IAW-Diskussionspapiere

Discussion Paper

25

Estimation of the Probit Model from Anonymized Micro Data

Gerd Ronning
Martin Rosemann

May 2006

ISSN: 1617-5654

Institut für Angewandte Wirtschaftsforschung
Ob dem Himmelreich 1
72074 Tübingen
T: (0 70 71) 98 96-0
F: (0 70 71) 98 96-99
E-Mail: iaw@iaw.edu
Internet: www.iaw.edu

IAW-Diskussionspapiere

Das Institut für Angewandte Wirtschaftsforschung (IAW) Tübingen ist ein unabhängiges außeruniversitäres Forschungsinstitut, das am 17. Juli 1957 auf Initiative von Professor Dr. Hans Peter gegründet wurde. Es hat die Aufgabe, Forschungsergebnisse aus dem Gebiet der Wirtschafts- und Sozialwissenschaften auf Fragen der Wirtschaft anzuwenden. Die Tätigkeit des Instituts konzentriert sich auf empirische Wirtschaftsforschung und Politikberatung.

Dieses **IAW-Diskussionspapier** können Sie auch von unserer IAW-Homepage als pdf-Datei herunterladen:

http://www.iaw.edu/Publikationen/IAW-Diskussionspapiere

ISSN 1617-5654

Weitere Publikationen des IAW:

- IAW-News (erscheinen 4x jährlich)
- IAW-Report (erscheinen 2x jährlich)
- IAW-Wohnungsmonitor Baden-Württemberg (erscheint 1x jährlich kostenlos)
- IAW-Forschungsberichte

Möchten Sie regelmäßig eine unserer Publikationen erhalten, dann wenden Sie sich bitte an uns:

IAW Tübingen, Ob dem Himmelreich 1, 72074 Tübingen,
Telefon 07071 / 98 96-0
Fax 07071 / 98 96-99
E-Mail: iaw@iaw.edu

Aktuelle Informationen finden Sie auch im Internet unter: http://www.iaw.edu

Estimation of the Probit Model From Anonymized Micro Data

Gerd Ronning* Martin Rosemann†

April 10, 2006

Abstract

The demand of scientists for confidential micro data from official sources has created discussion of how to anonymize these data in such a way that they can be given to the scientific community. We report results from a German project which exploits various options of anonymization for producing such "scientific-use- files". The main concern in the project however is whether estimation of stochastic models from these perturbed data is possible and – more importantly – leads to reliable results. In this paper we concentrate on estimation of the probit model under the assumption that only anonymized data are available. In particular we assume that the binary dependent variable has undergone post-randomization (PRAM) and that the set of explanatory variables has been perturbed by addition of noise. We employ a maximum likelihood estimator which is consistent if only the dependent variable has been anonymized by PRAM. The errors-in-variables structure of the regressors then is handled by the simulation extrapolation (SIMEX) estimation procedure where we compare performance of quadratic and nonlinear (rational) extrapolation.

KEYWORDS: anonymization, misclassification, noise addition, post-randomization, SIMEX procedure, statistical disclosure.

Acknowledgements: Research in this paper is related to the project "Faktische Anonymisierung wirtschaftsstatistischer Einzeldaten". Financial support from Bundesministerium für Bildung und Forschung is gratefully acknowledged. An earlier version has been presented at UNECE Work Session on Statistical Data Confidentiality, Geneva, 10 November 2005. We dedicate this paper to Reinhard Hujer on the occasion of his 65-th birthday. We thank Elena Biewen for helpful comments on earlier versions of this paper.

1 Introduction

Empirical research in economics has for a long time suffered from the unavailability of individual "micro" data and has forced econometricians to use (aggregate) time series data in order to estimate, for example, a consumption function. On the contrary other disciplines like psychology, sociology and, last not least, biometry have analyzed micro data already for decades. The software for microeconometric models has created growing demand for micro data in economic research, in particular data describing firm behaviour. However, such data

*Wirtschaftswissenschaftliche Fakultät, Universität Tübingen, Mohlstrasse 36, 72074 Tübingen (gerd.ronning@uni-tuebingen.de) .

†Institut für Angewandte Wirtschaftsforschung e.V. (IAW), Ob dem Himmelreich 1, D-72074 Tübingen, (martin.rosemann@iaw.edu)

are not easily available when collected by the Statistical Office because of confidentiality. On the other hand these data would be very useful for testing microeconomic models. This has been pointed out recently by KVI commission.[1] Therefore, the German Statistical Office initiated research on the question whether it is possible to produce scientific use files from these data which have to be anonymized in a way that re-identification is almost impossible and, at the same time, distributional properties of the data do not change too much. Results from this project have been published quite recently. See Ronning et al. (2005) where most known anonymization procedures have been rated both with regard to data protection and to informational content left after perturbation.

Published work on anonymization of micro data and its effects on the estimation of microeconometric models has concentrated on *continuous* variables where a variety of procedures is available. See, for example, Ronning and Gnoss (2003) for such procedures and the contribution by Lechner and Pohlmeier (2003) also for the effects on estimation when anonymizing data either by microaggregation or addition of noise. Discrete variables, however, mostly have been left aside in this discussion. The only stochastic-based procedure to anonymize discrete variables is post-randomization (PRAM) which switches categories with prescribed probability.

In this paper we concentrate on estimation of the probit model for which only anonymized data are available. In particular we assume that the binary dependent variable has undergone post-randomization (PRAM) and that the set of explanatory variables has been perturbed by addition of noise. We employ a maximum likelihood estimator which is consistent if only the dependent variable has been anonymized by PRAM. The errors-in-variables structure of the regressors then is handled by the simulation extrapolation (SIMEX) estimation procedure.

In Section 2 we consider the probit model. We assume that the binary dependent variable has been anonymized by PRAM whereas right-hand regressor variables have been left in original form. Consistent estimates are available from an adapted estimation procedure. We then turn to the situation that the continuous regressors have been anonymized by noise addition (section 3). An attractive procedure for handling such situations is the simulation extrapolation (SIMEX) estimator which will be briefly described. Section 4 then presents some estimation results for the probit model when both the dependent and the independent variables have been anonymized. We present results from a simulation study where the PRAM adapted probit estimator is combined with the SIMEX approach. Some concluding remarks are added in section 5.

2 The probit model under post randomization

2.1 The probit model

Consider the following linear model:[2]

$$Y^* \quad = \quad \alpha \, + \, \beta x \, + \, \varepsilon \tag{2-1}$$

with $E[\varepsilon] \; = \; 0$ and $V[\varepsilon] \; = \; \sigma_\varepsilon^2$. Here the $*$ indicates that the continuous variable Y is latent or unobservable. This model asserts that the conditional expectation of Y^* but not the corresponding conditional variance depends on some explanatory variable x.[3] However

[1] See KVI (2001).

[2] See, for example, Ronning (1991) or Greene (2000).

[3] x could also be interpreted as a vector representing a set of explanatory variables. However in this paper we stick to the simple case.

we observe only a binary variable Y which is related to the latent variable by the "threshold model":

$$Y = \begin{cases} 0 & \text{if } Y^* \leq \tau \\ 1 & \text{else} \end{cases} . \tag{2-2}$$

It can be shown that two of the four parameters α, β, σ_ε^2 and τ have to be fixed in order to attain identification of the two remaining ones. Usually we set $\tau = 0$ and $\sigma_\varepsilon^2 = 1$ assuming additionally that the error term ε is normally distributed. This is the famous probit model. Note that only the probability of observing $Y = 1$ for a given x can be determined. If we alternatively assume hat the error term follows a logistic distribution, we obtain the closely related binary logit model.

2.2 Randomized response and post randomization

Randomized response originally was introduced to avoid non-response in surveys containing sensitive questions on, e.g., drug consumption or AIDS disease. See Warner (1965). Särndal et al. (1992 p. 573) suggested use of this method "to protect the anonymity of individuals". A good description of the difference between the two (formally equivalent) approaches is given by van den Hout and van der Heijden (2002): In the randomized response setting the stochastic model has to be defined in advance of data collection whereas in post randomization this method will be applied to the data already obtained.

Randomization of the binary variable Y can be described as follows: Let Y^m denote the 'masked' variable obtained from post randomization. Then the transition probabilities can be defined by $p_{jk} \equiv P(Y^m = j \,|\, Y = k)$ with $j, k \,\varepsilon\, \{0, 1\}$ and $p_{j0} + p_{j1} = 1$ for $j = 0, 1$. If we define the two probabilities of no change by $p_{00} \equiv \pi_0$ and $p_{11} \equiv \pi_1$, respectively, the probability matrix can be written as follows:

$$\mathbf{P}_y = \begin{pmatrix} \pi_0 & 1 - \pi_0 \\ 1 - \pi_1 & \pi_1 \end{pmatrix}$$

Since the two probabilities of the post randomization procedure usually are known and there is no argument not to treat the two states symmetrically, in the following we will consider the special case

$$\pi_0 = \pi_1 . \tag{2-3}$$

When the variable Y has undergone randomization, we will have a sample with n observations y_i^m where y_i^m is the dichotomous variable obtained from y_i by the randomization procedure.

In the handbook on anonymization (Ronning et al. 2005) we also discuss the extension of PRAM to more than two categories. If the categories are ordered as, for example, in the case of ordinal variables or count data, switching probabilities for adjoining categories should be higher since otherwise the ordering would be totally destroyed. Of course, PRAM could also be extended to joint anonymization of two or more discrete variables.

2.3 Estimation of the model under PRAM

Under randomization of the dependent observed variable we have the following data generating process:

$$Y_i^m = \begin{cases} 1 & \text{with probability} \quad \Phi_i \, \pi + (1 - \Phi_i)(1 - \pi) \\ 0 & \text{with probability} \quad \Phi_i \,(1 - \pi) + (1 - \Phi_i)\, \pi \end{cases} \tag{2-4}$$

Here Φ_i denotes the conditional probability under the normal distribution that the unmasked dependent variable Y_i takes on the value 1 for given x_i, i.e. $\Phi_i \equiv \Phi(\alpha + \beta x_i) = P(Y_i^* > 0 \mid x_i)$.

>From (2-4) we obtain the following likelihood function:

$$\begin{aligned}&\mathcal{L}(\alpha, \beta|(y_i^m, x_i), i = 1, \dots, n)\\ &= \prod_{i=1}^{n} \left[\Phi_i \, \pi \;+\; (1-\Phi_i)\,(1-\pi)\right]^{y_i^m} \; \left[\Phi_i\,(1-\pi) \;+\; (1-\Phi_i)\,\pi\right]^{(1-y_i^m)} \quad .\end{aligned} \tag{2-5}$$

Global concavity of this function with respect to α and β may be checked by deriving first and second (partial) derivatives of the log-likelihood function. Ronning (2005) derives the Hessian matrix of partial derivatives. A simple formula for the information matrix can be derived from which it is immediately apparent that maximum likelihood estimation under randomization is consistent but implies an efficiency loss which is greatest for values of π near 0.5. See Ronning (2005) for detailed results.

3 Addition of noise and the simulation extrapolation approach

3.1 Data protection by addition of noise

Consider the linear model which we write in usual way as follows: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$. Let $\mathbf{e}_y$ be a vector of errors with expectation zero and positive variance corresponding to $\mathbf{y}$ and let $\mathbf{E}_X$ be a matrix of errors corresponding to $\mathbf{X}$. Addition of noise means that we have to estimate the unknown parameter vector from the model

$$\mathbf{y} + \mathbf{e}_y \;=\; (\mathbf{X} + \mathbf{E}_X)\,\boldsymbol{\beta} \;+\; \mathbf{u}\,. \tag{3-1}$$

This is the well-known errors-in-variables model for which anonymization of right-hand variables creates estimation problems whereas anonymization of the dependent variable only increases the error variance.[4] Lechner and Pohlmeier (2005) consider nonparametric regression models where the regressors are anonymized by addition of noise. They show that from the simulation-extrapolation method (SIMEX) reliable estimates can be obtained. For the logit model Cook and Stefanski (1994) present results regarding the effect of noise addition and the suitability of the SIMEX method if the dependent variable y is observed without error.

Additive errors have the disadvantage that greater values of a variable are less protected. Take as an example sales of firms. If one firm has sales of 1 million and another sales of 100 million then addition of an error of 1 doubles sales of the first but leaves nearly unchanged sales of the second firm. Therefore research has been done also for the case of multiplicative errors which in this case should have expectation one. Formally this leads to

$$\mathbf{y} \odot \mathbf{e}_y \;=\; (\mathbf{X} \odot \mathbf{E}_X)\,\boldsymbol{\beta} \;+\; \mathbf{u}$$

where $\odot$ denotes element-wise multiplication (Hadamard product). For results regarding estimation of this linear model see Ronning et al (2005). In the following we consider only the additive case.

[4] See Lechner and Pohlmeier (2003) for details. This should be compared with the case of microaggregation where (separate) anonymization of the *dependent* variable creates problems. See Ronning et al. (2005).

3.2 The SIMEX approach

We will only sketch the idea of this approach[5] for the simple linear regression model which is a special case of the linear model (3-1) considered above with only one regressor and a constant term. It is well known from econometric textbooks that estimation of the regression coefficient β by least squares leads to

$$plim\,\hat{\beta} \;=\; \beta\,\frac{\sigma_x^2}{\sigma_x^2+\sigma_e^2}. \tag{3-2}$$

if the regressor variable x can only be observed with error e_x where σ_x^2 is the variance of x and σ_e^2 is the variance of the error. Now assume that this variance is known and that another error $\lambda\, e_x$ with $\lambda > 0$ is added to the error affected regressor variable by purpose. Then we obtain

$$plim\,\hat{\beta}(\lambda) \;=\; \beta\,\frac{\sigma_x^2}{\sigma_x^2+(1+\lambda)\sigma_e^2} \tag{3-3}$$

so that a consistent estimator would be obtained for $\lambda = -1$.

Moreover, it can be shown that for *nonlinear* models this extrapolation approach is appropriate at least approximately! Of course also for these models $\hat{\beta}(\lambda)$ can be evaluated for any positive λ using simulation whereas results for $\lambda < 0$ have to be guessed. Cook and Stefanski (1994) suggested an extrapolation procedure which fits a curve to the various points and extrapolates it for $\lambda = -1$. In particular they considered alternatively a "quadratic" and a "nonlinear" extrapolation function. Both will also be used in this paper. See sections 4.2 and 4.3.

Usually M simulation runs are averaged for each λ so that

$$\overline{\hat{\beta}(\lambda)} = \frac{1}{M}\sum_{j=1}^{M}\hat{\beta}_j(\lambda)$$

is the estimate actually used. We follow this approach also in the present paper. Alternatively the median might be used. See Cook and Stefanski (1994).

In the simulation study on which we report in the next section we use the ML estimator suitable for the probit model under PRAM (see subsection 2.3) in the SIMEX routine thereby taking account of the measurement error in the regressor.[6] Therefore in the following $\hat{\beta}(\lambda)$ will be the PRAM-corrected maximum likelihood estimator considered in section 2.3.

4 Simulation results

4.1 Simulation design

In this section we will estimate the two parameters α and β of the probit model defined in (2-1) and (2-2) assuming that the dependent variable y has been anonymized by PRAM and that the regressor variable x has been protected by addition of noise which is normally

[5] For details see, for example, Carroll et al (1995).

[6] Quite recently Küchenhoff et al (2005) have proposed a "misclassification SIMEX" (MC SIMEX) estimator for the probit model under randomization which can bee seen as an alternative to the maximum likelihood approach used here. However the case of a continuous regressor observed with error is not discussed in that paper.

distributed with variance σ_e^2. We assume that the PRAM parameter π and the error variance σ_e^2 are known since in the anonymization approach these parameters will be controlled and released to users of the data.[7] Simulated data will be used for estimation. The two unknown parameters are given by $\alpha = -2.5$ and $\beta = 0.6$. The regressor variable is generated from a normal distribution $N(4.35\,; 1.75^2)$ and the error variable satisfies $\varepsilon \sim N(0; 1)$ the latter recognizing the identification constraint of the probit model.

Since both the PRAM parameter π and the error variance σ_e^2 cause estimation bias in the "naive" estimation approach[8] we will study the effect of both parameters on the estimation results using[9] $\pi \;\varepsilon\; \{1.0, 0.9, 0.8\}$ and $\sigma_e^2 \;\varepsilon\; \{0.1^2, 0.5^2, 0.7^2, 1.1^2\}$. The latter should be compared with the variation of the regressor given by $\sigma_x^2 = 1.75^2$ indicating a maximal measurement error of about 60 %! Furthermore we will vary the sample size using $n \;\varepsilon\; \{500, 1000\}$.

4.2 Quadratic extrapolation function

The maximum likelihood (ML) estimator of the probit model based on the likelihood function (2-4) is evaluated by a GAUSS programme written by the first author.[10] We use $R = 500$ iterations in this simulation study. In each iteration the ML estimator of the probit model is employed in the SIMEX procedure: First for each $\lambda\,\varepsilon\,\{0\,,\,0.5\,,\,1.0\,,\,1.5\,,\,2.0\}$ we computed $M = 250$ values of this estimator from which $\overline{\hat{\beta}(\lambda)}$ was determined. Using the five different estimates we then fitted a quadratic function to these five points and obtained the final estimate of both α and β from evaluating this function at $\lambda = -1$. From the $R = 500$ estimates we computed mean, standard deviation, median, skewness (abbreviated as skew.) , kurtosis (abbreviated as kurt.) and both the minimal and the maximal value which are presented in tables 4.1 and 4.2 which differ by the magnitude of the error variance.

The results in the two tables show that this approach is quite promising even for a substantial proportion of misclassified y-values and 'moderate' measurement errors in the regressor variable. See table 4.1. However for $\sigma_e^2 = 0.49$ the bias is notable (parts E. and F. of table 4.2) and becomes unacceptably large for $\sigma_e^2 = 1.21$ (parts F. and H. of this table). Interestingly the bias is smaller when π moves away from 1.0 indicating a countervailing effect induced by the PRAM corrected ML estimator. A larger sample size helps a lot if the measurement error is small as can be seen from a comparison of parts C. and D. in table 4.1 the latter showing much better results for the larger sample size of $n = 1,000$.

The performance of the SIMEX approach depends on the appropriateness of the quadratic extrapolation function which we used in our simulation. We therefore analyzed the scatter plots from our simulations. Figure 4/1 shows some examples. For each graph we ran a single simulation ($R = 1$) with $n = 1,000$ observations. From $M = 250$ values of the ML estimator $\overline{\hat{\beta}(\lambda)}$ was computed for each λ. The extrapolated value of the function at $\lambda = -1$

---

[7]It is possible to extend the estimation procedure to the case of unknown π. See Hausman et al. (1998) and Ronning (2005).

[8]Neuhaus (1999) presents a detailed discussion of bias from 'misclassification' in binary regression models. Table 2 in his paper gives formula of the (approximated) bias **factor** for the probit model although he considers the case of a *binary* regressor. His formula reads (in our terminology assuming $\pi > 0.5$)) as

$$\text{bias factor} = \frac{(2\pi - 1)\,\phi(\alpha)}{\phi\left[\Phi^{-1}\left\{(2\pi - 1)\,\Phi(\alpha) + 1 - \pi\right\}\right]} \quad .$$

Note that this expression contains the factor $2\pi - 1 \;<\; 1$. In particular the bias factor reduces to this expression if we set $\alpha = 0$ implying shrinkage towards zero.

[9]Since we know from earlier simulation experiments that values of the PRAM parameter π create computational problems if π is far away from 1.0 we confined simulation to the interval $\pi \;\varepsilon\; [0.8\,; 1.0]$.

[10]Many thanks to Sandra Lechner for providing us with a SIMEX routine!

Table 4.1: PRAM adapted ML estimation of the probit model combined with SIMEX procedure (quadratic extrapolation) - Small error variance

A. $n = 500$, $\sigma_e^2 = 0.01$								
π		estimate	std. dev.	min.	median	max.	skew.	kurt.
1.000	α	-2.5072	0.2288	-3.1847	-2.4950	-1.9039	-0.222	2.883
	β	0.6016	0.0509	0.4784	0.5975	0.7536	0.218	2.796
0.900	α	-2.5429	0.2598	-3.3810	-2.5366	-1.8287	-0.127	3.344
	β	0.6103	0.0598	0.4523	0.6069	0.8012	0.170	3.151
0.800	α	-2.5820	0.2861	-3.6368	-2.5735	-1.7812	-0.230	3.216
	β	0.6196	0.0659	0.4412	0.6176	0.8399	0.264	3.071
B. $n = 1,000$, $\sigma_e^2 = 0.01$								
π		estimate	std. dev.	min.	median	max.	skew.	kurt.
1.000	α	-2.5104	0.1589	-3.0866	-2.5044	-2.0670	-0.175	3.154
	β	0.6024	0.0354	0.5126	0.6012	0.7345	0.244	3.178
0.900	α	-2.5315	0.1837	-3.0447	-2.5416	-1.9115	-0.035	3.138
	β	0.6072	0.0425	0.4754	0.6077	0.7314	0.180	3.101
0.800	α	-2.5537	0.2039	-3.3882	-2.5384	-2.0183	-0.452	3.552
	β	0.6129	0.0463	0.4866	0.6093	0.7983	0.510	3.679
C. $n = 500$, $\sigma_e^2 = 0.25$								
π		estimate	std. dev.	min.	median	max.	skew.	kurt.
1.000	α	-2.5190	0.2538	-3.2960	-2.5096	-1.8565	-0.278	3.086
	β	0.6033	0.0575	0.4465	0.6011	0.7700	0.256	3.062
0.900	α	-2.5019	0.2729	-3.2696	-2.4691	-1.7340	-0.332	3.080
	β	0.6005	0.0616	0.4382	0.5933	0.7870	0.462	3.198
0.800	α	-2.5772	0.3287	-4.0162	-2.5571	-1.7685	-0.646	4.258
	β	0.6185	0.0751	0.4255	0.6127	0.9342	0.619	4.156
D. $n = 1,000$, $\sigma_e^2 = 0.25$								
π		estimate	std. dev.	min.	median	max.	skew.	kurt.
1.000	α	-2.4927	0.1737	-3.1709	-2.4878	-2.0356	-0.259	3.058
	β	0.5992	0.0392	0.5033	0.5966	0.7417	0.314	3.088
0.900	α	-2.5161	0.1970	-3.2549	-2.5097	-1.9824	-0.318	3.083
	β	0.6044	0.0440	0.4851	0.6032	0.7645	0.328	3.131
0.800	α	-2.5209	0.2080	-3.2921	-2.5128	-2.0215	-0.408	3.265
	β	0.6055	0.0476	0.4891	0.6028	0.7768	0.408	3.206
Remark: True parameter values: $\alpha = -2.50$, $\beta = 0.60$.								

was determined. Then $\overline{\hat{\beta}(\lambda)}$ was plotted against λ. In order to show the effect of the error variance we used $\sigma_e^2 = 0.49$ as we did in the simulations reported in table 4.2. The biasing effect of both increasing σ_e^2 and shifting π away from 1.0 can be clearly seen from this figure. And it is evident from these examples that a monotonic extrapolation function is quite adequate for this model!

Table 4.2: (Table 4.1 continued) PRAM adapted ML estimation of the probit model combined with SIMEX procedure (quadratic extrapolation) - Large error variance

E. $n = 500$, $\sigma_e^2 = 0.49$								
π		estimate	std. dev.	min.	median	max.	skew.	kurt.
1.000	α	-2.4434	0.2586	-3.2482	-2.4409	-1.8305	-0.134	2.778
	β	0.5873	0.0585	0.4425	0.5862	0.7781	0.141	2.866
0.900	α	-2.4538	0.2934	-4.1007	-2.4314	-1.6708	-0.676	4.724
	β	0.5901	0.0666	0.4319	0.5825	0.9640	0.730	4.820
0.800	α	-2.5097	0.3383	-3.7274	-2.4687	-1.6240	-0.450	3.137
	β	0.6027	0.0763	0.3978	0.5970	0.8563	0.387	2.913
F. $n = 1,000$, $\sigma_e^2 = 0.49$								
π		estimate	std. dev.	min.	median	max.	skew.	kurt.
1.000	α	-2.4268	0.1820	-3.0655	-2.4174	-1.9797	-0.225	3.074
	β	0.5823	0.0416	0.4676	0.5791	0.7465	0.336	3.351
0.900	α	-2.4472	0.1923	-3.1788	-2.4379	-1.9467	-0.382	3.391
	β	0.5883	0.0430	0.4718	0.5843	0.7216	0.359	3.248
0.800	α	-2.4517	0.2077	-3.5712	-2.4428	-1.9029	-0.351	4.272
	β	0.5894	0.0469	0.4601	0.5855	0.8463	0.387	4.509
G. $n = 500$, $\sigma_e^2 = 1.21$								
π		estimate	std. dev.	min.	median	max.	skew.	kurt.
1.000	α	-2.1507	0.2519	-2.9290	-2.1422	-1.2313	-0.218	3.115
	β	0.5180	0.0564	0.3018	0.5140	0.7250	0.282	3.537
0.900	α	-2.1815	0.2787	-4.2900	-2.1742	-1.2684	-1.040	9.401
	β	0.5257	0.0631	0.3165	0.5234	0.9720	0.984	8.166
0.800	α	-2.2205	0.3193	-3.4608	-2.1944	-1.4751	-0.648	3.683
	β	0.5352	0.0718	0.3641	0.5285	0.7900	0.631	3.551
H. $n = 1,000$, $\sigma_e^2 = 1.21$								
π		estimate	std. dev.	min.	median	max.	skew.	kurt.
1.000	α	-2.1465	0.1785	-2.6047	-2.1407	-1.7046	-0.137	2.574
	β	0.5176	0.0401	0.4174	0.5157	0.6249	0.123	2.539
0.900	α	-2.1641	0.1803	-2.9017	-2.1582	-1.7577	-0.367	3.241
	β	0.5214	0.0412	0.4194	0.5190	0.6711	0.253	3.087
0.800	α	-2.1767	0.1938	-3.1319	-2.1730	-1.6496	-0.313	3.881
	β	0.5247	0.0445	0.4093	0.5230	0.7441	0.367	3.831
Remark: True parameter values: $\alpha = -2.50$, $\beta = 0.60$.								

4.3 Nonlinear extrapolation function

Since the performance of the SIMEX estimator using quadratic extrapolation is not satisfactory for larger measurement errors, it seems worth to compare those results with results from the alternative 'nonlinear' function also proposed by Cook and Stefanski (1994 p. 1314) which is given by

$$f(\lambda) \quad = \quad \gamma + \frac{\delta}{\theta + \lambda} \tag{4-1}$$

where γ, δ and θ are the three parameters of the function. Note that this function can describe only monotonic behavior contrary to the quadratic function and tends towards (minus) infinity at $\lambda = -\theta$. Some more properties are discussed in appendix A.

Using the same simulation design as described in subsection 4.1 we now estimate the model using extrapolation function (4-1). However we restrict the fitting of the 3-parameter extrapolation function to only three values of λ which reduces considerably the numerical

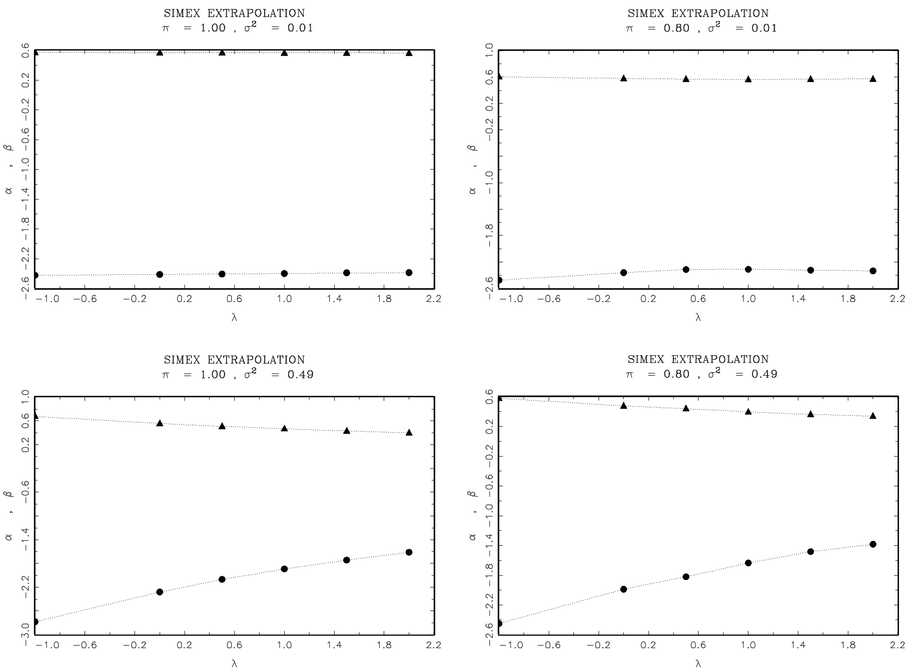

Figure 4/1: Quadratic SIMEX extrapolation for different values of π and σ_e^2. (▲ shows estimates of $\alpha = 0.60$ and • those of $\beta = -2.50$.)

effort. (Otherwise an iterative procedure has to be used.) Actually we choose $\lambda \,\varepsilon\, \{0, 1, 2\}$. The corresponding formulae are given in the appendix. We also consider only the larger sample size of $n = 1,000$. Results from our simulations are given in table 4.3.

Let us first look at results without post randomization ($\pi = 1.00$). Even for the largest error variance parameter estimates of the probit model show almost no bias at all contrary to results for the quadratic case! This corresponds to results in Cook and Stefanski (1994 section 3) for the logit model where nonlinear extrapolation outperformed quadratic extrapolation.[11] More importantly, the results are much more satisfying if post randomization is applied to the dependent variable: the bias is considerably lower than in the quadratic case especially for larger error variance (compare table 4.2).

However the standard errors of the estimators resulting from the nonlinear extrapolation function are considerably larger than for the quadratic case. This is especially pronounced for $\pi = 0.80$ indicating a substantial portion of randomization. Again this corresponds to (graphical) results presented in Cook and Stefanski (1994): For the logit model and a non-randomized dependent variable the nonlinear case implies larger variation than the quadratic case. See figures 5 and 6 in their paper.

[11] These authors consider *two* explanatory variables which follow a bivariate standard normal distribution. They also add the interaction of these two variables to the set of regressors. The error variance is set to $\sigma_e^2 = 0.5$ which corresponds to our case $\sigma_e^2 = 0.49$ and they choose a sample size of $n = 1500$ whereas we use $n \,\varepsilon\, \{500\,,\, 1000\}$.

Table 4.3: PRAM adapted ML estimation of the probit model combined with SIMEX procedure (nonlinear extrapolation)

B. $n = 1,000$, $\sigma_e^2 = 0.01$								
π		estimate	std. dev.	min.	median	max.	skew.	kurt.
1.000	α	-2.5203	0.1602	-3.0190	-2.5181	-2.0724	-0.087	2.879
	β	0.6047	0.0353	0.5008	0.6040	0.7209	0.046	3.003
0.900	α	-2.5097	0.2304	-4.0292	-2.5060	-0.2324	1.119	25.583
	β	0.5981	0.1949	-2.9282	0.6006	2.4475	-10.124	233.920
0.800	α	-2.4858	0.8993	-10.0619	-2.5220	10.9660	9.902	166.838
	β	0.5968	0.2391	-4.4924	0.6064	1.8826	-19.028	413.761
C. $n = 1,000$, $\sigma_e^2 = 0.25$								
π		estimate	std. dev.	min.	median	max.	skew.	kurt.
1.000	α	-2.5157	0.1781	-3.1270	-2.5081	-2.0497	-0.269	3.343
	β	0.6034	0.0410	0.4934	0.6031	0.7606	0.267	3.294
0.900	α	-2.5387	0.2009	-3.2369	-2.5341	-1.9225	-0.578	3.853
	β	0.6083	0.0460	0.4689	0.6072	0.7974	0.636	3.871
0.800	α	-2.5734	0.2847	-4.1278	-2.5214	-2.0660	-0.678	3.873
	β	0.6178	0.0636	0.4922	0.6072	0.8994	0.822	4.652
D. $n = 1,000$, $\sigma_e^2 = 0.49$								
π		estimate	std. dev.	min.	median	max.	skew.	kurt.
1.000	α	-2.5062	0.2050	-3.2516	-2.4991	-1.9661	-0.427	2.876
	β	0.6011	0.0460	0.4918	0.5989	0.7888	0.437	2.895
0.900	α	-2.5224	0.2304	-3.4580	-2.5059	-1.9129	-0.423	2.991
	β	0.6053	0.0532	0.4761	0.6016	0.8169	0.470	3.094
0.800	α	-2.5716	0.2768	-3.7070	-2.5359	-1.9782	-0.560	3.362
	β	0.6173	0.0644	0.4858	0.6117	0.9497	0.558	3.588
E. $n = 1,000$, $\sigma_e^2 = 1.21$								
π		estimate	std. dev.	min.	median	max.	skew.	kurt.
1.000	α	-2.4959	0.2278	-3.2437	-2.4697	-1.9493	-0.238	3.121
	β	0.5988	0.0520	0.4837	0.5918	0.7746	0.246	3.197
0.900	α	-2.5146	0.3119	-3.4572	-2.4754	-1.5631	-0.377	3.313
	β	0.6036	0.0713	0.3899	0.5938	0.8332	0.459	3.753
0.800	α	-2.5657	0.3687	-4.2288	-2.5077	-1.8329	-1.227	6.082
	β	0.6145	0.0839	0.4390	0.6056	1.0098	1.050	4.855
Remark: True parameter values: $\alpha = -2.50$, $\beta = 0.60$.								

The extremely large variation of the estimator for $\sigma_e^2 = 0.01$ in table 4.3 needs an extra comment: If the error variance tends towards zero, the fit of the points $(\lambda\,,\,\overline{\hat{\beta}(\lambda)})$ used for extrapolation tends towards a straight line which however cannot be represented by the nonlinear extrapolation function (4-1). From the results in the appendix it can be demonstrated that in such cases the behaviour of the nonlinear extrapolating function becomes very unsteady especially at $\lambda = -1$. Since the PRAM adapted ML estimator has larger variance for $\pi < 1$ this instability is greater for $\pi = 0.90$ and extreme for $\pi = 0.80$. See also figure 4/2 where we compare SIMEX estimates from both extrapolating functions; for the nonlinear case the nonlinearity is more pronounced! Note that Cook and Stefanski (1994) considered only one single – and rather large – value of the error variance. Therefore the nonlinear variant of the extrapolation function cannot be recommended in general.

We also note that in some cases the median of estimates is closer to the true value than the mean indicating asymmetric behavior of the distribution of estimators. However in most cases considered in our simulations skewness is only moderate and does not produce smaller

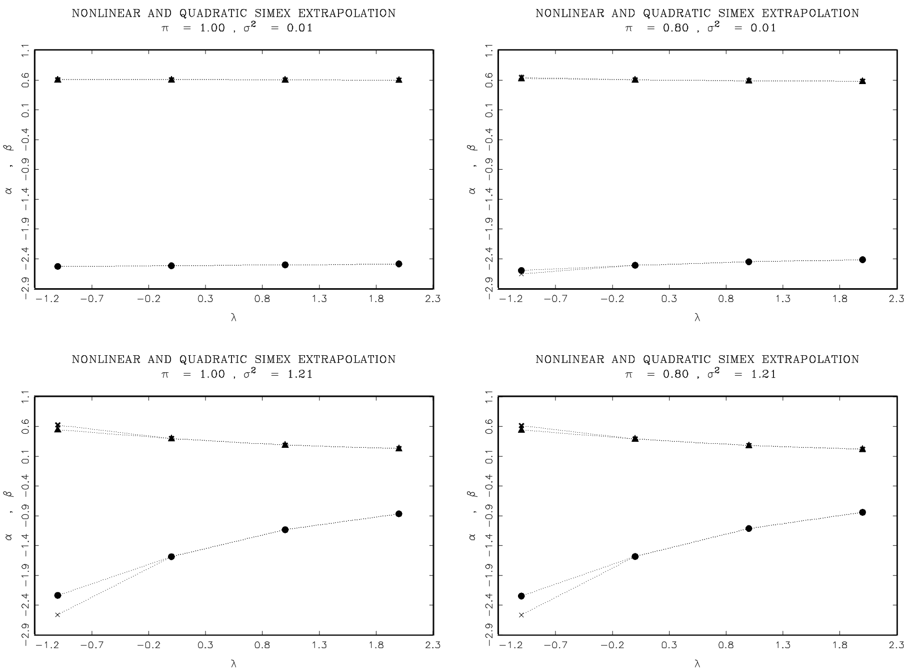

Figure 4/2: Nonlinear SIMEX extrapolation for different values of π and σ_e^2. (▲ shows estimates of $\alpha = 0.60$ and • those of $\beta = -2.50$ in the quadratic case, $\times$ indicates nonlinearly extrapolated values.)

bias on average. Moreover kurtosis indicates 'normality' in most cases. One exception is noted in part G. of table 4.2 which is due to to the relatively small sample size of $n = 500$. The other abnormal case is given in part E of table 4.3 for nonlinear extrapolation, large error variance and a high proportion of randomization. Finally, the extreme case of a very small error variance has been commented commented already above.

5 Concluding remarks

Our simulation results show that the PRAM-adapted ML estimator of the binary probit model facing misclassification can be used successfully in the SIMEX approach when the continuous regressor is observed with error. The performance of the estimators for our model is improved in terms of bias if the nonlinear extrapolation ist used. However this advantage is obtained at the cost of a sometimes much larger variation of the estimator. In particular this alternative cannot recommended if the error variance is small and the dependent variable has been randomized (misclassified).

Results from this paper can be used to estimate the probit model from anonymized data if PRAM and addition of noise are applied as anonymization procedures. We plan to extend our analysis to the case of an arbitrary number of regressors. In particular we want to study the case that binary regressors are included and are observed with errors as well.[12] This may lead to an approach which combines the traditional SIMEX with the Misclassification (MC) SIMEX recently proposed by Küchenhoff et al. (2005).

References

Carroll, R.J., Ruppert, D. and Stefanski, L.A., (1995). Measurement Error in Nonlinear Models. London: Chapman and Hall.

Cook, J.R. and Stefanski, L.A.(1994). Simulation-Extrapolation Estimation in Parametric Measurement Error Models. Journal of the American Statistical Association 89, 1314-1328.

Greene, W.H. (2000). Econometric Analysis. Upper Saddle River (NJ): Prentice Hall (fourth edition).

Frazis, H. und Loewenstein, M.A. (2003). Estimating Linear Regressions with Mismeasured, Possibly Endogenous, Binary Explanatory Variables . Journal of Econometrics, 117 , 151-178.

Hausman, J.A. Abrevaya, J. and Scott-Morton, F.M. (1998). Misclassification of the Dependent Variable in a Discrete-Response Setting. Journal of Econometrics, 87, 239-269.

Kommission zur Verbesserung der informationellen Infrastruktur (editor) (2001). Wege zu einer besseren informationellen Infrastruktur. Wiesbaden: Nomos (cited as KVI(2001)).

Küchenhoff, H., Mwalili, S.M., and Lesaffre, E. (2006). A general method for dealing with misclassification in regression: The misclassification SIMEX. Biometrics 62 , 85-96.

Lechner, S. and Pohlmeier, W. (2003). Schätzung ökonometrischer Modelle auf der Grundlage anonymisierter Daten. In: Gnoss, R. und Ronning, G. (editors). Anonymisierung wirtschaftsstatistischer Einzeldaten. Forum der Bundesstatistik, volume 42, 115-137.

Lechner, S. and Pohlmeier, W. (2005). Data Masking by Noise Addition and the Estimation of Nonparametric Regression Models. Jahrbücher für Nationalökonomie und Statistik, 225 , 517-528.

Neuhaus, J. (1999). Bias and efficiency loss due to misclassified responses in binary regression. Biometrika, 86, 843-855.

Ronning, G. (1991). Mikroökonometrie. Berlin: Springer.

Ronning, G. (2005). Randomized response and the binary probit model. Economics Letters, 86, 221-228.

Ronning, G. and Gnoss, R. (editors) (2003). Anonymisierung wirtschaftsstatistischer Einzeldaten. Statistisches Bundesamt. Forum der Bundesstatistik, volume 42, Wiesbaden.

Ronning, G., Sturm, R., Höhne, J., Lenz, J., Rosemann, M., Scheffler, M., Vorgrimler, D. (2005). Handbuch zur Anonymisierung wirtschaftsstatistischer Mikrodaten. Statistisches Bundesamt, Wiesbaden , series "Statistik und Wissenschaft", volume 4.

Särndal, C.-E., Swensson, B., and Wretman, J.(1992). Model Assisted Survey Sampling . New York: Springer.

[12] Usually this is called 'misclassification'. See Frazis and Loewenstein (2003) for most recent results in case of the linear regression model.

van den Hout, A. and van der Heijden, P.G.M.(2002). Randomized response, statistical disclosure control and misclassification: a review. International Statistical Review, 70, 2-69.

Warner, S.L.(1965). Randomized response: a survey technique for eliminating evasive answer bias. Journal of the American Statistical Association, 57, 622-627.

A Nonlinear Simex Extrapolation

A.1 Introduction

In this appendix we derive the solution for determining coefficients a, b and c of the extrapolation function

$$f(x) \quad = \quad a \; + \; \frac{b}{c \, + \, x} \tag{A-1}$$

if only three points are available (or used) and the point at $x = 0$ is included. (Note that in the SIMEX literature usually λ instead of x is used.) Motivation for this approach stems from the fact that this nonlinear functions can be seen as optimal in some stochastic models (see Cook and Stefanski 1994) and that usually only very few points are used in extrapolation. Since for three points an explicit solution is easily obtained and on the other hand the approximation of the function to more than three points asks for an iterative numerical procedure, this may be an acceptable compromise in finding a "good" approximation. Note that depending on the sign of b the function tends towards (minus) infinity at $x = -c$, i.e. it has a pole at this point. Moreover the function is always monotonically in/de-creasing except in the trivial case $b = 0$ whereas the quadratic function is able to describe non-monotonic behavior.

In section A.2 the solution is given and discussed. Section A.3 then contains the proof of the result.

A.2 Results

We assume that three points (y_0, x_0) , (y_1, x_1) , (y_2, x_2) are available for fitting the function and that in particular for the first point $x_0 \equiv 0$ holds. For easier exposition we set $x_1 = 1$ and $x_2 = 2$ although the solution could also be given in general terms of x_1 and x_2. Note that in the SIMEX approach these values can be fixed by the user. Our special choice corresponds to a subset of the usual SIMEX approach which considers $x \, \varepsilon \, \{0, 0.5, 1, 1.5, 2\}$.

The solution

$$\begin{aligned} c &= 2\,\frac{y_1 - y_2}{y_2 - 2y_1 + y_0} \\ a &= y_1 \, - \, (y_0 \, - \, y_1)\, c \\ b &= (y_0 \, - \, a)\, c \end{aligned}$$

or, equivalently, when given explicitly only in terms of y_0, y_1 and y_2

$$a \;=\; \frac{y_0(y_2 - y_1) \, - \, y_2(y_1 - y_0)}{y_2 - 2y_1 + y_0} \tag{A-2}$$

$$b \;=\; 2\,\frac{(y_1 - y_0)\,(y_0 - y_2)\,(y_2 - y_1)}{(y_2 - 2y_1 + y_0)^2} \tag{A-3}$$

$$c \;=\; 2\,\frac{y_1 - y_2}{y_2 - 2y_1 + y_0} \tag{A-4}$$

satisfies exactly the three points given above. Since the denominator of all three solutions equals zero if the three points lie on a straight line (and therefore $y_0 - y_1 \;=\; y_1 - y_2$ holds), this case has to be excluded!

Extrapolation of the function for $x = -1$ is the main concern of the SIMEX approach. The above solution leads to

$$f(-1) \quad = \quad a \;+\; \frac{b}{c\;-\;1}$$

or, equivalently,

$$\begin{aligned} f(-1) \quad &= \quad \frac{y_0(y_2 - y_1) \;-\; y_2(y_1 - y_0)}{y_2 - 2y_1 + y_0} \\ &\quad + \quad 2\,\frac{(y_1 - y_0)\,(y_0 - y_2)\,(y_2 - y_1)}{(y_2 - 2y_1 + y_0)(4y_1 - 3y_2 - y_0)} \end{aligned} \tag{A-5}$$

A.3 Proof

Assuming that an exact fit of the three points to the function (A-1) exists we set

$$y_j \quad = \quad f(x_j) \quad , \; j = 0, 1, 2.$$

For the special choice of the x_j's given above we obtain the following set of equations:

$$\begin{aligned} y_0 \quad &= \quad a \;+\; \frac{b}{c} \\ y_1 \quad &= \quad a \;+\; \frac{b}{c+1} \\ y_2 \quad &= \quad a \;+\; \frac{b}{c+2} \end{aligned} \tag{A-6}$$

>From the first two equations of (A-6) we obtain

$$\begin{aligned} y_1 - y_0 \quad &= \quad \tfrac{b}{c+1} - \tfrac{b}{c} \\ &= \quad -\tfrac{b}{c\,(c+1)} \end{aligned}$$

and from the last two equations

$$\begin{aligned} y_2 - y_1 \quad &= \quad \tfrac{b}{c+2} - \tfrac{b}{c+1} \\ &= \quad -\tfrac{b}{(c+1)\,(c+2)} \end{aligned}$$

and therefore

$$\begin{aligned} \tfrac{y_2 - y_1}{y_1 - y_0} \quad &= \quad \tfrac{(-b)}{(c+2)(c+1)} \left[\tfrac{(-b)}{(c+1)c}\right]^{-1} \\ &= \quad \tfrac{c}{c+2} \end{aligned}$$

from which the solution for coefficient c is obtained as follows: Rearranging the last equation we get

$$(y_2 - y_1)c + 2(y_2 - y_1) = (y_1 - y_0)c$$

and therefore

$$c \quad = \quad 2\,\frac{y_1 - y_2}{y_2 - 2y_1 + y_0} \tag{A-7}$$

which equals (A-4) in section A.2.

We now write the two first equations of (A-6) as follows:

$$\begin{array}{lcl} (y_0 - a)c & = & b \\ (y_1 - a)c \;+\; (y_1 - a) & = & b \end{array}$$

Subtracting the second equation from the first one results in

$$(y_0 - y_1)c = y_1 - a$$

and therefore using (A-7) we get

$$\begin{aligned} a &= y_1 \;-\; (y_0 \;-\; y_1)\,c \\ &= y_1 \;-\; 2\frac{(y_0 - y_1)(y_1 - y_2)}{y_2 - 2y_1 + y_0} \\ &= \frac{y_1\,[(y_2 - y_1) - (y_1 - y_0)] \;-\; 2(y_1 - y_0)(y_2 - y_1)}{y_2 - 2y_1 + y_0} \\ &= \frac{y_0(y_2 - y_1) \;-\; y_2(y_1 - y_0)}{y_2 - 2y_1 + y_0} \qquad \text{(A-8)} \end{aligned}$$

which equals (A-2) in section A.2.

Finally we obtain b from the first equation which we write again as

$$b \quad = \quad (y_0 - a)\,c \quad .$$

Inserting (A-7) and (A-8) we obtain

$$\begin{aligned} b &= \left[y_0 - y_1 \;+\; 2\frac{(y_0 - y_1)(y_1 - y_2)}{y_2 - 2y_1 + y_0} \right] \;\; 2\,\frac{y_1 - y_2}{y_2 - 2y_1 + y_0} \\ &= 2\,\frac{[(y_0 - y_1)(y_2 - 2y_1 + y_0) \;+\; 2(y_0 - y_1)(y_1 - y_2)]\;(y_1 - y_2)}{(y_2 - 2y_1 + y_0)^2} \\ &= 2\,\frac{(y_0 - y_1)\;[(y_2 - 2y_1 + y_0 + 2y_1 - 2y_2)]\;(y_1 - y_2)}{(y_2 - 2y_1 + y_0)^2} \\ &= 2\,\frac{(y_0 - y_1)\,(y_0 - y_2)\,(y_1 - y_2)}{(y_2 - 2y_1 + y_0)^2} \\ &= 2\,\frac{(y_1 - y_0)\,(y_0 - y_2)\,(y_2 - y_1)}{(y_2 - 2y_1 + y_0)^2} \qquad \text{(A-9)} \end{aligned}$$

which equals (A-3) in section A.2.

We now derive the expression for the function at $x = -1$. First note that

$$c - 1 \quad = \quad 2\,\frac{y_1 - y_2}{y_2 - 2y_1 + y_0} \;-\; 1 \quad = \quad \frac{4y_1 - 3y_2 - y_0}{y_2 - 2y_1 + y_0}$$

Therefore inserting the solutions of a, b and c into

$$f(-1) \quad = \quad a \;+\; \frac{b}{c - 1}$$

we get

$$\begin{aligned} f(-1) &= \frac{y_0(y_2 - y_1) \;-\; y_2(y_1 - y_0)}{y_2 - 2y_1 + y_0} \\ &\quad +\; 2\,\frac{(y_1 - y_0)\,(y_0 - y_2)\,(y_2 - y_1)}{(y_2 - 2y_1 + y_0)^2} \left[\frac{4y_1 - 3y_2 - y_0}{y_2 - 2y_1 + y_0} \right]^{-1} \\ &= \frac{y_0(y_2 - y_1) \;-\; y_2(y_1 - y_0)}{y_2 - 2y_1 + y_0} \\ &\quad +\; 2\,\frac{(y_1 - y_0)\,(y_0 - y_2)\,(y_2 - y_1)}{(y_2 - 2y_1 + y_0)(4y_1 - 3y_2 - y_0)} \end{aligned}$$

which equals (A-5) in section A.2.

IAW-Diskussionspapiere

Bisher erschienen:

Nr. 1
Das Einstiegsgeld – eine zielgruppenorientierte negative Einkommensteuer: Konzeption, Umsetzung und eine erste Zwischenbilanz nach 15 Monaten in Baden-Württemberg
Sabine Dann / Andrea Kirchmann / Alexander Spermann / Jürgen Volkert

Nr. 2
Die Einkommensteuerreform 1990 als natürliches Experiment. Methodische und konzeptionelle Aspekte zur Schätzung der Elastizität des zu versteuernden Einkommens
Peter Gottfried / Hannes Schellhorn

Nr. 3
Gut betreut in den Arbeitsmarkt? Eine mikroökonomische Evaluation der Mannheimer Arbeitsvermittlungsagentur
Jürgen Jerger / Christian Pohnke / Alexander Spermann

Nr. 4
Das IAW-Einkommenspanel und das Mikrosimulationsmodell SIMST
Peter Gottfried / Hannes Schellhorn

Nr. 5
A Microeconometric Characterisation of Household Consumption Using Quantile Regression
Niels Schulze / Gerd Ronning

Nr. 6
Determinanten des Überlebens von Neugründungen in der baden-württembergischen Industrie – eine empirische Survivalanalyse mit amtlichen Betriebsdaten
Harald Strotmann

Nr. 7
Die Baulandausweisungsumlage als ökonomisches Steuerungsinstrument einer nachhaltigkeitsorientierten Flächenpolitik
Raimund Krumm

Nr. 8
Making Work Pay: U.S. American Models for a German Context?
Laura Chadwick, Jürgen Volkert

Nr. 9
Erste Ergebnisse von vergleichenden Untersuchungen mit anonymisierten und nicht anonymisierten Einzeldaten am Beispiel der Kostenstrukturerhebung und der Umsatzsteuerstatistik
Martin Rosemann

Nr. 10
Randomized Response and the Binary Probit Model
Gerd Ronning

Nr. 11
Creating Firms for a New Century: Determinants of Firm Creation around 1900
Joerg Baten

Nr. 12
Das fiskalische BLAU-Konzept zur Begrenzung des Siedlungsflächenwachstums
Raimund Krumm

Nr. 13
Generelle Nichtdiskontierung als Bedingung für eine nachhaltige Entwicklung?
Stefan Bayer

Nr. 14
Die Elastizität des zu versteuernden Einkommens. Messung und erste Ergebnisse zur empirischen Evidenz für die Bundesrepublik Deutschland.
Peter Gottfried / Hannes Schellhorn

Nr. 15
Empirical Evidence on the Effects of Marginal Tax Rates on Income – The German Case
Peter Gottfried / Hannes Schellhorn

Nr. 16
Shadow Economies around the World: What do we really know?
Friedrich Schneider

Nr. 17
Firm Foundations in the Knowledge Intensive Business Service Sector. Results from a Comparative Empirical Study in Three German Regions
Andreas Koch / Thomas Stahlecker

Nr. 18
The impact of functional integration and spatial proximity on the post-entry performance of knowledge intensive business service firms
Andreas Koch / Harald Strotmann

Nr. 19
Legislative Malapportionment and the Politicization of Germany's Intergovernmental Transfer System
Hans Pitlik / Friedrich Schneider / Harald Strotmann

Nr. 20
Implementation ökonomischer Steuerungsansätze in die Raumplanung
Raimund Krumm

Nr. 21
Determinants of Innovative Activity in Newly Founded Knowledge Intensive Business Service Firms
Andreas Koch / Harald Strotmann

Nr. 22
Impact of Opening Clauses on Bargained Wages
Wolf Dieter Heinbach

Nr. 23
Hat die Einführung von Gewinnbeteiligungsmodellen kurzfristige positive Produktivitätswirkungen? – Ergebnisse eines Propensity-Score-Matching-Ansatzes
Harald Strotmann

Nr. 24
Who Goes East? The Impact of Enlargement on the Pattern of German FDI
Claudia M. Buch / Jörn Kleinert

Nr. 25
Estimation of the Probit Model from Anonymized Micro Data.
Gerd Ronning / Martin Rosemann