

# Misspecified heteroskedasticity in the panel probit model: A small sample comparison of GMM and SML estimators

Joachim Inkmann\*

University of Konstanz, Department of Economics and Statistics, 78457 Konstanz, Germany

## Abstract

This paper compares generalized method of moments (GMM) and simulated maximum likelihood (SML) approaches to the estimation of the panel probit model. Both techniques circumvent multiple integration of joint density functions without the need to restrict the error term variance-covariance matrix of the latent normal regression model. Particular attention is paid to a three-stage GMM estimator based on nonparametric estimation of the optimal instruments for given conditional moment functions. Monte Carlo experiments are carried out which focus on the small sample consequences of misspecification of the error term variance-covariance matrix. The correctly specified experiment reveals the asymptotic efficiency advantages of SML. The GMM estimators outperform SML in the presence of misspecification in terms of multiplicative heteroskedasticity. This holds in particular for the three-stage GMM estimator. Allowing for heteroskedasticity over time increases the robustness with respect to misspecification in terms of multiplicative heteroskedasticity. An application to the product innovation activities of German manufacturing firms is presented.

*Key words:* panel probit model; heteroskedasticity; conditional moment restrictions; optimal instruments; k-nearest neighbor estimation; GHK simulator

*JEL classification:* C14, C15, C23, C25

---

\* Financial support by the *Deutsche Forschungsgemeinschaft* (DFG) is gratefully acknowledged. I am particularly indebted to Bernd Fitzenberger, Michael Lechner, and Winfried Pohlmeier for valuable suggestions and discussion. I would like to thank my colleagues Frank Gerhard, Nikolaus Hautsch, Stefan Klotz, and Sikandar Siddiqui, as well as seminar participants in Konstanz and Mannheim, at the Panel Data Conference in Amsterdam 1996, the EC<sup>2</sup> Conference in Florence 1996, and the Econometric Society European Meeting in Istanbul 1996 for helpful comments. All errors are solely my responsibility.

## 1. Introduction

In recent years substantial progress has been achieved in estimating panel probit models without the need to assume restrictive variance-covariance specifications of the error term distribution. In particular, there are two promising approaches dating back to the early eighties that are now widely accepted as useful tools: the generalized method of moments (GMM) based on the seminal work of Hansen (1982) and applied to the panel probit model by Avery, Hansen and Hotz (1983) and the simulated maximum likelihood (SML) technique suggested by Lerman and Manski (1981). Both approaches solve the problem of multiple integration of the multivariate normal density function inherent in panel probit models. GMM estimates can be obtained from treating the panel probit model like a nonlinear regression model and restricting only the first conditional moment functions of the error terms which do not depend on off-diagonal elements of the error term variance-covariance matrix. SML estimates exploit the full information of the error term distribution but rely on Monte Carlo integration techniques to simulate the likelihood function.

Since their introduction both methods have been refined in detail. Chamberlain (1987) shows that the GMM estimates based on optimal instruments reach the lower asymptotic efficiency bound of any consistent and asymptotically normal estimator within the class of models based on conditional moment restrictions. Newey (1990) proposes a feasible estimation approach to this optimal GMM estimator in the case of conditional homoskedasticity involving nonparametric estimation techniques for those elements of the optimal instruments that depend on conditional expectations. Newey (1993) presents the corresponding estimator in the case of conditional heteroskedasticity. He suggests k-nearest neighbor estimation and series approximation to estimate the conditional variance-covariance matrix of the moment functions. The former nonparametric estimator is also used by Robinson (1987) and Delgado (1992) to derive feasible multiple equation generalized least squares estimators for linear and nonlinear regression models with heteroskedasticity of unknown form. First applications of the GMM estimator with nonparametrically estimated optimal instruments to the panel probit model include Inkmann and Pohlmeier (1995) and Bertschek and Lechner (1998).

The SML method has been improved by several authors with the introduction of more accurate and efficient simulation techniques for the likelihood function. Other criterion functions have been considered: McFadden (1989) introduces a method of simulated moments and Hajivassiliou and McFadden (1990) suggest simulation of the score function. Pakes and Pol-

lard (1989) provide a general central limit theorem that applies to a large class of simulation estimators with nonsmooth criterion functions. Hajivassiliou and Ruud (1994) present an excellent survey of the different simulation methods in models for limited dependent variables (LDV). These models usually rely on a latent normal regression model and require simulation of multivariate normal rectangle probabilities. An extensive comparison of 13 different simulation algorithms for this particular case is carried out by Hajivassiliou, McFadden, and Ruud (1996). The Geweke-Hajivassiliou-Keane (GHK) simulator turns out to be the most successful method with respect to root mean squared error performance in over 70 of their 84 experiments under consideration. For the simulation of the likelihood function of LDV models, the GHK simulator is applied by Börsch-Supan and Hajivassiliou (1993). The authors show that this SML estimator possesses desirable properties such as a smooth and unbiased simulated likelihood function. Applications of the GHK-SML estimator to the panel probit model include Keane (1994) and Mühleisen (1994).

This paper brings together the two existing lines of literature and focuses on the estimation of the panel probit model. The choice of this particular model is justified by its enormous practical importance and the fact that first applications of both GMM and SML methods include the panel probit model which highlights the value of the two approaches for this model. Significant attention is paid to the specification of the variance-covariance matrix of the error term in the latent normal regression model underlying the observed binary dependent variable. All aforementioned studies applying either GMM or SML to the panel probit model rest on the assumption of homoskedasticity over time and individuals. This paper presents GMM and SML estimators for the panel probit model that allow for heteroskedasticity over time. This extension makes it possible to test for the homoskedasticity assumption instead of just imposing the restriction. In fact, little attention seems to be paid to the homoskedasticity assumption in applied work. An exception is the work by Laisney, Lechner, and Pohlmeier (1992a) who apply minimum distance estimation to a random effects probit model of process innovations using a five year panel of manufacturing firms. The authors reject the hypothesis of constant error term variances over time if the unobserved individual effects are specified to be uncorrelated with the regressors.

The consequences of erroneously imposing homoskedasticity over time in the presence of heteroskedasticity become clear if one recalls the identification conditions for the unknown parameters in the panel probit model (cf. Heckman, 1981). The coefficients of the explanatory

variables are identified subject to a scale normalization in any arbitrarily chosen period. Given this normalization, the slope parameters and the scale coefficients are separately identified for the remaining periods. Imposing homoskedasticity implies not using the separate information and therefore creating bias if the estimated coefficients are still interpreted as slope coefficients. Depending on the variation of the error term variances over time, this bias may seriously affect any conclusions obtained from the model. In this sense, in particular probit models based on panel data covering a long period are likely to produce misleading results.

In a second attempt, this paper tries to evaluate the small sample consequences of misspecification in terms of multiplicative heteroskedasticity over individuals for the GMM and SML estimators under consideration. If not explicitly taken into account, multiplicative heteroskedasticity renders both classes of estimators inconsistent. However, the GMM estimators may show superior small sample properties compared to SML because their limited distributional assumptions should make them more robust with respect to misspecification. Both, GMM and SML estimators under the assumption of homoskedastic and heteroskedastic error terms over time are confronted with multiplicative heteroskedasticity over individuals. This is done to get an idea how less demanding identification conditions affect the robustness of the estimators with respect to misspecification.

The paper is organized as follows: the panel probit model is introduced in Section 2. The conditional moment approach to GMM estimation of this model is described in Section 3. Particular attention is paid to the optimal GMM estimator in the class of estimators using conditional moment functions as the only distributional assumption. An efficient estimator in the class of estimators exploiting full distributional information is presented in Section 4. This estimator simulates the likelihood function by means of the GHK algorithm. Section 5 presents the Monte Carlo experiments for the proposed estimators which focus on misspecification of the error term variance-covariance matrix of the latent normal regression model as described above. The data generating process implements a combined random effects and autocorrelated error term specification which is frequently used in applications of the GHK simulator to discrete choice panel data models (cf. Börsch-Supan et al., 1993, Mühleisen, 1994). It is shown, that the correctly specified experiment reveals the asymptotic efficiency advantages of the different estimators. Compared to the SML estimates, the GMM estimates turn out to be much more robust with respect to misspecification in form of multiplicative heteroskedasticity. This holds in particular for the GMM estimator that rests on the optimal

choice of instruments. Allowing for heteroskedasticity over time increases the robustness of all estimators.

To demonstrate the practicability of the proposed estimators in nonexperimental work, they are applied in Section 6 to the data set underlying the aforementioned empirical analysis of process innovations by Laisney et al. (1992a). The sample consists of 1,325 West German manufacturing firms observed over the period 1984 – 1988. In this paper, the data are used to estimate the probability of successful product innovations.<sup>1</sup> The results obtained by the different estimators turn out to be very similar. Section 7 concludes and suggests a possible direction for future research.

## 2. The panel probit model

Observed binary outcomes of utility maximizing behavior of individuals  $i$  at time  $t$  can be conveniently thought of as generated by a random utility model of the form

$$y_{it}^* = x_{it}'\beta + u_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (2.1)$$

where the latent variable  $y_{it}^*$  represents the net utility of the individual emanating from the choice of a particular state of interest, explained by a  $K \times 1$  vector of observed regressors  $x_{it}$  with a corresponding parameter vector  $\beta$  and the unobserved error term  $u_{it}$ . The individual occupies the state of interest in case of positive net utility indicated by the binary variable  $y_{it}$  according to the observability rule

$$y_{it} = \mathbf{1}[x_{it}'\beta + u_{it} \geq 0] \quad (2.2)$$

under the assumption that  $y_{it}^*$  is adequately scaled. This threshold crossing binary choice model is conveniently rewritten in matrix notation as  $y_i = \mathbf{1}[x_i\beta + u_i \geq 0]$ , where  $y_i$  is the  $T \times 1$  vector of dependent variables for observation  $i$ ,  $x_i = (x_{i1}, \dots, x_{iT})'$  is the  $T \times K$  matrix of explanatory variables, and  $u_i = (u_{i1}, \dots, u_{iT})'$  the  $T \times 1$  vector of error terms. The panel probit model results from the assumption of jointly normal error terms,  $u_i \sim \text{iid } N(0, \Sigma)$ , where an arbitrary main-diagonal element of  $\Sigma$  is set to unity to ensure identification of the

---

<sup>1</sup> Previous studies applying alternative panel probit estimation techniques to these data to estimate process and product innovation equations include Laisney et al. (1992a, b), König et al. (1994), Bertschek (1995), Inkmann and Pohlmeier (1995), and Bertschek and Lechner (1998).

slope parameters (cf. Heckman, 1981). It will be required throughout this paper that the  $u_i$  be distributed mutually independent of the explanatory variables which rules out the fixed effects specification but includes the random effects model. The data  $z_i = (y_i, x_i)$  are independently and identically distributed according to a joint probability distribution characterized by the true parameters  $\beta_0$  and  $\Sigma_0$ .

Usual maximum likelihood estimation of the unknown parameters requires the computation of a T-dimensional integral over a multivariate normal density function without further restriction of  $\Sigma$ . This task becomes computationally prohibitive even for moderate sizes of T. One possibility to circumvent this problem is to impose restrictions on the parameters of  $\Sigma$  as in the case of the popular random effects specification with equicorrelated residuals (cf. Hsiao, 1996) resulting from the error components specification

$$u_{it} = \tau_i + v_{it}. \quad (2.3)$$

Assuming that the random effect  $\tau_i \sim N(0, \sigma_\tau^2)$  and the white noise error  $v_{it} \sim N(0, \sigma_v^2)$  are mutually independent distributed and uncorrelated with the regressors  $x_{it}$ , the variance-covariance matrix  $\Sigma$  reduces to  $\Sigma_{RE} = \sigma_v^2 I + \sigma_\tau^2 11'$  and the computation of the joint probability of  $y_i$  conditional on  $\tau_i$  simplifies to an univariate integration problem.<sup>2</sup> Two alternative estimation approaches allowing for less demanding specifications of  $\Sigma$  are presented in the following two sections.

### 3. Generalized method of moments estimation

The first application of the GMM principle, introduced by Hansen (1982), to the estimation of panel probit models under consideration here is the work by Avery, Hansen, and Hotz (1983). Lechner and Breitung (1996) provide a comprehensive survey of GMM estimators for nonlinear panel data models. This section reviews the conditional moment approach to GMM estimation formulated by Newey (1990, 1993) for a general class of nonlinear regression models. The estimation principle can be described by introducing the following set of heteroskedastic conditional moment functions

---

<sup>2</sup> See Butler and Moffitt (1982) who introduce an efficient computation procedure for this case and Guilkey and Murphy (1993) for the small sample properties of this approach.

$$\rho_{it} = y_{it} - E[y_{it}|x_{it}], \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (3.1)$$

$$\text{where } E[y_{it}|x_{it}] = \Phi\left(\frac{x'_{it}\beta}{\sigma_t}\right)$$

in the panel probit case and  $\Phi(\cdot)$  denotes the c.d.f. of the standard normal distribution. As the only substantial distributional assumption it is required throughout the section that the true distribution satisfies the conditional moment restriction

$$E[\rho(z_i, \theta_0)|x_i] = 0, \quad (3.2)$$

$$\text{where } \rho(z_i, \theta) = (\rho_{i1}, \dots, \rho_{iT})'$$

and  $\theta$  includes the  $Q = K + T - 1$  unknown parameters  $\beta$  and  $\sigma_t^2$ , the main-diagonal elements of  $\Sigma$  for  $t = 2, \dots, T$ .<sup>3</sup> The estimation procedure makes use of the fact that the conditional moment functions are uncorrelated with any function of  $x_i$  which leads to unconditional moment functions suggesting a GMM estimator to obtain a consistent estimate of the true  $\theta_0$ . To be more specific, define a  $(T \times L)$ ,  $L \geq Q$ , matrix of suitable instruments  $A(x_i)$  that are functions of the explanatory variables and possibly of the parameters to be estimated. Then unconditional moment functions are obtained by

$$g(z_i, \theta) = A(x_i)' \rho(z_i, \theta), \quad (3.3)$$

where the orthogonality condition

$$E[g(z_i, \theta_0)] = E[A(x_i)' E[\rho(z_i, \theta_0)|x_i]] = 0 \quad (3.4)$$

is satisfied by (3.2) and the law of iterated expectations. For the case of strict exogeneity of the explanatory variables a straightforward matrix of instruments used in the estimation part of this paper can be constructed as  $A(x_i) = I_T \otimes a_{iT}$  where  $a_{iT} = (x_i^c; x_{i1}^v, \dots, x_{iT}^v)$ . Here  $x_i^c$  denotes the  $1 \times K^c$  subvector of the explanatory variables with time-constant regressors while  $x_{is}^v$  denotes the  $1 \times K^v$  subvector of time-varying variables at time  $s$ , such that  $K^c + K^v = K$ . This particular choice of instruments implies  $L - Q = (T - 1)(K^c - 1) + (T^2 - 1)K^v$  overidenti-

<sup>3</sup> Without loss of generality the variance for  $t=1$  is normalized to unity to ensure identification. Any other main-diagonal element of the variance-covariance could be normalized to one instead. To obtain simple analytical gradients it is more convenient to optimize with respect to the inverse standard deviations than with respect to the variances. The standard errors of the estimated variances can be retrieved afterwards by means of the delta method.

fyng restrictions. The GMM estimator of  $\theta_0$  using moment functions (3.3) minimizes a quadratic criterion function

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} \left[ \frac{1}{N} \sum_{i=1}^N g(z_i, \theta) \right]' P_N \left[ \frac{1}{N} \sum_{i=1}^N g(z_i, \theta) \right], \quad (3.5)$$

where  $P_N$  is a symmetric and positive definite weight matrix that converges almost surely to a nonrandom matrix  $P$  and the parameter space  $\Theta$  is a compact subspace of  $\mathbb{R}^q$ . The asymptotic distribution of  $\hat{\theta}$  under the assumption of certain regularity conditions (cf. Gouriéroux and Monfort, 1995, ch. 9.5) is given by

$$\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, (G'PG)^{-1} G'PVPG(G'PG)^{-1}\right), \quad (3.6)$$

$$\text{where } V = E\left[g(z_i, \theta_0)g(z_i, \theta_0)'\right],$$

$$G = E\left[\frac{\partial g(z_i, \theta_0)}{\partial \theta'}\right].$$

Hansen (1982) shows that the optimal choice of the weight matrix  $P_N$  converges to  $P = V^{-1}$  which leads to the asymptotically efficient GMM estimator *for a given set of unconditional moment restrictions*. A feasible estimator can be obtained from using a consistent estimate  $\hat{P}$  of the weight matrix calculated from a first estimation step in which (3.5) is solved for a weight matrix that is independent of  $\theta$ , e.g. the identity matrix. In this case the asymptotic distribution of the GMM estimator collapses to

$$\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, (G'V^{-1}G)^{-1}\right). \quad (3.7)$$

Throughout this paper, the estimates (3.5) are called one-step GMM (GMM1) estimates when using the identity matrix as a weight matrix and two-step GMM (GMM2) estimates if they are based on the optimal weight matrix computed from GMM1 in the first step.

It is obvious that the choice of the matrix of instruments  $A(x_i)$  is quite arbitrary since different choices lead to different GMM estimators that are not comparable with respect to their relative efficiency. However, optimal instruments are available that lead to the asymptotically efficient GMM estimator *for a given set of conditional moment restrictions*. In addition, Chamberlain (1987) shows that this estimator reaches the lower bound of the asymptotic



variance of any consistent and asymptotically normal estimator in the class of estimators using (3.2) as the only substantive distributional assumption. Under the assumptions of strictly exogenous explanatory variables and conditional heteroskedasticity, the matrix of optimal instruments is given by (cf. Newey, 1993)<sup>4</sup>

$$A^*(z_i, \beta_0, \Sigma_0) = \Omega(z_i, \beta_0, \Sigma_0)^{-1} G^*(z_i, \theta_0) F, \quad (3.8)$$

$$\text{where } \Omega(z_i, \beta_0, \Sigma_0) = E\left[\rho(z_i, \theta_0) \rho(z_i, \theta_0)' | x_i\right],$$

$$G^*(z_i, \theta_0) = E\left[\frac{\partial \rho(z_i, \theta_0)}{\partial \theta'} | x_i\right]$$

and  $F$  is any nonsingular matrix representing proportional transformations of  $A^*$ . GMM estimation is based on as many moment functions as parameters to be estimated

$$g^*(z_i, \theta) = A^*(z_i, \beta, \Sigma)' \rho(z_i, \theta) \quad (3.9)$$

leading to the estimator

$$\hat{\theta}^* = \underset{\theta \in \Theta}{\operatorname{argmin}} \left[ \frac{1}{N} \sum_{i=1}^N g^*(z_i, \theta) \right]' P_N^* \left[ \frac{1}{N} \sum_{i=1}^N g^*(z_i, \theta) \right]. \quad (3.10)$$

Under regularity conditions the asymptotic distribution of  $\hat{\theta}^*$  is given by

$$\sqrt{N}(\hat{\theta}^* - \theta_0) \xrightarrow{d} N\left(0, \left(E\left[G^*(x_i, \theta_0)' \Omega(z_i, \beta_0, \Sigma_0)^{-1} G^*(x_i, \theta_0)\right]\right)^{-1}\right) \quad (3.11)$$

and does not depend on the choice of the weight matrix  $P_N^*$ . In the presented form the optimal GMM estimator is not feasible because the optimal instruments are functions of the  $0.5 T(T-1)$  unknown off-diagonal elements of  $\Sigma_0$ . A feasible estimator, called three-step GMM (GMM3) estimator in the following, can be obtained by a three-step procedure in which  $A^*(z_i, \beta_0, \Sigma_0)$  is replaced by a consistent estimate  $\hat{A}^*$  without affecting the asymptotic distribution of the estimator. Newey (1993) suggests using  $\hat{A}^*$  to form the weight matrix

<sup>4</sup> See Newey (1990) for the optimal instruments in case of conditional homoskedasticity. Chamberlain (1992) and Hahn (1997) discuss the implementation of optimal instruments with weakly exogenous instruments implying a set of sequential moment restrictions.

$$P_N^* = \left[ \frac{1}{N} \sum_{i=1}^N \hat{A}^* \hat{A}^{*\prime} \right]^{-1} \quad (3.12)$$

in order to make the objective function invariant to nonsingular transformations of the optimal instruments and to improve the numerical properties in this way. The estimate  $\hat{A}^*$  results from substituting the two unknown conditional expectations in (3.8) by their estimates. Using the conditional mean function (3.1) for the panel probit model, it is obvious that  $G^*(z_i, \theta_0)$  does not depend on the endogenous variables and can be estimated consistently by replacing  $\theta_0$  with any consistent first step estimate  $\hat{\theta}_1$ , e.g. the GMM1 or GMM2 estimates introduced above.<sup>5</sup>

The estimation of the second expression in (3.8) involving conditional expectations,  $\Omega(z_i, \beta_0, \Sigma_0)$ , is more difficult because it depends on the endogenous variables. One simple way of obtaining a consistent estimate of  $\Omega(z_i, \beta_0, \Sigma_0)$  goes back to Heckman (1981) who suggests using 0.5 T (T - 1) bivariate probit estimations to estimate the unknown covariance elements of  $\Sigma_0$ . However, this method becomes inconvenient for larger Ts. Two alternative approaches are suggested by Newey (1990, 1993) which make use of nonparametric estimation. The first one is based upon a series approximation while the second one uses the k-nearest neighbor (k-nn) estimation technique. For the latter approach which may be easier to implement, the k-nn estimate of  $\Omega(z_i, \beta_0, \Sigma_0)$  computed in the second estimation stage of the GMM3 estimator is defined by<sup>6</sup>

$$\hat{\Omega}(z_i, \hat{\theta}_1, k) = \sum_{j=1}^N w_{ij} \rho(z_j, \hat{\theta}_1) \rho(z_j, \hat{\theta}_1)' \quad (3.13)$$

where the weights  $w_{ij}$  are positive for the nearest neighbors  $j \leq k$  and equal to zero for  $j > k$ . The most simple case is characterized by uniform weights  $w_{ij} = 1/k$  for  $j \leq k$ . Smoother weight functions, e.g. with quadratic or triangular weights, are discussed by Stone (1977). Following Newey (1990, 1993), the k-nearest neighbors for observation i are those of the sample that minimize a distance between the explanatory variables scaled by their respective standard deviations,  $sd(\cdot)$ , given by

<sup>5</sup> In the empirical section GMM2 estimates are used for this purpose.

<sup>6</sup> Applications of this semiparametric optimal GMM estimator to the panel probit model include Inkmann and Pohlmeier (1995) and Bertschek and Lechner (1998). Pohlmeier (1994) presents the corresponding estimator for count data panel models and Pohlmeier and Inkmann (1997) apply the principle to nonlinear SUR models and give an example to count data.

$$D_{ij} = d_{ij} d_{ij}' , \quad j = 1, \dots, N, \quad (3.14)$$

$$\text{with } d_{ij} = \left[ \left( \frac{x_{i1} - x_{j1}}{\text{sd}(x_{\cdot 1})} \right)', \dots, \left( \frac{x_{iT} - x_{jT}}{\text{sd}(x_{\cdot T})} \right)' \right]$$

A well known problem of the k-nearest neighbor approach is the choice of the smoothing parameter k. <sup>7</sup> Newey (1993) proposes a choice criterion based on the cross-validation function

$$C\hat{V}(k) = \text{tr} \left[ Q \sum_{i=1}^N R(z_i, \hat{\theta}_1) \hat{\Omega}(z_i, \hat{\theta}_1, k) R(z_i, \hat{\theta}_1)' \right], \quad (3.15)$$

$$\text{where } R(z_i, \hat{\theta}_1) = \left\{ \hat{A}^* \left[ \rho(z_i, \hat{\theta}_1) \rho(z_i, \hat{\theta}_1)' - \hat{\Omega}(z_i, \hat{\theta}_1, k) \right] \right\} \hat{\Omega}(z_i, \hat{\theta}_1, k)^{-1}$$

and Q represents any positive definite matrix. A simple way to obtain an optimal smoothing parameter is to compute (3.15) for a given sequence of k, e.g.  $k = 0.1 \cdot N, \dots, 0.9 \cdot N, N$ , and to select the particular k that minimizes the cross-validation function. This strategy is employed in the empirical part of the paper using the identity matrix for Q.

#### 4. Simulated maximum likelihood estimation

Simulation methods are required for maximum likelihood estimation of panel probit models to allow for an unrestricted correlation structure of the error terms over time. Extensive surveys of simulation methods for models of limited dependent variables are given by Hajivassiliou (1993) and Hajivassiliou and Ruud (1994). The articles by Gouriéroux and Monfort (1993) and Keane (1993) focus explicitly on applications of simulation methods to panel data models. This section reviews a simulated maximum likelihood estimator of the panel probit model that builds on the Geweke-Hajivassiliou-Keane (GHK) simulator.<sup>8</sup> Previous applications of this particular estimator include Keane (1994) and Mühleisen (1994). The choice of the GHK simulator is justified by the comprehensive comparison of 13 simulators for multivariate normal rectangle probabilities by Hajivassiliou, McFadden, and Ruud (1996). The

<sup>7</sup> The asymptotic mean squared error of uniform k-nearest neighbor estimates in the T-variate case is proportional to  $N^{4/(4+T)}$  (cf. Silverman, 1986, p.86). However, this knowledge does not help very much to determine k in practical work.

<sup>8</sup> See Börsch-Supan and Hajivassiliou (1993) who use the term 'Smooth, Recursive Conditioning Simulator'.

authors recommend the GHK simulator as the most reliable method. In 70 of their 84 experiments, GHK achieved the best root mean squared error performance of all methods.

To circumvent the indicator function in (2.2), define  $d_{it} = 2y_{it} - 1$  and write  $\mu_{it} = x'_{it}\beta$ . Then it always holds from (2.2) that  $-d_{it}u_{it} < d_{it}\mu_{it}$  or equivalently

$$-D_i u_i < D_i \mu_i, \quad (4.1)$$

where  $D_i$  is a matrix with diagonal elements  $(d_{i1}, \dots, d_{iT})$  and zeros elsewhere. Dropping the indices for convenience, the left hand side of (4.1) is distributed as multivariate  $N(0, D\Sigma D')$ . Define  $B = D\mu$  and let  $L$  be the lower triangular Cholesky factor of  $D\Sigma D'$  such that  $LL' = D\Sigma D'$ . Given a  $T \times 1$  vector  $\eta$  of standard normal random numbers it follows immediately that  $L\eta$  is distributed as well as multivariate  $N(0, D\Sigma D')$ . Instead of computing each individual likelihood contribution on the probability of (4.1), the GHK algorithm is based upon the probability

$$\Pr(L\eta < B) = \Pr \left( \begin{pmatrix} l_{11} & & & & \\ l_{21} & l_{22} & & & \\ l_{31} & l_{32} & l_{33} & & \\ & & \dots & & \\ l_{T1} & l_{T2} & l_{T3} & \dots & l_{TT} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \vdots \\ \eta_T \end{pmatrix} < \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_T \end{pmatrix} \right). \quad (4.2)$$

The triangular structure of this expression is exploited by drawing a sequence of univariate standard normal distributed random numbers subject to a truncation rule which is recursively determined according to

$$\begin{aligned} \eta_1 &< b_1 / l_{11} \\ &\vdots \\ \eta_t &< (b_t - l_{t1}\eta_1 - l_{t2}\eta_2 - \dots - l_{t,t-1}\eta_{t-1}) / l_{tt} \\ &\vdots \\ \eta_T &< (b_T - l_{T1}\eta_1 - l_{T2}\eta_2 - \dots - l_{T,T-1}\eta_{T-1}) / l_{TT}. \end{aligned} \quad (4.3)$$

These draws from the truncated normal density are obtained by an application of the probability integral transform theorem (cf. Hajivassiliou, 1993)

$$\eta_t = \Phi^{-1}[\Phi(ub_t) \cdot v_t], \quad (4.4)$$

where  $v_t$  is a uniform random variable from the interval  $[0,1]$ ,  $ub_t$  denotes the upper truncation bound on the right-hand side of (4.3), and  $\Phi^{-1}(\cdot)$ , the inverse of the standard normal

c.d.f. This sequential way of proceeding generally leads to biased draws of the corresponding multivariate truncated normal distribution in cases where  $L$  is not a diagonal matrix because the interdependence is ignored (cf. Börsch-Supan and Hajivassiliou, 1993, for a simple two-dimensional example). Nevertheless, the joint probability (4.2) is correctly simulated by the product of the single probabilities that each  $\eta_i$  falls in the respective truncated interval. This yields the individual simulated likelihood contribution

$$\begin{aligned} \ell(z, \beta, \Sigma) &= \Pr(\eta_1 < b_1 / 1_{11}) \cdot \dots \cdot \Pr(\eta_T < (b_T - 1_{11}\eta_1 - 1_{21}\eta_2 - \dots - 1_{T-1,1}\eta_{T-1}) / 1_{TT}) \\ &= \Phi(b_1 / 1_{11}) \cdot \dots \cdot \Phi((b_T - 1_{11}\eta_1 - 1_{21}\eta_2 - \dots - 1_{T-1,1}\eta_{T-1}) / 1_{TT}) \\ &= Q_1 \cdot Q_2(\eta_1) \cdot Q_3(\eta_1, \eta_2) \cdot \dots \cdot Q_T(\eta_1, \dots, \eta_{T-1}), \end{aligned} \quad (4.5)$$

where the factors  $Q$  are introduced to reflect the dependence on the random draws  $\eta_i$ .<sup>9</sup> The sampling variance of the probability simulator is reduced by increasing the number of replications  $R$  and using the average

$$\bar{\ell}(z, \beta, \Sigma, R) = \frac{1}{R} \sum_{r=1}^R \prod_{i=1}^T Q_i(\eta_i^r, \dots, \eta_{i-1}^r). \quad (4.6)$$

The simulated likelihood is a continuous and differentiable function of the parameters to be estimated. In addition, Börsch-Supan and Hajivassiliou (1993) prove that  $\bar{\ell}(z, \beta, \Sigma, R)$  is an unbiased simulator of the likelihood. However, as Hajivassiliou and Ruud (1994) point out, an unbiased simulation is neither necessary nor sufficient for consistent SML estimation. This is due to the nonlinearity introduced by the logarithmic transformation necessary in order to derive the simulated log-likelihood function. To solve this problem, the number of draws  $R$  has to increase sufficiently with the sample size. Under certain regularity conditions, a sufficient rate is  $R / \sqrt{N} \rightarrow \infty$  as  $N \rightarrow \infty$  to obtain consistent, asymptotically normal and efficient estimates (cf. Hajivassiliou and Ruud, 1994, Propositions 4 and 5).

It is common practice to impose a certain structure on the error term variance-covariance matrix in order to restrict the number of covariance parameters to be estimated. In comparison to GMM, this is obviously a drawback to the SML approach. However, compared to ordinary ML estimation of the panel probit model, much less restrictive specifications are possible.

---

<sup>9</sup> Note, that the last draw  $\eta_T$  is not necessary for the computation of the likelihood contribution.

## 5. Monte Carlo evidence

In order to evaluate the small sample properties of the proposed estimators some Monte Carlo simulation experiments were carried out. Recent Monte Carlo evidence by Bertschek and Lechner (1998) reveals a rather small efficiency loss by three-step GMM estimation (GMM3) compared to full information maximum likelihood estimation in the case of a correctly specified random effects panel probit model with equicorrelated error terms and homoskedasticity over time. The implementation of the SML estimator allows for a comparison of the GMM type estimators with suitable ML estimators for models with a more demanding specification of the error term process.

The underlying data generating process (DGP) is a slight variation of the one suggested by Keane (1994). The true model is

$$y_{it}^* = \alpha_0 + \beta_0 x_{it} + u_{it}, \quad i = 1, \dots, 500, \quad t = 1, \dots, 5, \quad (5.1)$$

where the parameters are set to  $\alpha_0 = -1$  and  $\beta_0 = 0.25$ . The  $x_{it}$  are sampled from the normal distribution with mean 5, an across individual variance of 3 and a within individual variation of 2.<sup>10</sup> It is assumed that the error term follows a combined random effects and autoregressive specification

$$u_{it} = \tau_i + v_{it} \quad \text{with} \quad v_{it} = \rho v_{i,t-1} + \varepsilon_{it}, \quad (5.2)$$

where  $\tau_i \sim N(0, \sigma_\tau^2)$  and  $\varepsilon_{it} \sim N(0, \sigma_{\varepsilon t}^2)$  independently of  $\tau_i$ . It is imposed that  $\sigma_{\varepsilon 1}^2 = 1 - \sigma_\tau^2$  to ensure that the identification conditions introduced below for the different specifications of the proposed estimators hold for the data generating process. Two processes are distinguished by their respective specification of  $\sigma_{\varepsilon t}^2$  for  $t = 2, \dots, 5$ . DGP1 creates heteroskedasticity over time while DGP2 leads in addition to multiplicative heteroskedasticity over individuals:

DGP 1 heteroskedasticity over t	DGP 2 heteroskedasticity over t and i
$\begin{pmatrix} \sigma_{\varepsilon 2}^2 \\ \sigma_{\varepsilon 3}^2 \\ \sigma_{\varepsilon 4}^2 \\ \sigma_{\varepsilon 5}^2 \end{pmatrix} = \begin{pmatrix} 0.96875 \\ 1.43750 \\ 0.90625 \\ 0.43750 \end{pmatrix}$	$\begin{pmatrix} \sigma_{\varepsilon 2}^2 \\ \sigma_{\varepsilon 3}^2 \\ \sigma_{\varepsilon 4}^2 \\ \sigma_{\varepsilon 5}^2 \end{pmatrix} = \begin{pmatrix} 0.96875 \cdot \bar{x}_{i2}^2 \\ 1.43750 \cdot \bar{x}_{i3}^2 \\ 0.90625 \cdot \bar{x}_{i4}^2 \\ 0.43750 \cdot \bar{x}_{i5}^2 \end{pmatrix}$

<sup>10</sup> That is:  $x_{it} = x_i + e_{it}$  where  $x_i \sim \text{iid } N(5, 3)$  and  $e_{it} \sim \text{iid } N(0, 2)$ , independently of  $x_i$ .

where  $\tilde{x}_{it}$  denotes the individual realization of the  $N(0,1)$  distributed standardized counterpart of the regressor  $x_{it}$ . This guarantees that the expected error term variances of DGP2 are equal to the corresponding variances of DGP1. In combination with  $\sigma_\tau^2 = 0.5$  and  $\rho = 0.25$ , both experiments imply an expected variance-covariance matrix of the error term  $u_i$  with main-diagonal elements  $(1 \ 1.5 \ 2 \ 1.5 \ 1)'$ .

For both data generating processes two specifications of the estimators under consideration are computed which differ in their identification conditions and the resulting set of parameters to be estimated:

Specification 1 (SPEC1) heteroskedasticity over t	Specification 2 (SPEC2) homoskedasticity over t
GMM	GMM
$E[u_{it}^2] = 1$ for identification	$E[u_{it}^2] = 1 \quad \forall t$
$E[u_{is}^2] = \sigma_{us}^2$ for $s = 2, \dots, T$	
parameter set: $\theta_{SPEC1}^{GMM} = (\alpha, \beta; \sigma_{u2}^2, \dots, \sigma_{u5}^2)$	parameter set: $\theta_{SPEC2}^{GMM} = (\alpha, \beta)$
SML	SML
$\sigma_{\epsilon 1}^2 = 1 - \sigma_\tau^2$ for identification	$\sigma_{\epsilon t}^2 = (1 - \sigma_\tau^2)(1 - \rho^2) \quad \forall t$
$E[u_{is}u_{it}] = \sigma_\tau^2 + \sum_{r=0}^{\min(s,t)-1} \sigma_{\epsilon, \min(s,t)-r}^2 \rho^{ s-t +2r} \quad \forall s, t$	$E[u_{is}u_{it}] = \sigma_\tau^2 + \frac{\sigma_{\epsilon t}^2}{1 - \rho^2} \rho^{ s-t } \quad \forall s, t$
parameter set: $\theta_{SPEC1}^{SML} = (\alpha, \beta, \rho, \sigma_\tau^2; \sigma_{\epsilon 2}^2, \dots, \sigma_{\epsilon 5}^2)$	parameter set: $\theta_{SPEC2}^{SML} = (\alpha, \beta, \rho, \sigma_\tau^2)$

Based on the SML estimates of  $(\rho, \sigma_\tau^2; \sigma_{\epsilon 2}^2, \dots, \sigma_{\epsilon 5}^2)$  in SPEC1 the corresponding estimates and estimated standard errors of  $(\sigma_{u2}^2, \dots, \sigma_{u5}^2)$  are derived by means of the delta method to ensure comparability with the GMM results. Note, that GMM only requires an assumption about the main-diagonal elements of the error term variance-covariance matrix  $\Sigma$  while the GHK estimator demands for a complete parametric representation of  $\Sigma$ . The particular specification of  $\Sigma$  for the SML estimator in case of SPEC2 has become the first choice in panel data applications of the GHK simulator (e.g., Keane, 1994, Mühleisen, 1994). The extension in SPEC1 to allow for heteroskedasticity over time has not yet been treated in the SML literature, at least not to the author's knowledge.

Four experiments emerge from all possible combinations of the two data generating processes and the two specifications. SPEC1 is correctly specified under DGP1 while SPEC2 suffers from the identification problem mentioned in the introduction in the sense that the slope

parameters are not identified subject to the homoskedasticity restriction because they cannot be separated from the impact of the scale coefficients. Both specifications imply inconsistency of the GMM and SML estimates under DGP2 because the assumption of mutual independence of the error terms and explanatory variables stated in Section 2 is violated. However, it should be interesting to observe the relative performance of GMM and SML in the presence of this type of misspecification. One should expect that GMM behaves more robust with respect to misspecification of the error term variance-covariance matrix of the latent normal model because GMM exploits only limited distributional information compared to full information ML.

The particular choice of the autocorrelation parameter  $\rho = 0.25$  allows to choose a small number of replications of the GHK algorithm for SML estimation. This is justified by previous Monte Carlo evidence by Keane (1994) who finds that the GHK simulator already works precisely with a small number of replications  $R$  if the coefficient of autocorrelation is small. Therefore the first SML estimator computed in all experiments rests on 5 replications of the GHK algorithm and is denoted SML5. This choice is in the magnitude of the number of replications chosen in a previous study by Börsch-Supan et al. (1993) who do not find any significant improvement from increasing  $R$  from 3 to 9 for the simulation of a multinomial panel probit model. However, the number of replications required for accurate simulation of the log-likelihood function depends on the shape of the particular likelihood function and the data under consideration which renders a comparison between different studies difficult. Therefore a second SML estimator is computed in all experiments that uses 25 replications of the GHK algorithm and is denoted SML25. The first 5 replications rest on the same random numbers used for SML5. The pseudo-random numbers for the GHK algorithm are created by the multiplicative congruential method implemented in the GAUSS procedure RNDU.<sup>11</sup> The inverse of the standard normal c.d.f. in (4.4) is computed by the polynomial approximation suggested by Odeh and Evans (1974) with a maximum approximation error of less than  $1.5 \cdot 10^{-8}$ .<sup>12</sup>

The GMM estimators under consideration include GMM1, GMM2, and GMM3. While all estimates are consistent in the correctly specified experiment (SPEC1 on DGP1), the as-

---

<sup>11</sup> Alternatively, Tausworthe's (1965) pseudo-random number generator was implemented without changing the results in any systematic matter. Therefore these results are omitted here.

<sup>12</sup> A Gauss procedure for this approximation written by the author is available online from the Gauss archives at the American University in Washington. In this application the GHK method turns out to be unaffected by larger approximation error sizes as shown by some experiments with the less precise approximations of the inverse standard normal distribution function suggested by Hastings (1955) and Bailey (1981).



ymptotic efficiency for this experiment increases in the order GMM1, GMM2, GMM3, SML as described in the previous two sections. The summary statistics of the estimation results for 1000 Monte Carlo replications of the four experiments are presented in Tables A1 to A4 in the appendix.<sup>13</sup> Apart from commonly used statistics such as BIAS, SE (standard error) and RMSE (root mean squared error), the statistic RELSE (relative standard error) is computed to analyze the precision of the estimated asymptotic standard errors of the intercept and slope estimates. RELSE is defined as the ratio of the average estimated standard error using the appropriate asymptotic variance-covariance matrix of the respective estimator and the empirical standard deviation of the estimator over the number of completed Monte Carlo replications. The latter should converge towards the true standard errors of the respective estimates as the number of replications increases. The estimated asymptotic standard errors are biased upwards when RELSE is greater than one and they are underestimated on average for RELSE less than one.

A more convenient way to look at the Monte Carlo estimation results is offered by the box plot technique. Figures 1 and 2 display box plots for the empirical distribution of the bias of the estimated slope-intercept ratios obtained over 1000 Monte Carlo replications. Figure 1 contains the estimation results for the first specification of the respective estimators that accounts for heteroskedasticity over time (SPEC1) while the results depicted in Figure 2 rest on the homoskedastic specification (SPEC2). To take care of the different variance normalization of the two specifications, the bias of the ratio of the two estimated coefficients is given instead of the bias of the two separate estimates. The ratios are comparable for all experiments while the separate estimates are only comparable for estimators based on SPEC1. The horizontal line within each box indicates the median of the empirical distribution. The boxes are bounded at the 25th and 75th percentiles of the empirical distributions. The vertical lines (fences) originating from the boxes are bounded at the last estimate that falls into the 1.5-fold interquartile range. Dots outside the fences denote outliers in the tails of the distribution. Note, that the one-step GMM estimation results are omitted in both figures because of their weak performance. However, they are included in the tables given in the appendix.

---

<sup>13</sup> All calculations were performed on two IBM RS 6000 workstations using GAUSS version 3.2.37.

Figure 1.  
Empirical Distribution of the Bias of the Estimated Slope-Intercept Ratio  
SPEC1: Heteroskedasticity over Time is Taken into Account

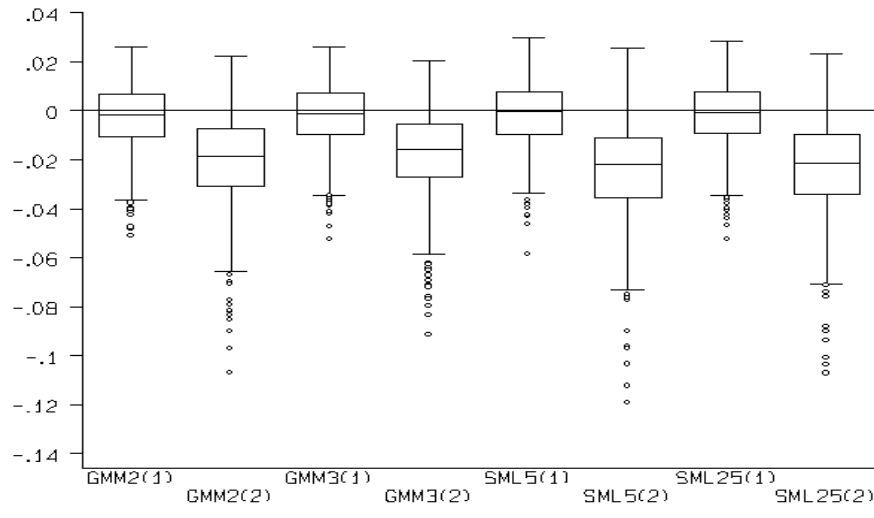
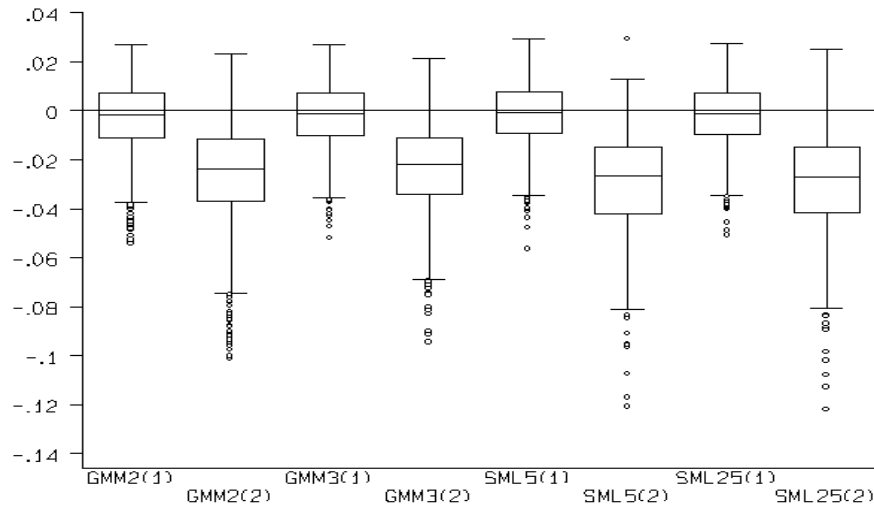


Figure 2.  
Empirical Distribution of the Bias of the Estimated Slope-Intercept Ratio  
SPEC2: Heteroskedasticity over Time is not Taken into Account



Note: The numbers in parentheses refer to the data generating process: DGP1 leads to heteroskedasticity over time and DGP2 in addition to heteroskedasticity over individuals. GMM2 denotes the two-step GMM estimator with suboptimal instruments but optimal weighting of the moment functions. GMM3 is the three-step GMM estimator with optimal instruments. SML5 and SML25 are simulated maximum likelihood estimators with 5 and 25 replications of the GHK algorithm.

As expected, all estimators under consideration come close to the true parameters of the correctly specified experiment, SPEC1 on DGP1. Table A1 indicates that the three-step GMM estimator exhibits the smallest bias of the three GMM estimators with respect to the estimates of the intercept and slope coefficient. The GMM3 estimator also outperforms SML5 with respect to this criterion and meets the performance of the SML estimator with 25 replications of the GHK algorithm. It is not surprising that increasing the number of GHK replications reduces the average bias of the SML estimator, at least for the intercept and slope parameter. The described performance of the respective estimators hold as well for the bias of the ratio of the slope parameter and the intercept for which the complete empirical distribution is depicted in Figure 1. The skewness indicated by the presence of outliers in the left tail of each estimator's distribution should not be overemphasized because the reflection property described by Chesher and Peters (1994) holds for the covariate design of the experiments. In short, this property states that a skewed distribution obtained from a Monte Carlo experiment for the estimated slope coefficient of a regressor that is generated from a symmetric distribution can be reversed by reversing the sign of the true coefficient. The skewness can be eliminated by setting the true coefficient to zero. This property holds for a large class of M-estimators. However, the relative shape of the empirical distribution of the different estimates is important for the experiments presented in this section. In the correctly specified experiment the box plots for the estimated slope-intercept ratios are nearly indistinguishable, having slight advantages of GMM3 over GMM2 and of SML25 over SML5 with respect to the width of the interquartile range.

In general, the magnitude of the empirical standard errors reveals the asymptotic efficiency advantages of the respective estimators. In particular, GMM1 turns out to be less efficient than GMM2 and GMM3 while both SML estimators are more efficient than the two- and three-step GMM estimators. The efficiency gains from using GMM2 or GMM3 instead of GMM1 are remarkably high while using full information ML yields only modest efficiency gains over GMM2 and GMM3. The only surprising result concerns the relative precision of GMM2 and GMM3. Recall that GMM2 is asymptotically efficient for given unconditional moment functions and GMM3 for given conditional moment functions. Therefore GMM3 should be more efficient than GMM2. The small sample results represented by the standard errors for the intercept and slope coefficient indicate the opposite: GMM2 slightly outper-

forms GMM3 with respect to both efficiency and root mean squared error performance. However, this advantage is not visible before the fourth digit after the decimal place.<sup>14</sup>

More importantly, the results concerning the relative precision of the estimated asymptotic standard errors compared to the empirical standard errors over the Monte Carlo replications (RELSE) indicate that GMM2 suffers from a serious downward bias of 9 % while GMM3 slightly overestimates the asymptotic standard errors. Therefore, inference based on GMM2 may lead to serious misinterpretations in applied work. This result is not new, of course, it confirms previous findings by e.g. Arellano and Bond (1991) for the dynamic linear panel data model with overidentifying restrictions. The poor performance of GMM2 with respect to the precision of the estimated standard errors is usually explained by the first estimation step which is necessary to obtain an estimate of the variance-covariance matrix of the unconditional moment functions serving as a weight matrix for GMM2.<sup>15</sup> Having this explanation in mind, it is surprising that GMM3, which rests on an estimate of the variance-covariance matrix of the conditional moment functions, does not display a similar behavior and produces accurate estimates of the asymptotic standard errors in small samples.

Turning to the variances of the error term distribution over time,  $\sigma_{u_2}^2, \dots, \sigma_{u_5}^2$ , the performance of the different estimators described for the intercept and slope coefficient is, in general, replicated. However, some differences are present. Firstly, the small sample properties of GMM1 are extremely poor. The one-step GMM estimator with identity weighting matrix leads to substantial bias and large standard errors compared to the other GMM estimators. Both errors are much more severe for the scale parameters than for the slope coefficients. Secondly, corresponding to the large sample theory, GMM3 now produces smaller empirical standard errors than GMM2 in three of the four cases. Thirdly, contrary to the aforementioned results for  $\alpha$  and  $\beta$ , increasing the number of replications of the GHK algorithm does not help to reduce the bias of the scale parameters in the small sample. Surprisingly, SML5 turns out to be less biased than SML25 for all estimated variances.

If the specification of the estimators erroneously imposes homoskedastic error terms in the presence of heteroskedasticity over time (SPEC2 on DGP1), the estimators still successfully reveal the true ratio of the slope and intercept coefficients as Figure 2 shows. Compared

---

<sup>14</sup> Recent evidence by Bertschek and Lechner (1998) suggests that the small sample properties of GMM3 can be improved by scaling the conditional moment functions by their standard deviation.

<sup>15</sup> These findings have led to a number of alternative information theoretic estimation approaches that reach the same asymptotic distribution as the GMM2 estimator but do not require the first estimation step, cf. Imbens (1997), Kitamura and Stutzer (1997), and Imbens et al. (1998).

to the correctly specified estimators in Figure 1, the distributions of the bias remain almost unaffected for all estimators depicted in Figure 2. However, the estimators underestimate the absolute values of the separate coefficients by about 15 %, as indicated in Table A3. This bias reflects the deviation of the imposed unity scale from the average true standard deviation of the error terms over time which is equal to 1.17. Surprisingly, GMM1 seems to reveal the true coefficients which, due to the impossibility of identification of the true parameters, does not point in favor of this estimator. The small bias of GMM1 is artificially created by the empirical distribution of the GMM1 estimates that is characterized by extremely large standard errors. The standard errors are about 20 times larger than the ones obtained for the competing estimators. Apart from this striking exception, all methods yield lower empirical standard errors for the estimates of  $\alpha$  and  $\beta$  which can be explained by the reduced number of parameters to be estimated,

The asymptotic standard errors of the GMM1 estimates are underestimated by more than 50 % as indicated by the RELSE statistics in Table A1. GMM2 standard errors are biased downwards by almost 20 %. Keeping in mind that in comparison with GMM2, the one-step GMM estimator is usually praised for the accurateness of its estimated asymptotic standard errors, this result casts serious doubts on the robustness of this advantage of GMM1. At least for this particular Monte Carlo design of the panel probit model, the precision of the estimated GMM1 standard errors crucially depends on the correct specification of the error term variation over time. The GMM3 and SML estimators produce very reliable estimates of the asymptotic standard errors regardless of the underlying misspecified identification condition.

Table A2 presents summary statistics for the experiment involving the heteroskedastic specification of the estimators (SPEC1) in combination with the data generating process which creates multiplicative heteroskedasticity over time and individuals (DGP2). The latter property renders all parameter estimates inconsistent which raises the question of the relative robustness of the different estimators regarding this kind of misspecification. With respect to the bias of the intercept and slope coefficient, GMM1 performs the worst and underestimates the true parameters by about 25 %. The two SML estimators are slightly superior to GMM1 showing an average bias of about 20 %. Increasing the number of replications of the GHK algorithm leads, as expected, to an improvement of the estimation bias. However, the SML estimators are clearly outperformed by the two- and three-step GMM estimators. GMM2 and GMM3 underestimate  $\alpha$  by about 10 % and  $\beta$  by only 5 % with advantages to GMM3 in

both cases. Regarding the ratio of the two parameters, a large part of the biases of the two separate estimates cancel each other out, but the general performance order remains unchanged. It becomes obvious from Figure 1 that the median of the GMM3 bias distribution, in the case of DGP2, is the closest to the ideal zero bias line. A loss of estimation precision is reflected by an increase in the standard errors of the estimates depicted in Figure 1. This holds in particular for the SML estimators. The root mean squared error performance of the different estimators is dominated by GMM3. GMM2 comes close to GMM3 while the other estimators are clearly inferior.

With respect to the accuracy of the estimated asymptotic standard errors (RELSE), GMM2 falls again short of GMM1 and GMM3 as previously shown in Table A1. Compared to the first experiment, the downward bias of the estimated standard errors is almost doubled for GMM2. GMM1 and GMM3 still produce very reliable results. The contrary can be said of SML. On average the two SML estimators underestimate the asymptotic standard errors by about 20 % with clear advantages of SML25 over SML5.

Concerning the estimation of the variances of the error terms over time, the SML estimates show a superior performance over GMM as indicated in Table A2. This replicates the aforementioned results for the correctly specified experiment in Table A1. With the exception of GMM1, all estimators produce larger biases and standard errors in the presence of multiplicative heteroskedasticity. Somewhat puzzling is that SML5 once again exhibits smaller biases for three of the four estimated error term variances than SML25.

The two figures reveal that the differences between the two distributions of the respective estimates under the first (Figure 1) and second (Figure 2) specification are emphasized in case of DGP2. The medians of the empirical bias distributions of the different estimators are closer to the ideal zero bias if the specification allows for heteroskedasticity over time. This specification increases the robustness with respect to multiplicative heteroskedasticity over individuals. The root mean squared errors given in Table A4 for the ratio of the slope coefficient and the intercept under DGP2 always exceed their counterparts under DGP1 given in Table A2. Regarding the two separate parameters, imposing the homoskedasticity assumption under DGP2 implies similar consequences as were previously seen under DGP1. In particular, the estimated standard errors of the GMM1 estimator are again underestimated, in this case by 70 %. The downward bias of the GMM2 standard errors is again in the magnitude of 20 %, while GMM3 produces more reliable standard errors. The latter also holds for the SML standard

errors in Table A4 which are almost estimated without bias. Recalling the large downward bias of the SML standard errors under the heteroskedastic specification (Table A2), this result is surprising.

Tables A5 – A7 in the appendix contain additional results from the Monte Carlo experiments. The empirical distributions of the cross-validated number of nearest neighbors  $k$  used for the nonparametric estimation of the variance-covariance matrix of the conditional moment functions in the second estimation step of GMM3 are given in Table A5. It turns out that there is a tendency in all experiments considered here to chose a large neighborhood for the nonparametric estimation. This corresponds to the experiments of Bertschek and Lechner (1998) but is in opposition to the work of Delgado (1992) and Pohlmeier and Inkmann (1997) on efficient estimation of nonlinear SUR models for count data, who obtain relative small numbers. However, the cross-validation results are driven by the data and do not allow a simple comparison with other studies.

Table A6 displays percentages of prediction success of the different estimators for all experiments. GMM3 outperforms all competing estimators in each experiment, but the differences are small. Finally, relative computation costs are given in Table A7 in percentages of the SML5 computation time. A relative measure of computational costs is given instead of absolute computation times in order to account for different levels of CPU exhaustion during the Monte Carlo replications. The computation times for GMM1 and GMM2 turn out to be negligible in comparison to SML5 and vary between 1.37 % and 3.37 %. GMM3, having ten fully iterated evaluations of the cross-validation criterion, uses about 67 % of the SML5 time in the case of SPEC1 and about 163 % in the case of SPEC2, when heteroskedasticity over time is not specified. Compared to SPEC1 all GMM estimators show a much slower convergence under SPEC2 in relation to SML. Because the implementation of the GHK algorithm rests on matrix operation, the increase of computation time associated with an increase of  $R$  from 5 to 25 is less than proportional. SML25 requires about 220 % of the SML5 time.

The main results from the Monte Carlo experiments can be summarized as follows. In general, the correctly specified experiment reveals the asymptotic efficiency order of the estimators under consideration. The efficiency gains from using GMM2 or GMM3 instead of GMM1 are remarkably high, while using full information ML yields only modest improvements over the two- and three-step GMM estimators. Contrary to GMM2, GMM3 produces reliable estimates of the asymptotic standard errors. Erroneously imposing homoskedasticity

in the presence of heteroskedasticity over time has serious consequences for the GMM1 estimator. In this case, the standard errors of the empirical distribution of the GMM1 estimates are 20 times as large as the ones obtained for the other GMM estimates. With the exception of GMM1, all other estimators exhibit the expected magnitude of bias if the identification condition of unity error term variances does not hold for the true model. Allowing for heteroskedasticity over time increases the robustness of all estimators against misspecification in terms of multiplicative heteroskedasticity over individuals. The two-step and in particular the three-step GMM estimates are much less affected by this kind of misspecification as the SML estimates and show a superior root mean squared error performance. Increasing the number of replications of the GHK algorithm for SML estimation yields only small improvements, in some cases even worsens the SML performance.

## **6. Application to product innovations**

This section provides an application of the proposed estimators to real data. The primary purpose of this section is proving the practicability of the extended heteroskedastic specification of the different estimators for the panel probit in nonexperimental work. The example refers to the huge number of empirical literature on the relationship between market structure and innovation (cf. Kamien and Schwartz, 1982). The data source is based on the German survey '*ifo Konjunkturtest*' conducted by the *ifo* institute in Munich. The sample contains annual observations of 1,325 West German manufacturing firms for the period 1984 – 1988. The binary indicator of interest takes the value 'one' if the firm has successfully established a product innovation within the year before the interview. The explanatory variables and descriptive statistics are given in Table A8 in the appendix.

This particular example can be regarded as an attractive application of the heteroskedastic panel probit estimators because previous work by Laisney et al. (1992a, b) and König et al. (1994) using the same data casts some doubt on the usual homoskedasticity over time assumption. This hypothesis is clearly rejected for the panel probit model with uncorrelated random effects (Laisney et al., 1992a). The normality assumption of the panel probit model cannot be rejected, putting trust in the model specification (1992b). However, these studies cover process innovations instead of product innovations and use a slightly different set of regressors emerging from a structural model of firm behavior. Therefore the results do not



have to hold for the current application. Product innovation equations are estimated by Bertschek (1995), Inkmann and Pohlmeier (1995), and Bertschek and Lechner (1998) using the same data. The latter two studies also apply GMM2 and GMM3 estimators but impose the homoskedasticity assumption. SML estimators have not been previously applied to the data.

Table 1 contains the estimation results for the GMM2, GMM3, SML5 and SML25 estimators using the same heteroskedastic specification (SPEC1) introduced in the Monte Carlo experiments. Because of its weak performance in the previous section, GMM1 is omitted here. The smoothing parameter for the k-nn estimation of the variance-covariance matrix of the conditional moment functions in the second estimation step of GMM3 is chosen again by cross-validation of the set of fully iterated estimates obtained for  $k = 0.1 \cdot N, \dots, 0.9 \cdot N, N$ . The optimal number of nearest neighbors for the current application turns out to be  $k = 0.9 \cdot N$ .

All estimated coefficients in Table 1 have the expected signs and are in accordance with previous work. In general, the differences between the four estimators are very small. Regarding the magnitude of the estimated slope coefficients, GMM3 and the two SML estimators are nearly indistinguishable while GMM2 produces small deviations in some cases. The results of the estimated standard errors of the slope coefficients reveal the asymptotic efficiency order of the different estimators. GMM2 leads to slightly less precise estimates than GMM3 which itself is slightly outperformed by the two SML estimators. Looking again at the results of the Monte Carlo section, the striking similarity of GMM3 and SML suggests that the model may be correctly specified in the sense that heteroskedasticity over individuals is not likely to be present in the current application. However, specification tests are necessary, of course, in order to confirm this, but they are beyond the scope of this section.

The SML estimates of the AR(1) parameter  $\rho$  and the variance of the random effects  $\sigma_v^2$  nicely correspond to the true parameters of the data generating process used in the previous section. Contrary to the Monte Carlo results, the SML estimates of the error term variances exhibit larger standard errors than the GMM estimates. The magnitude of the estimated variances suggests a test of homoskedasticity over time. Table 1 contains both Wald tests and likelihood ratio (LR) respective D tests<sup>16</sup> for the hypothesis of unity error term variances in all periods. The hypothesis cannot be rejected by the Wald and D tests, but is clearly rejected by the LR tests for the two SML estimators.

---

<sup>16</sup> The D test is the GMM counterpart of the LR test (cf. Newey and West, 1987). This test is not available for the exactly identified GMM3 estimator which always obtains a zero value for the minimum of the objective function.

Table 1.  
Panel Probit Estimation of Successful Product Innovations (N=1325, T=5)

regressor	GMM2		GMM3		SML5		SML25	
	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.
intercept	-0.44	0.11	-0.67	0.12	-0.73	0.11	-0.71	0.11
<i>industry specific variables</i>								
investment industry	-0.37	0.08	-0.22	0.08	-0.23	0.08	-0.23	0.08
raw material industry	-0.67	0.10	-0.52	0.10	-0.51	0.09	-0.52	0.10
R&D intensity	0.24	0.03	0.16	0.03	0.19	0.03	0.18	0.03
concentration ratio	-0.10	0.20	-0.01	0.19	-0.03	0.19	-0.04	0.19
value added	0.29	0.11	0.25	0.10	0.32	0.10	0.29	0.11
import to sales ratio	0.24	0.09	0.28	0.09	0.29	0.08	0.25	0.08
<i>firm specific variables</i>								
internationally operating	0.32	0.09	0.42	0.09	0.43	0.08	0.45	0.08
rel. plant size in company	0.22	0.07	0.27	0.07	0.31	0.07	0.29	0.07
rel. company size in industry	0.49	0.12	0.28	0.07	0.26	0.05	0.26	0.05
increasing demand expected	0.37	0.04	0.28	0.04	0.31	0.04	0.30	0.04
$\rho$					0.17	0.04	0.23	0.05
$\sigma_{\tau}^2$					0.51	0.04	0.51	0.04
$\sigma_{u2}^2$	1.28	0.19	1.33	0.28	1.20	0.18	1.20	0.20
$\sigma_{u3}^2$	0.93	0.14	0.88	0.18	1.04	0.29	0.99	0.33
$\sigma_{u4}^2$	1.01	0.16	0.92	0.19	1.06	0.45	1.04	0.52
$\sigma_{u5}^2$	1.22	0.19	0.95	0.20	1.21	0.67	1.21	0.77
prediction success in %	65.13		65.10		65.18		64.78	
<i>tests of homoskedasticity</i>								
Wald test of $H_0 : \sigma_{ut}^2 = 1, \forall t$	$\chi^2(4)$	p-val.	$\chi^2(4)$	p-val.	$\chi^2(4)$	p-val.	$\chi^2(4)$	p-val.
D/LR test of $H_0 : \sigma_{ut}^2 = 1, \forall t$	6.31	0.18	4.00	0.41	3.91	0.42	6.50	0.17
	4.92	0.30	n.a.		23.77	0.00	19.56	0.00

Note: GMM2 denotes the two-step GMM estimator with suboptimal instruments but optimal weighting of the moment functions. GMM3 is the three-step GMM estimator with optimal instruments, SML5 and SML25 are simulated maximum likelihood estimators with 5 and 25 replications of the GHK algorithm.

## 7. Conclusions

This paper presents small sample results for GMM and SML estimators of the panel probit model. Compared to more traditional ML approaches to the estimation of panel probit models, the SML and GMM estimators share the important advantage of solving the problem of multiple integration without the need to impose strong restrictions on the variance-covariance matrix of the error terms of the latent normal regression model. Three GMM estimators are presented in this paper which rely on the same set of conditional moment restrictions. The one-step GMM estimator (GMM1) minimizes the GMM objective function using a suboptimal weight matrix for given instruments. The two-step GMM estimator (GMM2) exploits the optimal weight matrix and is asymptotically efficient in the class of GMM estimators using the same instruments for the particular set of conditional moment restrictions. Finally, the three-step GMM estimator (GMM3) relies on the optimal choice of instruments and reaches the asymptotic efficiency bound of any consistent and asymptotically normal estimator within the class of estimators using the same set of conditional moment restrictions (cf. Chamberlain, 1987). Parts of these optimal instruments that involve unknown conditional expectations are estimated by nonparametric k-nearest neighbor techniques as proposed by Newey (1990, 1993). The three GMM estimators are compared with an efficient full information ML estimator obtained by simulating the likelihood function by means of the Geweke-Hajivassiliou-Keane (GHK) algorithm.

While the asymptotic properties of the different estimators are well understood, little is known about their relative performance in small samples, in particular in the case of misspecification. To shed some light on this area, Monte Carlo experiments, which allow for misspecified heteroskedasticity over time and individuals, are carried out. The main results from these experiments can be summarized as follows. In general, the correctly specified experiment reveals the asymptotic efficiency order of the estimators under consideration. The efficiency gains from using GMM2 or GMM3 instead of GMM1 are remarkably high, while using full information ML yields only modest improvements over the two- and three-step GMM estimators. The GMM2 estimator suffers from the well known problem of underestimating the asymptotic standard errors in small samples. GMM3 solves this problem and shows a comparable small sample efficiency. Therefore this estimator unifies the best properties of GMM1 and GMM2 and should be regarded as a strong competitor to the recently developed informational theoretic one-step estimation approaches which share the asymptotic properties of

GMM2 (cf. Imbens, 1997, Kitamura and Stutzer, 1997, and Imbens et al., 1998). A small sample comparison of these methods with the GMM3 estimator should be an interesting route for future research.

Erroneously imposing homoskedasticity in the presence of heteroskedasticity over time has serious consequences for the GMM1 estimator. In this case, the standard errors of the empirical distribution of the GMM1 estimates are 20 times as large as the ones obtained for the other GMM estimates. With the exception of GMM1, all other estimators exhibit the expected magnitude of bias if the identification condition of unity error term variances does not hold for the true model. Allowing for heteroskedasticity over time increases the robustness of all estimators against misspecification in terms of multiplicative heteroskedasticity over individuals. The two-step and in particular the three-step GMM estimates are, by this kind of misspecification, much less affected than the SML estimates and show a superior root mean squared error performance. GMM1 and SML exhibit a considerable amount of bias in the presence of multiplicative heteroskedasticity. Increasing the number of replications of the GHK algorithm for SML estimation yields only small improvements, in some cases even worsens the SML performance.

An application to product innovation equations using a panel of West German manufacturing firms proves the practicability of the extended heteroskedastic specification of the different estimators for the panel probit model in nonexperimental work. The results generally confirm the findings of the correctly specified Monte Carlo experiment. The similarity of the estimation results obtained by SML and GMM3 is striking.

**Appendix: Tables**

Table A1.  
DGP1 - Heteroskedasticity over Time  
SPEC1 - Heteroskedasticity over Time Taken into Account  
Summary Statistics of 1000 Monte Carlo Replications (N=500, T=5)

		GMM1	GMM2	GMM3	SML5	SML25
$\alpha$ (= -1.00)	MEAN	-0.9832	-1.0085	-0.9995	-0.9843	-0.9990
	BIAS	0.0168	-0.0085	0.0005	0.0157	0.0010
	SE	0.2004	0.1330	0.1333	0.1184	0.1189
	RMSE	0.2010	0.1332	0.1333	0.1194	0.1189
	RELSE	0.9809	0.9130	1.0140	0.9700	0.9837
$\beta$ (= 0.25)	MEAN	0.2481	0.2539	0.2510	0.2466	0.2504
	BIAS	-0.0019	0.0039	0.0010	-0.0034	0.0004
	SE	0.0432	0.0286	0.0288	0.0249	0.0248
	RMSE	0.0432	0.0289	0.0288	0.0251	0.0247
	RELSE	0.9912	0.9151	1.0474	0.9820	0.9927
$\beta/\alpha$ (= -0.25)	MEAN	-0.2541	-0.2526	-0.2520	-0.2514	-0.2515
	BIAS	-0.0041	-0.0026	-0.0020	-0.0014	-0.0015
	SE	0.0153	0.0126	0.0124	0.0123	0.0123
	RMSE	0.0159	0.0129	0.0125	0.0124	0.0124
	V[u <sub>2</sub> ] (= 1.50)	MEAN	1.7085	1.5708	1.5728	1.5398
BIAS		0.2085	0.0708	0.0728	0.0398	0.0763
SE		0.9092	0.4890	0.4851	0.3627	0.3586
RMSE		0.9324	0.4938	0.4903	0.3647	0.3664
V[u <sub>3</sub> ] (= 2.00)		MEAN	2.5189	2.1419	2.1474	2.0096
	BIAS	0.5189	0.1419	0.1474	0.0096	0.0774
	SE	3.3025	0.7869	0.7966	0.5659	0.5590
	RMSE	3.3414	0.7992	0.8098	0.5657	0.5641
	V[u <sub>4</sub> ] (= 1.50)	MEAN	1.7554	1.5659	1.5642	1.4792
BIAS		0.2554	0.0659	0.0642	-0.0208	0.0344
SE		1.1130	0.4985	0.4974	0.3821	0.3750
RMSE		1.1414	0.5026	0.5013	0.3825	0.3764
V[u <sub>5</sub> ] (= 1.00)		MEAN	1.1302	1.0503	1.0479	0.9796
	BIAS	0.1302	0.0503	0.0479	-0.0204	0.0209
	SE	0.5747	0.3074	0.3100	0.2272	0.2298
	RMSE	0.5890	0.3113	0.3135	0.2280	0.2307

Note: GMM2 denotes the two-step GMM estimator with suboptimal instruments but optimal weighting of the moment functions. GMM3 is the three-step GMM estimator with optimal instruments. SML5 and SML25 are simulated maximum likelihood estimators with 5 and 25 replications of the GHK algorithm.

Table A2.  
DGP2 - Heteroskedasticity over Time and Individuals  
SPEC1 - Heteroskedasticity over Time Taken into Account  
Summary Statistics of 1000 Monte Carlo Replications (N=500, T=5)

		GMM1	GMM2	GMM3	SML5	SML25
$\alpha$ (= -1.00)	MEAN	-0.7358	-0.8784	-0.9021	-0.7622	-0.7921
	BIAS	0.2642	0.1216	0.0979	0.2378	0.2079
	SE	0.1587	0.1363	0.1370	0.1373	0.1334
	RMSE	0.3082	0.1826	0.1683	0.2746	0.2470
	RELSE	0.9638	0.8292	0.9141	0.7831	0.8464
$\beta$ (= 0.25)	MEAN	0.2015	0.2354	0.2393	0.2074	0.2146
	BIAS	-0.0485	-0.0146	-0.0107	-0.0426	-0.0354
	SE	0.0336	0.0277	0.0279	0.0294	0.0280
	RMSE	0.0590	0.0313	0.0298	0.0518	0.0451
	RELSE	0.9810	0.8691	1.0046	0.7703	0.8369
$\beta/\alpha$ (= -0.25)	MEAN	-0.2775	-0.2700	-0.2671	-0.2744	-0.2732
	BIAS	-0.0275	-0.0200	-0.0171	-0.0244	-0.0232
	SE	0.0244	0.0183	0.0171	0.0194	0.0189
	RMSE	0.0367	0.0271	0.0242	0.0312	0.0299
	V[u <sub>2</sub> ] (= 1.50)	MEAN	1.3528	1.8401	1.7421	1.5598
BIAS		-0.1472	0.3401	0.2421	0.0598	0.1411
SE		0.5843	0.6345	0.6182	0.5040	0.4666
RMSE		0.6023	0.7196	0.6636	0.5073	0.4872
V[u <sub>3</sub> ] (= 2.00)		MEAN	1.7686	2.4732	2.3467	1.8580
	BIAS	-0.2314	0.4732	0.3467	-0.1420	-0.0260
	SE	1.0386	1.0288	0.9942	0.6279	0.5930
	RMSE	1.0636	1.1320	1.0524	0.6434	0.5932
	V[u <sub>4</sub> ] (= 1.50)	MEAN	1.3895	1.8350	1.7137	1.4966
BIAS		-0.1105	0.3350	0.2137	-0.0034	0.1186
SE		0.6965	0.6429	0.6079	0.4871	0.4769
RMSE		0.7048	0.7247	0.6441	0.4869	0.4912
V[u <sub>5</sub> ] (= 1.00)		MEAN	1.0108	1.2483	1.1511	1.0994
	BIAS	0.0108	0.2483	0.1511	0.0994	0.1996
	SE	0.4384	0.3848	0.3524	0.3462	0.3566
	RMSE	0.4383	0.4578	0.3832	0.3600	0.4085

Note: Cf. Table A1.

Table A3.  
DGP1 - Heteroskedasticity over Time  
SPEC2 - Heteroskedasticity over Time not Taken into Account  
Summary Statistics of 1000 Monte Carlo Replications (N=500, T=5)

		GMM1	GMM2	GMM3	SML5	SML25
$\alpha$ (= -1.00)	MEAN	-1.0681	-0.8894	-0.8693	-0.8586	-0.8596
	BIAS	-0.0681	0.1106	0.1307	0.1414	0.1404
	SE	1.8488	0.0959	0.0860	0.0817	0.0814
	RMSE	1.8491	0.1464	0.1564	0.1633	0.1623
	RELSE	0.4662	0.8248	0.9744	0.9519	0.9804
$\beta$ (= 0.25)	MEAN	0.2568	0.2239	0.2183	0.2153	0.2157
	BIAS	0.0068	-0.0261	-0.0317	-0.0347	-0.0343
	SE	0.3672	0.0177	0.0158	0.0147	0.0145
	RMSE	0.3671	0.0315	0.0354	0.0377	0.0373
	RELSE	0.4586	0.8130	1.0167	0.9723	0.9974
$\beta/\alpha$ (= -0.25)	MEAN	-0.2624	-0.2528	-0.2520	-0.2516	-0.2518
	BIAS	-0.0124	-0.0028	-0.0020	-0.0016	-0.0018
	SE	0.1376	0.0134	0.0127	0.0125	0.0125
	RMSE	0.1380	0.0137	0.0128	0.0126	0.0126

Note: Cf. Table A1.

Table A4.  
DGP2 - Heteroskedasticity over Time and Individuals  
SPEC2 - Heteroskedasticity over Time not Taken into Account  
Summary Statistics of 1000 Monte Carlo Replications (N=500, T=5)

		GMM1	GMM2	GMM3	SML5	SML25
$\alpha$ (= -1.00)	MEAN	-0.8338	-0.7241	-0.7488	-0.6397	-0.6424
	BIAS	0.1662	0.2759	0.2512	0.3603	0.3576
	SE	2.2764	0.0949	0.0913	0.0781	0.0781
	RMSE	2.2813	0.2917	0.2673	0.3687	0.3660
	RELSE	0.3096	0.8127	0.8738	0.9735	1.0034
$\beta$ (= 0.25)	MEAN	0.2165	0.1983	0.2038	0.1773	0.1781
	BIAS	-0.0335	-0.0517	-0.0462	-0.0727	-0.0719
	SE	0.4432	0.0175	0.0176	0.0137	0.0138
	RMSE	0.4443	0.0546	0.0495	0.0740	0.0732
	RELSE	0.3055	0.7975	0.8622	0.9696	0.9842
$\beta/\alpha$ (= -0.25)	MEAN	-0.3198	-0.2759	-0.2737	-0.2791	-0.2792
	BIAS	-0.0698	-0.0259	-0.0237	-0.0291	-0.0292
	SE	2.2915	0.0202	0.0179	0.0205	0.0202
	RMSE	2.2914	0.0329	0.0297	0.0356	0.0355

Note: Cf. Table A1.

Table A5.  
Distribution of the Cross-Validated Number of Nearest Neighbors

Experiment		k=50	100	150	200	250	300	350	400	450	500=N
SPEC1	DGP1	0.00	0.01	0.03	0.06	0.15	0.33	0.63	0.91	1.00	1.00
SPEC1	DGP2	0.00	0.02	0.05	0.08	0.10	0.14	0.23	0.47	0.83	1.00
SPEC2	DGP1	0.00	0.01	0.03	0.07	0.14	0.31	0.58	0.87	0.99	1.00
SPEC2	DGP2	0.01	0.04	0.08	0.13	0.19	0.29	0.45	0.70	0.94	1.00

Note: Number of nearest neighbors for k-nn estimation of the variance-covariance matrix of the conditional moment functions in the second estimation step of GMM3. Results from 1000 replications of the Monte Carlo experiments. DGP1 leads to heteroskedasticity over time and DGP2 in addition to heteroskedasticity over individuals. SPEC1 specifies heteroskedasticity over time, SPEC2 assumes homoskedasticity.

Table A6.  
Percentages of Correct Prediction of the Different Estimators

Experiment		GMM1	GMM2	GMM3	SML5	SML25
SPEC1	DGP1	65.58	65.61	65.62	65.61	65.61
SPEC1	DGP2	65.67	65.82	65.85	65.76	65.78
SPEC2	DGP1	65.27	65.61	65.62	65.61	65.61
SPEC2	DGP2	65.30	65.74	65.78	65.71	65.71

Note: Cf. Tables A1 and A5.

Table A7.  
Percentages of Relative Computation Costs of the Different Estimators

Experiment		GMM1	GMM2	GMM3	SML5	SML25
SPEC1	DGP1	1.75	1.37	67.16	100.00	222.24
SPEC1	DGP2	1.78	1.36	66.39	100.00	226.03
SPEC2	DGP1	3.37	2.78	162.55	100.00	220.69
SPEC2	DGP2	3.35	2.98	163.31	100.00	224.17

Note: Cf. Tables A1 and A5. Mean computation times relative to SML5. Values for GMM2 (GMM3) do not include computation times of first step estimates obtained by GMM1 (GMM2). Values for GMM3 include cross-validation of ten estimates obtained from varying the number of nearest neighbors.



Table A8.  
Descriptive Statistics (N=1325, T=5)

variables	Mean	St. dev.	Min.	Max.
<i>dependent variable</i>				
product innovation realized	0.59	0.49	0.00	1.00
<i>industry specific variables</i>				
investment industry (time-invariant)	0.48	0.50	0.00	1.00
raw material industry (time-invariant)	0.11	0.31	0.00	1.00
ln(R&D intensity)	0.14	1.29	-2.25	2.11
concentration ratio	0.21	0.16	0.06	0.83
value added ( $\cdot 10^5$ )	0.42	0.36	0.00	1.04
import to sales ratio	0.67	0.32	0.14	1.93
<i>firm specific variables</i>				
internationally operating	0.42	0.36	0.00	1.00
relative plant size in company	0.64	0.33	0.00	1.00
relative company size in industry ( $\cdot 10^2$ )	0.26	1.00	0.00	31.92
positive demand expectation	0.47	0.50	0.00	1.00

Note: The data source is the 'ifo Konjunkturtest', 1984-1988. A detailed description can be found in Section 3 in Laisney, Lechner and Pohlmeier (1992a). The R&D intensity variable defined as the ratio of R&D expenditures to sales at the 2-digit industry level was derived from the information for the years 1985, 87, and 89 given in Legler et al. (1992, p. 140). From these values the missing observations for the years 1984, 86, and 88 were computed by OLS. The logarithm of the resulting fitted variable was used in the regressions in order to circumvent linear dependencies of moment restrictions.

## References

- Arellano, M., and S. Bond (1991): 'Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations', *Review of Economic Studies*, 58, 277-297.
- Avery, R., L. P. Hansen, and V. Hotz (1983): 'Multiperiod Probit Models and Orthogonality Condition Estimation', *International Economic Review*, 24, 21-35.
- Bailey, B. J. R. (1981): 'Alternatives to Hastings' Approximation to the Inverse of the Normal Cumulative Distribution Function', *Applied Statistics*, 30, 275-276.
- Bertschek, I. (1995): 'Product and Process Innovation as a Response to Increasing Imports and Foreign Direct Investments', *Journal of Industrial Economics*, 43, 341-357.
- Bertschek, I., and M. Lechner (1998): 'Convenient Estimators for the Panel Probit Model', *Journal of Econometrics*, 87, 329-371.
- Börsch-Supan, A., and V. A. Hajivassiliou (1993): 'Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models', *Journal of Econometrics*, 58, 347-368.
- Börsch-Supan, A., V. A. Hajivassiliou, L. J. Kotlikoff, and J. N. Morris (1993): 'Health, Children, and Elderly Living Arrangements: A Multiperiod-Multinomial Probit Model with Unobserved Heterogeneity and Autocorrelated Errors', NBER working paper no 3343.
- Butler, J. S., and R. Moffitt (1982): 'A Computationally Efficient Quadrature Procedure for the One Factor Multinomial Probit Model', *Econometrica*, 50, 761-764.
- Chamberlain, G. S. (1984): 'Panel Data', in: Z. Griliches and M. D. Intriligator (eds.), *Handbook of Econometrics*, vol. 2, Amsterdam: North-Holland.
- Chamberlain, G. S. (1987): 'Asymptotic Efficiency in Estimation with Conditional Moment Restrictions', *Journal of Econometrics*, 34, 305-334.
- Chamberlain, G. S. (1992): 'Comment: Sequential Moment Restrictions in Panel Data', *Journal of Business & Economic Statistics*, 10 (1), 20-26.
- Chesher, A., and S. Peters (1994): 'Symmetry, Regression Design, and Sampling Distributions', *Econometric Theory*, 10, 116-129.
- Delgado, M. (1992): 'Semiparametric Generalized Least Squares in the Multivariate Nonlinear Regression Model', *Econometric Theory*, 8, 203-222.
- Gouriéroux, C., and A. Monfort (1993): 'Simulation-Based Inference: A Survey with Special Reference to Panel Data Models', *Journal of Econometrics*, 59, 5-33.
- Gouriéroux, C., and A. Monfort (1995): *Statistics and Econometric Models, Volume 1: General Concepts, Estimation, Prediction, and Algorithms*, Cambridge: Cambridge University Press.
- Guilkey, D. K., and J. L. Murphy (1993): 'Estimation and Testing in the Random Effects Probit Model', *Journal of Econometrics*, 59, 301-317.
- Hahn, J. (1997): 'Efficient Estimation of Panel Data Models with Sequential Moment Restrictions', *Journal of Econometrics*, 79, 1-21.

- Hajivassiliou, V. A. (1993): 'Simulation Estimation Methods for Limited Dependent Variable Models', in: G. S. Maddala, C. R. Rao and H. D. Vinod (eds.), *Handbook of Statistics*, vol. 11, ch. 19, Amsterdam: North-Holland.
- Hajivassiliou, V. A., and D. L. McFadden (1990): 'The Method of Simulated Scores for the Estimation of LDV Models with an Application to External Debt Crises', Yale Cowles Foundation Discussion Paper nr. 967.
- Hajivassiliou, V. A., D. L. McFadden, and P. A. Ruud (1996): 'Simulation of Multivariate Normal Rectangle Probabilities and their Derivatives: Theoretical and Computational Results', *Journal of Econometrics*, 72, 85-134.
- Hajivassiliou, V. A., and P. A. Ruud (1994): 'Classical Estimation Methods for LDV Models Using Simulation', in: R. F. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, ch. 40, Amsterdam: North-Holland.
- Hansen, L. P. (1982): 'Large Sample Properties of Generalized Method of Moments Estimators', *Econometrica*, 50, 1029-1054.
- Hastings, C. (1955): *Approximations for Digital Computers*, Princeton: Princeton University Press.
- Heckman, J. J. (1981): 'Statistical Models for Discrete Panel Data', in: C. F. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, ch. 3, Cambridge: MIT Press.
- Hsiao, C. (1996): 'Logit and Probit Models', in: L. Mátyás and P. Sevestre (eds.), *The Econometrics of Panel Data*, 2nd edition, ch. 16, Dordrecht: Kluwer.
- Imbens, G. W. (1997): 'One-Step Estimators for Over-Identified Generalized Method of Moments Models', *Review of Economic Studies*, 64, 359-383.
- Imbens, G. W., R. H. Spady, and P. Johnson (1998): 'Information-Theoretic Approaches to Inference in Moment Condition Models', *Econometrica*, 66 (2), 333-358.
- Inkmann, J., and W. Pohlmeier (1995): 'R&D Spillovers, Technological Distance and Innovative Success', presented at the IFS conference on 'R&D, Innovation and Productivity', London.
- Kamien, M. I., and N. L. Schwartz (1982): *Market Structure and Innovation*, Cambridge: Cambridge University Press.
- Keane, M. P. (1993): 'Simulation Estimation for Panel Data with Limited Dependent Variables', in G. S. Maddala, C. R. Rao, and H. D. Vinod (eds.), *Handbook of Statistics*, vol. 11, ch. 20, Amsterdam: North-Holland.
- Keane, M. P. (1994): 'A Computationally Practical Simulation Estimator for Panel Data', *Econometrica*, 62, 95-116.
- Kitamura, Y., and M. Stutzer (1997): 'An Information-Theoretic Alternative to Generalized Method of Moments Estimation', *Econometrica*, 65 (4), 861-874.
- König, H., F. Laisney, M. Lechner, and W. Pohlmeier (1994): 'On Dynamics of Process Innovative Activity: An Empirical Investigation Using Panel Data', in: K. H. Oppenländer and G. Poser (eds.), *The Explanatory Power of Business Cycle Surveys*, 245-262, Avebury: Aldershot.

- Laisney, F., M. Lechner, and W. Pohlmeier (1992a): 'Innovative Activity and Firm Heterogeneity: Empirical Evidence from West Germany', *Structural Change and Economic Dynamics*, 3, 301-320.
- Laisney, F., M. Lechner, and W. Pohlmeier (1992b): 'Semi-Nonparametric Estimation of Binary Choice Models using Panel Data: Application to Innovative Activity of German Firms', *Recherches Economiques de Louvain*, 58, 329-347.
- Lechner, M., and J. Breitung (1996): 'Some GMM Estimation Methods and Specification Tests for Nonlinear Models', in: L. Mátyás and P. Sevestre (eds.), *The Econometrics of Panel Data*, 2nd edition, ch. 22, Dordrecht: Kluwer.
- Legler, H., H. Grupp, B. Gehrke, and U. Schasse (1992): *Innovationspotential und Hochtechnologie*, Heidelberg: Physica.
- Lerman, S. R., and C. F. Manski (1981): 'On the Use of Simulated Frequencies to Approximate Choice Probabilities', in: C. F. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, ch. 7, Cambridge: MIT Press.
- McFadden, D. (1989): 'A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration', *Econometrica*, 57 (5), 995-1026.
- Mühleisen, M. (1994): *Human Capital Decay and Persistence: A Simulation Approach to German Unemployment*, Studies in Labor Economics, Frankfurt: Campus.
- Newey, W. K. (1990): 'Efficient Instrumental Variables Estimation of Nonlinear Models', *Econometrica*, 58 (4), 809-837.
- Newey, W. K. (1993): 'Efficient Estimation of Models with Conditional Moment Restrictions', in G. S. Maddala, C. R. Rao, and H. D. Vinod (eds.), *Handbook of Statistics*, vol. 11, ch. 16, Amsterdam: North-Holland.
- Newey, W. K., and K. D. West (1987): 'Hypothesis Testing with Efficient Method of Moments Estimation', *International Economic Review*, 28, 777-787.
- Odeh, R. E., and J. O. Evans (1974): 'The Percentage Points of the Normal Distribution', *Applied Statistics*, 23, 96-97.
- Pakes, A., and D. Pollard (1989): 'Simulation and the Asymptotics of Optimization Estimators', *Econometrica*, 57 (5), 1026-1057.
- Pohlmeier, W. (1994): 'Panelökonometrische Modelle für Zähldaten: Einige neuere Schätzverfahren', *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*, 130 (3), 553-574.
- Pohlmeier, W., and J. Inkmann (1997): 'Efficient Conditional Moment Estimation of Nonlinear SUR Models: Theory and Application to Count Data', working paper presented at the European Econometric Society Meeting in Toulouse.
- Robinson, P. M. (1987): 'Asymptotically Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form', *Econometrica*, 55, 875-891.
- Silverman, B. W. (1986): *Density Estimation*, London: Chapman & Hall.
- Stone, C. J. (1977): 'Consistent Nonparametric Estimation', *The Annals of Statistics*, 5 (4), 595-645.
- Tausworthe, R. C. (1965): 'Random Numbers Generated by Linear Recurrence Modulo Two', *Mathematics of Computation*, 19, 201-209.