GMOD+: An Innovative Tax-Benefit Microsimulation Modeling Tool

Gerhard Wagenhals and Jürgen H. Buck
Department of Statistics and Econometrics
University of Hohenheim
70593 Stuttgart, Germany
E-Mail: wagenhals@uni-hohenheim.de

KEYWORDS

GMOD, Microsimulation Modeling, Propensity Score Matching, Statistical Matching, Tax-Benefit Simulation

ABSTRACT

This paper presents GMOD+, an innovative modeling tool for behavioral tax-benefit microsimulation in Germany. GMOD+ is an evolution of the GMOD project. This new suite includes an extended data base and a state of the art microeconometric behavioral simulation model for analyzing the impact of complex tax-benefit reforms in Germany on the labor market.

IMPORTANCE OF MICROSIMULATION MODELING FOR POLICY DECISION MAKERS

For a long time, political decision makers have been interested in predicting fiscal and social impacts of policy decisions. A huge number of models has been developed, some with a macroeconomic focus, others centered on microeconomic theory. Due to the availability of both individual and household micro data and increasing computing power, microsimulation models have become more and more popular. The key advantage of microsimulation modeling is that micro models actually map the decision level in society: the individual person in a household context.

One field of political decisions affects always every citizen and is therefore subject to controversial discussions in most economies: the tax and benefit system. In the past, politicians were often surprised not only by the fiscal impact of changes in tax and benefit legislature, but also by their distributional effects. Therefore, there is a quite urgent need for improving tax-benefit forecasts. The quality of forecasts is very much influenced by the available data basis. As collection of tax relevant data is subject to legal restrictions, the available data often lack the richness of details needed for complex simulations.

GMOD: A TAX-BENEFIT MICROSIMULATION TOOL

Tax-benefit simulation models are increasingly used as an input to the process of policy making in Germany. Such models have gained an active role in the public discussion and evaluation of tax-benefit policies. The models, their databases, operating horizons and range of applications vary widely (see [20] for a survey).

GMOD is one of these tax-benefit models. It is the only German tax-benefit microsimulation model which has fully capitalized on a panel data structure which spans more than two decades. It is the only German tax-benefit microsimulation model that can be used to estimate and test truly dynamic intertemporal panel data models.

Since the late 1980s, GMOD has been developed in several stages from a simpleGAUSS program reflecting the personal tax scale according to §32a of the German income tax law *(EkSt, Einkommensteuergesetz)* and the most important statutory deductions to a very detailed mapping of the complex rules and interactions of the German tax, social security and benefit system.

GMOD+ extends GMOD in two ways ways. (1) It extends the 1998 database using statistical matching techniques and (2) it routinely includes a state of the art cross-section microeconometric discrete choice tax-benefit model.

Model

GMOD consists of a suite of program and help files written in the highly flexible statistical programming language STATA. These modules simulate the policy rules of the major federal tax, social security and transfer programs, including eligibility, entitlement and interaction. The model simulates the payment of personal income taxes and social security contributions and the receipt of public transfers. It applies detailed statutory provisions to a database of income units representing the German population.

Database

The GMOD basefile comprises income units, the individuals of which are a representative sample of the German population. This database is constructed using GSOEP, the German Socio-Economic Panel (see [4]), and extensive juridical and other information. Each record in the basefile represents an income unit

Proceedings 20th European Conference on Modelling and Simulation
Wolfgang Borutzky, Alessandra Orsoni, Richard Zobel © ECMS, 2006
ISBN 0-9553018-0-7 / ISBN 0-9553018-1-5 (CD)

(mostly, but not always, a household) and contains information about the income unit as a whole and information on each person in the income unit. In 2002, the basefile consisted of 12,692 households with 23,892 persons.

Operating Time

Currently, GMOD operates on data from 1983 to 2002. It allows to analyze changes in labor supply and wage structures over two decades using panel data methods. Recently, the model was extended and improved especially with respect to housing and social assistance benefit code, and with tax reform policy proposals up to 2010, so that the impact of many tax-benefit and social security reforms can be simulated.

Individual Behavior

GMOD did not include behavioral equations routinely because behavioral simulation results depend on additional sources of error such as sample variation or specification errors in behavioral relationships. GMOD+, however, includes a state of the art discrete choice behavioral microsimulation model that allows the simulation of a rich variety of behavioral patterns.

Applications

A series of GMOD based papers has been published. Based on a variety of microeconometric models, these analyses include the impact of tax reform acts and proposals on the labor supply of married women ([12], [13], [15], [16], [17], [18], [19]), tax and benefit reforms on the labor force participation and welfare dependence of single mothers ([10], [6], income taxation on the intertemporal labor supply of married women ([7]), and a potential removal of the income tax splitting rule for married couples ([11], [14], [21]). It has been shown (see [1]) that extrapolating GMOD results to population totals compares well with the benchmark figures published by the Federal Statistical Office [3].

STATISTICAL MATCHING IN TAX-BENEFIT MICROSIMULATION MODELING

Statistical matching is used to combine two files $(\mathbf{X},\mathbf{Z})$ (usually called recipient sample) and $(\mathbf{Y},\mathbf{Z})$ (usually called donor sample). These files include common variables $\mathbf{Z}$, whereas the variables $\mathbf{X}$ and $\mathbf{Y}$ are never jointly observed and therefore only available in one of the files. The basic idea of a matching algorithm is to find the most similar observation in the donor sample for each observation in the recipient sample using the information in the common variablen $\mathbf{Z}$. The corresponding values for $\mathbf{Y}$ are then added to the files, thus creating artificial observations containing all the variables $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$.

If the observations in the donor and the recipient file refer to the same individuals, the problem is quite easy from a statistical point of view. In this case the terms exact matching or record linkage are usually used. The difficulties that have to be overcome are more technical than statistical, e.g. there are different strategies how to cope with inconsistent data in the two files. If the individuals in both files are not *identical*, but *similar*, the problem is usually referred to as statistical matching.

In the past, researchers often tried to solve the problem using regression methods. In a first step, a regression was carried out using the common files $\mathbf{Z}$ in the donor file as regressors, the variables $\mathbf{Y}$ being explained within the regression. In a second step, the regression results were used to generate estimated values $\hat{\mathbf{Y}}$ for the variables missing in the recipient sample using the common files $\mathbf{Z}$ as explaining variables. This method has some disadvantages that question its usability. First, in order to obtain reliable results for $\hat{\mathbf{Y}}$ there has to be a strong correlation of $\mathbf{Y}$ and the common variables $\mathbf{Z}$ which is not present in many practical cases. Second, the use of a regression needs the specification of a functional form, thus reducing the flexibility of the approach in comparison to nonparametric methods. Third, the use of regression generates an artificial dataset that has very limited variance of the variable $\hat{\mathbf{Y}}$ compared to the original data $\mathbf{Y}$ because all the variance that is in the error terms of the regression model is eliminated. In conclusion, regression methods usually do not provide acceptable results for microsimulation modeling.

Statistical matching methods have been developed to overcome this problem. There are two main methodical clusters: (1) Methods resulting in a file with conditional independence of $\mathbf{X}$ and $\mathbf{Y}$, given the common variables $\mathbf{Z}$ and (2) methods that use additional information on the $(\mathbf{X},\mathbf{Y})$ distribution from a third party source, thus resulting in a matched file without conditional independence.

Theoretical arguments and simulation experiments show that methods of cluster (1) result in a matched file with conditional independence of the variables never jointly observed $\mathbf{X}$ and $\mathbf{Y}$ given the common variables $\mathbf{Z}$ regardless conditional independence being present in the original data. There are two main approaches that have many steps in common.

The most often used approach is nearest neighbor matching. Those algorithms determine the most similar observation in the donor file for each observation in the recipient file using a distance function like the Euclidian distance or the Mahalanobis distance. The calculation of the distance function is quite simple as various algorithms developed for the classical transport problem in operations research can be used. Prior to calculating the distance function it can be necessary to add weights to the common variables reflecting the importance of each variable and adjusting different scales or variances. The Mahalanobis distance has the advantage of implicitly correcting for differences in variance within the used variables.

Another interesting algorithm that has rarely been used so far is propensity score matching. Here, in a first step a binary dummy variable is introduced indicating if the observation belongs to the recipient file (dummy value 1) or the donor file (dummy value 0). Then a logit or probit regression is calculated using the common variables as regressors. The outcome of the regression – which can be interpreted as the probability of being part of the recipient file for each observation given a set of common variables – for each observation is called the propensity score. The distance function then reduces to a simple calculation of the difference of the propensity scores. Matching is then performed again using a nearest neighbor technique: for each observation in the recipient file the observation in the donor file with the minimum distance is determined. The advantage of propensity score matching is that via the logit or probit estimation each common variable's influence is integrated in calculation of the propensity score, therefore implicitly finding an appropriate weight for each common variable. [2] contains a detailed description of those algorithms.

The methods of cluster (2) take into account additional information about the common distribution of the variables never jointly observed. In practice, this information is very rarely available. One algorithm was first introduced by Kadane and further developed by Moriarity and Scheuren. They use regression methods to add estimates for the variables not observed in the samples and then calculate the minimum distance based on all variables. For a detailed description, see [5] and [8]. This is a very interesting approach, but we do not know of any practical application yet. Another way of handling the problem that was recently suggested is the interpretation of the statistical matching task as a missing data problem and the use of multiple imputation methods. Those methods have currently been fully implemented only for univariate variables, see [9]. There are some doubts whether it is a good idea to use multiple imputation to generate datasets for microsimulation models. The main criticism is that multiple imputation algorithms usually give different outcomes (typically 5–10), so the researcher would have to run the simulation on various datasets.

Matching algorithms can be easily modified by introducing additional restrictions, such as limiting how often observations in the donor files are allowed to be used for matching. If there are huge structural differences in the data, it can also be very useful to restrict matching to a subset of total data (e. g. by sex, by region, etc.).

Regardless of which algorithm is used, it should be kept in mind that statistical matching does not add any additional information about the common distribution of the variables that are never jointly observed. The only link between those variables is established via the common variables or (if type (2) algorithms are used) by a priori assumptions that are made. If the only analytical target is the $(\mathbf{X},\mathbf{Y})$ distribution, statistical matching is not needed. All information can be obtained by analyzing the $(\mathbf{X},\mathbf{Z})$ distribution and the $(\mathbf{Y},\mathbf{Z})$ distribution separately. [2] gives a review of the different approaches, their advantages, and limitation from a microsimulation point of view. It also contains a discussion of the statistic properties of the various algorithms.

THE GMOD+ DATASET

From a tax-benefit microsimulation point of view, the GSOEP lacks the richness of tax related details desired by microsimulation modelers. Among the missing tax details, the most important are the tax deductible income-related expenses (*Werbungskosten*) and the information whether individuals are member of religious organizations. The latter is very important because German tax laws allow a tax deduction of payments to the church (*Kirchensteuer*). There are other missing items like tax deductible donation to welfare organizations or tax counseling expenses which are not that important because of their relatively small amounts. The effect of the missing tax details is that a tax simulation that is only based on GSOEP data is not as accurate as it could be. In general, there is a tendency to obtain estimates that give a net income that is too low, thus overestimating the amount of tax.

For 1998, German tax authorities have published a very useful source of data for the first time, the FAST 98 dataset. This is a 10 percent sample of all income tax declarations (*Steuererklärung*) in an anonymized form. This dataset contains all tax related details. Despite this richness of tax details, this dataset cannot be used as single basis for some tax simulation analyses because it only contains details about the persons that actually do their taxes. So this sample is not representative for the whole population and therefore cannot be used for labor participation studies or for studies about the effect of changes in social transfer systems. The 1998 data are the most recent data available. The data is currently collected every three years, so the next FAST dataset will be FAST 2001 which is expected to appear in autumn 2006.

This led to the idea of adding information from the FAST 98 dataset to the GSOEP in order to construct a merged database for tax simulation for the year 1998. Based on in depth analysis of the data we found very little evidence that there is a significant functional relationship between the common variables and the variables never jointly observed. We therefore tried to replicate the distribution of tax related details only available in the FAST dataset in the constructed dataset. The implicit assumption made is that the official tax data are the "real" data when it comes to taxation issues.

As there was no economic theory or evidence found suggesting that the conditional independence assumption does not hold here and as there was no additional source of information available, we decided to use propensity score matching to add the relevant information from FAST data to the constructed dataset based on GSOEP. The tax details added to the GSOEP data were income-related expenses, tax counseling expenses, donations to welfare organizations, and religious affiliation (the only relevant information here is whether or not a taxpayer is member of an official church and therefore pays church tax).

Table I shows an example of the results: a comparison between the key distribution statistics of the original data FAST 98 and the matched dataset. Please note that in 1998, every taxpayer was allowed to deduct at least 1 023 EUR (*Werbungskosten-pauschale*), a deduction of higher expenses was only allowed

		FAST 98	Matched data	Dev.
Quantiles	1 %	1 023 EUR	1 023 EUR	0.0 %
	5 %	1 023 EUR	1 023 EUR	0.0 %
	10 %	1 023 EUR	1 023 EUR	0.0 %
	25 %	1 023 EUR	1 023 EUR	0.0 %
	50 %	1 023 EUR	1 084 EUR	+5.9 %
	75 %	2 205 EUR	2 207 EUR	+0.1 %
	90 %	3 844 EUR	4 189 EUR	+9.0 %
	95 %	5 240 EUR	5 682 EUR	+8.4 %
	99 %	8 847 EUR	9 299 EUR	+5.1 %
Mean		1 928.40 EUR	2 030.19 EUR	+5.7 %

TABLE I: Matching results for income-related expenses

if receipts summing up to a higher amount were shown to the authorities. [2] contains a very detailed description of the algorithms used and all the other results obtained. In general it can be stated that the distribution of the FAST data could be mapped astonishingly well to the matched data.

THE GMOD+ BEHAVIORAL MODEL

Due to data restrictions, the GMOD+ operating time is currently restricted to 1998. Though this is a temporary disadvantage, it allows us to apply state of the art discrete choice modeling techniques which are still impossible to apply to the complete GMOD 1983–2002 panel-data basefile due to the "curse of dimensionality".

Motivation: The Random Utility Model

We start by assuming that individual (for singles) or household preferences (for couples) may be represented by a stochastic utility function (unit i, alternative j)

$$U_{ij} = V_{ij} + \varepsilon_{ij} \quad (i = 1, \dots, n; j = 0, 1, 2, \dots, J) \tag{1}$$

V_{ij} denotes the observable (up to a finite set of parameters) "representative" utility, and ε_{ij} the unobservable random part of utility. We obtain a logit model by assuming that ε_{ij} is i.i.d. Type I extreme value distributed. As the difference between two Type I extreme value distributed random variables follows a logistic distribution we can apply McFadden's theorem: A household acts such as to maximize of the stochastic utility function (1) if and only if ε_{ij} are i.i.d. with $F(\varepsilon_{ij}) = \exp[-\exp(-\varepsilon_{ij})]$. This yields the response probability

$$P(y_i = j|x_i) = \frac{\exp(x_{ij}\beta)}{\sum_{h=0}^{J} \exp(x_{ih}\beta)} \quad (j = 0, 1, \dots, J) \tag{2}$$

where y_i is a choice vector of unit i and x is a set of conditioning observable variables. $\{0, 1, \dots, J\}$ is the set of feasible choices. Maximizing the conditional likelihood function

$$\max_{\beta} \sum_{i=1}^{n} \sum_{j=0}^{J} 1[y_i = j] \ln[P(y_i = j|x_i, \beta)] \tag{3}$$

yields estimates for the unknown coefficients β which under general regularity assumptions are consistent and asymptotically normal. The likelihood function is globally concave. In our case $y = (c, l_1, \dots, l_k)$ where where c denotes household consumption and l is the vector of individual leisures $l = T - h$ (i.e. vector of time endowments T minus vector of individual hours of work h) of all k household members.

A Discrete Choice Model of Labor Supply

We apply the random utility approach in choosing a quadratic direct representative utility function (omitting the above indices to avoid cluttering of notation)

$$U(c,l) = \beta_{cc}c^2 + \beta_{ll}l^2 + \beta_{cl}cl + \beta_c c + \beta_l l$$

Observed and unobserved heterogeneity enters through the linear utility parameters. We condition on demographics, potential wage, demand side characteristics policy variables and account for net income in different states of employment and out of employment. Each element of the vector of hours is chosen from a set of discrete alternatives. Household net income is derived for each element of the set of choice alternatives. We allow for the fact that individuals account for the decision of other household members when they select themselves into a consumption-leisure regime. More precisely, we assume that household members select themselves into one of six regimes: non-participation or one of five employment states $h = (0, 10, 20, 30, 40, 50)$ (the elements denoting hours per week). For couples we consider 36 states: $h \times h$. Actual working hours observed in the data are rounded to fit the elements in the restricted choice set. State specific household net incomes are calculated based on GMOD starting from the GMOD+ database. We derive individual gross earnings assuming state invariant gross wage rates, and derive the corresponding net income.

We deal with unobserved wages of nonworkers by estimating a MINCER type wage function based on the selected sample of workers. We use these estimates to draw wage rates for nonworkers, conditional on observed characteristics.

Supplementing representative household utility we add stochastic terms accounting for state specific errors and random preference heterogeneity and finally derive the probability of choosing any consumption-leisure combination in the set of feasible household decisions.

Summing Up

GMOD+ accounts for personal income taxes based on the sum of all seven types of income according to the German Income Tax Act: from agriculture and forestry to miscellaneous income which consists mainly in capital gains from private old-age pensions. It accounts for tax exemptions, and – for the first time in German tax simulation modeling – it accounts for allowances, deductions and itemized special expenses without assuming that all tax payers claim flat rates deductions only. This may improve the quality and precision of tax policy simulations considerably.

Furthermore, GMOD+ allows a detailed simulation of consumption, labor force participation and hours of work decisions of individuals and households. Estimation of household utility functions provides an opportunity to simulate the standard economic welfare measures such as the HICKSian Compensating Variation, the HICKSian Equivalent Variation or deadweight losses. GMOD+ also allows to decompose total labor supply, welfare and distribution effects into those induced by statutory changes ("cash gains") and those induced by behavioral changes. Finally, grossing up to represent population numbers is easily done.

CONCLUSIONS

Today, microeconometric tax-benefit microsimulation models are used extensively throughout the industrialized countries, often for predicting immediate revenue and distributional impacts of tax-benefit policy changes. Such models often provide detailed information about fiscal, distributional and behavioral impacts of tax-benefit reforms, but their success is often hampered by an insufficient database. In Germany, up to now only regression techniques have been used as a matching tool for tax-benefit microsimulations. For a long time, this has been known as inadequate due to the loss of variance. For the first time, we have applied modern statistical matching algorithms to generate a new database GMOD+ and extended the 1998 version of GMOD by a new discrete choice microsimulation model which allows complex interactions of the behavior of household members.

From a methodological point of view, our results show that the augmentation of a database for a microsimulation model is a very useful application of statistical matching. Propensity score matching proves to be a very flexible way of matching in microsimulation context. Nevertheless, the researcher should always keep in mind the assumptions that are implicitly made when using such algorithms, especially the fact that statistical matching does not add any information about the relationship of the variables that are never jointly observed.

From a practical point of view, our results allow far more detailed tax-benefit microsimulations for Germany than ever before. This has been recognized by the German Federal Ministry of Finance which has commissioned the authors of this paper to apply GMOD+ to assess the impacts of changes of the tax deductibility for income-related expenses, a hot topic in the current German tax reform discussion.

REFERENCES

[1] BRAUCHLE, C. M. Auf dem Sozio-ökonomischen Panel basierende Simulation und Hochrechnung steuerrelevanter Einzeldaten und Vergleich mit den Eckwerten der Lohn- und Einkommensteuerstatistik. Diplomarbeit, Universität Hohenheim, Stuttgart, April 1998.

[2] BUCK, J. *Datenfusion und Steuersimulation. Theorie und Empirie im Rahmen des Mikrosimulationsmodells GMOD.* Shaker, Aachen, 2006.

[3] FEDERAL STATISTICAL OFFICE. Finanzen und Steuern. Fachserie 14, Reihe 7.1.

[4] HOLST, E., LILLARD, D. R., AND DIPRETE, T. A. The German Socio-Economic Panel (GSOEP) after more than 15 years – Overview. *Vierteljahrshefte zur Wirtschaftsforschung (Quarterly Journal of Economic Research) 70*, 1 (2001), 7–14.

[5] KADANE, J. Some statistical problems in merging data files. *Compendium of Tax Research, U. S. Department of Treasury* (1978), p. 159–171.

[6] LAISNEY, F., LECHNER, M., STAAT, M., AND WAGENHALS, G. Labour force and welfare participation of lone mothers in Germany. *Économie Publique 2* (1999), 111–144.

[7] LAISNEY, F., LECHNER, M., VAN SOEST, A., AND WAGENHALS, G. A life cycle labour supply model with taxes estimated on German panel data: The case of parallel preferences. *The Economic and Social Review 24*, 4 (1993), 335–368.

[8] MORIARITY, C., AND SCHEUREN, F. Statistical matching: A paradigm for assessing the uncertainty in the procedure. *Journal of Official Statistics 17*, 3 (2001), p. 423–433.

[9] RÄSSLER, S. *Statistical Matching*. Springer, New York, 2002.

[10] STAAT, M., AND WAGENHALS, G. The labour supply of German single mothers: A bivariate probit model. *Vierteljahreshefte zur Wirtschaftsforschung 1/2* (1994), 113–118.

[11] STRØM, S., AND WAGENHALS, G. Female labour supply in the Federal Republic. *Jahrbücher für Nationalökonomie und Statistik 208*, 6 (1991), 575–595.

[12] WAGENHALS, G. Einkommensbesteuerung und Frauenerwerbstätigkeit. In *Bevölkerung und Wirtschaft*, B. Felderer, Ed. Duncker und Humblodt, Berlin, 1990, pp. 473–492. Schriften des Vereins für Socialpolitik.

[13] WAGENHALS, G. Income tax reform in Germany : A welfare analysis. In *Models and Measurement of Welfare and Inequality*, W. Eichhorn, Ed. Springer-Verlag, Berlin, Heidelberg, New York, 1994, pp. 419–432.

[14] WAGENHALS, G. Auswirkungen des Ehegattensplitting in der Bundesrepublik Deutschland. Ergebnisse einer mikroökonometrischen Analyse. In *Frauenpolitische Aspekte im Einkommensteuerrecht*, B. Seel, Ed. Hessisches Ministerium für Frauen, Arbeit und Sozialordnung, Wiesbaden, 1996, pp. 159–183.

[15] WAGENHALS, G. Wohlfahrt und Besteuerung. In *Wohlfahrtsmessung. Aufgabe der Statistik im gesellschaftlichen Wandel*, Statistisches Bundesamt, Ed. Metzler-Poeschel, Stuttgart, 1996, pp. 97–120.

[16] WAGENHALS, G. Arbeitsangebotseffekte des Steuer- und Transfersystems in der Bundesrepublik Deutschland. *Jahrbücher für Nationalökonomie und Statistik 220/2* (2000), 191–213.

[17] WAGENHALS, G. Incentive and redistribution effects of the German tax reform 2000. *FinanzArchiv 57* (2001), 316–332.

[18] WAGENHALS, G. Incentive and redistribution effects of the "Karlsruher Entwurf zur Reform des Einkommensteuergesetzes". The case of married couples. *Schmollers Jahrbuch, Journal for Applied Social Science Studies 4* (2001), 425–437.

[19] WAGENHALS, G. Ökonomische Wirkungen der Steuerreform 2000. *Materialien und Berichte 28* (2001), 5–11. Beiträge zum vierten Statistischen Kolloquium des Statistischen Landesamtes mit Vertretern badenwürttembergischer Universitäten.

[20] WAGENHALS, G. Tax-benefit microsimulation models for Germany: A survey. *IAW-Report 32*, 1 (2004), 55–74.

[21] WAGENHALS, G., AND KRAUS, M. *Neuansätze des Familienlastenausgleichs*. Hessisches Ministerium für Frauen, Arbeit und Sozialordnung, Wiesbaden, 1998.

AUTHOR BIOGRAPHIES

Prof. Dr. rer. pol. habil. **GERHARD WAGENHALS** is Full Professor of Statistics and Econometrics at the University of Hohenheim, Germany, and Research Fellow at the Institute for the study of Labor (IZA) in Bonn. He received his diploma in economics from the University of Tübingen, Germany, in 1976; his doctoral and habilitation degrees were received from the University of Heidelberg, in 1980 and 1984. He worked as a postdoc at the Department of Economics, University of Pennsylvania (1980–1982), as an Associate Professor (C2) at the University of Heidelberg (1986–1988), as a Visiting Professor at the Department of Economics, University of Bern (Switzerland) (1989–1990), as a Professor of Computational Economics at the Department of Economics, University of Paderborn (1990–1992), and since then as Full Professor of Statistics and Econometrics at the University of Hohenheim. His key research interests are microsimulation modeling and microeconometrics.

Dr. oec. **JÜRGEN BUCK** studied economics at the University of Hohenheim in Stuttgart, Germany. He obtained his diploma in economics in 2002. He has been working as a management consultant in McKinsey's Stuttgart Office serving primarily public sector clients. He is also doing research in econometrics at Hohenheim university together with G. Wagenhals where he obtained his PhD equivalent (Dr. oec.) in econometrics in 2006 working on statistical matching for microsimulation models. His key research interests are practical uses of statistical matching and microsimulation based tax modeling.