Institut für
Volkswirtschaftslehre
und Statistik

No. 528-95

Convenient Estimators for the Panel Probit Model

Irene Bertschek und Michael Lechner

Beiträge zur angewandten Wirtschaftsforschung

Universität Mannheim
A5, 6
D-68131 Mannheim

Convenient Estimators for the Panel Probit Model

Irene Bertschek*)‡ Michael Lechner**)‡

*) Université Catholique de Louvain, Louvain-la-Neuve
**) University of Mannheim

Revised: May 1997

Abstract

The paper shows that several estimators for the panel probit model suggested in the literature belong to a common class of GMM estimators. They are relatively easy to compute because they are based on conditional moment restrictions involving univariate moments of the binary dependent variable only. Applying nonparametric methods we discuss an estimator that is optimal in this class. A Monte Carlo study shows that a particular variant of this estimator has good small sample properties and that the efficiency loss compared to maximum likelihood is small. An application to the product innovation decisions of German firms reveals the expected efficiency gains.

KEY WORDS: panel probit model, GMM, conditional moment restrictions, k-nearest neighbor estimation.
JEL-classification: C14, C23, C25

---

‡We gratefully acknowledge financial support by the "Human Capital and Mobility" programme, ERBCHBICT 941032 (first author) and by the Deutsche Forschungsgemeinschaft (DFG) (second author). Parts of the paper were written when the first author was visiting Humboldt University, Berlin, and the second author was at the Center for European Studies, Harvard University. Therefore, we thank the SFB373 and the CES, respectively, for financial support. Thanks also to François Laisney, Jörg Breitung, Gary Chamberlain, Bruno Crépon, Wolfgang Härdle, Guido Imbens, Sigbert Klinke, Byeong Park, Gerd Ronning and the participants of econometric seminars at Harvard, Konstanz, Mannheim, Tilburg and Humboldt University, Berlin, for helpful comments and suggestions on a previous draft of the paper.

Correspondence to: Michael Lechner, Fakultät für Volkswirtschaftslehre, Universität Mannheim, 68131 Mannheim, Germany, e-mail: lechner@haavelmo.vwl.uni-mannheim.de, http://www.vwl.uni-mannheim.de/lehrst/ls-oek/lechner.

1 Introduction

The probit model is a popular model in applied microeconometric work. In cross-section analysis, when the error terms of the observations are assumed to be identically and independently distributed (iid), maximum-likelihood (ML) is typically the chosen estimation method. It is easy and fast to compute and asymptotically efficient. However, using ML on panel data is burdensome unless one adopts the unattractive assumption of iid error terms, which rules out any persistent or idiosyncratic components in the errors of the same unit (firm, individual) over time.[4] As a consequence, the joint T-variate probability distribution over time needs to be specified. In cases when no analytical expressions exist for the individual likelihood contributions — such as for probit and tobit models — numerical evaluations of cumulative distribution functions could be a problem. In addition, there may be a possibly large number of nuisance parameters resulting from the intertemporal error covariance matrix.

Several solutions to that problem are discussed in the literature: one group of methods focuses on approximating the integrals by simulation. Although substantial progress has been achieved recently — see for example the survey by Hajivassiliou and Ruud (1994) — these methods are still computationally expensive. At least for a larger number of time periods, they require a specification of the error process with a limited number of covariance parameters. Another group of estimators restricts the error terms to have a random effects specification whereby the computation of the ML estimator is considerably simplified (see for instance Butler and Moffitt, 1982). This can be generalized by introducing one-factor or multi-factor schemes to allow for a more flexible error structure as proposed for example by Heckman (1981). However, this generalization comes at the cost of having to estimate more parameters of the covariance matrix, and — in the case of multi-factor schemes — increases the dimension of integration. Finally, a third group of estimators is based on the generalized method of moments (GMM) using moment restrictions that do not depend on parameters of the intertemporal error covariance matrix (Avery, Hansen and Hotz, 1983). An evaluation of the joint T-variate cumulative distribution function is not necessary.

The main focus of the paper is on this third group of estimators. We show that several often-used and conveniently computable estimators, such as pooled probit,

[4] We consider the case of a large number of independent units (N) observed for a finite number of periods (T).

Chamberlain's (1980, 1984) sequential estimator or several variants of other suggested GMM estimators belong to a class of GMM estimators using the same conditional moment restrictions. An asymptotic efficiency ranking of these and other related GMM estimators is established. Applying the asymptotic efficiency results for instrumental variable estimation of nonlinear models established by Newey (1990) and Chamberlain (1987), we discuss several feasible estimators that are asymptotically efficient in that class of GMM estimators. Although the estimators use nonparametric estimation to obtain the asymptotically efficient instruments, they retain the basic simplicity, feasibility and robustness to arbitrary error structures that are the great advantages of the previously discussed GMM estimators. An extensive Monte Carlo study shows that a particular variant has good small sample properties. Furthermore, the efficiency loss compared to full information maximum likelihood appears to be rather small. Finally, various estimators are applied to an example taken from industrial economics. Firms' product innovative activity is analysed using a panel data set that contains 1270 firms of the German manufacturing industry observed over five periods (years). The suggested estimator performs well in practice, and the efficiency gains compared to the other estimators turn out to be important for the economic interpretation of the estimation results.

The following section motivates the econometric discussion by introducing the economic example. Furthermore, it establishes the necessary notation and the statistical assumptions underlying the analysis. Section 3 briefly discusses restricted and unrestricted maximum likelihood estimation and points out some of the problems that could appear in the context considered here. The first part of section 4 gives a compact summary of the theory of GMM estimation with conditional moment restrictions. An asymptotic efficiency ranking of several estimators using this framework is established in the second part. The third part discusses the implementation of an estimator that exploits the information of these conditional moment restrictions optimally. The Monte Carlo results are presented in section 5. In particular, the data generating processes are described in the first subsection. Subsection 2 discusses the implementation of the various estimators. Their asymptotic distributions are compared in subsection 3. Finally, subsection 4 addresses the finite sample properties. The application is given in section 6. Section 7 concludes. Appendix A gives a useful lemma concerning the asymptotic equivalence of several types of GMM estimators. Appendix B contains additional Monte Carlo results and Appendix C gives more details on the data used for the application.

2 Empirical Example, Notation and Basic Assumptions

An empirical example for our discussion of panel probit models is the analysis of firms' innovative activity as a response to imports and foreign direct investment (FDI) as considered in Bertschek (1995). The main hypothesis put forward in that paper is that imports and inward FDI have positive effects on the innovative activity of domestic firms. The intuition for this effect is that imports and FDI represent a competitive threat to domestic firms. Competition on the domestic market is enhanced and the profitability of the domestic firms might be reduced. As a consequence, these firms have to produce more efficiently. Increasing the innovative activity is one possibility to react to this competitive threat and to maintain the market position.

The dependent variable available in the data takes the value one if a product innovation has been realized within the last year and the value zero otherwise. The binary character of this variable leads us to formulate the model in terms of a latent variable y_{ti}^* that represents for instance the firms' unobservable expenditures for innovation. y_{ti}^* is linearly related to the explanatory variables x_{ti}. The vector β^0 contains K deterministic coefficients. u_{ti} is a scalar error term controlling effects that are not captured by the regressors:

$$y_{ti}^* = x_{ti}\beta^0 + u_{ti}. \tag{1}$$

The observation rule is:

$$y_{ti} = \mathbb{1}(y_{ti}^* > 0), \quad i = 1, ..., N,\ t = 1, ..., T, \tag{2}$$

where the indicator function $\mathbb{1}(\cdot)$ equals one if the expression in brackets is true and zero otherwise. For each individual i we collect T observations such that $y_i = (y_{1i}, ..., y_{Ti})'$ is a $T \times 1$ vector and $x_i = (x'_{1i}, ..., x'_{Ti})'$ represents a $T \times K$ matrix of regressors.

The following standard assumptions are made: We observe N independent random draws $(y_i, x_i) = z_i$ in the joint distribution of the random variables $(Y, X) = Z$. Thus, z_i has dimension $T \times (K + 1)$. In the data set, z_i is observed for $N = 1270$ firms from 1984 to 1988 ($T = 5$). This is compatible with the assumption of fixed T and increasing N forming the basis of the following asymptotic arguments.

The error terms $u_i = (u_{1i}, ..., u_{Ti})'$ are assumed to be jointly normally distributed with mean zero and covariance matrix Σ and to be independent of the explanatory variables which implies the strict exogeneity of the latter. They are uncorrelated over firms but may be correlated over time for the same firm. One main-diagonal element of Σ has to be set to unity because identification of β^0 is only up to scale.[5] The off-diagonal elements of Σ are not of interest in the empirical study and therefore, they are considered as nuisance parameters.

3 Maximum Likelihood Estimation

The typical approach for estimating probit models in applied microeconometric work based on single cross-sections is maximum likelihood (cf. Maddala, 1983, for many examples). Due to the availability of fast and accurate methods to evaluate the univariate normal cumulative distribution function (cdf) — for which no analytical formula is available — and due to the global concavity of the log likelihood function, this is a useful approach implemented in many software packages. However, in the case of panel data there are several issues that make ML estimation less attractive: first of all, the likelihood function depends on $T(T-1)/2$ unknown off-diagonal elements of Σ that have to be estimated. Secondly and probably even more important in practice, the computation time of the T-variate cdf instead of the univariate cdf is prohibitively high for $T > 4$ or 5 even on very powerful computers (cf. Hajivassiliou and Ruud, 1994). Finally, the possible lack of global concavity may represent a problem as well.

While the last issue is widely ignored, many papers appeared recently in the literature suggesting that the dimensionality problem with respect to integration can be overcome by using suitable simulation methods to approximate the multidimensional integral. For details on these issues the reader is referred to the excellent survey by Hajivassiliou and Ruud (1994). However, although estimates based on simulation methods are easier to compute than exact ML, there are drawbacks with this approach as well: firstly, a sufficiently accurate estimation of the multivariate probabilities may still be computationally expensive. Secondly, Σ is typically restricted by assuming parametric error processes, mostly AR or MA processes, sometimes

[5]In order to simplify the exposition we normalize all variances ($\sigma_{11} = \sigma_{tt} = \sigma_{TT} = 1$, for all t). However, the basic structure of the results remains unchanged if the following reparameterization is made: $\theta = (\theta_1', \theta_2')', \theta_1 = \beta/\sigma_1, \theta_2 = (\theta_{22}, ..., \theta_{2T})', \theta_{2t} = \sigma_1/\sigma_t$ and $\rho_{ts} = \sigma_{ts}/(\sigma_s\sigma_t)$ where σ_{ts} denotes EU_tU_s and $\sigma_t = \sqrt{\sigma_{tt}}$. θ is then used in place of β in sections 4 and 5. The exact results using this extensive notation are contained in a previous version of this paper that is available on request from the authors.

combined with random effects specifications. Due to these restrictions the number of parameters is considerably reduced and simulation estimation becomes feasible.

Some restrictions also drastically reduce the dimension of integration. The most widely used restriction is the assumption that the error terms are equicorrelated over time, i.e. the error terms can be decomposed into two mutually independent components: a time constant random effect c_i and a remainder term ε_{ti} (assumed to be independent over time) such that $u_{ti} = \delta c_i + \varepsilon_{ti}$. δ is a positive constant to be estimated and c_i and ε_{ti} follow a standard normal distribution.[6] In this case the log likelihood function is given by:

$$
\begin{aligned}
L(y, x; \beta, \delta) \quad = \quad & \frac{1}{N} \sum_{i=1}^{N} \ln \int_{-\infty}^{+\infty} \sum_{t=1}^{T} \{\Phi(x_{ti}\beta + \delta c)^{y_{ti}} \\
& [1 - \Phi(x_{ti}\beta + \delta c)]^{(1-y_{ti})}\} \phi(c) dc, \qquad (3)
\end{aligned}
$$

$\Phi(\cdot)$ and $\phi(\cdot)$ denote the cumulative distribution function and the probability distribution function (pdf) of the univariate standard normal distribution, respectively. Butler and Moffitt (1982) suggest an efficient method to evaluate the integral numerically by Hermite integration (e.g. Stroud and Secrest, 1966, p. 22). The estimator is then given by:

$$
\begin{aligned}
\begin{pmatrix} \hat{\beta}_N \\ \hat{\delta}_N \end{pmatrix} \quad = \quad & \arg\max_{\beta,\delta} \frac{1}{N} \sum_{i=1}^{N} \ln \sum_{v=1}^{V} \sum_{t=1}^{T} \{\Phi(x_{ti}\beta + \delta c_v)^{y_{ti}} \\
& [1 - \Phi(x_{ti}\beta + \delta c_v)]^{(1-y_{ti})}\} w_v, \qquad (4)
\end{aligned}
$$

where c_v and w_v are the respective evaluation points and weights. Their values have been tabulated by various authors, e.g. Abramowitz and Stegun (1966, p.924) and Stroud and Secrest (1966, Table 5).

A comparison of restricted 'exact' and simulated ML for this error decomposition can be found in Guilkey and Murphy (1993). The error structure can be made more flexible by allowing δ to vary over time and by the possible introduction of more than one factor (cf. Heckman, 1981). However, this will again increase the number of parameters to be estimated and the dimension of integration.

Using the framework of generalized method of moments (GMM) estimation Avery et al. (1983) show that ML estimation under the even more restrictive assumption of independent errors over time leads to consistent estimates for β^0.[7] The 'pooled'

[6] For notational convenience we change the normalization of σ_{tt} in this case, by setting $\sigma^2_{\varepsilon_t} = 1$.

[7] See also Robinson (1982) for the more general proof that the probit-ML assuming independent errors represents a consistent estimator even when the true errors are not independent.

pseudo-log likelihood function is given by:

$$\hat{\beta}_N = \arg\max_{\beta \in B} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} y_{ti} \ln \Phi(x_{ti}\beta) + (1 - y_{ti}) \ln[1 - \Phi(x_{ti}\beta)]. \qquad (5)$$

The advantage of the pooled estimator is its simplicity, global concavity and lack of nuisance parameters. However, the disadvantage is its inefficiency and that the estimated asymptotic standard errors assuming the pseudo-log likelihood function to be the true log likelihood are inconsistent, if the errors are in fact correlated over time. The following section will show that with GMM estimation the advantages of the pooled probit can be retained and that the disadvantages can be overcome.

4 GMM Estimation

The following subsection introduces the GMM framework based on estimation with conditional moment restriction. It summarizes some of the results of Chamberlain (1987) and Newey (1990, 1993) and adapts them to the panel probit model. The second subsection shows how this framework can be used to obtain an efficiency ranking of several panel probit estimators that have been suggested in the literature. Finally, the estimation of the asymptotically optimal instruments is discussed.

4.1 Conditional Moment Restrictions and Asymptotic Efficiency

The model presented in section 2 implies the following moment conditions:

$$\begin{aligned} E[M(Z;\beta^0)|X] &= 0, \\ M(Z;\beta) &= [m_1(Z_1;\beta), ..., m_t(Z_t;\beta), ..., m_T(Z_T;\beta)]', \\ m_t(Z_t;\beta) &= Y_t - \Phi(x_{ti}\beta). \end{aligned} \qquad (6)$$

The use of these conditional moments for estimation has the advantage that their evaluation does not require multidimensional integration and that they do not depend on the $T(T-1)/2$ off-diagonal elements of Σ.

Given these conditional moment restrictions GMM can be used for estimation (Hansen, 1982). The following exposition borrows from the excellent survey by Newey (1993), who summarizes the results for such GMM estimators.

Based on eq. (6) the unconditional moment restriction to be used for the estimation is obtained by observing that $M(Z;\beta^0)$ will be uncorrelated with all functions of X, hence:

$$EA(X)M(Z;\beta^0) = 0. \tag{7}$$

$A(X)$ is a $p \times T$ "instrument matrix". An estimate $\hat{\beta}_N$ of β^0 is obtained by setting a quadratic form of the sample analogues

$$g_N(\beta) = \frac{1}{N}\sum_{i=1}^{N} A(x_i)M(z_i;\beta) \tag{8}$$

close to zero, such that

$$\hat{\beta}_N = \arg\min_{\beta} \; g_N(\beta)' P g_N(\beta). \tag{9}$$

Under suitable regularity conditions on g_N and with the positive semi-definite matrix P, $\hat{\beta}_N$ is $\sqrt{N}$-consistent and asymptotically normal:

$$\sqrt{N}(\hat{\beta}_N - \beta^0) \xrightarrow{d} N(0,\Lambda), \tag{10}$$

where

$$\begin{array}{rcll}
\Lambda & := & (G'PG)^{-1}G'PVPG(G'PG)^{-1}, & [K \times K] \\
G & := & E\left[A(X)\frac{\partial M(Z;\beta^0)}{\partial \beta'}\right] = EA(X)D(X), & [p \times K] \\
V & := & E[A(X)M(Z;\beta^0)M(Z;\beta^0)'A(X)'] & \\
 & = & E[A(X)\Omega(X)A(X)'], & [p \times p] \\
D(X) & = & E\frac{\partial M(Z;\beta^0)}{\partial \beta'}|X, & [T \times K] \\
\Omega(X) & = & EM(Z;\beta^0)M(Z;\beta^0)'|X, & [T \times T].
\end{array} \tag{11}$$

A consistent estimate of the covariance matrix Λ can be obtained by replacing expectations by sample means and β^0 by $\hat{\beta}_N$. The tools to minimize the asymptotic variance of this estimator are the optimal choice of the instruments $A(X)$ and of the weighting matrix P. As shown by Hansen (1982) in a more general setting, the optimal choice of P is V^{-1} or any consistent estimator of it. Chamberlain (1987) and Newey (1990) derived the optimal instrument matrix A^*:

$$A^*(X) = C\; D(X)'\Omega(X)^{-1}, \tag{12}$$

where C is any nonsingular $K \times K$ matrix. The column-dimension of A^* equals K, so the choice of P is irrelevant. To obtain a feasible estimator, P, $D(X)$ and $\Omega(X)$ may be substituted by consistent estimates without affecting the asymptotic distribution of $\hat{\beta}_N$. For the GMM estimator using the optimal instruments, the covariance matrix simplifies to

$$\Lambda^* = \{E[D(X)'\Omega(X)^{-1}D(X)]\}^{-1}. \tag{13}$$

For the probit model and observation 'i', $D(x_i)$ with typical row d_{ti}, and $\Omega(x_i)$ with typical element ω_{tsi} have the following form:

$$d_{ti} \;=\; -\phi(x_{ti}\beta)x_{ti}. \tag{14}$$

For notational convenience let $\Phi_{ti} := \Phi(x_{ti}\beta^0)$ and $\Phi^{(2)}_{tsi} := \Phi^{(2)}(x_{ti}\beta^0, x_{si}\beta^0, \rho^0_{ts})$. $\Phi^{(2)}(\cdot)$ denotes the cumulative distribution function of the bivariate standardized normal distribution with correlation coefficient ρ^0_{ts}. Hence, we obtain:

$$\omega_{tsi} = [E(Y_t - \Phi_{ti})(Y_s - \Phi_{si})|X = x_i] = \begin{cases} \Phi_{ti}(1-\Phi_{ti}) & \text{if } t = s, \\ \Phi^{(2)}_{tsi} - \Phi_{ti}\Phi_{si} & \text{if } t \neq s. \end{cases} \tag{15}$$

Note that $\omega_{tsi}(x_i)$ has the same sign as ρ_{ts} and that $\omega_{tsi} = 0$ if $\rho_{ts} = 0$. The estimation of the optimal GMM-estimator is still difficult, because it depends on the unknown correlation coefficients of Σ through the terms $\Phi^{(2)}_{tsi}$ in eq. (15). A possible solution would be to replace these unknown coefficients by consistent estimates obtained from $(T-1)T/2$ separate bivariate probits. But this is cumbersome for large T.[8]

To circumvent these problems, Newey (1990, 1993) suggests the use of nonparametric methods, such as nearest neighbor estimation and series approximations to obtain consistent estimates of $\Omega(x_i)$. He derives the conditions necessary for these methods to result in consistent and asymptotically efficient estimates of β (Newey, 1993, theorems 1 and 2). We will come back to this issue in section 4.3.

Before doing so, we will discuss the asymptotic properties of various other sub-optimal GMM estimators suggested in the literature. They should be considered as competitors to the asymptotically optimal GMM estimator because of their computational simplicity, and because some of them require weaker conditions with respect to the exogeneity of the regressors. Note that the conditioning in the T-dimensional moment function given in eq. (6) is on $X = (X_1, ..., X_T)$. This is the so-called strict exogeneity restriction. The results with respect to consistency and asymptotic efficiency put forward in section 4 do require this assumption to hold. A weaker

[8] An alternative would be to set up another GMM estimator based on $\frac{T(T-1)}{2}$ moment conditions like $y_{ti}y_{si} - \Phi^{(2)}_{tsi}$ with unknown β^0 replaced by a consistent estimate $\hat{\beta}_N$. In this case these moment conditions would suffice to estimate the unknown correlations.

assumption would be to require only $E[m_t(Z_t;\beta^0)|X_t] = 0$ (weak exogeneity).[9] The consequence is that only functions of X_t are valid instruments for the t'th-element of M. Inspecting the form of the instrument matrix $A(X)$ suggested in this section, it can be seen that only the pooled probit and the sequential estimator are not affected by this weakening of the exogeneity assumption. This might be an important consideration in practice, since in particular the 'optimal' instrument matrix given in eq. (12) is no longer valid if errors are correlated over time.

4.2 Efficiency Ranking of Several Estimators

All GMM estimators to be discussed in the following are consistent regardless of the true covariance matrix of the error terms, but differ in their asymptotic variance. We use the GMM framework to readily obtain asymptotic efficiency comparisons.

Avery et al. (1983) observe that the scores of the pooled ML estimator imply moment conditions that can be used for GMM estimation. Using the Lemma in Appendix A leads to an asymptotically equivalent GMM estimator of the following form:

$$
\begin{aligned}
g_N^{PP1}(z;\beta) &= \frac{1}{N}\sum_{i=1}^{N} A_i^{PP1}(x_i)M(z_i;\beta) \qquad [K \times 1], \qquad (16)\\
A_i^{PP1}(x_i) &= D_i'(x_i)[\Omega_i^{PP}(x_i)]^{-1},
\end{aligned}
$$

$$
\Omega_i^{PP}(x_i) = \begin{pmatrix} \Phi(x_{1i}\beta^0)[1-\Phi(x_{1i}\beta^0)] & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Phi(x_{Ti}\beta^0)[1-\Phi(x_{Ti}\beta^0)] \end{pmatrix} \quad (T \times T).
$$

Since there are no overidentifying restrictions, the choice of the weighting matrix P does not matter. It is clear from the structure of Ω_i^{PP} that the efficiency loss of the pooled probit estimator is due to the ignorance of possible nonzero off-diagonal elements in Ω. In order to compare the pooled ML estimator with other GMM estimators, it is useful to rewrite this estimator in an equivalent representation:

$$
g_N{}^{PP2}(z;\beta) = \frac{1}{N}\sum_{i=1}^{N} A_i^{PP2}(x_i)M(z_i;\beta) \qquad [TK \times 1], \qquad (17)
$$

[9] Note that when X contains lagged dependent variables this condition holds only when errors are assumed to be independent over time.

$$A_i^{PP2}(x_i) = \begin{pmatrix} \frac{-\phi(x_{1i}\beta^0)}{\Phi(x_{1i}\beta^0)[1-\Phi(x_{1i}\beta^0)]}x'_{1i} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{-\phi(x_{Ti}\beta^0)}{\Phi(x_{Ti}\beta^0)[1-\Phi(x_{Ti}\beta^0)]}x'_{Ti} \end{pmatrix} (TK \times T). \tag{18}$$

Note that by *stacking* the moment conditions there are overidentifying restrictions and the particular choice of the weighting matrix P matters. We see that pooled probit is equivalent to a GMM estimator with an inefficient weighting matrix P^{PP2}:

$$P^{PP2} = RR', \quad R = \begin{pmatrix} I_K \\ \vdots \\ I_K \end{pmatrix} (TK \times K), \quad I_K = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \end{pmatrix} (K \times K).$$

Therefore using the optimal P^{PP2^*} instead of P^{PP2} leads to an asymptotically more efficient estimator:

$$\begin{aligned} P^{PP2^*} &= EA^{PP2}(X)M(Z;\beta^0)M(Z;\beta^0)'A^{PP2}(X)' \\ &= EA^{PP2}(X)\Omega(X)A^{PP2}(X)'. \end{aligned} \tag{19}$$

However, as Breitung and Lechner (1997) show in a Monte Carlo study, the estimator based on P^{PP2^*} may actually perform worse than pooled probit in small and medium sized samples. Other estimators suggested by Avery et al. (1983) are in the same spirit, i.e. the efficiency gains depend on an expansion of the instrument set together with the use of the optimal weighting matrix. Hence, they share the same problem, namely that a large number of observations is needed to get a sufficiently accurate estimate of this high-dimensional matrix.

Another popular and convenient estimator is the *sequential* estimator suggested by Chamberlain (1980, 1984). The idea is as follows: In a first step a probit is estimated for each cross-section. After computing the joint covariance matrix of all first step estimates, a minimum distance procedure is used to impose the coefficient restrictions due to the panel structure to obtain more efficient estimates. Using the scores of the probit for each period and employing similar reformulations as for the pooled probit the first step of this estimator can be expressed in our framework:

$$\begin{aligned} g_N^S(z;\beta_1,...,\beta_T) &= \frac{1}{N}\sum_{i=1}^{N} D_i^S(x_i)'(\Omega_i^P)^{-1}M(z_i;\beta_1,...,\beta_T) \\ &= \frac{1}{N}\sum_{i=1}^{N} A_i^{PP2}(x_i)M(z_i;\beta_1,...,\beta_T), \end{aligned} \tag{20}$$

$$D_i^S = \begin{pmatrix} -\phi(x_{1i}\beta^0)x_{1i} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & -\phi(x_{Ti}\beta^0)x_{Ti} \end{pmatrix} \quad (T \times KT).$$

Breitung and Lechner (1997) show that when the minimum distance step is based on the optimal weighting matrix, this estimator is asymptotically more efficient than pooled ML, because it is equivalent to the GMM estimator of eq. (17) using the optimal weighting matrix P^{PP2^*}.

Since Monte Carlo evidence suggests that increasing efficiency by increasing the number of instruments may lead to small sample problems, we now turn to a different way of increasing efficiency compared to pooled ML that avoids this dilemma.

The idea is to use the moment function given in eq. (8) together with the optimal instruments (eq. 12) leading to an estimator without overidentifying restrictions, so that a high-dimensional weighting matrix does not need to be estimated. The problem is then to find a consistent estimator of $\Omega(X)$. One approach is to assume an error distribution that is plausible in many cases and leads to an $\Omega(X)$ that is easy to compute. When this assumption is wrong, then the estimator is not efficient but still consistent. In this vein is a suggestion by Breitung and Lechner (1997). They assume a random effects (equicorrelation) structure with a 'small' variance of the random effect (σ_c^2). A Taylor-expansion of the moment condition (conditional on c) around σ_c^2, leads to an approximation of $\Omega(X)$, denoted by $\Omega^{SS}(X)$, that should be particularly good when the true error structure is small random effects:

$$\begin{aligned} \Omega^{SS}(x_i) &= \begin{pmatrix} \Phi_{1i}(1-\Phi_{1i}) + \sigma_c^2\phi_{1i}^2 & & & \\ \sigma_c^2\phi_{1i}\phi_{2i} & \ddots & & \\ \vdots & & \ddots & \\ \sigma_c^2\phi_{1i}\phi_{Ti} & \cdots & \sigma_c^2\phi_{T-1,i}\phi_{Ti} & \Phi_{Ti}(1-\Phi_{Ti}) + \sigma_c^2\phi_{Ti}^2 \end{pmatrix}. \\ \phi_{ti} &= \phi(x_{ti}\beta^0). \end{aligned}$$

There is only one unknown parameter, σ_c, in addition to β^0, and it can be estimated for example by OLS with the following regressions:

$$(y_{ti} - \tilde{\Phi}_{ti})(y_{si} - \tilde{\Phi}_{si}) = \sigma_c^2\tilde{\phi}_{ti}\tilde{\phi}_{si} + \text{error}, \quad t, s = 1, ..., T;\ t \neq s, \tag{21}$$

where ˜ denotes quantities evaluated with a consistent $\sqrt{N}$-normal first step estimate $\tilde{\beta}_N$ of β^0. This estimator (GMM-SS) is easy and fast to compute. However, the dependence of the potential efficiency gains on the validity of the rather restrictive assumption about the true error covariance is not very satisfactory. Therefore, the next subsection discusses a simple nonparametric estimate of the optimal instruments.

4.3 Nonparametric Estimation of $\Omega(x_i)$

In the following we focus on the k-nearest neighbor (k-NN) approach to estimate $\Omega(x_i)$, because of its simplicity. k-NN averages locally over functions of the data of those observations belonging to the k nearest neighbors. Under regularity conditions (Newey, 1993), this gives consistent estimates of $\Omega(x_i)$ evaluated at $\tilde{\beta}_N$ and denoted by $\tilde{\Omega}(x_i)$ for each observation 'i' without the need for estimating ρ_{ts}. Thus, an element of $\Omega(x_i)$ is estimated by:

$$\tilde{\omega}_{tsi}(x_i) = \sum_{j=1}^{N} W_{tsij} m_t(z_{tj}; \tilde{\beta}_N) m_s(z_{sj}; \tilde{\beta}_N), \tag{22}$$

where W_{tsij} represents a weight function.

In order to determine the neighbors of observation 'i' it is necessary to define a distance or similarity measure. Since the elements of the diagonal of $\Omega(x_i)$, ($\omega_{tsi}, t = s$) depend only on the individual indices of one time-period ($x_{ti}\beta$), we face a simple one-dimensional estimation problem. The off-diagonal elements, ω_{tsi}, $t \neq s$, depend on the linear indices of two time-periods, $\Psi_{tsij} = [(x_{ti} - x_{tj})\beta, (x_{si} - x_{sj})\beta]$.[10]

Hence, the distance used to define the neighbors should refer to those two. Here, another possibility is considered: The distance between observations 'i' and 'j' refers to the indices of all periods $\Psi_{ij} = [(x_{1i} - x_{1j})\beta, ..., (x_{Ti} - x_{Tj})\beta]$. Thus, the distance is defined either by $\xi_{tsij} = \Psi_{tsij}\Gamma^{x\beta}_{ts}\Psi'_{tsij}$ in case of estimating each single element of $\Omega(x_i)$ *individually* or by $\xi_{ij} = \Psi_{ij}\Gamma^{x\beta}\Psi'_{ij}$ when all elements of $\Omega(x_i)$ are estimated *jointly* ($W_{tsij} = W_{ij}$ for all t, s). If not indicated otherwise, the weighting matrices $\Gamma^{x\beta}$ and $\Gamma^{x\beta}_{ts}$, respectively, are set to unity in the Monte Carlo study. However, different choices of positive definite matrices can be used as well, such as the inverse of the covariance matrix of the linear indices ($S^{-1}_{x\beta}$) or only the main diagonal of $S^{-1}_{x\beta}$.

As a consequence, the nonparametric estimators can either be constructed according to (22) which estimates every single element of $\Omega(x_i)$ individually (*indiv*) or alternatively, by estimating all elements of $\Omega(x_i)$ jointly (*joint*). The *joint* estimator

[10]Note that the dimension is lower than for the measure suggested by Newey (1993) who uses the single elements of x_{ti} (x^k_{ti}) instead of $x_{ti}\beta$ to define similarity for more general models.

is given by:

$$\tilde{\Omega}^j(x_i) = \sum_{j=1}^{N} W_{ij} M(z_j;\tilde{\beta}_N) M(z_j;\tilde{\beta}_N)'. \tag{23}$$

The second procedure is much faster to compute, because the observations have to be sorted only once for the estimation of all elements of $\Omega(x_i)$. Therefore, the larger T the faster is *joint* relative to *indiv*.

A weight function assigns positive weights to those observations belonging to the k nearest neighbors ($k \leq N$), but zero weights to all other observations and the observation i itself. The weights sum to unity. Stone (1977, p. 600) suggests several weight functions that fulfil the necessary regularity conditions. Suppose that the observations are ordered according to their distance to observation 'i', where '$j = 1$' denotes observation 'i' itself. The uniform weight function (*uniform*) is then given by:

$$W_{tsij} = \begin{cases} 1/k & 2 \leq j \leq k, \\ 0 & j = 1, j > k. \end{cases} \tag{24}$$

A smoother version is for example the quadratic (*quadr*) weight function:

$$W_{tsij} \;=\; \begin{cases} \left[k^2 - (j-1)^2\right] / \left[k(k+1)(4k-1)/6\right] & 2 \leq j \leq k, \\ 0 & j = 1, j > k. \end{cases} \tag{25}$$

It remains to choose the smoothing parameter k. In our Monte Carlo study we follow Newey (1993) in applying cross-validation for a data-driven choice of k. He shows that cross-validation can be based on the difference between estimated and true moment functions:

$$\sum_{i=1}^{N} \{\tilde{A}^*(x_i) - A^*(x_i)\} M(Z;\beta^0)/\sqrt{N}. \tag{26}$$

Suppose that $\tilde{A}^*(x_i)$ denotes a consistent estimate of $A^*(x_i)$ evaluated at $\tilde{\beta}_N$. Then the resulting cross-validation function to be minimized is as follows (Newey, 1993, p. 433):

$$\hat{CV}(k) = \mathrm{tr}\left[Q \sum_{i=1}^{N} \tilde{R}(x_i)\tilde{\Omega}(x_i)\tilde{R}(x_i)'\right], \tag{27}$$

$$\tilde{R}(x_i) = \left\{\tilde{A}^*(x_i)[M(z_i,\tilde{\beta}_N)M(z_i,\tilde{\beta}_N)' - \tilde{\Omega}(x_i)]\right\} \tilde{\Omega}(x_i)^{-1}.$$

Q is a positive definite matrix. In the following Monte Carlo study and the application we choose $Q = \sum_{i=1}^{N} D(x_i)D(x_i)'$ according to Newey (1993).

Another possibility for nonparametric estimation of $\Omega(x_i)$ is to use kernel regression (Carroll, 1982; Härdle, 1990, for example). However, two drawbacks of kernel regression in our context are the increased complexity of the estimation and the problem of random denominators which might produce erratic behavior (see Robinson, 1987). This problem could be solved by trimming, but this leads to a loss of efficiency by reducing the number of observations.

5 Monte Carlo Study

The following subsection describes the data generating processes (DGPs) used. Following this, we discuss the implementation of the various estimators. We compare the standard errors of their asymptotic distributions in subsection 3. Finally, their finite sample properties are addressed in subsection 4.[11]

5.1 Data Generating Processes

The data generating processes (DGP) considered crudely mimic situations common in applied microeconometric work, e.g. regressors and error terms are both correlated over time, and dummy explanatory variables are part of the regressors. All DGPs can be summarized in the following equations:

$$\begin{aligned}
& y_{ti} = \mathbb{1}(\beta^C + \beta^D x_{ti}^D + \beta^N x_{ti}^N + u_{ti} > 0), && \\
& x_{ti}^D = \mathbb{1}(\tilde{x}_{ti}^D > 0), && P(\tilde{x}_{ti}^D > 0) = 0.5, \\
& x_{ti}^N = \gamma^x x_{t-1,i}^N + \gamma^t t + \tilde{x}_{ti}^U, && \tilde{x}_{ti}^U \sim \text{uniform}(-1,1), \\
& u_{ti} = \delta c_i + \varepsilon_{ti}, && c_i \sim N(0,1), \\
& \varepsilon_{ti} = \alpha \varepsilon_{t-1,i} + \sigma \tilde{\varepsilon}_{ti}, && \tilde{\varepsilon}_{ti} \sim N(0,1), \\
\text{or } \quad & u_{ti} = 0.5(\tilde{\varepsilon}_{ti} + \tilde{\varepsilon}_{t-1,i}), && \\
& i = 1, ..., N, \quad t = 1, ..., T. &&
\end{aligned}$$

$(\beta^C, \beta^D, \beta^N, \gamma^x, \gamma^t, \delta, \alpha, \sigma)$ are fixed coefficients. Initial values for the dynamic processes are discussed below. All random numbers are drawn independently over time and individuals.[12] The first regressor is an indicator variable that is uncorrelated over time, whereas the second regressor is a smooth variable with bounded support. The dependence on lagged values and on a time trend induces a correlation over

[11] In order to save space we did not include all simulation results in the following tables. The excluded results are available on request from the authors.

[12] We used the random number generators RNDN and RNDU implemented in GAUSS 3.1 and GAUSS 3.2.

time. This type of regressor is suggested by Nerlove (1971). The error terms may exhibit correlations over time due to an individual specific effect as well as a first order autoregression or a moving average process. To diminish the impact of initial conditions, the dynamic processes start at $t = -10$ with $x^N_{t-11,i} = \varepsilon_{t-11,i} = 0$. T is set to 5 and 10, and N to 100, 400 and 1600 in order to study the behavior in fairly small and large samples. Since all estimators are $\sqrt{N}$-consistent, the standard errors for the larger sample size should be approximately half the size for the next smaller sample. In addition to these finite sample, we use these DGPs to derive the asymptotic covariance matrices of the estimators as described below.

Tables 1 and 2 contain some statistics for the DGPs used in the estimations as well as the chosen coefficient values. All DGPs have the common feature that the unconditional mean of the dependent indicator variable is close to 0.5 in order to obtain maximum variance and thus to contain maximum information about the underlying latent variable. For ease of presentation let $\mu_{ti} = \beta^C + x^D_{ti}\beta^D + x^N_{ti}\beta^N$. Table 1 gives some summary statistics for the part of the DGP related to the regressors. The coefficients γ^x and γ^t are used to generate different correlation patterns of μ_{ti} over time. Here, we focus only on a 'medium' case. The results for other DGPs with $\gamma^x = 0$ and $\gamma^x = 0.9$ are available on request from the authors. Table 2 contains similar statistics for the error terms. The first error process (DGP 1) generalizes the equicorrelation pattern of DGP 2 by adding a first order autoregressive process. DGP 3 is a moving average process where correlation patterns die out after one period. Finally, there is an AR(1) process (DGP 4) with a negative coefficient so that the signs of the correlations alternate. Depending on T, 500 ($T = 5$) or 1000 ($T = 10$) replications (R) have been performed.

[insert Tables 1 and 2 about here]

5.2 Estimators

The following estimators already discussed in sections 3 and 4 are analysed: maximum likelihood with random effects according to eq. (4), *ML-RE*. We use the algorithm suggested by Butler and Moffitt (1982). The number of evaluation points V is set to 5 as a compromise between computational speed and numerical accuracy (see Guilkey and Murphy, 1993, for more Monte Carlo results). When the assumed error structure is the true one this estimator is consistent and asymptotically efficient for (β, δ). In the Monte Carlo study the standard errors are computed in the 'robust' way suggested by White (1982). Furthermore, there are the *pooled* estimator and the *sequential* estimator as presented in eqs. (5) and (18), respectively. The latter uses the optimal weighting matrix, given by the inverse of the joint covariance matrix of the first step estimates, to obtain the final estimates in the second step.

The estimator based on the instruments $A^*(x_i)$ and computed under the assumption of random effects with a small variance (*small sigma*) of the random effect relative to the total error variance (*GMM-SS*, cf. eq. (21)), dominated the other GMM estimators (for details see Breitung and Lechner, 1997). Hence, we report only results for this 'best' estimator to see whether it can be improved by the use of nonparametric methods. *GMM-SS* and all nonparametric estimators depend on a preliminary consistent estimate $\tilde{\beta}_N$. In the Monte Carlo study this is always the *pooled* probit estimate.

Several different variants of nonparametric estimators are considered. Instead of using the conditional moments given in eq. (6) and denoted by *NP*, one could also use the following scaled moments (*WNP*):

$$m_t^W(Z_t;\beta) = \frac{m_t(Z_t;\beta)}{\sqrt{E[m_t(Z_t;\beta)^2|X]}} \tag{28}$$

$$E[m_t^W(Z_t;\beta^0)|X=x_i] = \frac{E[m_t(Z_t;\beta^0)|X=x_i]}{\sqrt{\Phi_{ti}(1-\Phi_{ti})}} = 0.$$

The conditional variance of the moments given by eq. (6) is heteroscedastic across individuals, because it depends on explanatory variables, whereas the version given by eq. (28) leads to the following conditional covariance of the moments:

$$\omega_{tsi} = [E(m_t^W m_s^{W\prime})|X=x_i] = \begin{cases} 1 & \text{if } t=s, \\ \frac{\Phi_{tsi}^{(2)} - \Phi_{ti}\Phi_{si}}{\sqrt{\Phi_{ti}(1-\Phi_{ti})\Phi_{si}(1-\Phi_{si})}} & \text{if } t \neq s. \end{cases} \tag{29}$$

The latter is homoscedastic on the main diagonal and this could lead to small sample improvements. The matrix $D(\cdot)$ has to be changed accordingly. We also consider versions of *NP* and *WNP* for which only the off-diagonal elements of $\Omega(x_i)$ are estimated nonparametrically, and the main diagonal elements are set either to unity (*WNP*) or to $\Phi_{ti}(1-\Phi_{ti})$ (*NP*). However, this leads to numerical problems in some cases of *NP*, because some eigenvalues of $\tilde{\Omega}(x_i)$ are very small or even negative.

For the cross-validation a grid with eight values of k equally spaced in the interval $N^{0.67}$, $N^{0.97}$ is chosen. When the estimate of $\Omega(x_i)$ is not positive definite, the smoothing parameter k is increased according to this grid of ks until a positive definite estimate of $\Omega(x_i)$ is obtained. If the use of the largest possible k still does not produce a positive definite estimate, the values of the main diagonal are doubled until the matrix becomes positive definite. The latter correction is also applied to the versions of *NP* and *WNP* which only estimate the off-diagonal elements of $\Omega(x_i)$. We also analysed the case $k = N$, however, the results turned out to be worse than those for the above mentioned ks smaller than N (results available on request).

Finally, we computed a conditional moment estimator based on the true values of $\Omega(x_i)$ and $D(x_i)$ which are known in a Monte Carlo study. This estimator (*infeasible GMM-IV*) is generally infeasible and is used only as a benchmark for what could be achieved with an estimator optimal in that class and free of any variability coming from a first step estimation.

5.3 Asymptotic Comparisons

Before comparing the finite sample performance of the various estimators, we address the issue on how informative the particular moments are asymptotically. This will give us an indication about the efficiency gains that might be achievable in finite samples under these DGP's.[13]

Since the analytical compution of the asymptotic covariances proved intractable, we employed the following simulation strategy. For the GMM estimators (including the pooled probit), eq. (11) is the appropriate variance formula. Using the information about the DGP we computed $D(X)$, $A(X)$, $\Omega(X)$ analytically. Then, we drew 50.000 ($T = 5$) or 25.000 ($T = 10$) realizations from the distribution of X. The unconditional expectations appearing in eq. (11) (and also in eq. (13)) are estimated by averaging the respective quantities using this large number of independent draws.

For the ML-RE obtaining analytical expressions for the information matrix proved intractable as well. Therefore, we drew realizations from the joint distribution of Y and X to simulate the asymptotic covariance matrix. To account for possible additional simulation variance introduced by drawing also the binary indicator Y, the number of draws are doubled to 100 000 and 50 000, respectively. Additionally, to reduce the influence of having to approximate the integral appearing in eq. (3), we increased the number of evaluation points to 136 (taken from Stroud and Secrest, 1966, pp.250–252).[14] For the ML-RE two other issues arise: first, it is only for the case of (pure) random effects (DGP 2) that it is clearly consistent. Hence, we do not give asymptotic standard errors for the other DGP's. Second,

[13]We thank an Associate Editor of this journal for proposing this comparison to us.

[14]However, there is virtually no difference for the asymptotic standard deviation when 20 points are used instead. Similarly, the results based on using either the OPG, the Hessian, or White-version of the covariance matrix for these simulations are virtually identical.

there are two versions of this estimator that are interesting as a benchmark for the GMM estimators. One treats the standard deviation of the random effect (δ) as a known quantity. No $\sqrt{N}$- consistent GMM estimator for β can be more efficient than this ML-RE that attains the Cramér-Rao bound for β^0. The other (feasible) ML-RE that has to be used in practice is the one that treats δ as an additional coefficient to be estimated. Because of this additional coefficient, it is not clear whether this estimator is always more efficient than the GMM estimators (it attains the Cramér-Rao bound for (β', δ)'). The asymptotic covariance matrix of the GMM estimators for β is the same, irrespectively whether the instruments that may depend on the covariance structure of the errors are known or consistently estimated. However, for the ML-RE estimating δ implies a different asymptotic covariance matrix for β than treating δ as known.

Table 3 contains the results of these computations.[15] The overall efficiency ranking is as expected from the theoretical results presented in the previous section. Comparing the *pooled* estimator, that has the largest variance of the estimators presented, to the *sequential* estimator, we find that the efficiency gains from using the latter one are tiny. *GMM-SS* is always more efficient than *sequential*, however — as expected — the efficiency gains depend very much on the particular DGP. This feature is not true for the *Optimal GMM-IV*. *Optimal GMM-IV* denotes the limit for all the various versions of GMM estimators that are asymptotically efficient, such as the *Infeasible GMM-IV*, and the various versions of *WNP's* and *NP's*. To see the magnitude of the efficiency gains, it is useful to ask how many additional observations would be necessary so that inference — for example — with the *pooled* probit is as efficient as with an *Optimal GMM-IV*. For β^N and $T = 5$, a *pooled* probit analysis based on about 30% (DGP 1), 17% (DGP 2), 13% (DGP 3) or 32% (DGP 4) more observations would be as precise as an analysis based on an *Optimal GMM-IV*. These effficiency gains rise for $T = 10$, but the magnitude of the rise depends on the DGP's. The corresponding numbers are 32% (DGP 1), 21% (DGP 2), 15% (DGP 3), and 34% (DGP 4). It appears that such efficiency gains are well worth pursuing by using an asymptotically *Optimal GMM-IV*. Before we analyse whether these gains materialise in finite samples, a comparison of the GMM's with the ML-RE is informative. For β^N we find an efficiency loss of the *Optimal GMM-IV* of about 4% for $T = 5$ and about 11% for $T = 10$ compared to the (infeasible) ML-RE with known correlation coefficients. However, when making the more relevant

[15]Since in binary choice models identification is only up to scale, the ratio of estimated coefficients is also of interest.

comparison to the feasible ML-RE we find the efficiency loss to be non-existent. For $T = 5$, it appears that the *Optimal GMM-IV* is even more efficient than ML-RE. For $T = 10$, there is no clear ranking.

[insert Table 3 about here]

5.4 Finite Sample Results

Table 4 describes the measures for the accuracy of the estimates used in this section. They are the root mean square error and the bias of the estimated asymptotic standard errors of the coefficient estimates. $\hat{\beta}_r$ denotes the estimate of the true value β^0 in replication 'r', and $asstd(\hat{\beta}_r)$ denotes the estimated asymptotic standard error in replication 'r'. Since there may be concerns that the expectation and the variance of the estimates may not exist in finite samples for all estimators used, we also consider the median absolute error. Additionally, Table B.1 presents the upper and lower bounds as well as the width of the central 95% quantile of estimates based on the Monte Carlo simulations and on the asymptotic normal approximation using the average of the estimated asymptotic covariance matrices of the coefficients. For the sake of brevity, the statistics related to the constant term are omitted.

[insert Table 4 about here]

For the DGP with an AR(1) process combined with random effects (Tables 5 and B.1), several GMM-estimators with nonparametric estimation of the covariance matrix are computed for $N = 100, 400$.[16] The results do not reveal much difference within the groups of estimators *NP* and *WNP*, however, there are substantial differences between these groups (although not reported in the tables, all estimates are almost unbiased). The estimators with scaled moments (*WNP*) have lower root mean squared errors (RMSE) and median absolute error (MAE). *WNP* is sometimes observed with a somewhat larger bias of the asymptotic standard errors. Increasing the sample size N, the performance of the estimators is generally improving, becoming closer to that of the asymptotically optimal GMM-estimates given by the benchmark of the *infeasible GMM-IV*. Increasing the time-dimension T generally improves the results with respect to the RMSE, the MAE, and the bias of asymptotic standard errors. A notable exception on the latter criterion is *WNP-indiv-quadr*.

[16]The same range of estimators has been computed for the case of independent errors. Since the obtained results are qualitatively the same as for the AR(1) process they are omitted from the tables. The results are available from the authors on request.

Compared to choosing between *NP* and *WNP* the ranking according to the choice of the weight function for the k-NN approach, *uniform* or *quadr*atic, is inconclusive with the exception of *WNP-indiv-quadr*. The results for the triangular weights are omitted from the tables since they turned out to be nearly indistinguishable from those for *quadr*atic weights. Furthermore, it does not seem to matter if the distance is measured *indiv*idually or *joint*ly over the indices $x_{ti}\beta$, although the results of the latter are more stable in the small sample.

For each replication of the Monte Carlo study the smoothing parameter k is chosen via cross-validation. The distributions of the choice of k are depicted in Tables B.2 to B.5. For DGP 1 there is a tendency to choose a relatively large k (see Table B.2). This tendency is slightly stronger for the *joint* than for the *individual* measure of distance. The same is true when increasing the time periods from 5 to 10. Both cases represent an increase in the dimension of the nonparametric estimation. *WNP* has an even stronger tendency to choose a large k than *NP*. An intuitive explanation is that the diagonal elements of the covariance matrix for *WNP* are estimated without bias, because they do not vary across 'i'. With increasing k the variance of the estimation is reduced. If these elements have strong weights in the cross-validation, the effect of decreasing the variance by large k might dominate the bias-variance trade-off important for the off-diagonal elements, and hence, a large (or the largest) k minimizes the cross-validation function. However, replacing alternatively the diagonal elements of the covariance matrix by their parametrically estimated values and using only the off-diagonal elements for the cross-validation (*WNP-joint-uniform-no.d.*) does not seem to matter for $N = 400$. For $N = 100$ the results are mixed: for $T = 5$ it is the best for all criteria, but for $T = 10$ several *WNP*-estimators perform better. Introducing the covariance matrix of the linear indices $S_{x\beta}^{-1}$ as the weight matrix for measuring the distance does not seem to bring clear-cut efficiency gains. This is supported by the results for independent errors.[17] Since there does not appear to be any significant difference between using $S_{x\beta}^{-1}$ or its main diagonal, the latter is omitted from the tables.

Comparing the results of Table 5 and Table B.1 we find that there are no substantial qualitative differences. Furthermore, the confidence bounds are symmetric around the true value, which implies that, at least for its tail, there is no concern about a severely asymmetric distribution.

[17] The results are available on request from the authors

Given these results, we choose within the class of *WNP* estimators the estimator with uniform weights and joint distance measure (*WNP-joint-uniform*) as the preferred version, because it has the simplest form of weights, it seems to be more robust with respect to the asymptotic standard errors than for instance the *WNP-indiv-quadr*, and saves computation time compared to an estimator with an individual distance measure.

Turning now to the fully parametric estimators, the following results are obtained: for the DGPs that have several forms of correlation of the error terms, *ML-RE* and *GMM-SS* show best results with respect to the RMSE and MAE. For the (nonreported) case of independent errors the *small sigma* method is already in small samples as good as *pooled* probit, although the latter is the maximum likelihood estimator for this DGP. In some cases *ML-RE* shows a relatively high downward bias of the asymptotic standard errors for $N = 100$ (see for instance Table 5). The calculation of *ML-RE* is time-consuming compared to the other methods, and often convergence cannot be reached, especially for $T = 10$.

Chamberlain's *sequential* estimator produces quite large RMSE and MAE compared to the other parametric methods for the sample size of $N = 100$. Although it improves with increasing N, it performs still worse than *ML-RE* and *GMM-SS*. A striking feature is the large bias of the asymptotic standard errors for $N = 100$, which is even larger for $T = 10$ than for $T = 5$. Thus, *sequential* needs quite a large sample size N in order to obtain good results. The relatively bad performance of *sequential* is also illustrated in Figures 1 and 2 where kernel plots of the distribution of the various estimators based on DGP 1 are depicted.[18] Compared to the other estimators the distribution of *sequential* turns out to be flatter and right-skewed, even stronger for $T = 10$ than for $T = 5$. Note that the asymptotic efficiency gains for this estimator compared to *pooled*, for example, depend on the accurate estimation of the covariance matrix in the first step estimates, which then has to be inverted. In our Monte Carlo example this matrix has the dimension 15 for $T = 5$ and 30 for $T = 10$. However, the asymptotic standard errors ignore the fact that this inverse may exhibit considerable variability in small samples.

Comparing now *GMM-SS* and *WNP-joint-uniform*, one observes that for the DGPs presented in Tables 5 and 6 (and also for independent errors) *GMM-SS* is better than *WNP-joint-uniform* for $N = 100$. For $N = 400$ and $N = 1600$ the differences in RMSE and bias of asymptotic standard errors are very small and the

[18] We present kernel plots for β_N, but only for DGP 1 and $N = 100, 400$. For $N = 1600$, the sample distributions are very close to the asymptotic ones. The results for the other coefficients as well as for the other DGPs are qualitatively the same as for DGP 1.

ranking is inconclusive. The latter result is supported by the corresponding kernel plots in Figures 1 and 2 except for $N = 100$ and $T = 5$ where the distribution of *GMM-SS* is left-skewed compared to that of *WNP-joint-uniform* with a flat region at the top of the distribution. For the other cases the distributions of the two estimators are very close to each other.

When considering covariance structures that are very different from a random effects structure implying equicorrelated errors, as presented by the MA(1) process or the AR(1) process with alternating signs of the correlation over time (Tables 7 and 8), *WNP-joint-uniform* is clearly superior to *GMM-SS* (as it is asymptotically, see Table 3).

Thus, since in real-world applications the true DGP is unknown, *WNP-joint-uniform* with nonparametric estimation of the covariance matrix can be recommended for applications.

[insert Tables 5 to 8 and Figures 1 and 2 about here]

6 Application

Now we come back to the empirical example that motivated our discussion. The main hypothesis to be tested is that imports and foreign direct investment (FDI) have positive effects on the innovative activity of domestic firms. The firm-level data have been collected by the Ifo-Institute, Munich ('Ifo-Konjunkturtest') and have been merged with official statistics from the German Statistical Yearbooks. The binary dependent variable indicates whether a firm reports having realized a product innovation within the last year or not. The independent variables refer to the market structure, in particular the market size of the industry **ln(sales)**, the shares of imports and FDI in the supply on the domestic market (**import share** and **FDI-share**), the **productivity** as a measure of the competitiveness of the industry as well as two variables indicating whether a firm belongs to the **raw materials** or to the **investment goods** industry. Moreover, including the **relative firm size** allows to take account of the innovation — firm size relation often discussed in the literature. Hence, all variables with exception of the firm size are measured at the industry-level (for descriptive statistics see Appendix C).

The estimators applied to the example include the one used in Bertschek (1995), (*sequential*), the simplest one (*pooled*), the maximum likelihood estimator under the assumption of random effects with five evaluation points as used in the Monte Carlo study (*ML-RE 5*) and with 10 evaluation points (*ML-RE 10*), the best parametric

GMM estimator (*GMM-SS*) and finally, the simplest (*WNP-joint-uniform*) of the estimators with weighted nonparametric estimation of the covariance. Furthermore, we compute two versions of the asymptotic t-values for the pooled estimator. The first (denoted by *t-val*) ignore the possible correlations over time. They are the same that would be obtained by using a standard software package for cross-section probit estimation. The second are computed using the correct GMM-formula as given in eq. (16). The latter are comparable to the covariance matrices used in the Monte Carlo study.

The estimation results are presented in Tables 9 and 10. Let us start by comparing the outcomes implied by the two different ways to compute the covariance matrices of *pooled*: not surprisingly, the t-values ignoring the intertemporal correlations are generally larger than the t-values taking account of these correlations.

The results of the various parametric estimators are quite similar and lead to the same conclusions (first part of Table 9). Both **import share** and **FDI-share** have positive and significant effects on product innovative activity. The **firm size** variable has a positive and significant impact (except for *sequential*) — a result that supports the Schumpeterian hypothesis that large firms are more innovative than small firms. The **productivity** coefficient is insignificant in most estimations. The estimates for *ML-RE 5* and *ML-RE 10* are very close to each other except of the **productivity** that is significant only for *ML-RE 10*. Increasing the number of evaluation points to 20 produces almost the same results as for 10 evaluation points. Hence, they are omitted from the table.

Turning now to the results of *WNP-joint-uniform* (second part of Table 9) one can find that the estimates as well as the t-values are quite robust for $k = 100, 263, 880, 1270$. For $k = 263$ the cross-validation function (Figure 3) has a local minimum whereas its global minimum is reached for $k = 880$. For all these k's the **productivity** coefficient is negative and significant. In contrast to the results of the sequential estimator the **firm size** always keeps its positive and significant coefficient. For $k = 20$ and $k = 50$ the dummy for the **raw materials** industry has an insignificant coefficient. In the first case all coefficients and t-values differ considerably from those of the other k's, however, the value of the cross-validation function indicates that $k = 20$ is far from being *optimal*.

A look at the estimated standard errors (Table 10) reveals again that *pooled* when ignoring possible correlations of the error over time generally leads to downward biased standard errors and thus to upward biased t-values. The standard errors of *GMM-SS* and *sequential* are in general lower than those of *pooled*. The standard errors of *WNP-joint-uniform* differ only slightly over $k = 100$ to $k = 1270$. The large values for $k = 20$ accompanied by a very large value for the cross-validation function confirm the inaccuracy of the estimates for this size of undersmoothing. For $k = 880$ the standard errors are the smallest for all variables.

Hence, *WNP-joint-uniform* clearly dominates the other estimators with respect to its efficiency for most coefficients. There are efficiency gains compared to *pooled* probit, *GMM-SS* and *sequential* concerning the variables firm size and productivity. The standard errors for the FDI-share coefficient decrease compared to *pooled* and *GMM-SS*. These results are in particular important with respect to the variable **productivity**. While its coefficient is negative and insignificant for all parametric estimators (except *ML-RE 10*), it now turns out to be significant allowing to draw more reliable conclusions. Firms belonging to more productive industries put less effort in product innovative activity. At first glance, this result seems surprising since one expects that higher productivity implies more innovative activity. However, a closer look at the data reveals that industries with extremely high levels of labour-productivity are those of the raw materials industry. These industries produce with a high capital-intensity. Since raw materials such as non-ferrous metals or paper generally cannot be changed much — as compared to consumer goods, for example — by innovations the negative sign of the productivity coefficient seems plausible.

The estimations have been performed on a personal computer with a Pentium 100MHz processor. The computation times are: 30 seconds for *pooled* probit, 10 minutes for *GMM-SS* and for *WNP-joint-uniform* and 15 minutes for *ML-RE 10*. The GAUSS-program is available on request from the authors.

[insert Tables 9 and 10 and Figure 3 about here]

7 Conclusions

The paper shows that often-used and conveniently computable estimators, such as pooled probit, Chamberlain's (1980, 1984) sequential estimators or several variants of GMM estimators, belong to a class of GMM estimators using the same conditional moment restrictions. An asymptotic efficiency ranking of these and other related estimators is established. Additionally, using nonparametric methods a feasible estimator that is asymptotically efficient in that class but retains most of the simplicity of the other class members is suggested. This estimator — fast to compute and efficient — represents an attractive alternative to simulation or restricted ML methods. The Monte Carlo study shows that this estimator has good small sample properties and only a tiny efficiency loss compared to ML occurs. Finally, when analysing the determinants of product innovation of 1270 German firms, the suggested nonparametric estimators perform very well in practice. The efficiency gains, compared to the other estimators, turn out to be important for the economic interpretation of the estimation results.

One aim of future research should be to analyse if the same results can be obtained for other nonlinear models such as tobit. Another topic of interest might be the comparison of other nonparametric methods with the k-nearest neighbor approach within this context as well as to consider other approaches for choosing the smoothing parameter.

Appendix A: Lemma

The following lemma shows a useful invariance property of GMM estimators based on conditional moments. It shows that we can consider functions of X alone either as part of the instrument matrix A or as part of the conditional moment function M without changing the asymptotic distribution of the estimator.

Lemma: Asymptotic Invariance

Let H be a square matrix of full rank and let $\tilde{\beta}_N$ denote a consistent and square root normal estimator of β. Define a new estimator $\hat{\beta}^+$ based on minimizing the following modified conditional moments and modified instrument. All quantities are defined as in section 4.1, but those related to $\hat{\beta}^+_M$ are marked with a '+'.

$$A^+(x_i) = A(x_i; \tilde{\beta}_N)[H(x_i; \tilde{\beta}_N)]^{-1}, \ \ \sqrt{N}(\tilde{\beta}_N - \beta^0) \longrightarrow N(0, \Gamma).$$

$$M^+(z_i; \beta) = H(x_i; \beta)M(z_i; \beta)$$

$$E\left[M^+(Z; \beta^0)|X = x_i\right] = H(x_i; \beta^0)E\left[M(Z; \beta^0)|X = x_i\right] = 0$$

$$g_N^+(z; \beta) := \frac{1}{N}\sum_{i=1}^{N} A^+(x_i)M^+(z_i; \beta),$$

$$\hat{\beta}_N^+ = \arg\min_{\beta\in B} g_N^+(z; \beta)'Pg_N^+(z; \beta),$$

$$\sqrt{N}(\hat{\beta}_N^+ - \beta^0) \longrightarrow N(0, \Lambda^+),$$

$$\Lambda^+ = (G^{+\prime}PG^+)^{-1}G^{+\prime}PV^+PG^+(G^{+\prime}PG^+)^{-1},$$

$$G^+ = E\left[A^+\frac{\partial M^+}{\partial \beta'}(Z; \beta^0)\right],$$

$$V^+ = E(A^+M^+M^{+\prime}A^{+\prime}).$$

Now compare this estimator to the one consisting of instruments A and moments M with covariance matrix Λ. The claim is that $\Lambda = \Lambda^+$ so that the estimators are asymptotically identical. The proof proceeds in two steps. First, note that the respective weighting matrices are identical:

$$V^+ = E\left\{A[H(X; \beta^0)]^{-1}H(X; \beta^0)M(Z; \beta^0)M'H'H^{-1\prime}A'\right\} = V.$$

Second, the expected values of the derivatives of the moments are also the same.

$$G^{+} = E\left\{A(X)[H(X;\beta^0)]^{-1}\frac{\partial(HM)}{\partial\beta'}(Z;\beta^0)\right\} = E\left\{AE\left[\frac{\partial(M)}{\partial\beta'}|X = x_i\right]\right\} = G.$$

The last equality can be verified by considering a typical element of

$$\begin{aligned} E\left[\left(h_{\tau t}\frac{\partial m_t}{\partial\beta_k} + m_t\frac{\partial h_{\tau t}}{\partial\beta_k}\right)|X = x_i\right] &= E\left(h_{\tau t}\frac{\partial m_t}{\partial\beta_k}|X = x_i\right) + E(m_t|X = x_i)\frac{\partial h_{\tau t}}{\partial\beta_k} \\ &= E\left(h_{\tau t}\frac{\partial m_t}{\partial\beta_k}|X = x_i\right) \\ &= h_{\tau t}E\left(\frac{\partial m_t}{\partial\beta_k}|X = x_i\right) \end{aligned}$$

$$E\left[\frac{\partial(HM)}{\partial\beta'}|X = x_i\right] = HE\left[\frac{\partial M}{\partial\beta'}|X = x_i\right] \ \Rightarrow G^* = G \ \Rightarrow \Lambda^* = \Lambda. \qquad \text{q.e.d.}$$

Appendix B: Additional Monte Carlo Results

In Table B.1 the bounds ($\beta^{2.5}, \beta^{97.5}$) and the widths ($W$) of the 95% confidence intervals for one of the data generating processes (DGP 1, Table 5) are given. Results under the heading 'Monte Carlo' are computed using the Monte Carlo distributions of the estimates, whereas the heading 'Asymptotic-W' refers to the widths computed using normality and the estimated asymptotic covariance matrices. The latter widths are averaged over the sample.

[insert Tables B.1 to B.5 about here]

Appendix C: Descriptive Statistics

[insert Table C about here]

References

Abramowitz, M. and Stegun, I. A. (1966). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, Dover Publications, New York.

Avery, R., Hansen, L. and Hotz, V. (1983). Multiperiod probit models and orthogonality condition estimation, *International Economic Review* **24**: 21–35.

Bertschek, I. (1995). Product and process innovation as a response to increasing imports and foreign direct investment, *Journal of Industrial Economics* **43**(4): 341–357.

Breitung, J. and Lechner, M. (1997). Estimation de modèles non linéaires sur données de panel par la méthode des moments généralisés, *Economie et Prévision,* forthcoming. (The English version is available as SFB-discussion paper 67, Humboldt-University, Berlin, 1995).

Butler, I. and Moffitt, R. (1982). A computationally efficient quadrature procedure for the one-factor multinomial probit model, *Econometrica* **50**: 761–764.

Carroll, R. J. (1982). Adapting for heteroscedasticity in linear models, *The Annals of Statistics* **10**: 1224–1233.

Chamberlain, G. (1980). Analysis of covariance with qualitative data, *Review of Economic Studies* **47**: 225–238.

Chamberlain, G. (1984). Panel data, *in* Z. Griliches and M. Intrilligator (eds), *Handbook of Econometrics*, Vol. 2, chap. 22, Elsevier Science Publishers, New York.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions, *Journal of Econometrics* **34**: 305–334.

Guilkey, D. and Murphy, J. (1993). Estimation and testing in the random effects probit model, *Journal of Econometrics* **59**: 301–317.

Hajivassiliou, V. A. and Ruud, P. A. (1994). Classical estimation methods for LDV models using simulation, *in* R. F. Engle and D. L. McFadden (eds), *Handbook of Econometrics*, Vol. 4, chap. 40, Elsevier Science Publishers, New York.

Hansen, L. (1982). Large sample properties of generalized methods of moments estimators, *Econometrica* **50**: 1029–1055.

Härdle, W. (1990). *Applied Nonparametric Regression*, Vol. 19 of *Econometric Society Monographs*, Cambridge University Press, Cambridge.

Heckman, J. (1981). Statistical models for discrete panel data, *in* C. Manski and D. McFadden (eds), *Structural Analysis of Discrete Data*, MIT Press, Cambridge.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*, Vol. 3 of *Econometric Society Monographs*, Cambridge University Press, Cambridge.

Nerlove, M. (1971). Further evidence on the estimation of dynamic economic relations from a time series of cross sections, *Econometrica* **39**: 359–383.

Newey, W. K. (1990). Efficient instrumental variables estimation of nonlinear models, *Econometrica* **59**: 809–837.

Newey, W. K. (1993). Efficient estimation of models with conditional moment restrictions, *in* G. Maddala, C. Rao and H. Vinod (eds), *Handbook of Statistics*, Vol. 11, North-Holland, Amsterdam, chapter 16.

Robinson, P. M. (1982). On the asymptotic properties of estimators of models containing limited dependent variables, *Econometrica* **50**: 27–41.

Robinson, P. M. (1987). Asymptotically efficient estimation in the presence of heteroscedasticity of unknown form, *Econometrica* **55**: 875–891.

Stone, C. J. (1977). Consistent nonparametric regression, *The Annals of Statistics* **5**: 595–645.

Stroud, A. and Secrest, D. (1966). *Gaussian Quadrature Formulas*, Englewood Cliffs: Prentice Hall.

White, H. (1982). Maximum likelihood estimation of misspecified models, *Econometrica* **50**: 1–25.

Table 9: Innovation probit: Estimated coefficients and t-values

	pooled			ML-RE 5		ML-RE 10		GMM-SS		sequential	
variable	coef	*t-val*	t-val	coef	t-val	coef	t-val	coef	t-val	coef	t-val
intercept	**-1.96**	*8.2*	5.2	**-2.91**	4.6	**-2.84**	5.0	**-1.79**	4.8	**-1.80**	5.1
ln(sales)	**0.18**	*7.2*	4.8	**0.25**	4.1	**0.25**	4.6	**0.16**	4.4	**0.15**	4.4
rel. firm size	**1.07**	*5.2*	3.5	**1.56**	2.5	**1.52**	3.9	**0.88**	3.5	0.26	1.4
import share	**1.13**	*7.4*	4.6	**1.77**	4.8	**1.78**	4.8	**1.17**	4.9	**1.27**	5.4
FDI-share	**2.85**	*6.1*	4.2	**3.77**	3.7	**3.65**	3.8	**2.54**	3.9	**2.52**	3.9
productivity	-2.34	*2.1*	1.8	-2.21	1.5	**-2.30**	2.0	-1.50	1.8	-0.43	0.4
raw materials	**-0.28**	*2.9*	2.1	**-0.48**	2.6	**-0.48**	2.7	**-0.33**	2.8	**-0.28**	2.3
investment	**0.19**	*4.7*	3.0	**0.33**	3.5	**0.33**	3.4	**0.21**	3.3	**0.21**	3.3

	WNP-joint-uniform											
	k=20		k=50		k=100		k=263		k=880		k=1270	
variable	coef	t-val	coef	t-val	coef	t-val	coef	t-val	coef	t-val	coef	t-val
intercept	-3.82	0.5	**-1.70**	4.4	**-1.81**	4.8	**-1.74**	4.7	**-1.74**	4.8	**-1.79**	4.9
ln(sales)	0.44	0.7	**0.16**	4.4	**0.16**	4.6	**0.15**	4.4	**0.15**	4.5	**0.16**	4.6
rel. firm size	0.005	0.03	**0.91**	4.2	**0.97**	4.8	**1.00**	4.9	**0.95**	4.7	**0.98**	4.8
import share	-3.12	0.7	**1.02**	4.1	**1.10**	4.4	**1.13**	4.7	**1.14**	4.8	**1.15**	4.8
FDI-share	**14.09**	2.9	**2.64**	4.2	**2.59**	4.4	**2.56**	4.3	**2.59**	4.4	**2.57**	4.4
productivity	**-7.32**	5.2	**-2.48**	2.6	**-1.76**	2.0	**-1.89**	2.2	**-1.91**	2.3	**-1.92**	2.3
raw materials	0.50	0.4	-0.19	1.5	**-0.27**	2.2	**-0.28**	2.4	**-0.28**	2.4	**-0.29**	2.5
investment	0.10	0.2	**0.22**	3.4	**0.24**	3.7	**0.22**	3.4	**0.21**	3.4	**0.21**	3.3
cv-value	$99*10^7$		21.00		17.36		15.51		14.80		15.12	

Note: dependent variable: product innovation, N=1270, T=5. *t-val:* t-values of pooled probit assuming independent errors over time. Bold letters if t-values are larger than 1.96.

Table 10: Innovation probit: Estimated standard errors

	pooled		SS	seq.	WNP-joint-uniform with k =					
variable					20	50	100	263	880	1270
intercept	*0.21*	0.38	0.37	0.35	7.30	0.38	0.37	0.37	0.37	0.37
ln(sales)	*0.025*	0.037	0.035	0.034	0.61	0.036	0.035	0.034	0.034	0.034
firm size	*0.21*	0.31	0.25	0.19	0.16	0.22	0.20	0.20	0.20	0.20
imp. share	*0.15*	0.24	0.24	0.24	4.63	0.25	0.25	0.24	0.24	0.24
FDI-share	*0.47*	0.68	0.65	0.64	4.94	0.63	0.59	0.60	0.59	0.59
product.	*1.10*	1.32	0.84	1.05	1.39	0.96	0.87	0.84	0.82	0.83
raw mat.	*0.097*	0.13	0.12	0.12	1.34	0.12	0.12	0.12	0.12	0.12
invest.	*0.040*	0.063	0.063	0.062	0.63	0.065	0.064	0.063	0.063	0.063

Note: see Table 8. ML-RE uses different error variance normalization, therefore, it is omitted from the table.

Figure 1: Kernel plots of the estimators for β_N (DGP 1, T=5)

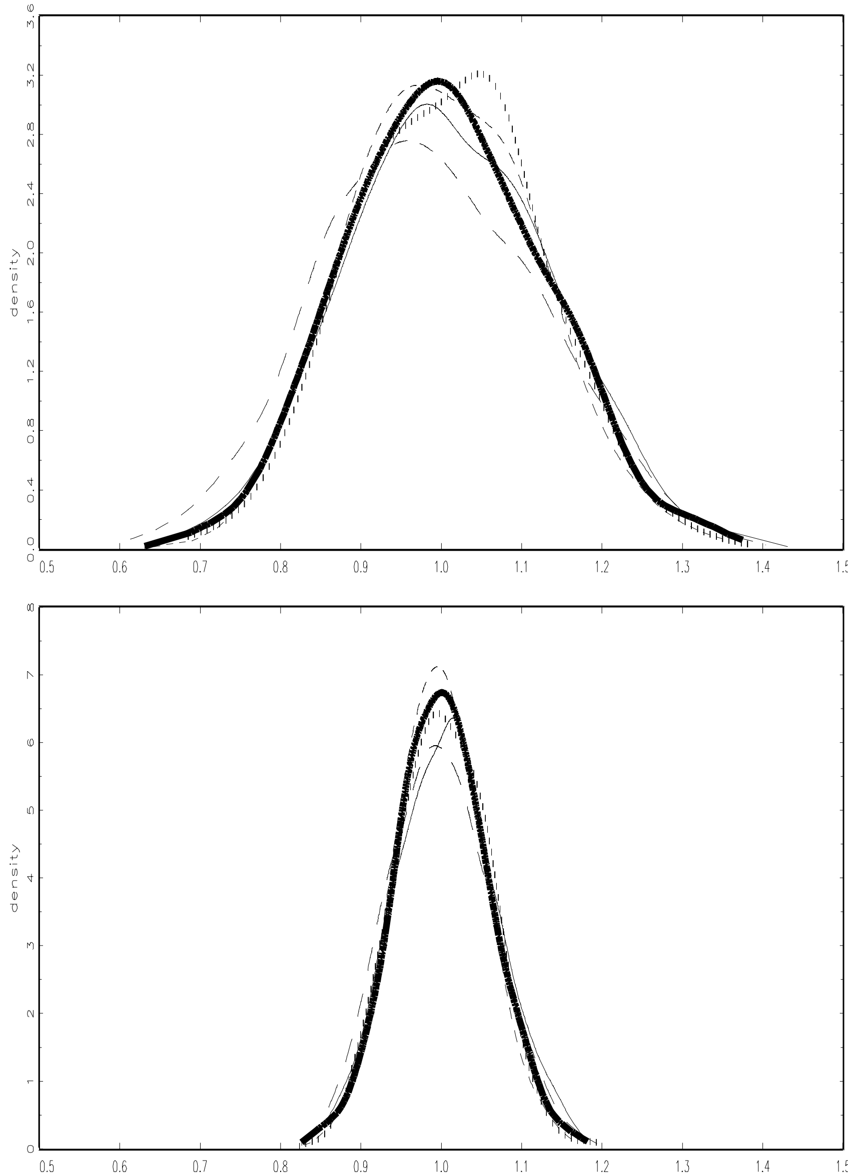
Note: upper window: N=100, lower window: N=400. WNP-joint-uniform (thick solid line), ML-RE (short dashes), pooled (thin solid line), sequential (long dashes), GMM-SS (dots). The bandwidths are 0.033 and 0.016 using the Gaussian kernel.

Figure 2: Kernel plots of the estimators for β_N (DGP 1, T=10)

Note: upper window: N=100, lower window: N=400. WNP-joint-uniform (thick solid line), pooled (thin solid line), sequential (long dashes), GMM-SS (dots). The bandwidths are 0.029 and 0.014 using the Gaussian kernel.

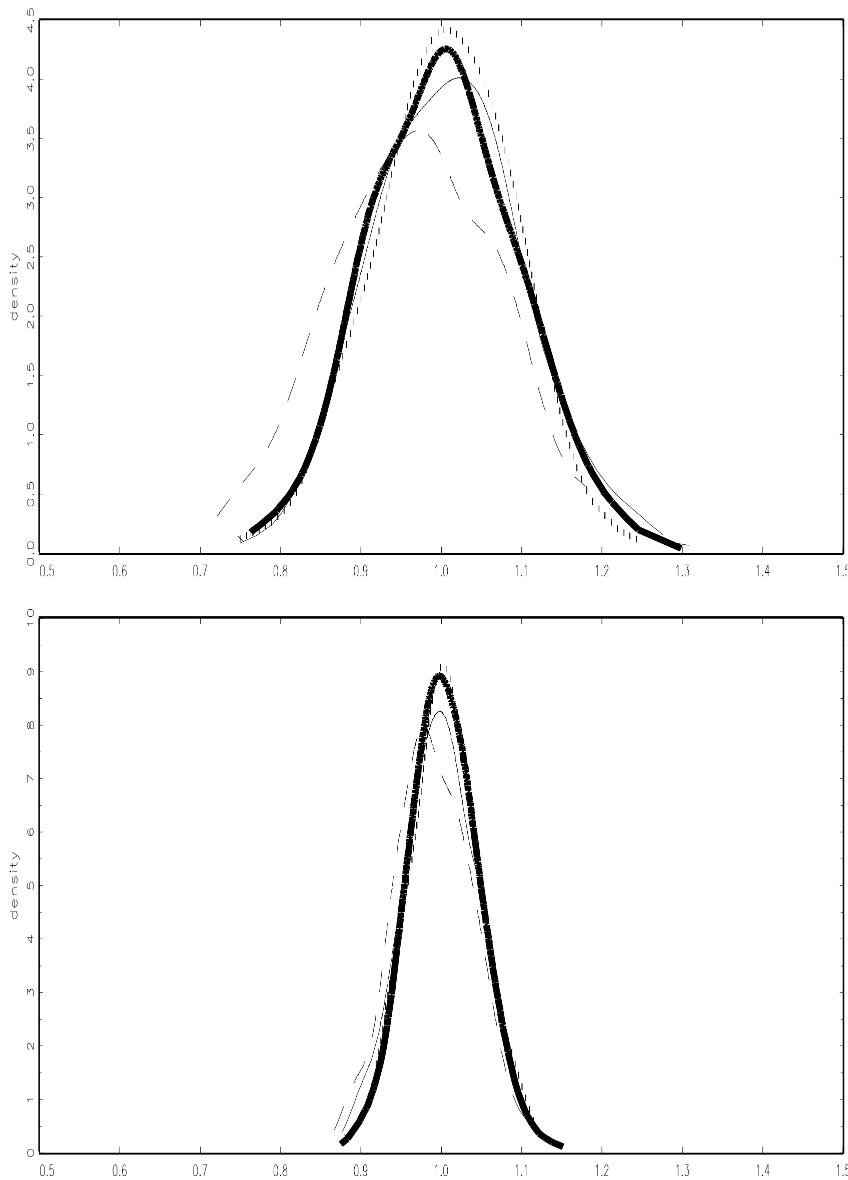
Figure 3: Cross-validation function for innovation probit estimated with WNP-joint-uniform

Figure 1: Kernel plots of the estimators for β_N (DGP 1, T=5)



Note: upper window: N=100, lower window: N=400. WNP-joint-uniform (thick solid line), ML-RE (short dashes), pooled (thin solid line), sequential (long dashes), GMM-SS (dots). The bandwidths are 0.033 and 0.016 using the Gaussian kernel.

Figure 2: Kernel plots of the estimators for β_N (DGP 1, T=10)



Note: upper window: N=100, lower window: N=400. WNP-joint-uniform (thick solid line), pooled (thin solid line), sequential (long dashes), GMM-SS (dots). The bandwidths are 0.029 and 0.014 using the Gaussian kernel.

Figure 3: Cross-validation function for innovation probit estimated with WNP-joint-uniform

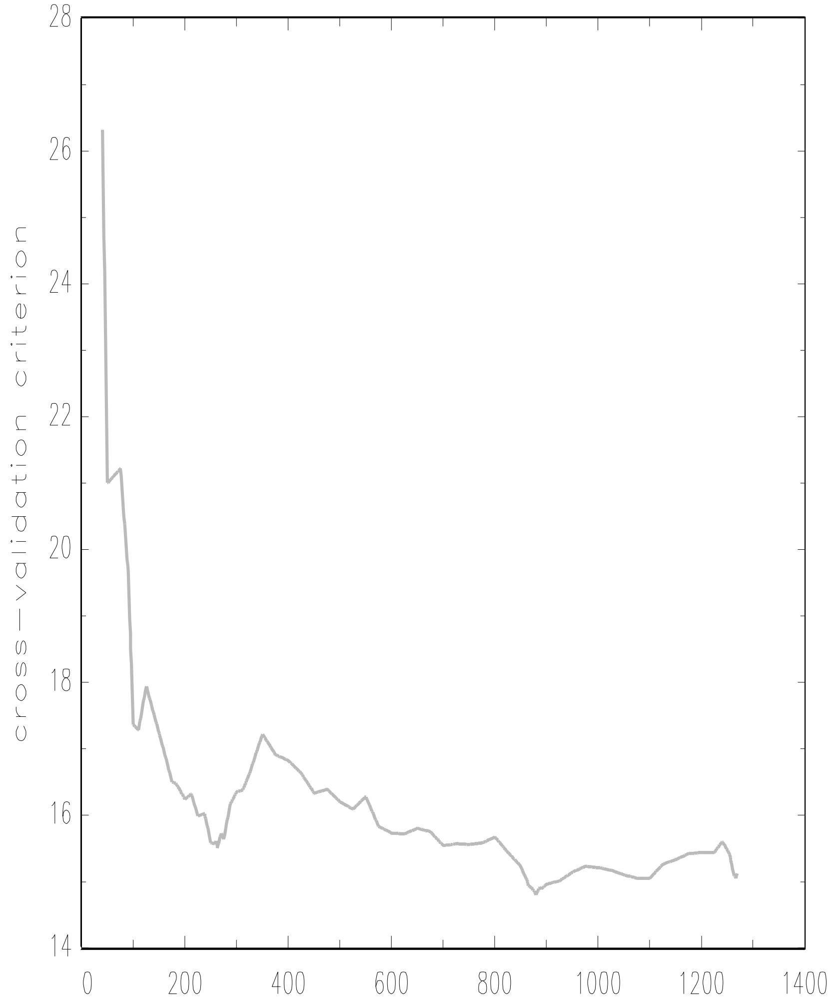


Table 1: Summary statistics for different DGP's of $\mu_{ti} = x_{ti}\beta^0$

	mean(μ_{ti})		$std(\mu_{ti})$	$corr(\mu_{ti},\mu_{si})$	
	min	max	mean	min	max
T = 5	-0.2	0.2	0.8	0.1	0.3
T = 10	-0.2	0.6	0.8	0.0	0.3

Note: $\beta^C = -0.75, \beta^D = \beta^N = 1, \gamma^x = 0.5, \gamma^t = 0.05$.

Table 2: Summary statistics for different DGP's of u

		$std(u_{ti})$		$corr(u_{ti}, u_{si})$	
	DGP	min	max	min	max
$\delta = \sqrt{0.5}, \alpha = 0, \sigma_t = \sqrt{0.5}$	2	1.0	1.0	0.5	0.5
T = 5					
$\delta = 0.2, \alpha = 0.8, \sigma_t = 0.5$	1	0.9	0.9	0.4	0.8
$\delta = 0, \alpha = 0, \sigma_t = \sqrt{2}, u_{ti} = 0.5(\varepsilon_{ti} + \varepsilon_{t-1i})$	3	1.0	1.0	0	0.5
$\delta = 0, \alpha = -0.8, \sigma_t = 0.6$	4	1.0	1.0	-0.8	0.6
T = 10					
$\delta = 0.2, \alpha = 0.8, \sigma_t = 0.5$	1	0.9	0.9	0.2	0.8
$\delta = 0, \alpha = 0, \sigma_t = \sqrt{2}, u_{ti} = 0.5(\varepsilon_{ti} + \varepsilon_{t-1i})$	3	1.0	1.0	0	0.5
$\delta = 0, \alpha = -0.8, \sigma_t = 0.6$	4	1.0	1.0	-0.8	0.6

Note: See note below Table 1.

Table 3: Asymptotic standard errors (x 10)

	$T=5$				$T=10$			
	β_c	β_D	β_N	β_D/β_N	β_c	β_D	β_N	β_D/β_N
DGP 1	AR (1) ($\alpha=0.8$) and random effects ($\delta=0.2$)							
Pooled	13.40	13.74	14.69	13.75	11.21	10.08	10.92	9.86
Sequential	13.36	13.65	14.68	13.70	11.13	10.02	10.91	9.83
GMM-SS	12.94	12.35	13.31	11.57	10.87	9.30	10.07	8.61
Optimal GMM-IV	12.72	12.00	12.88	10.94	10.57	8.76	9.51	7.67
DGP 2	random effects ($\alpha=0, \delta=\sqrt{0.5}$)							
ML-RE – known δ	11.40	10.51	12.06	13.59	9.71	7.01	8.55	9.18
ML-RE – unknown δ	12.12	11.88	13.26	13.59	9.99	7.58	9.02	9.18
Pooled	11.89	11.90	13.31	15.38	10.43	8.63	9.90	10.83
Sequential	11.88	11.37	13.31	15.34	10.41	8.50	9.90	10.72
GMM-SS	11.57	10.86	12.36	13.75	10.07	7.65	9.05	9.26
Optimal GMM-IV	11.53	10.80	12.31	13.68	10.04	7.60	9.00	9.19
DGP 3	MA(1)							
Pooled	10.63	11.33	13.07	15.16	8.13	7.91	9.40	10.72
Sequential	10.60	11.31	13.06	15.15	8.11	7.90	9.40	10.71
GMM-SS	10.56	11.20	12.85	14.89	8.10	7.89	9.32	10.64
Optimal GMM-IV	10.31	10.85	12.28	14.15	7.87	7.57	8.77	9.94
DGP 4	AR (1) ($\alpha=-0.8, \delta=0$)							
Pooled	9.42	10.40	13.59	13.90	7.13	7.40	9.94	9.76
Sequential	9.34	10.38	13.59	13.89	7.09	7.37	9.93	9.97
GMM-SS	9.31	10.45	13.33	13.64	7.08	7.47	9.84	9.70
Optimal GMM-IV	8.52	9.68	11.82	11.59	6.42	6.93	8.59	8.04

Note: Asymptotic standard errors of $\sqrt{N}(\hat\beta-\beta^0)$. Asymptotic variance of $\sqrt{N}(\hat\beta_D/\hat\beta_N-\beta_D^0/\beta_N^0)$ is computed by the delta method.

Table 4: Measures of accuracy used in the Monte Carlo study

Root mean square error x 10	$[\frac{100}{R}\sum_{r=1}^{R}(\hat\beta_r-\beta^\circ)^2]^{1/2}$
Median of absolute error x 10	median$_r$ $\lvert\hat\beta_r-\beta^\circ\rvert$
Bias of asymptotic standard error in %	$\frac{100}{R}\sum_{r=1}^{R}\frac{as\,std(\hat\beta_r)-std\,err(\hat\beta)}{std\,err(\hat\beta)}$

Note: $\hat\beta_r$ and $asstd(\hat\beta_r)$ denotes the cofficient estimate and the estimated asymptotic standard error obtained in replication *r*. $std(\hat\beta)$ denotes the empirical standard error of the estimated cofficients in the Monte Carlo study.

Table 5: Simulation results for AR (1) ($\alpha = 0.8$) and random effects ($\delta = 0.2$) (DGP 1)

	root MSE x10			median absolute error x 10			bias of as. std. err. in %	
	β_D	β_N	β_D/β_N	β_D	β_N	β_D/β_N	β_D	β_N
N = 100				$T = 5$, 1000 replications				
Infeasible GMM-IV	1.18	1.30	1.17	0.78	0.89	0.77	2.7	-0.8
ML-RE	1.26	1.34	1.19	0.81	0.96	0.79	-4.6	-3.7
Pooled	1.41	1.49	1.51	0.95	1.00	0.99	-3.1	-2.0
Sequential	1.51	1.63	1.66	1.03	1.16	1.04	-17.6	-15.1
GMM-SS	1.21	1.35	1.24	0.80	0.95	0.80	2.2	-1.4
NP-joint-uniform	1.44	1.48	1.24	0.89	1.05	0.82	-2.1	-0.5
NP-indiv-uniform	1.44	1.48	1.24	0.90	1.02	0.84	-2.5	-1.0
NP-joint-quadr.	1.44	1.47	1.24	0.90	1.04	0.83	-1.8	0.4
NP-indiv-quadr.	1.41	1.47	1.23	0.90	1.04	0.83	-1.1	-0.7
NP-joint -unif.- $S_{x\beta}^{-1}$	1.43	1.47	1.24	0.91	1.02	0.81	-1.5	0.2
WNP-joint-uniform	1.31	1.42	1.24	0.87	0.97	0.83	-2.2	-3.1
WNP-indiv-uniform	1.38	1.47	1.34	0.88	0.98	0.87	-4.6	-3.7
WNP-joint-quadr.	1.30	1.42	1.24	0.86	0.96	0.82	-1.2	-2.6
WNP-indiv-quadr.	1.32	1.48	1.29	0.88	1.02	0.82	-1.0	-4.1
WNP-joint-unif.- $S_{x\beta}^{-1}$	1.30	1.41	1.24	0.87	0.99	0.83	-1.5	-2.5
WNP-joint-unif.-no d.	1.29	1.37	1.21	0.86	0.86	0.78	-0.4	1.7
N = 400				$T = 5$, 1000 replications				
Infeasible GMM-IV	0.61	0.66	0.56	0.41	0.45	0.38	-0.2	-1.3
ML-RE	0.61	0.64	0.56	0.40	0.44	0.39	4.0	2.2
Pooled	0.70	0.73	0.69	0.47	0.51	0.47	-1.3	0.9
Sequential	0.72	0.75	0.70	0.47	0.51	0.47	-6.5	-3.0
GMM-SS	0.63	0.67	0.60	0.42	0.47	0.39	-2.6	-0.3
NP-joint-uniform	0.67	0.71	0.58	0.44	0.46	0.38	2.0	1.2
NP-indiv-uniform	0.66	0.71	0.58	0.44	0.46	0.39	2.4	0.9
NP-joint-quadr.	0.67	0.71	0.58	0.44	0.49	0.39	1.2	0.4
NP-indiv-quadr.	0.66	0.70	0.58	0.43	0.45	0.38	2.6	1.0
NP-joint-unif.- $S_{x\beta}^{-1}$	0.67	0.71	0.59	0.44	0.50	0.39	-0.7	-0.5
WNP-joint-uniform	0.61	0.68	0.58	0.41	0.46	0.38	2.2	-0.2
WNP-indiv-uniform	0.62	0.68	0.58	0.41	0.45	0.39	2.2	-0.1
WNP-joint-quadr.	0.61	0.68	0.58	0.41	0.46	0.38	2.3	0.0
WNP-indiv-quadr.	0.71	0.72	0.65	0.41	0.45	0.38	-10.7	-5.2
WNP-joint-unif.- $S_{x\beta}^{-1}$	0.61	0.67	0.58	0.41	0.45	0.38	2.2	0.1
WNP-joint-unif.-no d.	0.60	0.68	0.59	0.41	0.46	0.39	2.9	-1.1
N = 1600 [§]				$T = 5$, 1000 replications				
Infeasible GMM-IV	0.30	0.33	0.27	0.20	0.22	0.19	2.2	-0.6
Pooled	0.34	0.37	0.34	0.23	0.26	0.23	2.0	1.0
Sequential	0.35	0.39	0.35	0.22	0.25	0.23	1.3	-3.2
GMM-SS	0.31	0.34	0.29	0.21	0.22	0.19	1.8	0.3
WNP-joint-uniform	0.31	0.34	0.28	0.20	0.23	0.18	1.3	2.0

Table 5 to be continued

Table 5: continued

N = 100 [§] [$]				$T = 10$, 500 replications				
Infeasible GMM-IV	0.90	0.94	0.77	0.61	0.61	0.53	-3.2	0.5
Pooled	1.05	1.12	1.02	0.73	0.74	0.67	-5.2	-2.6
Sequential	1.24	1.33	1.27	0.87	0.93	0.81	-32.4	-27.7
GMM-SS	0.97	0.98	0.88	0.66	0.67	0.60	-5.5	1.1
NP-joint-uniform	1.05	1.09	0.87	0.70	0.76	0.61	-2.8	0.6
NP-indiv-uniform	1.07	1.10	0.87	0.73	0.74	0.58	-2.5	1.3
NP-joint-quadr.	1.05	1.11	0.87	0.74	0.74	0.62	0.3	1.4
NP-indiv-quadr.	1.09	1.12	0.86	0.72	0.74	0.58	-2.8	0.6
NP-joint-unif.- $S_{x\beta}^{-1}$	1.04	1.09	0.87	0.67	0.74	0.60	0.2	2.2
WNP-joint-uniform	0.99	1.06	0.86	0.68	0.75	0.59	-2.1	0.7
WNP-joint-quadr.	1.00	1.08	0.87	0.70	0.75	0.60	-2.5	-1.9
WNP-indiv-quadr.	1.05	1.39	2.93	0.75	0.79	0.64	-2.5	-18.1
WNP-joint-unif- $S_{x\beta}^{-1}$	0.98	1.06	0.86	0.66	0.73	0.58	-1.6	-1.0
WNP-joint-unif.-no d.	1.00	1.17	1.00	0.64	0.84	0.66	2.3	-3.0
N = 400 [§]				$T = 10$, 500 replications				
Infeasible GMM-IV	0.44	0.49	0.39	0.30	0.33	0.25	0.2	-3.9
Pooled	0.53	0.56	0.51	0.34	0.38	0.35	-3.7	-1.1
Sequential	0.56	0.60	0.54	0.37	0.39	0.37	-12.0	-9.1
GMM-SS	0.47	0.52	0.43	0.32	0.34	0.29	-0.2	-3.1
NP-joint-uniform	0.51	0.55	0.43	0.34	0.33	0.30	-2.3	-2.1
NP-indiv-uniform	0.51	0.54	0.42	0.34	0.34	0.30	-1.8	-1.8
NP-joint-quadr.	0.51	0.55	0.43	0.36	0.34	0.30	-1.7	-2.0
NP-indiv-quadr.	0.50	0.53	0.42	0.33	0.33	0.30	-1.4	-1.3
NP-joint-unif- $S_{x\beta}^{-1}$	0.51	0.55	0.42	0.34	0.33	0.30	-2.3	-2.1
WNP-joint-uniform	0.46	0.50	0.41	0.30	0.32	0.29	1.3	0.1
WNP-indiv-uniform	0.46	0.51	0.41	0.30	0.32	0.30	1.7	0.2
WNP-joint-quadr.	0.46	0.50	0.41	0.30	0.33	0.29	2.0	0.4
WNP-indiv-quadr.	0.46	0.50	0.42	0.31	0.33	0.31	1.4	0.7
WNP-joint-unif- $S_{x\beta}^{-1}$	0.46	0.50	0.41	0.31	0.32	0.30	1.1	0.0
WNP-joint-unif.-no d.	0.45	0.51	0.42	0.31	0.32	0.29	1.4	-0.5
N = 1600 [§]				$T = 10$, 500 replications				
Infeasible GMM-IV	0.24	0.25	0.19	0.16	0.15	0.12	-4.2	1.5
Pooled	0.28	0.29	0.25	0.18	0.18	0.18	-2.3	-1.0
Sequential	0.29	0.30	0.25	0.17	0.19	0.17	-3.7	-2.1
GMM-SS	0.25	0.27	0.22	0.16	0.17	0.14	-4.0	-0.4
WNP-joint-uniform	0.25	0.28	0.20	0.17	0.18	0.14	-5.5	-9.5

Note: [§] ML-RE not available, because variance of random effect (δ^2) approached zero;
[$] WNP-indiv-uniform not available, because of singular weighting matrix.

Table 6: Simulation results for pure random effects ($\alpha = 0, \delta = \sqrt{0.5}$) (DGP=2)

	root MSE x10			median absolute error x 10			bias of as. std. err. in %	
	β_D	β_N	β_D/β_N	β_D	β_N	β_D/β_N	β_D	β_N
N = 100				$T=5$, 1000 replications				
Infeasible GMM-IV	1.09	1.29	1.43	0.73	0.84	0.92	0.2	-3.9
ML-RE	1.09	1.22	1.42	0.69	0.78	0.92	-5.4	0.4
Pooled	1.21	1.34	1.66	0.75	0.85	1.10	-1.4	-0.3
Sequential	1.29	1.41	1.78	0.85	0.89	1.10	-14.6	-10.6
GMM-SS	1.10	1.29	1.43	0.72	0.84	0.95	-0.3	-3.8
WNP-joint-uniform	1.14	1.29	1.50	0.76	0.86	0.94	-0.7	-0.3
N = 400				$T=5$, 1000 replications				
Infeasible GMM-IV	0.55	0.60	0.68	0.37	0.39	0.44	-0.5	2.4
ML-RE	0.52	0.61	0.69	0.35	0.42	0.46	0.8	-2.2
Pooled	0.58	0.67	0.79	0.39	0.45	0.54	2.0	-1.0
Sequential	0.59	0.69	0.81	0.40	0.46	0.55	-2.0	-4.6
GMM-SS	0.55	0.60	0.69	0.37	0.39	0.45	-0.7	2.5
WNP-joint-uniform	0.54	0.63	0.71	0.36	0.43	0.48	1.9	-0.6
N = 1600				$T=5$, 1000 replications				
Infeasible GMM-IV	0.27	0.30	0.35	0.19	0.21	0.24	-1.1	2.6
ML-RE	0.28	0.30	0.34	0.18	0.21	0.23	-2.4	1.9
Pooled	0.31	0.34	0.39	0.20	0.23	0.26	-3.0	-2.2
Sequential	0.29	0.34	0.40	0.19	0.24	0.27	1.2	-2.8
GMM-SS	0.27	0.30	0.35	0.20	0.21	0.23	-0.7	2.5
WNP-joint-uniform	0.27	0.31	0.34	0.18	0.21	0.23	1.9	-0.6
N = 100				$T=10$, 500 replications				
Infeasible GMM-IV	0.80	0.89	0.93	0.55	0.61	0.65	-4.8	1.1
ML-RE	0.74	0.85	0.96	0.52	0.53	0.65	1.2	5.7
Pooled	0.86	0.94	1.17	0.58	0.61	0.75	-0.5	5.0
Sequential	1.08	1.20	1.44	0.72	0.80	0.93	-34.3	-28.9
GMM-SS	0.81	0.89	0.93	0.53	0.60	0.64	-5.2	2.0
WNP-joint-uniform	0.84	0.96	1.03	0.58	0.60	0.70	-2.4	1.2
N = 400				$T=10$, 500 replications				
Infeasible GMM-IV	0.38	0.45	0.47	0.24	0.31	0.34	0.2	-0.6
ML-RE	0.37	0.45	0.45	0.24	0.30	0.31	1.4	-0.3
Pooled	0.41	0.49	0.51	0.28	0.34	0.35	3.9	0.2
Sequential	0.44	0.52	0.52	0.31	0.36	0.36	-6.2	-7.2
GMM-SS	0.38	0.45	0.47	0.25	0.32	0.33	0.4	-1.0
WNP-joint-uniform	0.40	0.47	0.47	0.26	0.32	0.36	-1.4	-1.7
N = 1600				$T=10$, 500 replications				
Infeasible GMM-IV	0.19	0.21	0.22	0.13	0.13	0.16	2.6	6.2
ML-RE[§]	0.18	0.23	0.22	0.12	0.15	0.13	6.4	2.4
Pooled	0.21	0.24	0.26	0.14	0.16	0.17	3.0	1.5
Sequential	0.21	0.25	0.27	0.15	0.17	0.18	-0.9	-0.6
GMM-SS	0.19	0.21	0.22	0.13	0.13	0.15	3.1	6.1
WNP-joint-uniform	0.18	0.23	0.24	0.11	0.15	0.16	7.9	-0.8

Note: [§] Only 165 replications, because of convergence problems.

Table 7: Simulation results for MA(1) (DGP=3)

	root MSE x10			median absolute error x 10			bias of as. std. err. in %	
	β_D	β_N	β_D/β_N	β_D	β_N	β_D/β_N	β_D	β_N
N = 100 [§]				$T=5$, 1000 replications				
Infeasible GMM-IV	1.12	1.23	1.52	0.75	0.84	0.96	-2.2	0.8
Pooled	1.18	1.33	1.57	0.76	0.88	1.00	-2.5	-1.2
Sequential	1.22	1.38	1.70	0.76	0.96	1.11	-13.6	-10.1
GMM-SS	1.16	1.29	1.65	0.79	0.88	0.98	-2.8	0.4
WNP-joint-uniform	1.18	1.34	1.57	0.78	0.91	1.00	-5.1	1.6
N = 400				$T=5$, 1000 replications				
Infeasible GMM-IV	0.53	0.61	0.70	0.34	0.41	0.48	1.9	-0.1
ML-RE	0.56	0.61	0.77	0.37	0.39	0.48	-1.2	3.7
Pooled	0.57	0.63	0.79	0.39	0.42	0.51	-1.2	4.1
Sequential	0.59	0.63	0.80	0.40	0.41	0.51	-5.2	1.9
GMM-SS	0.55	0.65	0.76	0.36	0.45	0.50	2.3	-0.6
WNP-joint-uniform	0.55	0.60	0.73	0.37	0.40	0.48	-0.3	4.4
N = 1600				$T=5$, 1000 replications				
Infeasible GMM-IV	0.26	0.31	0.35	0.18	0.20	0.23	4.2	0.4
ML-RE	0.28	0.32	0.37	0.19	0.21	0.23	-2.5	-2.2
Pooled	0.29	0.33	0.38	0.20	0.22	0.25	-0.7	-0.1
Sequential	0.28	0.33	0.38	0.20	0.22	0.25	-0.9	-1.5
GMM-SS	0.27	0.31	0.37	0.19	0.21	0.25	3.7	2.3
WNP-joint-uniform	0.28	0.31	0.36	0.20	0.21	0.23	-2.4	-1.7
N = 100 [§]				$T=10$, 500 replications				
Infeasible GMM-IV	0.75	0.91	1.02	0.53	0.60	0.68	0.6	-3.2
Pooled	0.75	0.93	1.10	0.49	0.61	0.69	4.6	-0.2
Sequential	0.92	1.12	1.37	0.56	0.74	0.85	-27.8	-27.5
GMM-SS	0.79	0.96	1.09	0.53	0.63	0.74	-0.7	-2.3
WNP-joint-uniform	0.77	0.96	1.07	0.52	0.65	0.67	6.1	-2.1
N = 400 [§]				$T=10$, 500 replications				
Infeasible GMM-IV	0.36	0.43	0.52	0.24	0.28	0.35	5.0	0.9
Pooled	0.42	0.47	0.55	0.28	0.30	0.36	-5.6	1.1
Sequential	0.43	0.48	0.56	0.30	0.28	0.36	-11.6	-6.0
GMM-SS	0.38	0.46	0.57	0.26	0.31	0.39	2.3	0.7
WNP-joint-uniform	0.39	0.42	0.48	0.27	0.28	0.34	0.1	6.6
N = 1600 [§]				$T=10$, 500 replications				
Infeasible GMM-IV	0.18	0.22	0.24	0.12	0.14	0.16	2.7	-1.5
Pooled	0.20	0.24	0.27	0.14	0.16	0.17	0.2	-1.2
Sequential	0.20	0.24	0.27	0.13	0.16	0.17	-2.1	-3.3
GMM-SS	0.20	0.24	0.26	0.13	0.16	0.17	0.7	-3.9
WNP-joint-uniform	0.19	0.22	0.25	0.13	0.15	0.16	-0.2	0.1

Note: [§] ML-RE not available, because variance of random effect (δ^2) approached zero.

Table 8: Simulation results for AR (1) ($\alpha = -0.8, \delta = 0$) (DGP 4)

	root MSE x10			median absolute error x 10			bias of as. std. err. in %	
	β_D	β_N	β_D/β_N	β_D	β_N	β_D/β_N	β_D	β_N
N = 100 [§]				$T = 5$, 1000 replications				
Infeasible GMM-IV	1.02	1.22	1.22	0.72	0.83	0.78	-4.6	-3.0
Pooled	1.06	1.38	1.49	0.68	0.92	0.90	-1.1	-1.4
Sequential	1.12	1.45	1.58	0.77	0.98	0.98	-12.1	-10.6
GMM-SS	1.14	1.39	1.52	0.76	0.88	0.93	-8.0	-4.5
WNP-joint-uniform	1.05	1.25	1.27	0.71	0.82	0.85	-2.5	-0.2
N = 400 [§]				$T = 5$, 1000 replications				
Infeasible GMM-IV	0.48	0.57	0.58	0.32	0.39	0.39	1.0	3.9
Pooled	0.53	0.70	0.72	0.36	0.46	0.48	-2.2	-2.5
Sequential	0.54	0.71	0.73	0.36	0.47	0.50	-4.1	-5.2
GMM-SS	0.52	0.64	0.69	0.34	0.45	0.46	1.0	3.8
WNP-joint-uniform	0.51	0.62	0.61	0.36	0.40	0.40	-1.6	-1.2
N = 1600 [§]				$T = 5$, 1000 replications				
Infeasible GMM-IV	0.24	0.30	0.29	0.16	0.19	0.19	-0.1	-0.8
Pooled	0.26	0.34	0.35	0.17	0.22	0.24	0.2	0.2
Sequential	0.26	0.35	0.35	0.18	0.22	0.24	-1.6	-0.8
GMM-SS	0.26	0.34	0.35	0.17	0.23	0.23	-0.3	-1.8
WNP-joint-uniform	0.25	0.31	0.30	0.17	0.21	0.19	2.0	-1.1
N = 100 [§]				$T = 10$, 500 replications				
Infeasible GMM-IV	0.70	0.85	0.80	0.47	0.54	0.57	0.2	0.4
Pooled	0.77	1.02	0.97	0.52	0.73	0.64	-4.2	-2.0
Sequential	0.92	1.21	1.16	0.60	0.84	0.86	-31.6	-29.2
GMM-SS	0.75	0.97	0.96	0.50	0.65	0.65	-0.2	1.5
WNP-joint-uniform	0.75	0.95	0.93	0.53	0.69	0.65	-0.9	-1.1
N = 400 [§]				$T = 10$, 500 replications				
Infeasible GMM-IV	0.34	0.43	0.40	0.23	0.26	0.26	3.4	0.2
Pooled	0.39	0.50	0.47	0.27	0.36	0.29	-5.2	0.7
Sequential	0.40	0.51	0.48	0.27	0.34	0.30	-11.0	-4.8
GMM-SS	0.36	0.49	0.50	0.24	0.30	0.35	2.7	1.0
WNP-joint-uniform	0.39	0.45	0.42	0.24	0.27	0.29	-6.4	0.2
N = 1600 [§]				$T = 10$, 500 replications				
Infeasible GMM-IV	0.18	0.22	0.20	0.12	0.14	0.13	-3.1	-2.8
Pooled	0.20	0.26	0.24	0.14	0.17	0.16	-5.5	-3.0
Sequential	0.20	0.26	0.24	0.14	0.17	0.16	-7.5	-4.8
GMM-SS	0.19	0.25	0.24	0.14	0.17	0.16	-2.4	-2.1
WNP-joint-uniform	0.18	0.22	0.22	0.12	0.14	0.14	1.3	2.2

Note: [§] ML-RE not available, because variance of random effect (δ^2) approached zero.

Table B.1: Simulation results for AR(1) ($\alpha = 0.8$) and random effects ($\delta = 0.2$) (DGP 1): upper and lower bounds and width of 95 % confidence intervals §

	Monte Carlo						Asymptotic - W	
	$\beta_D^{2.5}$	$\beta_D^{97.5}$	*W*	$\beta_N^{2.5}$	$\beta_N^{97.5}$	*W*	β_D	β_N
N = 100			$T = 5$, 1000 replications					
Infeasible GMM-IV	-0.18	0.22	0.40	-0.21	0.23	0.43	0.40	0.43
ML-RE	-0.20	0.22	0.42	-0.22	0.23	0.45	0.41	0.43
Pooled	-0.21	0.25	0.46	-0.24	0.24	0.48	0.45	0.49
Sequential	-0.24	0.24	0.49	-0.27	0.24	0.52	0.41	0.45
GMM-SS	-0.17	0.23	0.40	-0.21	0.23	0.44	0.41	0.44
NP-joint-uniform	-0.23	0.26	0.49	-0.22	0.26	0.48	0.47	0.49
NP-indiv-uniform	-0.22	0.26	0.49	-0.23	0.26	0.49	0.47	0.49
NP-joint-quadr.	-0.23	0.26	0.48	-0.23	0.25	0.48	0.47	0.49
NP-indiv-quadr.	-0.22	0.26	0.48	-0.24	0.26	0.49	0.47	0.49
NP-joint –unif.- $S_{x\beta}^{-1}$	-0.22	0.25	0.47	-0.23	0.26	0.49	0.47	0.49
WNP-joint-uniform	-0.20	0.23	0.43	-0.23	0.24	0.46	0.43	0.46
WNP-indiv-uniform	-0.22	0.25	0.46	-0.23	0.25	0.48	0.44	0.47
WNP-joint-quadr.	-0.21	0.23	0.43	-0.23	0.24	0.46	0.43	0.46
WNP-indiv-quadr.	-0.21	0.24	0.45	-0.24	0.25	0.49	0.44	0.47
WNP-joint-unif.- $S_{x\beta}^{-1}$	-0.20	0.23	0.43	-0.22	0.23	0.45	0.43	0.46
WNP-joint-unif.-no d.	-0.20	0.23	0.43	-0.22	0.24	0.46	0.43	0.46
N = 400			$T = 5$, 1000 replications					
Infeasible GMM-IV	-0.11	0.10	0.21	-0.11	0.10	0.21	0.20	0.21
ML-RE	-0.11	0.09	0.20	-0.11	0.10	0.21	0.21	0.21
Pooled	-0.12	0.11	0.23	-0.12	0.13	0.25	0.23	0.24
Sequential	-0.13	0.11	0.23	-0.13	0.12	0.24	0.22	0.24
GMM-SS	-0.11	0.11	0.21	-0.11	0.10	0.21	0.21	0.22
NP-joint-uniform	-0.11	0.11	0.23	-0.12	0.11	0.24	0.23	0.24
NP-indiv-uniform	-0.11	0.11	0.22	-0.12	0.11	0.24	0.22	0.24
NP-joint-quadr.	-0.11	0.11	0.23	-0.12	0.11	0.24	0.22	0.24
NP-indiv-quadr.	-0.11	0.11	0.23	-0.12	0.11	0.24	0.22	0.24
NP-joint-unif.- $S_{x\beta}^{-1}$	-0.12	0.12	0.23	-0.12	0.12	0.24	0.23	0.24
WNP-joint-uniform	-0.11	0.10	0.21	-0.12	0.11	0.23	0.21	0.22
WNP-indiv-uniform	-0.11	0.11	0.22	-0.12	0.11	0.23	0.21	0.22
WNP-joint-quadr.	-0.11	0.10	0.21	-0.12	0.11	0.22	0.21	0.22
WNP-indiv-quadr.	-0.11	0.10	0.22	-0.12	0.11	0.23	0.21	0.22
WNP-joint-unif.- S_X^{-1}	-0.11	0.11	0.22	-0.12	0.11	0.23	0.21	0.22
WNP-joint-unif.-no d.	-0.10	0.10	0.20	-0.12	0.11	0.23	0.21	0.22
N = 1600			$T = 5$, 1000 replications					
Infeasible GMM-IV	-0.05	0.05	0.10	-0.06	0.05	0.11	0.10	0.11
Pooled	-0.06	0.05	0.11	-0.07	0.05	0.12	0.11	0.12
Sequential	-0.06	0.06	0.12	-0.07	0.05	0.12	0.11	0.12
GMM-SS	-0.05	0.05	0.10	-0.06	0.05	0.11	0.10	0.11
WNP-joint-uniform	-0.06	0.05	0.10	-0.06	0.05	0.11	0.10	0.11

Table B.1 to be continued.

Table B.1: continued

N = 100				$T = 10$, 500 replications				
Infeasible GMM-IV	-0.13	0.16	0.29	-0.17	0.15	0.31	0.29	0.31
Pooled	-0.16	0.19	0.35	-0.18	0.20	0.38	0.33	0.36
GMM-SS	-0.15	0.16	0.30	-0.17	0.15	0.32	0.31	0.33
NP-joint-uniform	-0.17	0.18	0.35	-0.17	0.18	0.35	0.35	0.37
NP-indiv-uniform	-0.16	0.20	0.36	-0.17	0.19	0.36	0.35	0.37
NP-joint-quadr.	-0.17	0.17	0.34	-0.17	0.19	0.37	0.35	0.38
NP-indiv-quadr.	-0.17	0.19	0.37	-0.18	0.20	0.38	0.35	0.38
NP-joint-unif.- $S_{x\beta}^{-1}$	-0.17	0.18	0.34	-0.18	0.19	0.37	0.35	0.37
WNP-joint-uniform	-0.17	0.16	0.32	-0.18	0.18	0.36	0.32	0.35
WNP-joint-quadr.	-0.17	0.16	0.33	-0.18	0.19	0.36	0.32	0.35
WNP-indiv-quadr.	-0.17	0.18	0.35	-0.19	0.20	0.39	0.34	0.38
WNP-joint-unif- $S_{x\beta}^{-1}$	-0.17	0.16	0.33	-0.19	0.18	0.36	0.32	0.35
WNP-joint-unif.-no d.	-0.14	0.18	0.32	-0.17	0.19	0.36	0.34	0.38
N = 400				$T = 10$, 500 replications				
Infeasible GMM-IV	-0.07	0.07	0.14	-0.08	0.08	0.17	0.15	0.16
Pooled	-0.09	0.08	0.17	-0.10	0.09	0.19	0.18	0.17
Sequential	-0.11	0.08	0.18	-0.11	0.09	0.20	0.16	0.17
GMM-SS	-0.08	0.08	0.15	-0.09	0.09	0.18	0.15	0.17
NP-joint-uniform	-0.09	0.08	0.16	-0.09	0.09	0.18	0.17	0.18
NP-indiv-uniform	-0.09	0.07	0.16	-0.09	0.09	0.18	0.16	0.18
NP-joint-quadr.	-0.09	0.07	0.16	-0.09	0.09	0.18	0.17	0.18
NP-indiv-quadr.	-0.08	0.07	0.16	-0.09	0.08	0.17	0.16	0.18
NP-joint-unif- $S_{x\beta}^{-1}$	-0.09	0.08	0.16	-0.09	0.09	0.18	0.17	0.18
WNP-joint-uniform	-0.08	0.06	0.14	-0.09	0.08	0.17	0.15	0.17
WNP-indiv-uniform	-0.08	0.07	0.15	-0.09	0.08	0.17	0.15	0.17
WNP-joint-quadr.	-0.08	0.06	0.15	-0.09	0.08	0.17	0.15	0.17
WNP-indiv-quadr.	-0.08	0.06	0.15	-0.09	0.08	0.17	0.15	0.17
WNP-joint-unif- $S_{x\beta}^{-1}$	-0.08	0.06	0.14	-0.09	0.08	0.18	0.15	0.17
WNP-joint-unif.-no d.	-0.08	0.07	0.15	-0.09	0.09	0.17	0.15	0.17
N = 1600				$T = 10$, 500 replications				
Infeasible GMM-IV	-0.04	0.03	0.076	-0.05	0.03	0.079	0.073	0.079
Pooled	-0.04	0.04	0.084	-0.05	0.04	0.089	0.084	0.091
Sequential	-0.04	0.04	0.087	-0.06	0.03	0.089	0.082	0.090
GMM-SS	-0.04	0.04	0.083	-0.05	0.04	0.086	0.077	0.084
WNP-joint-uniform	-0.04	0.03	0.081	-0.05	0.04	0.091	0.075	0.082

Note: [§] True value of coefficient subtracted.

Table B.2: Simulation results for AR ($\alpha = 0.8$) and random effects ($\delta = 0.2$) (DGP 1): distribution of number of neighbours (k)

	$T=5$, 1000 replications							$T=10$, 500 replications						
N = 100	31	40	49	58	67	76	85	31	40	49	58	67	76	85
NP-joint-uniform		0.1	2.6	7.0	13.3	22.3	54.7						4.8	95.2
NP-indiv-uniform				7.4	38.6	39.9	14.1				0.2	3.2	44.8	51.8
NP-joint-quadr.			0.1	0.7	2.3	4.7	92.2							100
NP-indiv-quadr.						12.0	88.0							100
NP-joint –unif.- $S_{x\beta}^{-1}$		0.9	2.4	6.0	10.9	17.0	62.8					0.4	3.8	95.8
WNP-joint-uniform			0.4	1.4	3.5	13.2	81.5				0.4	0.8	5.8	93.0
WNP-indiv-uniform				0.3	2.3	16.0	81.4	-	-	-	-	-	-	-
WNP-joint-quadr.				0.1		0.8	99.1						0.2	99.8
WNP-indiv-quadr.						0.6	99.4							100
WNP-joint-unif.- $S_{x\beta}^{-1}$		0.2	0.4	2.1	5.1	9.6	82.6				0.2	0.2	5.8	93.8
WNP-joint-unif.-no d.	0.5*	1.2	1.6	1.0	2.7	5.1	87.4	1.8**	2.0	4.2	3.6	1.6	2.6	82.8
N = 400	95	135	175	215	255	295	335	95	135	175	215	255	295	335
NP-joint-uniform	2.3	27.4	34.8	20.0	10.0	3.8	1.7			2.2	16.1	33.6	32.4	15.4
NP-indiv-uniform				2.6	81.5	15.9						52.0	48.0	
NP-joint-quadr.		3.2	27.5	32.7	21.1	10.3	5.2				0.2	5.4	24.8	69.6
NP-indiv-quadr.						43.3	56.7						1.6	98.4
NP-joint-unif.- $S_{x\beta}^{-1}$	11.3	42.6	26.9	10.0	4.5	1.7	3.0			1.4	15.0	26.4	27.2	30.0
WNP-joint-uniform	0.1	0.4	0.3	0.6	1.1	6.5	91.0			0.2	0.2	0.8	5.4	93.4
WNP-indiv-uniform			0.2	0.5	1.6	9.3	88.4				0.2	0.2	4.4	95.2
WNP-joint-quadr.		0.3	0.3	0.1	0.2	0.3	98.8						0.2	99.8
WNP-indiv-quadr.				0.1	0.2	0.9	98.8					0.2		99.8
WNP-joint-unif.- $S_{x\beta}^{-1}$	0.2	0.4	0.5	1.0	1.8	7.6	88.5			0.2	0.4	0.6	4.6	94.2
WNP-joint-unif.-no d.	0.3	0.5	0.3	0.5	1.7	5.2	91.5	0.4	1.4	1.6	0.6	0.8	4.2	91.0
N = 1600	303	466	629	792	955	1118	1281	303	466	629	792	955	1118	1281
WNP-joint-uniform	0.3	0.4	0.2	0.1	0.1	1.1	97.9		0.2	0.4	0.4	0.8	2.4	95.8

Note: * 0.5 % for k = 28; ** 1.4% for k = 28. See also note below Table 5.

Table B.3: Simulation results for pure random effects ($\alpha = 0, \delta = \sqrt{0.5}$) (DGP 2): distribution of number of neighbours (k)

	$T=5$, 1000 replications							$T=10$, 500 replications						
N = 100	31	40	49	58	67	76	85	31	40	49	58	67	76	85
WNP-joint-uniform			0.1	0.6	3.7	13.6	82.0						3.6	96.4
N = 400	95	135	175	215	255	295	335	95	135	175	215	255	295	335
WNP-joint-uniform			0.1	1.2	3.2	9.1	86.4					1.8	4.6	93.6
N = 1600	303	466	629	792	955	1118	1281	303	466	629	792	955	1118	1281
WNP-joint-uniform	0.2	0.8	0.6	1.5	2.3	3.8	90.8		0.2	0.6	2.2	3.0	7.4	86.6

Table B.4: Simulation results for MA(1) (DGP 3): distribution of number of neighbours (k)

	$T = 5$, 1000 replications							$T = 10$, 500 replications						
N = 100	31	40	49	58	67	76	85	31	40	49	58	67	76	85
WNP-joint-uniform			0.1	0.9	4.3	15.9	78.8				0.2	0.4	2.6	96.8
N = 400	95	135	175	215	255	295	335	95	135	175	215	255	295	335
WNP-joint-uniform			0.3	0.7	3.3	10.4	85.3					0.4	3.4	96.2
N = 1600	303	466	629	792	955	1118	1281	303	466	629	792	955	1118	1281
WNP-joint-uniform			0.1	0.5	1.6	5.4	92.4					0.6	3.2	96.2

Table B.5: Simulation results for AR (1) ($\alpha = -0.8, \delta = 0$) (DGP 4): distribution of number of neighbours (k)

	$T = 5$, 1000 replications							$T = 10$, 500 replications						
N = 100	31	40	49	58	67	76	85	31	40	49	58	67	76	85
WNP-joint-uniform				0.1	0.9	4.2	94.8					0.2	2.0	97.8
N = 400	95	135	175	215	255	295	335	95	135	175	215	255	295	335
WNP-joint-uniform	0.1	0.1	0.1	0.1	0.1	0.3	99.1					0.4	1.4	98.2
N = 1600	303	466	629	792	955	1118	1281	303	466	629	792	955	1118	1281
WNP-joint-uniform	0.5	2.0	0.1				97.4			0.8	0.4		0.4	98.4

Table C: Descriptive statistics for variables used in application

		mean	std.
ln (sales)	ln of industry sales in DM	10.54	1.0
rel. firm size	ratio of employment in business unit to employment in industry (* 30)	0.074	0.29
import share	ratio of industry imports to (industry sales + imports)	0.25	0.12
FDJ-share	ratio of industry foreign direct investment to (industry sales + imports)	0.046	0.047
Productivity	ratio of industry value added to industry employment	0.090	0.033
raw mat.	= 1 if firm in sector ´raw materials´	0.087	
Investment	= 1 if firm in sector ´investment goods´	0.50	
Dependent variable	= 1 if product innovation realized	0.60	
Observations: 1270 business units of German firms observed from 1984 to 1988		1270	