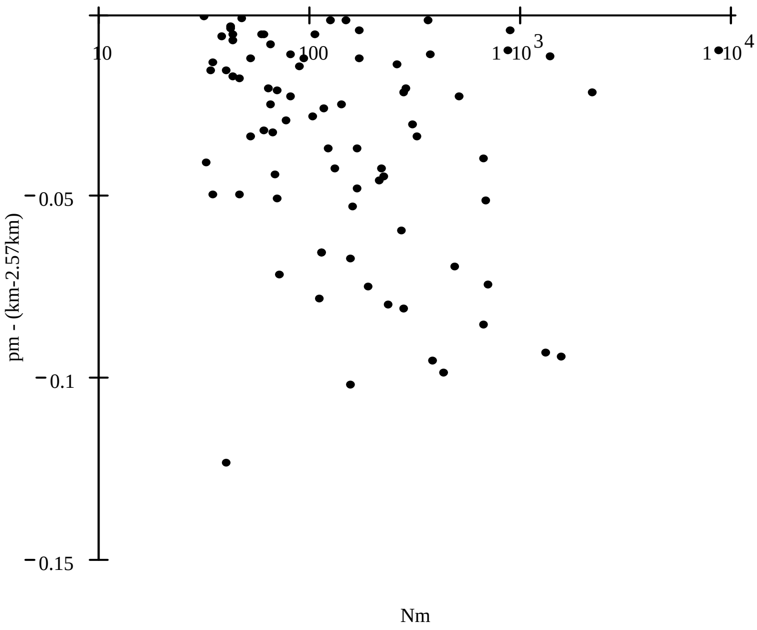
b) The points represent combinations of Baysian and Probit estimators for birth-rate for all 1849 Micro-types with $N_m \geq 30$. In addition to Figure 5a), approximately 90 % confidence intervals for the Bayesian estimators are marked by vertical lines. It can clearly be seen that the deviations of the Probit estimators are significant in numerous cases.



c) $p_m - k_m$, the difference of Probit and Bayesian estimators, is shown in dependency of case numbers N_m of the micro-types.



d) The difference between the Probit estimators and the upper boundary of a 99 %- percent confidence interval is shown for all micro-types, for which it is positive, in dependency on case numbers N_m.



e) The difference of the Probit estimators and the lower bound of a 99 %- percent confidence interval is shown for all micro-types, for which it is negative, in dependency on case numbers N_m.

Couples 1123, 1158, 1203 with average birth rates are all married and a little older than the others. Professional status and income, nationality, region and income assume variant values. The last three couples with very high birth-rates have in common that the wives are relatively young (and that they are legally married), but they also differ from each other with respect to some other characteristics.

Table 5 illustrates that a list of micro-types can provide detailed information, but does not permit the recognition of some more general trends. To detect those, one has to look at the signs of the coefficients in Table 4, based on a Probit estimation. Could it be that micro-types with high birth rates in Table 5 are explained mainly by a superposition of favourable characteristics according to Table 4.? In order to answer this question, we consider Figure 5 which shows that, at least for dozens of micro-types, there are significant differences between the Probit and Bayesian estimators of birth-rates. As we know from Section 4, the Bayesian estimators are the more reliable ones. Hence, the birth-rates derived from Boolean regression convey more information than those from Probit. They cannot simply be interpreted as a simple superposition of general trends as can already be seen in Tab. 4.

Let us approach this point from still another direction. Imagine, we are concerned with the question of how to increase birth-rates in Germany. Looking at the Probit estimates, this question can easily be answered at first glance, simply by setting the variables with positive coefficients in Tab. to 1 and those with negative coefficients to 0. However, this approach is not very useful. We cannot make German people younger or older at will. Also, altering incomes is not a very plausible solution, unless one considers the association of income with profession and education. Accordingly, we must simulate the situation more carefully. The basic ideas are firstly, that some variables must not change at all. Secondly, variables cannot change in an isolated manner. Instead, only a change from one micro-type to another that is empirically plausible, is possible. Therefore, we allow only micro-types as the aim of a transition that can be found in our data.

Accordingly, we proceed as follows. The following variables must not change:

ALTERF1, ALTERF2, ALTERF3, ALTERF4, EHE, ALTERM1, ALTERM2, DEUTSCHF, DEUTSCHM, NBL.

By fixing the age-variables, we avoid meaningless results as that birth-rates in Germany would rise drastically if women were younger. Marital state plays an important role for birth-rates. If we permitted the variable EHE to change from 0 (unmarried) to 1 (married), we could also simulate an enormous rise in birth-rates. However, it seems more likely that people do not get married because they do not want to have children.[34] Therefore, we do not permit the variable EHE to change. That the average birth-rates in Germany would increase if East-Germans were West-Germans, would represent a simplistic insight. Rather, we are interested in the reasons why East-Germans have such low birth-rates. Therefore, the variable NBL is not permitted to change. Finally, past experience demonstrated, and it is to be expected for the future that immigrants adapt to German behavioural patterns. For this reason, the variables DEUTSCHF and DEUTSCHM are also not permitted to change. The remaining variables (compare with Table 5) reflect mainly socio-economic reasons. Hence, the proposed simulation concentrates on socioeconomic causes of low or high fertility.

In the next step, we set a parameter w, the number of variables which are allowed to change. We then make a simulation as follows. Each micro-type T_m may change to a micro-type T_m', where T_m' is that type with the highest possible estimated birth rate that can be achieved with not more than w variables changing. It should be emphasised that a change is only allowed between micro-types that exist empirically, which refers, in our case, to micro-types T_m, T_m' that occur with case numbers $N_m>0$ and $N_m'>0$ in our data set. We define the relative increment of birth rates by RIBR:

$$\text{(5-1)} \quad RIBR = \frac{Simulated\ number\ of\ births}{Number\ of\ births\ before} - 1 \quad .$$

Sample size is 420 758 and number of births before is 70 008 if we allow for all micro-types with $N_m>0$. Sample size is 232 489 and number of births before is 35 847 if $N_m\geq30$ is requested. The simulated number of births can be yielded either by Bayesian estimates or by Probit estimates. A comparison of Bayesian-based and Probit-based results for RIBR is given in Table 6, depending on the different values of w, for the sample of micro-types with $N_m\geq30$.

Tab. 6 Simulated relative increments in births (RIBR) in the sample with $N_m\geq30$ depending on the numbers of allowed variable changes w

w	1	2	3	4	5	10
Bayesian est.	24,5 %	47.3 %	61,9 %	74,4 %	84.7 %	113.5 %
Probit est.	7,2 %	22,5 %	30,0 %	34,3 %	39,1 %	59,7 %
Lower bound Bay. est.	3,2 %	9,5 %	13,2 %	17,8 %	22,6 %	42,8 %

It can clearly be seen that the simulated increment of birth rates is greater if we work with Bayesian estimators rather than Probit estimators. Using the differences in micro-types in fitting detail yields higher total birth rates than only using general trends based on Probit results. That high or low birth rates are not just the additive outcome of general trends can also be seen from Table 7.

[34] Evidence to support this supposition is given in Höhn et al. (2006, pp. 32-33). Brose (2006, p. 275), however, argues that being married leads to greater fertility.

Tab. 7 Number of variables switching during a simulation

	Column N°.	1	2	3	4	5	6
		w=5					
		Bayesian			Probit		
N°.	**Variable**	**#(0→1)**	**#(1→0)**	**# net**	**#(0→1)**	**#(1→0)**	**# net**
8	MITHM	0	0	0	0	0	0
9	UNTM	42	77	-35	47	34	13
10	ICHAGM	23	33	-10	138	0	138
11	ARBM	207	274	-67	131	210	-79
12	NETM	74	80	-6	44	72	-28
13	ANGM	353	177	176	167	131	36
14	BAUER	57	12	45	167	3	164
15	BEAM	39	89	-50	0	72	-72
16	CLASSM1	34	21	13	45	15	30
17	CLASSM2	50	321	-271	3	492	-489
18	CLASSM3	275	245	30	298	485	-187
19	CLASSM4	272	78	194	380	103	277
20	CLASSM5	38	46	-8	211	3	208
21	AZUBIM	0	5	-5	0	5	-5
22	UNIF	201	55	146	441	3	438
23	SOLDATM	0	3	-3	0	3	-3
24	ASUM	78	64	14	56	45	11
25	FHSF	178	11	167	9	27	-18
26	TETM	0	0	0	0	0	0
27	FSF	27	156	-129	0	130	-130
28	LEHREF	214	377	-163	294	239	55
31	BEFRM	28	41	-13	0	42	-42
32	ABIF	430	42	388	457	19	438
33	UNIM	96	91	5	497	2	495
34	WALCM	135	167	-32	101	63	38
35	FSM	77	190	-133	33	219	-186
36	LEHREM	158	339	-181	145	269	-124
37	POLYTECM	28	76	-48	7	171	-164
39	BOOM	365	257	128	778	19	759
	Σ	3479	3327	152	4449	2876	1573
	Σ | |			2612			6200

#(0→1) refers to the number of cases in which the variable changes from 0 to 1 during the simulation. #(1→0) refers to the number of cases in which the variable changes from 1 to 0 during the simulation. # net = #(0→1) - #(1→0). w is the greatest number of variable changes allowed during the simulation when changing one micro-type T_m to another micro-type T_m'.

During the simulation described above, a number v(m,m') ≤w of variables changes for each transition $T_m \rightarrow T_m$'. The number of cases that a certain variable changes in total during all these transitions is listed in Table 7. Columns 1 and 4 show the number of changes from 0 to 1, Columns 2 and 5, the number of changes from 1 to 0. Columns 3, 6 and 9 show the corresponding net effects. For the simulation based on the Probit estimators, one would expect the variable to move only in one direction, i.e. either Column 4 or Column 5 should be zero for each variable. However, not all variables are permitted to change and not all combinations of variables have an empirically detected counterpart, so that, even for Probit estimators, the variables do not switch in a unique sense.

The same observation holds for the Bayesian estimators, for the same reasons and for a further reason as well:
A simulation based on the Bayesian estimators makes use of local, globally unobservable differences in birth-rates. Looking at Columns 3 and 6 of Table 7, we can see that global measures for improving birth-rates, as suggested, for example, by the sign of the Probit coefficients in Table 4, might work in the expected direction, but that they are probably not the most efficient way to bring about greater birth-rates. Interventions specific to single micro-types will have a greater effect. Finally, by comparing the sums of variable changes, and the sum of absolute values of variable changes in Table 7, it becomes evident that such interventions would need significantly less variable switching if they were based on the Bayesian estimators. Accordingly, using them would be more efficient in the sense just described.

To conclude, differences in birth-rate levels that are not globally predictable, have a remarkable magnitude. It is worth estimating the relationships by means of many Bayesian estimators and not only through a simplifying Probit procedure.

Now let us look again at the suggested increments in birth-rates in Table 6. For Bayesian estimators with w=5 or w=10, we could produce a relative increment of 84.7 % or 113.5 %. If we took the current German fertility rate of about 1.4 as a base , we would attain an impressive fertility rate of 2.6 or even 3, last time seen in Germany at the peaks of the post-war baby booms. Compare to this Figure 1.

The numbers shown in the first line of Table 6 should rather be taken as an upper bound of what might be possible. There are two reasons for this supposition. The first is that we certainly should take real world frictions into account. The second reason stems from stochastics. The latter requires some more detailed explanation.

For the simulation, we use birth-rate estimators $\widehat{k}_m$. These are random variables. Approximating the Beta-distribution by means of a normal distribution, which seems to be justified for $N_{m\geq}30$, this can be represented by Equation (5-2).

$$\text{(5-2)} \quad \widehat{k}_m = b_m + \sigma_m \cdot \varepsilon \qquad\qquad \varepsilon \cong N(0,1)$$

with the “true” value b_m of birth-rate in micro-type T_m and a standard normal distributed random variable ε. During the simulation, for a given micro-type T_m, we seek the greatest attainable $\widehat{k}_m'$. It must be expected that, by selecting a $\widehat{k}_m'$ this way, we not only find a micro-type T_m’ with a high true value b_m’, but also one for which ε assumes a high value. We must therefore conclude that b_m’ is generally lower than $\widehat{k}_m'$. Assume, for a given micro-type T_m, that there are n micro-types with a $\widehat{k} > \widehat{k}_m$. The greatest of these is $\widehat{k}_m'$. To find a careful correction of $\widehat{k}_m'$, we assume that the ε belonging to $\widehat{k}_m'$ is also the greatest of all. The greatest value $\bar{\varepsilon}$ of n realisations of a standard normal distributed random variable follows an extreme value distribution with the density function:

$$\text{(5-3)} \quad f(\bar{\varepsilon}) = n \cdot \varphi(\bar{\varepsilon}) \cdot \Phi(\bar{\varepsilon})^{n-1} \ .$$

Let be ε' the mode of the density function (5-3). Using (5-2), we obtain a plausible lower bound for the b_m' by:

(5-4) $b'_m = \widehat{k}'_m - \varepsilon' \cdot \sigma'_m$.

If the simulation is redone using (5-4), we obtain the results shown in the last line of Table 6. As expected, we obtain lower simulated birth-rates. However, they are still considerable. A rise in German birth-rates of 42.8 %, as predicted for w=10, would lead to a fertility rate of 2, enough for reproduction through own births. As the last line of Table 6 must be considered rather as the lower bound of what is possible, our simulation shows that fertility rates above than 2 could be obtained if economic conditions improved and if there was a fine tuning of measures according to the micro-types in German society.

Tab. 8 Simulated relative increments of birth-rates RIBR in the sample with $N_m \geq 1$, depending on the numbers of permitted variable changes w

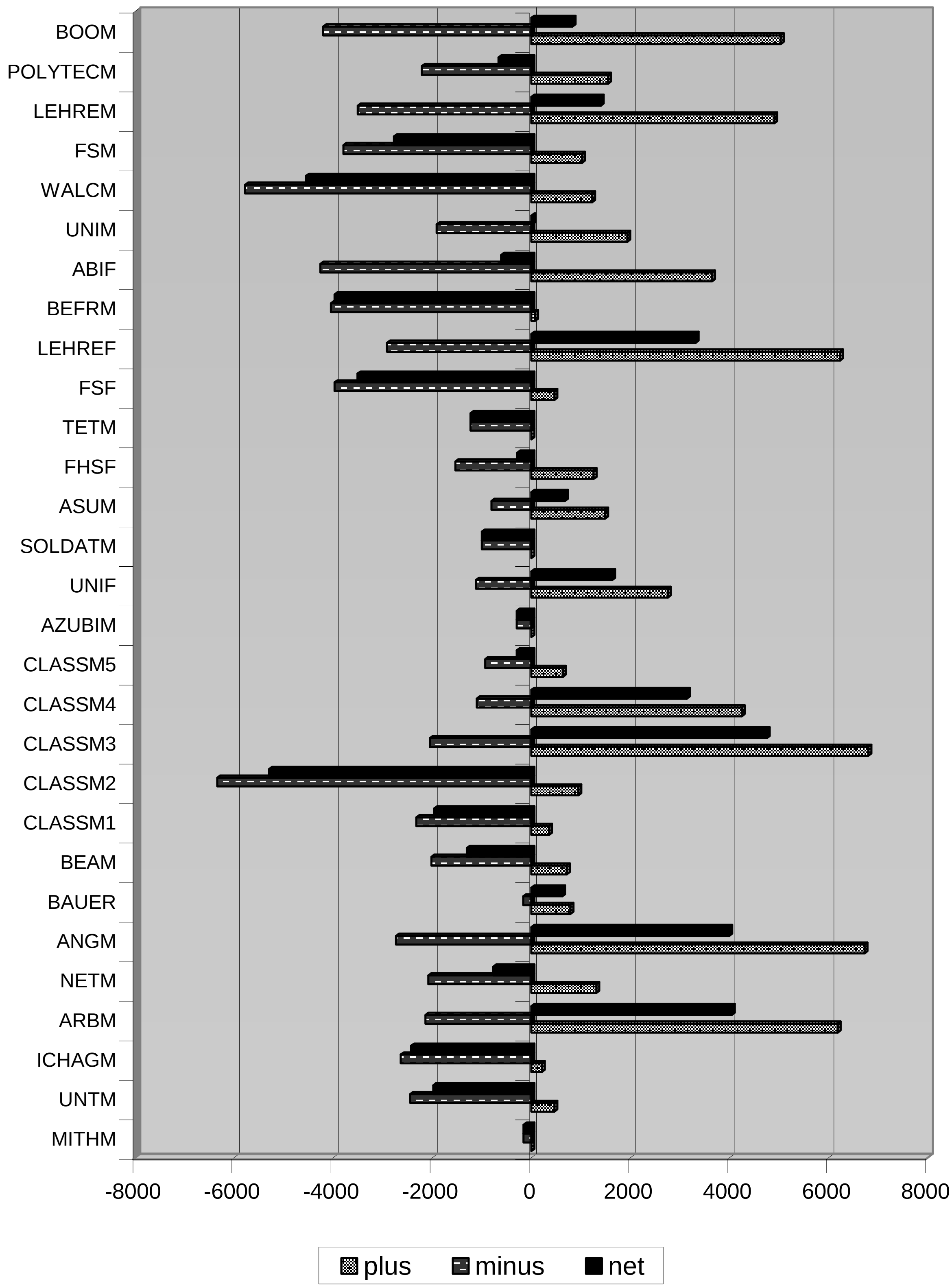
w	5	10	15
Lower bound hybrid estimator	38,7 %	60,1 %	60,1 %

We now investigate this in more detail. In order to do so, we use the entire data set of all micro-types with $N_m>0$. There are 72935 micro-types with $N_m>0$. However, a change in one micro-type to a better one is only allowed if $N_m \geq 30$ holds for the target micro-type. There are 1849 micro-types with $N_m \geq 30$. The results for the RIBR, depending on the numbers of changes permitted for w, are shown in Table 8. We use the hybrid estimators h described by Equation (4-7). The estimated birth rates of the target micro-types are adjusted downwards as described by equation (5-4). It can be seen that, for more than w=10 permitted changes in the variables, no more gains in the estimated overall birth-rates are yielded. A RIBR of about 40 % would raise German birth-rates to a value of about 2. Therefore, in the analysis below, we restrict ourselves to the case w=5, and outline the results based on this assumption.

Figure 6 or Table 9 show the number of changes in each variable during the simulation with w=5. The first remarkable fact is that none of the variables changes only into one direction. This is again an indication that the Probit results given in Table 4 might be misleading.

Furthermore, it is clearly evident that the level of income shifts from lower to higher classes. Birth rates in Germany would be higher if men earned more, at first glance, a definite confirmation of Malthus' ideas. However, Figure 6 reveals more. The shift in professional status is not in the same direction. The groups of entrepreneurs and civil servants who normally earn fairly well, will decline in numbers, and the groups of workers and employees gain members. High earnings also normally presuppose high levels of human capital. However, there is no evidence of this among men; vocational training gains (LEHREM), those with elaborate vocational training decline in numbers, the group with university degree (UNIM) remain unchanged. Another way to increase income is to increase the number of hours worked. Figure 6, however, provides evidence to the contrary. The group of men working more than 45 hours per week (WALCM) yields impressive net losses. On the other hand, working too few hours also leads to lower birth rates, as can be seen by looking onto the net balance of the variables TETM and NETM, part time employed and non working men.

Fig. 6 Number of variables switching during a simulation with all micro-types with $N_m \geq 1$ and all micro-types with $N_m \geq 30$ as the target. Number of permitted changes w=5.



Tab. 9 Number of variables switching during a simulation with all micro-types with $N_m \geq 1$ and all micro-types with $N_m \geq 30$ as the target. Number of permitted changes w=5.

N°.	Variable	#(0→1)	#(1→0)	# net
8	MITHM	0	159	-159
9	UNTM	465	2474	-1982
10	ICHAGM	209	2636	-2427
11	ARBM	6185	2140	4045
12	NETM	1310	2080	-770
13	ANGM	6726	2735	3991
14	BAUER	789	166	623
15	BEAM	714	2020	-1306
16	CLASSM1	357	2321	-1964
17	CLASSM2	948	6343	-5295
18	CLASSM3	6803	2052	4751
19	CLASSM4	4248	1106	3142
20	CLASSM5	639	933	-294
21	AZUBIM	0	296	-296
22	UNIF	2754	1122	1632
23	SOLDATM	0	1000	-1000
24	ASUM	1486	807	679
25	FHSF	1255	1537	-282
26	TETM	0	1228	-1228
27	FSF	465	3974	-3509
28	LEHREF	6233	2916	3317
31	BEFRM	73	4047	-3974
32	ABIF	3648	4259	-611
33	UNIM	1936	1912	24
34	WALCM	1224	5778	-4554
35	FSM	1023	3795	-2772
36	LEHREM	4904	3505	1399
37	POLYTECM	1547	2210	-663
39	BOOM	5031	4202	829

However, again, a man looking for work (ASUM) is favourable for fertility, whereas a man having insecure employment discourages fertility. The phase of economic cycle (BOOM), also exerts a rather modest influence on birth-rates. It seems reasonable that short-term fluctuations cannot be decisive for a “long-termed investment” as a child.

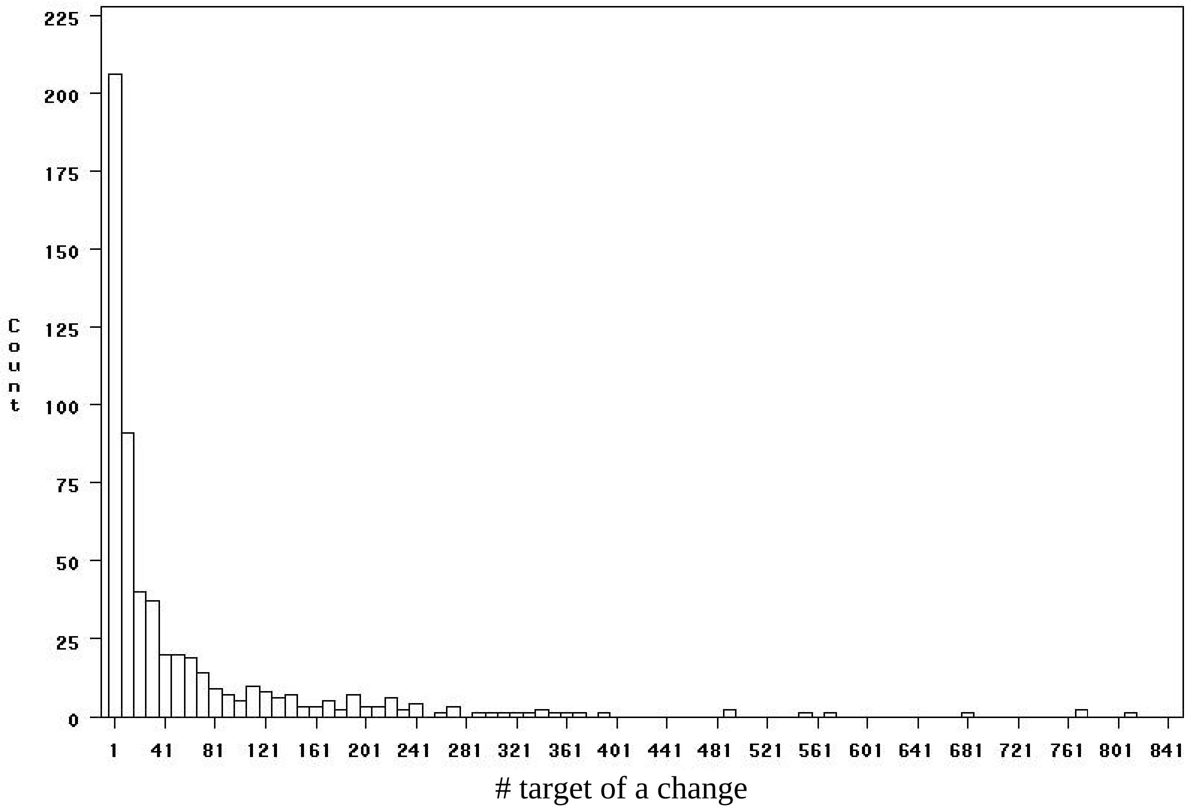
Therefore, we gain an approximate picture of the characteristics men should have if they are to be chosen by women as possible fathers of their children. They should be busy, but not too much so. They should also have secure employment. They should earn well, but without working too hard, either in terms of actual working hours, or past investment in human capital. In short: children are born where and when life is easy and the prospects for the future are good – largely in line with the ideas of Malthus, Easterlin and Colinvaux.

What about Becker's position that women will have fewer children, the greater their level of human capital? Instead of this direct inverse connection, we believe rather that it would be U-shaped. The low degree LEHREF and the high degree UNIF increase in number, the more average degrees FSF, ABIF, FHSF decrease in number, or remain more or less equal. It can be supposed that the income effect on children outweighs the price effect on children for high levels of human capital.[35]

Not only an overview of the variables that change during the simulation is of interest, but also a look at the micro-types that are the objectives of changing from one micro-type to another. There are 560 micro-types that occur as attractors during our simulation. Many of them attract only a few of other micro-types. The measure "# target of a change" indicates how many micro-types changed to the indicated micro-type. Figure 7 shows that some micro-types attract several hundreds of other micro-types. Table 10 shows the 12 most "attractive attractors" during the simulation described above.

The first potentially striking observation one can make is that not only micro-types with high birth-rates are found among the most attractive ones. Types A, C, F, K and L have birth-rates far below the average. What they all have in common is that the mother is older than 40. Therefore, from a demographic point of view, our attention is directed to the fact that birth rates would be increased, if more people beyond the thirties became parents. Surely, from a doctor's point of view, this might not be the best advice. Therefore, this point is left open in the context of this paper. Nevertheless, our method yields some rare insights.

Fig. 7 Histogram of absolute frequencies of attractors during simulation by # target of a change



[35] See again Hufnagel (2008a).

Tab. 10 The 12 most attractive micro-types during simulation

	Micro-type	A	B	C	D	E	F	G	H	I	K	L	M
	m	32424	72372	24087	66333	72199	26671	49781	72912	72823[1)]	31144	26669	72807
	Birth-rate	.10	.60	.07	.46	.59	.08	.25	.70	.65	.10	.08	.64
	# target of a change	806	772	767	685	573	555	495	492	387	370	356	351
1	ALTERF3	0	1	0	0	0	0	0	0	1	0	0	1
2	ALTERF2	0	0	0	0	1	0	0	1	0	0	0	0
3	ALTERF1	0	0	0	0	0	0	0	0	0	0	0	0
4	ALTERF4	0	0	0	1	0	0	1	0	0	1	0	0
5	EHE	1	1	1	1	1	1	1	1	1	0	1	1
6	ALTERM2	0	1	0	1	0	0	0	1	1	1	0	0
7	ALTERM1	0	0	0	0	1	0	0	0	0	0	1	1
8	MITHM	0	0	0	0	0	0	0	0	0	0	0	0
9	UNTM	0	0	0	0	0	0	0	0	1	0	0	0
10	ICHAGM	0	0	0	0	0	0	0	0	0	0	0	0
11	ARBM	0	0	1	0	1	1	0	0	0	1	0	0
12	NETM	0	0	0	0	0	0	0	0	0	0	0	0
13	ANGM	1	1	0	1	0	0	1	1	0	0	1	0
14	BAUER	0	0	0	0	0	0	0	0	1	0	0	0
15	BEAM	0	0	0	0	0	0	0	0	0	0	0	1
16	CLASSM1	0	0	0	0	0	0	0	0	0	0	0	0
17	CLASSM2	0	0	0	0	0	0	0	0	0	0	0	0
18	CLASSM3	1	0	1	0	1	0	0	0	0	1	1	1
19	CLASSM4	0	1	0	1	0	1	1	1	0	0	0	0
20	CLASSM5	0	0	0	0	0	0	0	0	0	0	0	0
21	AZUBIM	0	0	0	0	0	0	0	0	0	0	0	0
22	UNIF	0	0	0	0	0	0	1	0	0	0	0	0
23	SOLDATM	0	0	0	0	0	0	0	0	0	0	0	0
24	ASUM	0	0	0	0	0	0	0	0	0	0	0	0
25	FHSF	1	0	0	1	0	0	0	0	0	0	0	0
26	TETM	0	0	0	0	0	0	0	0	0	0	0	0
27	FSF	0	0	0	0	0	0	0	0	0	0	0	0
28	LEHREF	0	1	0	0	1	0	0	1	1	0	0	1
29	NBL	0	0	1	0	0	0	0	0	0	0	0	0
30	DEUTSCHF	1	1	1	1	1	1	1	1	1	1	1	1
31	BEFRM	0	0	0	0	0	0	0	0	0	0	0	0
32	ABIF	1	1	0	1	0	0	1	0	0	0	0	0
33	UNIM	0	0	0	0	0	0	1	0	0	0	0	0
34	WALCM	0	0	0	0	0	0	0	0	1	0	1	0
35	FSM	0	0	0	0	0	0	0	0	1	0	0	0
36	LEHREM	0	0	0	0	1	0	0	1	0	1	1	1
37	POLYTECM	0	0	0	0	0	0	0	0	0	0	0	0
38	DEUTSCHM	1	1	1	1	1	1	1	1	1	1	1	1
39	BOOM	0	1	0	1	0	0	1	1	0	0	1	0

[1)] The German Micro-census does not collect data on the income of farmers.

The other attractors in Table 10 have very high birth-rates. Among them is Type I, a farming couple. As Becker (1981) pointed out, from a theoretical point of view, it is very likely that farmers would have many children, because food is abundant and children can be a big help around the farm. In a direct sense, birth-rates in Germany would then logically be higher if the society were more agrarian. Micro-type I may not seem very helpful, but in the fairly indirect sense that birth-rates would be higher if rearing children were cheaper, it surely is.

Looking at the attractors in Tab. 10, we can identify two general trends.

Firstly, life is relatively secure. Nearly all couples are married, no man is a soldier, and none of the men, besides the farmer, is an entrepreneur. None is looking for work or temporarily employed.

Secondly, income is above average and earned without large investments in current or past work. Note that all couples belong to Income Classes 3 or 4, that only in two cases the weekly hours of work are greater than 45 and that the level of human capital is relatively low, in relation to the above-average incomes that are earned. There is only one couple with a university degree among the list of Table 10, but there are many which have just completed some vocational training.

Therefore, all in all, we obtain the same result as suggested by the earlier interpretation of Table 9. Birth rates in Germany would be increased, if people were economically better off, if their lives were more secure and if they were likely to become more prosperous over time.

The results of this section will be considered again in the following final section, the conclusions of this paper. In addition, the methodological and practical implications will be analysed.

6. Conclusions

In this concluding section, we briefly resume the discussion on the possible consequences for social policy in Germany which were outlined at the end of Section 5. We then discuss the consequences for methodology that arise from the insights gained in Sections 4 and 5 in Subsections 6.2 and 6.3.

6.1 How to Increase German Birth Rates – Economic Determinants

The Social-Democrat/Ecologists Government, elected in 1998 and again in 2002, redefined family policy fundamentally. Family policy and gender policy were mixed. Issues such as the extension of public child-care facilities and reconciling the needs of the job to the needs of the family were seen as central instruments of family and population policy. Politicians could base these ideas on expertises as BMFSFJ (2005) and Höhn et al. (2006). The notion was proposed of abolishing the advantages couples have due to the German income tax system in order to create resources to finance the planned new form of family politics. The Christian/Social-Democrat government, which was elected in 2005, is currently pursuing this approach. Again, conflicts on the distribution of resources and money are involved, as the more social-conservative wing of the Christian democrats is demanding compensations in the form of higher direct subsides to families, which do not wish to use the new public child care facilities which are to be established.

It might be well true that delays in German gender policy are one cause of the low fertility rates in Germany. That it cannot be the only cause can be seen by simply looking again at Figure 1. The collapse of birth-rates in Eastern Germany during the Nineties is as substantial as the collapses after the World Wars and during the Great Depression, even though public child care facilities were available, at least to a much greater extent than in Western Germany. Therefore, it is easy to suspect that economic conditions count as well. However, to try to provide an econometric proof with SOEP-data is neither easy nor particularly convincing.[36] In the present investigation, it was possible to demonstrate the correlation between economic conditions and fertility rates. On the one hand, this is due to the recently available data set, on the other hand to a new method, that of combining Probit and BR. An impressive increment in German fertility rates would be possible, if people could gain more by means of less disutility. In terms of this view, low fertility rates in Germany are not only due to the rising levels of female education and work force participation, but also to the economic stagnation since the 1980'. From a macro-economic point of view, declining birth rates in Germany are just part of a broader downward process, with respect to which it is difficult to discern cause and effect. Modest economic conditions lead to stagnating incomes, which lowers demand, in turn lowering the number of births, which again lowers demand, leading to the wrong size of infrastructure and a less innovative society. This lowers the number of net contributors to the social system, what raises the costs of labour, thus increasing the number of jobless again and so on.

The key result of our investigation is hence that fertility is at least related to the same degree to macro-economic conditions as it is related to family-labour constellations. Therefore, family and population policy should not only focus on the Becker framework and its applications, but also on the Malthus-Easterlin framework and the interrelations between population economics and macroeconomics.

[36] See Hufnagel (2004) and Witte, Wagner (1995).

Accordingly, the results of this paper seem to be of considerable practical interest. We should meticulously revise the methodology used, without the "interruptions" caused by the tables and equations contained in Sections 4 and 5 of this paper.

6.2 Consequences for Methodology

Causal explanations are generally formalized using the language of mathematics. Mathematical equations describe invariants, that is, relationships which do not change. Knowing invariants is central for technological interest. It is often the basis for deriving rules for intervention. Take, for example, Ohm's law:

(6-1) $$R = \frac{U}{I}.$$

This tells us in words, that the quotient of tension and current intensity is constant as long as resistance is constant. A simple transposition yields:

(6-2) $$I = \frac{U}{R}.$$

This equation tells us that, for a given tension, electric current I can be regulated by resistance, because there is an inverse proportionality. On the basis of this invariant, we are able to construct a dimmer, for example. Now, from a mathematical point of view, Equation (1) is a very simple rule. In general, mathematics has established that true sentences can be simple or more complicated and finally, that there is no upper bound for the complexity of true sentences.[37]

The aim of empirical investigation is to find invariants in reality. Mathematical invariants can be less or more complex. Therefore, we should suppose that - at least a priori - empirical invariants can also be more or less complex. Certainly, we would prefer simple invariants, but if we cannot have both simultaneously, we should nevertheless accept per se an invariant that has been found. Although we tend to expect that laws are easy to communicate, this must not seduce us into rejecting instrumentally worthy results for their complexity. In the introduction of this paper, we mentioned that BR seems the natural way to perceive complex invariants in empirical observations, but that there are specific difficulties in application. In particular, we needed a) a measure of how well the estimated function fits the data; b) an algorithm to find the best fitting function; c) a measure of the significance of the estimated coefficients a_m.

Concerning b), the calculations involved are feasible. On a 1,5 GHz-PC, they were conducted within a few hours, including data preparation and the simulations at the end of Section 5.

Concerning a), the first available choice is to try to maximize hit-rates. For this purpose, the probabilities that for a given micro-type T_m, the depending variable y will take the value 1, i.e. π_m=Prob(y=1|T_m=1), must be estimated as accurately as possible. We tried out three kinds of estimators: Laplacian, Bayesian and Probit. It turned out that which one is best, depends on N_m, the number of cases with T_m=1. We found that there is a specific threshold TH. If N_m is lower TH, Probit estimators are superior to Bayesian ones and these are again better than Laplacian ones. If N_m is greater TH, there is little difference between Laplacian and Bayesian

[37] For more on this topic, see, for example, Barrow (1992) or (1994).

estimators, but both are better than Probit. It is a simple consequence that a hybrid estimator, relying on Probit for a low N_m and on Bayesian (or Laplacian) estimators for a high N_m, will yield the most reliable values for Prob($y=1|T_m=1$). Given these estimators and their probability density function, Problem c) can be solved. We set $a_m=1$, if $\pi_m \geq ½$, and the measure of significance will be α_m, the probability that $\pi_m <½$ in the basic sample, although it was greater than ½ in the random sample under investigation. Knowing the probability density of π_m, α_m is found easily by means of integration.

Thus, the insights gained in this paper suggest proceeding as follows to exert a Boolean regression.

I. Problem

Let be given N observations of n BV x_i (regressors) and of one depending BV y. Find the 2^n Boolean coefficients a_m so that the number of observations with $\varepsilon^{[v]}=0$ in

(2-14) $$y^{[\nu]} = \bigoplus_m a_m \cdot T_m(x_1^{[\nu]},\dots,x_n^{[\nu]}) \oplus \varepsilon^{[\nu]} \qquad \nu = 1,\dots,N$$

is maximized, i.e. number of hits is maximized.[38]

II. Split the data set into two halves. For the first half, evaluate the Bayesian, Laplacian and Probit estimators of π_m.

III. Use the π_m to predict actual frequencies in the second half of the data set. Determine a threshold TH, where the Laplacian and Bayesian estimators start to become superior to Probit estimators.

IV. Knowing TH, evaluate hybrid estimators for π_m and their levels of significance α_m for the entire data set.

V. Put $a_m=1$ if $\pi_m \geq ½$.

The method suggested and its results require a few comments.

<u>1. Large data sets are needed</u>

Splitting the data set (Step II) and the determination of reliable Laplacian estimators require a large set of observations. Therefore, the natural field of application of BR seems to be micro-census data and other very large data sets now becoming available for socio-economic research. These include particularly the data sets of the German Institut für Arbeitsmarkt- und Berufsforschung (IAB), with millions of observations, and finally, the German Census planned for the year 2011.

[38] As shown in section 2.2, this is equivalent to a Maximum-Likelihood-approach.

2. Determination of a threshold

If the number N_m of observations for which a micro-type T_m assumes the value 1 is low, Probit yields predicitions that are superior to Laplacian or Bayesian estimators. There is a threshold TH for which Laplacian and Bayesian estimators get better than Probit estimators. For the data used in this investigation, this threshold turned out to be about 30. It was argued in Section 4 that this number must have an empirical content. The fact that the standard deviations of the estimators decrease with increasing N_m cannot not solely be responsible for this phenomenon. The magnitude of TH should depend on the smoothness of the landscape under investigation. Concerning this point, further investigations are necessary, using further data sets as mentioned above and further selections of depending variables (other than BIRTH in the present investigation). We need to wait until a catalogue of thresholds has been produced, to be able to make some more general propositions on this theme. Assuming we do indeed obtain similar results, this will be an important insight from an epistemological point of view. Probit-like neuronal nets, such as animals use it, and the way fuzzy logic works, are not only simpler and faster, moreover, there is a further advantage. If the number of observations is relatively small, it is better to rely on the interpolation of general trends for special cases, instead of on predictions of frequencies based on only a handful of observations. Given the theme of this investigation, we should restrict these extrapolations to the social sense of man. However, we do find that people often behave as suggested by our findings. People often have to act and react in social situations that they encounter the first time in a given specific constellation. In such moments, they rely on their social intuition (or their pre-judgements). It would require dozens of observations to establish judgements on the frequencies of one's own observations. Thus, it is perhaps this combination of the knowledge of many cases of different social types, gathered over long professional experience with a natural talent for summarizing social trends, that constitutes the superior expertise of outstanding psychologists, religious ministers or priests and social scientists (?).

For social science, this would mean that the now prevailing quantitative methods, relying mainly on multivariate analysis (as Probit or similar regressions) can only constitute one fundamental pillar. The other would be the field of studying types and micro-types, referred to as qualitative social science, or Institutionalism (old and new), or Historism (older and newer, German and American) often – surely depending on the personal point of view – dismissed as a heterodox substream. However, the old arguments against Historicism, the reproach of Inductivism and its incapacity to formulate a general theory from the many available case studies have to be considered. We will again refer to these items in the next section and in Section 6.3.

3. Avoiding Inductivism

It might be criticized that merely linking variables to micro-types and counting or estimating frequencies is a simple inductivist approach which inherits all its familiar weaknesses.[39]

In particular, simply looking at the common emergence of phenomena over time gives no indication as to the causes and consequences. Even worse, correlations may be caused by further intervening variables, so that the instrumental utility of the results found in this way could be dubious. Therefore, the regressors in (2-14) cannot be chosen at will or at random. The author therefore proceeded as follows. First, based on common fertility theory, only those

[39] For a comparison, see the Chapter 4 of Chalmers (2006).

variables of the micro-census which probably exerted an influence on birth-rates were selected out of all variables available in the micro-census. Subsequently, one has to examine these variables if they are really causes and omit those which are not clear in this respect. For the example given, the list of variables in Table 1 does not include women's income. It might be that there was a correlation, specifically, the lower her income, the higher her birth-rate. In this case, however, the direction of causality was not clear. Is her income low because she does not work, because there is a newly born child in the family, or had her low income merely strengthened her determination to stay at home and rear children? For this reason, this variable is omitted and instead, human capital instruments are used to capture the wage potential of women. Another example is the variable "number of the children in the household". This could be an indicator that a given family loves children and that, therefore, the probability of another birth event is high.[40] On the other hand, a family that has many children might be less inclined to have more. Because we could not clarify this with the data on hand, the variable "number of children in the household" is not included in the set of regressors. Secondly, with the remaining set of regressors, a Probit was conducted. Variables that turned out to be insignificant were omitted. This was done for technical reasons, because the number of regressors in BR must be kept below a reasonable magnitude, to assure that enough micro-types remain with reasonably large N_ms. However, it is also an additional filter for avoiding the occurrence of random results when doing BR.

Assuming that the described clearing of the preliminary regressor set has been done before starting Step I, BR does not constitute crude and blind inductive data mining, but a method within the realm of sophisticated falsificationism, as it acknowledges and relies on the importance of pre-knowledge.

4 Direct and probabilistic interpretation

As pointed out in Section 2, there are two ways of interpreting BR. We can use (2-14) for direct predictions. For the data set under investigation, this approach delivered poor results. It was more fruitful to interpret BR in a probabilistic manner. This means predicting not the event of y=1, but its probability π_m. For static predictions, the interpretational difference is unproblematic, because as stated above, we need large numbers of observations and hence can assume that probabilities emerge as frequencies in sufficiently large ensembles. In a dynamic context, when simulating iterations as given in equations (1-3) in the introduction, direct prediction would lead to completely meaningless results, because incorrect predictions cumulate over the time-path. This is considered in detail by Hufnagel (2008b). However, the latter paper also demonstrates that working with the dynamics of vectors of probabilities is likely to yield useful results.

Accordingly, the method for socio-economic investigations introduced and referred to as BR in this paper has potential weaknesses that have to be overcome. It should rather be applied to very large data sets, and anchored in existing theoretical frameworks. Using the results for direct predictions will yield disappointing results, but using them for probabilistic predictions allows for reliable simulations. On the other hand, it could be shown that the process offers some specifics. In Section 4 evidence was provided that, if the number of observations is large enough, BR is superior to Probit. In Section 5, it could be shown that observing the detailed results of BR could increase the value of the depending variable (in this case BIRTH) significantly more than relying on global regressor variations as suggested by Probit.

[40] Concerning attitudes towards children, see again Brose (2006).

However, the interpretation of results turned out to be difficult. An inspection of thousands of micro-types cannot be done by one human mind. Computer programmes were also needed for interpretation, using concepts such as simulation and the determination of attractors. The result of a common multivariate analysis is easy to understand. The estimated coefficients of the regressors tell us what will happen if this regressor variable is changed. For a BR as presented in this paper, such an easy interpretation is no longer possible. We have to accept that we can see phenomena we could not see before, but only if we learn to interpret them, and we have to learn new methods of interpretation. In this sense, the methods proposed are a new instrument of social observation, comparable to the introduction of telescopes, microscopes, X-ray cameras or Geiger counters. In the following section, we discuss the potential role this method could play in socio-economics.

6.3 Future Directions for the Social Sciences

A superior principle of peripatetic physics was to rely on the information provided by the human senses, especially the visual sense. As Thomas Aquino stressed, we can rely on our senses, because their natural function is to help us to survive.[41] One central specific of the scientific revolution of the 16th and 17th century was that Galilei and Kepler relied not only on the senses to proof their theses and to disprove medieval points of view, but they also used a new instrument, the telescope. A central question during the debate was whether scientists can rely on what they can only see through a telescope, for example, the moons of Jupiter. Maybe what Galilei saw was an optical illusion, caused by the instrument and not something really out there in the universe? As Chalmers describes it, it took a long process to convince most astronomers to accept Galilei's and Kepler's observation. The proof, not absolutely compelling, but plausible, was that one could direct the telescope towards a distant object such as a tower, and convince oneself afterwards that that what one had seen through the telescope conformed to reality.

Consider the role of this paper in a similar way. Applying BR to huge data sets with the help of fast computers and programs, interpreting the results by simulation, constitutes a new instrument for socio-economics.[42] We can see phenomena that we had not seen before, especially that Germany could achieve a considerably higher birth rate than the current one if life was easier.

One could say that this *merely* confirms the theories of Malthus, Easterlin and Colinvaux. Indeed, one main function of this paper should be to show that the new "telescope" reproduces well-known objects correctly. The author would like to contest the word "merely" in the above phrase, however. Section 5 leaves us with more knowledge than we had before, especially concerning the potential magnitude of the income effect on births, which is much higher than the estimates of the *Bundesministerium für Familie, Senioren, Frauen und Jugend* (2005).

We follow essentially the ideas of Holland (1995) and Kauffman if we rely on simulation when interpreting the results of our BR. One might ask whether there is perhaps a more readily understandable way of interpreting the results of BR. One way to do this is as follows.

---

[41] For some further insights into all these issues, see Chalmers (2006, p. 132 ff.).
[42] To illustrate what is meant, think back to the introduction of punch card computers into socio-economic research about 60 years ago, which for the first time, enabled the application of regressions to hundreds and thousand of observations (Pesaran 1991).

Take two micro-types T_m and T_m' with estimated frequencies π_m and π_m'. Test whether $\pi_m=\pi_m$' in the basic sample. Then T_m and T_m' can b e joined to a shorter micro-type T_m''=$T_m \vee T_m$'.

E.g.:

$$T_m = x_1 \cdot (1 - x_2) \cdot x_3 \qquad T_m' = x_1 \cdot x_2 \cdot x_3 \qquad \rightarrow \quad T_m'' = x_1 \cdot x_3 \,.$$

Depending on how hard the test criterion is chosen, the number of micro-types will reduce to a greater or lesser extent. In order to understand the result, it would be comforting to end with only a handful of micro-types. However, it is necessary to clarify how much information would be lost when proceeding in this manner.

The author's conviction is that to conceive complexity, new methods are needed, such as those tried out in Section 5 of this paper. Recently published monographs suggest thinking of a world whose laws cannot be reduced to or deduced from a handful of simple principles, be it in the field of science (Laughlin 2005) or of economics (Werner 2003). If we continue thinking along these lines, there might be more than a handful of regular occurrences, but an enormous number remain, waiting to be detected and prepared for instrumental applications. Difficulties such as those encountered in this paper, show that the historists who thought along similar lines and to whom Werner explicitly refers, could not achieve their objectives, because the appropriate instruments were not available to them, namely large data sets and computers.

In this context, the present paper is surely but a first step. However, the results seem promising enough to try out in a further step in some similar investigations. If successful, we will have more examples, also working on known themes with large data sets, that will not only reproduce known results but yield new or greater insights. Such a calibration of the new instrument might be necessary to broaden its use in socio-economics. Useful and productive results, which enhance our understanding of social life and enable instrumental applications in social policy, may be expected.

References

Arthur, B., Durlauf, St., Lane, D. (1997, eds.), The Economy as an Evolving Complex System; Santa Fe.

Baltagi, B. (2002), Econometrics; 3rd Edtion, Berlin et al.

Barrow, J. (1992), Pi in the Sky; Oxford.

Barrow, J. (1994), Ein Himmel voller Zahlen; Heidelberg, German traduction of Barrow (1992).

Becker, G. (1981), A Treatise on the Family; Harvard.

E. Boros, Hammer, P., Hooker, J. (1995), Boolean Regression; Annals of Operations Research 58, pp. 201-226.

Braumoeller, B. (2003), Causal Complexity and the Study of Politics; Political Analysis 11, pp. 209-233.

Braumoeller, B. (2004), Boolean logit and probit in Stata; The Stata Journal 4/4, pp. 361-366.

Brose, N. (2006), Gegen den Strom der Zeit? – Vom Einfluss der religiösen Zugehörigkeit und Religiosität auf die Geburt von Kindern und die Wahrnehmung des Kindernutzens; Zeitschrift für Bevölkerungswissenschaft 31, pp. 257-282.

Bundesministerium für Familie, Senioren, Frauen und Jugend (2005, ed.), Perspektive für eine nachhaltige Familienpolitik – Ergebnisse des Gutachtens von Prof. Dr. Eckart Bomsdorf; Berlin.

Chalmers, A. (2006), Wege der Wissenschaft; 6th ed. German translation of the 3rd ed. of "What is this thing called science?", Queensland 1999, Berlin et al. Traduced by N. Bergemann and C. Altstötter-Gleich.

Colinvaux, P. (1980), The Fates of Nations; Harmondsworth et al.

Easterlin, R. (1973), Relative Economic Status and the American Fertility Swing; *Sheldon* (ed.), Family Economic Behavior: Problems and Prospects; Toronto et al., pp. 170-223.

Eichner, A. (1986), Can Economics Become a Science? Challenge 29/5, pp .4-12.

Evans, M., Hastings, N., Peacock, B. (2000), Beta Distribution; Statistical Distributions, Wiley, New York, 3rd ed., pp. 34-42.

Frenken, K. (2006), Innovation, Evolution and Complexity Theory; Cheltenham et al.

Höhn, C., Ette, A., Ruckdeschel, K. (2006), Kinderwünsche in Deutschland, Konsequenzen für eine nachhaltige Familienpolitik; herausgegeben vom Bundesinstitut für Bevölkerungsforschung und der Robert Bosch Stiftung.

Holland, J. (1995), Hidden Order; Reading.

Hufnagel, R. (2000), Die Emergenz von Ordnung im Haushalt; Hauswirtschaft und Wissenschaft 2/2000, pp. 110-122.

Hufnagel, R. (2002), Die Kosten von Kindern und die Kosten einer egalitären Elternschaft; DIW Vierteljahreshefte zur Wirtschaftsforschung 1/2002, pp. 114-125.

Hufnagel, R. (2004), Leere Wiegen Ost; *Hufnagel, R., Simon, T.* (eds.), Problemfall deutsche Einheit - Interdisziplinäre Betrachtungen zu gesamtdeutschen Fragestellungen; Wiesbaden, pp. 147-154.

Hufnagel, R. (2007), Fertilitätsentscheidungen in Mikro- und Makroperspektive - Das Individuum zwischen Genen und Institutionen; *Höflacher, S., Hufnagel, R., Jaquemoth, M., Piorkowsky, M.* (eds.), Oikos 2010 – Haushalte und Familien im Modernisierungsprozess; Bonn, Göttingen, pp. 159-222.

Hufnagel, R. (2008a), Kinderwunsch und Partnerwahl in Deutschland; Hauswirtschaft und Wissenschaft 1/2008, pp. 5-23.

Hufnagel, R. (2008b), Entropy and Stability in the Time Use of Private Households; Electronic International Journal of Time Use Research 5, to be published soon.

Judge, G, Hill, C., Griffiths, W., Lütkepohl, H., Lee, T. (1988), Introduction to the Theory and Practice of Econometrics; 2nd edition, New York et al.

Kauffman, S. (1995), At Home in the Universe; New York.

Laughlin, R. (2005), A Different Universe – Reinventing Physics from the Bottom Down; New York.

Macunovich, D. (1998), Fertility and the Easterlin Hypothesis: An Assessment of the Literature; Journal of Population Economics 11, pp. 53-111.

Malthus, T., (1803), An Essay on the Principle of Population; 1803, 7th edition 1872, Everyman's Library Edition, London 1973.

Mincer, J. (1974), Schooling, Experience and Earnings; New York.

Pesaran, H. (1991), The ET Interview: Professor Sir Richard Stone; Econometric Theory 7, pp. 85-123.

Quine, W. (1962), Mathematical Logic; (Revised edition), San Francisco.

Quine, W. (1974), Grundzüge der Logik; Frankfurt a. M.

Ruczinski, I, Kooperberg, C., Leblanc M. (2002) Logic Regression; *Denison, D., Hansen, M., Holmes, C., Mallik, B., Yu, B.* (eds.), Methods and Software; Proceedings of the MSRI Workshop on Nonlinear Estimation and Classification, Berlin et al., pp. 333-344.

Rudeanu, S. (1974), Boolean Functions and Equations; Amsterdam et al.

Spitzer, M (2000), Geist im Netz. Modelle für Lernen, Denken und Handeln; Heidelberg et al.

Schultz, P. (1973), A Preliminary Survey of Economic Analyses of Fertility; American Economic Review 63, No. 2, Papers and Proceedings of the Eighty-fifth Annual Meeting of the American Economic Association (May, 1973), pp. 71-78.

Tucholsky, K. (2002), Kurzer Abriß der Nationalökonomie; *Gerold-Tucholsky, M.* (ed.), Kurt Tucholsky – Panter, Tiger & Co. Eine neue Auswahl aus seinen Schriften und Gedichten; Hamburg, 53th edition, pp. 194-197.

Werner, A. (2003), Princes of the Yen: Japan's Central Bankers and Monetary Policy Making; Armonk.

Witte, J., Wagner, G. (1995), Declining fertility in East Germany after unification: A demographic response to socioeconomic change; Population and Development Review 21, pp. 387-397.

Woolridge, J. (2006), Introductory Econometrics – A Modern Approach; 3rd edition, Mason.

Zimmerman, H., Zadeh, L., Gaines, B. (eds.,1984), Fuzzy Sets and Decision Analysis; Amsterdam et al.

Diskussionsbeiträge des Institutes für Ökonomische Bildung

(bisher erschienen)

Diskussionsreihe Ökonomische Bildung

Diskussionsbeitrag Nr. 1: Der sozialökonomische Beitrag zur Umweltbildung
Gerd-Jan Krol/ Jan Karpe/ Andreas Zoerner
Februar 1998

Diskussionsbeitrag Nr. 2: Moderne Ökonomik und Moderne Kunst – Ein Beitrag zur ökonomischen und ästhetischen Werteentwicklung moderner Kunst
Jan Karpe/ Mirco Derpmann
Mai 1998

Diskussionsbeitrag Nr. 3: Der sozialökonomische Ansatz der Umweltbildung – Grundlagen und Praxis der Umsetzung im Unterricht
Gerd-Jan Krol/ Thomas Hönemann
(Mit Unterrichtsbeispielen für die Sekundarstufen I und II von Anne Zumkley und Thomas Hönemann)
November 1998

Diskussionsbeitrag Nr. 4: Freiwilligkeit – Zwang – Anreize. Bemerkungen zu strategischen Ansatzpunkten der Umweltpolitik
Gerd-Jan Krol
November 1998

Diskussionsbeitrag Nr. 5: Enviromental Problems, Morals und Incentives in Modern Societies
Gerd-Jan Krol
Januar 2000

Diskussionsbeitrag Nr. 6 Ökonomische Bildung in der modernen Gesellschaft
Gerd-Jan Krol/ Jan Karpe/ Andreas Zoerner
August 2001

Diskussionsbeitrag Nr. 7 Eine ökonomische Analyse des Marktes für Klima- und Lüftungstechnik – Aktueller Stand und Zukunftsperspektiven in der Baubranche
Alexander Herrmann
März 2002

Ab 2005:

IÖB-Diskussionspapiere

IÖB-Diskussionspapier 1/05: Eliteuniversität Münster?!
Alexander Dilger
März 2005

IÖB-Diskussionspapier 2/05: Macro-Analysis of Transfer Fees and Investments in Sports
Alexander Dilger
Dezember 2005

IÖB-Diskussionspapier 1/06: Standards für die Lehrerbildung in der ökonomischen Bildung
Gerd-Jan Krol / Dirk Loerwald / Andreas Zoerner.
April 2006

IÖB-Diskussionspapier 2/06: Jugend – Ernährungsstil – Bildung: Zu den Perspektiven einer lebensstilorientierten Didaktik
Konstantin v. Norman
April 2006

IÖB-Diskussionspapier 3/06: Taschengeld und Sparverhalten bei Grundschulkindern
Nicole Dubbert / Rainer Hufnagel
Juli 2006

IÖB-Diskussionspapier 4/06: Theorie optimaler Währungsräume vor dem Hintergrund der EU-Erweiterung
Heiko Peters
Juli 2006

IÖB-Diskussionspapier 5/06: Höhere Geburtenraten in Deutschland – Die Rolle des Systems frühkindlicher Bildung und Betreuung
Antje Funcke
Juli 2006

IÖB-Diskussionspapier 6/06: Geldpolitische Strategien der neuen EU-Mitgliedsländer bis zur Euroeinführung
Heiko Peters
August 2006

IÖB-Diskussionspapier 1/07: German Universities as State-sponsored Co-operatives
Alexander Dilger
Januar 2007

IÖB-Diskussionspapier 2/07: Why Brennan and Buchanan are wrong (after all)
Thomas Apolte
März 2007

IÖB-Diskussionspapier 3/07: Empirische Ermittlung von Haushaltsführungsstilen mit Daten der Zeitbudgeterhebung 1991/92
Stéphanie Grossmann
Mai 2007

IÖB-Diskussionspapier 4/07: Managerverschanzung durch spezifische Investitionen
Benjamin Balsmeier
Mai 2007

IÖB-Diskussionspapier 5/07: Münsteraner Ernährungsmuster- und Lebensstilstudie bei Schulkindern „MEALS“ 2006
Konstantin von Normann
August 2007

IÖB-Diskussionspapier 6/07: Personelle Unternehmensverflechtung und Vorstandsgehälter
Benjamin Balsmeier / Heiko Peters
Dezember 2007

IÖB-Diskussionspapier 1/08: Theoretische Analyse der Strategienwahl unter der Zwei- und Drei-Punkte-Regel im Fußball
Hannah Geyer
Januar 2008

IÖB-Diskussionspapier 2/08: Welche Immigranten kehren Deutschland den Rücken? Eine Paneldatenuntersuchung für Deutschland
Sebastian Gundel / Heiko Peters
Februar 2008

IÖB-Diskussionspapier 3/08: Predicting Birth-Rates Through German Micro-Census Data.
A Comparison of Probit and Boolean Regression
Rainer Hufnagel
März 2008

Westfälische Wilhelms-Universität Münster
Wirtschaftswissenschaftliche Fakultät
Institut für Ökonomische Bildung
Scharnhorststraße 100
48151 Münster

Tel: 0251/ 83-24303 (Sekretariat)
E-Mail: iob2@uni-muenster.de
Internet: www.wiwi.uni-muenster.de/ioeb

Herausgeber:
Institut für Ökonomische Bildung
Westfälische Wilhelms-Universität Münster
Scharnhorststraße 100
48151 Münster
Tel: 0251/ 83-24303
Fax: 0251/ 83-28429

www.wiwi.uni-muenster.de/ioeb