Accounting for Nonresponse Heterogeneity in Panel Data

Joachim Inkmann*

Department of Economics and Center of Finance and Econometrics (CoFE)
University of Konstanz, Box D124, 78457 Konstanz, Germany

Abstract: The paper proposes a technique for the estimation of possibly nonlinear panel data models in the presence of heterogeneous unit nonresponse. Attrition or unit nonresponse in panel data usually renders parameter estimators inconsistent unless the unavailable information is missing completely at random. For moment based estimators this problem can be expressed in terms of the impossibility to construct the sample equivalents of the population moments of interest. However, if the attrition process is conditionally mean independent of the variables of interest then the sample equivalents of the population moments can be recovered by weighting the moment functions with the conditional response probability (or propensity score). The latter is usually unknown and has to be estimated. In the presence of nonresponse heterogeneity the propensity score can be estimated by conventional parametric estimation methods like the multinomial logit or probit model. The technique proposed in this paper leads to a moment estimator which simultaneously exploits the weighted moment functions of interest and the score function of the multinomial choice model. The use of simulated moments is discussed for applications with many nonresponse reasons. An applications of the estimator to firm level data is presented where the variables of interest are R&D investments related to product and process innovations.

Keywords: Attrition, selection on observables, missing at random, conditional mean independence, propensity score, multinomial choice, method of simulated moments

JEL Class.: C33, C42

---

* This paper has benefited from discussions with Irene Bertschek, Michael Lechner, Winfried Pohlmeier, Rainer Schnell and participants of the Econometric Society European Winter Meeting 2001 in Venice. I am grateful to the Centre for European Economic Research (ZEW) in Mannheim for making available the data source and to Sandra Gottschalk for supporting the data preparation. Financial support by the *Deutsche Forschungsgemeinschaft* (DFG) is gratefully acknowledged

1. Introduction

With an increasing length of the large-N panel data sets which have stimulated the development of panel data estimation techniques during the previous three decades, the attrition or unit nonresponse problem has attracted increasing interest of statisticians and econometricians. Unit (non-)response can be thought of as a particular sample selection problem which is well known to imply inconsistent parameter estimators (cf. Heckman, 1979). For moment based estimators, which are central in the current paper, the sample selection problem can be explained by the impossibility to construct a sample equivalent of the population moment of interest from the sample selected by the attrition process.

The existing approaches to the correction of the attrition bias have in common that they focus on a single nonresponse category: a sample unit either responds or not. This binary indicator prohibits any distinction between different nonresponse reasons. As an illustration consider the work by Harhoff, Stahl and Woywode (1998) who estimate firm survival and employment growth equations using firm panel data. The authors have information on two different causes of firm exit: voluntary liquidation and bankruptcy. While they use this distinction in estimating competing risk models of the instantaneous firm exit probability (hazard), they merge these two nonresponse reasons into a single exit category in estimating the employment growth equations. This was done although the competing risk model revealed substantial differences between the determinants of voluntary liquidation and bankruptcy. An obvious explanation for neglecting the different nonresponse mechanisms is the lack of appropriate panel data estimation techniques which account for nonresponse heterogeneity.

This paper attempts to provide one possible solution to the problem of estimating potentially nonlinear panel data models in the presence of nonresponse heterogeneity. This solution is based on the "selection on observables" assumption which originates in the statistics literature where the associated conditions eliminating the sample selection bias are known as the "missing at random" (Rubin, 1976) or "conditional independence assumption" (Rubin, 1977). The distinction between selection on observables and selection on unobservables is emphasized by Heckman and Robb (1985) and Heckman and Hotz (1989) in the context of evaluating the impact of treatments where program participation serves as a selection mechanism. Fitzgerald, Gottschalk, and Moffitt (1998) adopt this classification for conditions eliminating nonresponse bias. According to their definition, selection on observables occurs if the variable of interest is independent of the attrition process conditional on the explanatory variables and a set of "auxiliary" variables or "attributes" in the language of Holland (1986). Selection on unobservables occurs if this conditional independence does not hold which is the case in the parametric sample selection models in the tradition of Heckman (1979) which rely on a specification of the joint error term distribution for the response equation and the equation of interest (e.g. Hausman and Wise, 1979, and Ridder, 1990).

The popularity of the missing at random or conditional independence assumption in the statistics literature is due to a result by Rosenbaum and Rubin (1983) who show that this assumption implies that the set of conditioning variables can be replaced by the conditional selection probability (or propensity score) which is the conditional response probability in the application considered here. This result reduces the dimension of the conditioning set to one and facilitates the development of estimators implementing the conditional independence assumption. For moment based estimators calculated from the selected sample it can be shown that a sample equivalent of the population moment of interest can be generated by scaling each contribution with its conditional response probability. This result, which essentially goes back to Horvitz and Thompson (1952), is used throughout the nonresponse literature, e.g. by Cassel, Särndal, and Wretman (1983), Little (1986, 1988), Little and Rubin (1987, p. 58), Czaja, Hirabayashi, Little and Rubin (1992), Little and Schenker (1995), Robins, Rotnitzky and Zhao (1995), Abowd, Crepon and Kramarz (1997), Fitzgerald, Gottschalk, and Moffitt (1998), Horowitz and Manski (1998) and Hirano, Imbens, Ridder, and Rubin (1999).

Imbens (2000) and Lechner (2001) introduce an extension of the Rosenbaum and Rubin result for the evaluation of multiple treatments which cause selection problems which are similar to those implied by nonresponse heterogeneity. Their results are applied by Lechner (1999b) and Gerfin and Lechner (2000) to evaluate the impact of heterogeneous active labor market policy programs in Switzerland and will be used below.

Before returning to the details of the identification and estimation approach it is convenient to introduce the basic notation and to define nonresponse heterogeneity. Assume that the dependent and explanatory variables of interest are collected in the $(k_Y + k_X) \times T$ matrix $Z^1 = (Y', X')'$ where T is the maximum number of available waves of the panel. Values of k_Y exceeding one are allowed to maintain the possibility of estimating panel data models with multiple contemporaneous equations, e.g. systems of simultaneous or seemingly unrelated equations. The components of the nonresponse process are summarized by the $(1+k_W) \times T$ matrix $Z^2 = (A', W')'$ where A denotes a row vector of attrition indicators to be described below and W a set of conditioning variables. Define $Z^3 = \left(Z'^1, Z'^2\right)'$. Note that all matrices are defined in such a way that the t-th column, indicated by the subscript t, always contains the information for period t $\left(t = 1, \ldots, T\right)$. Suppose the data $\left\{Z_i^3 : i = 1, \ldots, N\right\}$ consists of N independent draws from the probability distribution of the random vector $\left(Z_1'^3, \ldots, Z_T'^3\right)'$. However, most arguments will be made on the population level.

Nonresponse heterogeneity is defined as $J > 1$ where J is the number of different attrition categories. Hence, the attrition indicator A_t for period $t = 1, \ldots, T$ is defined as follows

$$A_t = \begin{cases} 0 & \text{response in period t,} \\ 1 & \text{nonresponse in period t because of reason "1",} \\ \vdots & \vdots \\ J & \text{nonresponse in period t because of reason "J".} \end{cases} \qquad (1)$$

Occasionally, it is more convenient to work alternatively with a $\left(J+1\right)\times 1$ vector of dummy variables $D_t = \left(D_t^0, \ldots, D_t^J\right)'$ where $D_t^j = 1\left(A_t = j\right)$ for $j = 0, \ldots, J$.

Suppose the interest focuses on estimating the conditional expectation $E\left[Y_t \mid X_t\right]$.[1] The identification problem in the presence of nonresponse results from the impossibility to construct a sample counterpart of this conditional expectation in the sample selected by $A_t = 0$. Without further assumptions it is only possible to identify bounds for $E\left[Y_t \mid X_t\right]$ as suggested by Horowitz and Manski (1998). To see this, write $E\left[Y_t \mid X_t\right] = \Pr\left[A_t = 0 \mid X_t\right] \cdot E\left[Y_t \mid X_t, A_t = 0\right] + \Pr\left[A_t \neq 0 \mid X_t\right] \cdot E\left[Y_t \mid X_t, A_t \neq 0\right]$. If a priori knowledge allows restricting the domain of Y_t to an interval $\left[B_l, B_u\right]$, then $E\left[Y_t \mid X_t, A_t \neq 0\right] \in \left[B_l, B_u\right]$ and

$$\begin{aligned} \Pr\left[A_t = 0 \mid X_t\right] \cdot E\left[Y_t \mid X_t, A_t = 0\right] + \Pr\left[A_t \neq 0 \mid X_t\right] \cdot B_l \\ \leq E\left[Y_t \mid X_t\right] \leq \qquad\qquad \\ \Pr\left[A_t = 0 \mid X_t\right] \cdot E\left[Y_t \mid X_t, A_t = 0\right] + \Pr\left[A_t \neq 0 \mid X_t\right] \cdot B_u. \end{aligned} \qquad (2)$$

These bounds can be readily estimated using the selected sample.[2]

In the subsequent two sections alternative identifying assumptions are proposed which are sufficient to estimate the unknown parameters of a parametric specification of $E\left[Y_t \mid X_t\right]$ by a conditional moment approach to GMM estimation in the presence of nonresponse heterogeneity. The first assumption, MCAR (missing completely at random), leads to a GMM estimator which exploits unconditional moment functions implied by the conditional moment

1 For simplicity, it is assumed in expression (2) that the conditioning variables can be observed despite attrition (which would be the case, e.g., if they consist of lagged information).

2 See Lechner (1999a) for an application to the evaluation of treatment effects where the bounds are estimated by nonparametric estimation techniques.

restrictions without any modification. The second assumption, MAR (missing at random), requires scaling the unconditional moment functions by the conditional response probability or propensity score which can be estimated simultaneously using GMM. Parametric specifications of the propensity score are discussed in Section 4. Section 5 presents some Monte Carlo experiments for a GMM estimator which implements the MAR assumption proposed in Section 3 using a multinomial probit specification of the propensity score. The different GMM estimators are applied in Section 6 to estimate R&D equations using a panel of West German manufacturing firms. Section 7 concludes.

2. Missing Completely at Random (MCAR): Mean Independence

Following the conditional moment approach to GMM estimation of nonlinear panel data models (see e.g. Lechner and Breitung, 1996, Breitung and Lechner, 1999, and Inkmann 2000a, 2001) assume that $E\left[Y_t \mid X_t\right]=\mu\left(X_t,\beta_0\right)$. Then a $r_1\times 1$ vector of unconditional moment functions

$$\psi_1\left(Z_t^1,\beta\right)=B\left(X_t\right)\cdot\left(Y_t-\mu\left(X_t,\beta\right)\right) \tag{3}$$

can be obtained using any $r_1\times k_Y$ matrix of instruments $B\left(X_t\right)$, e.g. $B\left(X_t\right)=X_t$ which implies $r_1=k_X$[3]. The population moment of interest is the orthogonality condition

$$E\left[\psi_1\left(Z_t^1,\beta_0\right)\right]=0 \tag{4}$$

which holds by the law of iterated expectations. Without further assumptions a counterpart of (4) can not be identified from the sample selected by $A_t=0$. The first identifying assumption to be considered is MCAR (missing completely at random) which states that

$$Z_t^1 \amalg D_t^0 \tag{5}$$

where $a \amalg b$ means that a is independent of b. For the current purpose a weaker mean independence assumption can be imposed

$$E\left[\psi_1\left(Z_t^1,\beta_0\right)\mid A_t=0\right]=E\left[\psi_1\left(Z_t^1,\beta_0\right)\mid A_t\neq 0\right] \tag{6}$$

[3] For notational simplicity only contemporaneous instruments are considered. If the explanatory variables are (weakly) strictly exogenous then overidentifying restrictions can be gained from using the regressors from (past) all periods as additional instruments. These overidentifying restrictions will lead to asymptotic efficiency gains for the associated GMM estimator but are likely to increase its small sample bias (see the discussion in Inkmann (2001, ch. 6/7). Note also, that lagged explanatory variables are not available for all units in the presence of temporary nonresponse. For these reasons the restriction to contemporary instruments seems to be reasonable.

which is sufficient to ensure that (4) can be identified from the selected sample without any modification of the unconditional moment functions (3). Obviously, the conditions (5) and (6) are extremely restrictive and will be weakened therefore in the next section. Note, however, that MCAR is implicitly assumed to hold in all panel data applications where the nonresponse issue is not addressed.

Before turning to the MAR assumption, the GMM estimator for the panel data model described by some general moment function $\psi(Z_t, \theta)$, with an underlying true $q \times 1$ parameter vector θ_0, will be reviewed in short. The moment functions for each period are collected in a single $r \times 1$ (with $r \geq q$ for identification) vector of moment functions

$$\psi(Z, \theta) = \left(\psi(Z_1, \theta)', \ldots, \psi(Z_T, \theta)' \right)' \tag{7}$$

which serves as a basis for the GMM estimator $\hat{\theta}$ of θ_0

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \left(\tfrac{1}{N} \sum_{i=1}^{N} \psi(Z_i, \theta) \right)' \hat{W} \left(\tfrac{1}{N} \sum_{i=1}^{N} \psi(Z_i, \theta) \right) \tag{8}$$

where $\hat{W}$ is some positive semidefinite weight matrix and Θ denotes the space of possible parameters (Hansen, 1982).[4] Under some regularity conditions (cf. Newey and McFadden, 1994), $\operatorname{plim} \hat{\theta} = \theta_0$ and $\sqrt{N}\left(\hat{\theta} - \theta_0\right) \xrightarrow{a} N(0, \Lambda)$, where $\Lambda = \left(G_0' V_0^{-1} G_0\right)^{-1}$ provided that $\hat{W}$ satisfies $\operatorname{plim} \hat{W} = V_0^{-1}$, and denoting $G_0 = E\left[\partial\psi(Z, \theta_0)/\partial\theta'\right]$, $V_0 = E\left[\psi(Z, \theta_0)\psi(Z, \theta_0)'\right]$.

The GMM estimator of β_0 in (4) can be calculated according to (8) using the moment functions $\psi(Z_t, \theta) \equiv \psi_1\left(Z_t^1, \beta\right) \cdot D_t^0$ in (7). This unbalanced panel estimator should be more efficient than the balanced panel estimator which is based on those observations satisfying $D_t^0 = 1, \ \forall t$. However, both estimators are consistent under MCAR.

3. Missing at Random (MAR): Conditional Mean Independence

The second identifying assumption to be considered is MAR (missing at random)

$$Z_t^1 \amalg D_t^0 \mid W_t \tag{9}$$

where $a \amalg b \mid c$ means that a is independent of b conditional on c (cf. Dawid, 1979, Angrist, 1997). Again, (9) can be weakened to a conditional mean independence assumption

---

4 This paper focuses on GMM because it is the most frequently applied method of moments. A number of alternative estimators using moment functions of the type (7) can be employed instead without changing the basic procedure, e.g. the continuous updating estimator, the empirical likelihood estimator or the exponential tilting estimator (cf. Imbens, Spady and Johnson, 1998, or Inkmann, 2000b).

$$E\left[\psi_1\left(Z_t^1,\beta_0\right) \mid W_t, A_t = 0\right] = E\left[\psi_1\left(Z_t^1,\beta_0\right) \mid W_t, A_t \neq 0\right] \tag{10}$$

which is sufficient in the GMM context. The difference between MCAR and MAR is the set of conditioning variables W_t not entering the set of explanatory variables in $\mu\left(X_t,\beta\right)$. Note that the conditioning variables or attributes have to be observed for both the respondents and nonrespondents. This suggests using lagged values of Z_t^1 as possible candidates for W_t. Also information on lagged item nonresponse (e.g. to a question related to earnings) should be useful in determining unit nonresponse as pointed out by Lechner (1995) and Rendtel (1995). In any case, a reasonable application of the identification conditions (9) and (10) requires rich data sets containing information on nonrespondents. In answering the requests of researchers (e.g. Schnell, 1997), special nonrespondent surveys are carried out on a more regular basis today by the institutions collecting the data (e.g. by the ZEW, Mannheim, responsible for the Mannheim Innovation Panel, MIP, used in Section 6), which suggests that MAR will become more and more useful for analyzing nonresponse behavior.

To implement MAR within the GMM framework define the propensity score as the conditional response probability

$$P\left(W_t\right) = E\left[D_t^0 \mid W_t\right] = \Pr\left[A_t = 0 \mid W_t\right]. \tag{11}$$

The following properties of the propensity score are derived by Rosenbaum and Rubin (1983) for $J = 1$ and by Imbens (2000) and Lechner (2001) for $J > 1$:

$$W_t \amalg D_t^0 \mid P\left(W_t\right) \tag{12}$$

$$Z_t^1 \amalg D_t^0 \mid P\left(W_t\right). \tag{13}$$

The first result states that given the same response probability the attributes are independently distributed of the response indicator. This balancing property of the propensity score follows mechanically from (11) and is sometimes used to evaluate the quality of an estimator of $P\left(W_t\right)$ in applied work (e.g. Dehejia and Wahba, 1999). The second result requires MAR to hold and allows reducing the dimension of the conditioning set from k_W in (9) to one in (13).

The propensity score will be used in the following to show that the population moment of interest (4) has a sample counterpart if the original moment functions $\psi_1\left(Z_t^1,\beta\right)$ are replaced by the modified moment functions

$$\tilde{\psi}_1\left(Z_t^3,\beta\right) = \psi_1\left(Z_t^1,\beta\right)\cdot\frac{D_t^0}{P\left(W_t\right)} \tag{14}$$

which can be calculated from the selected sample. Therefore it has to be shown that

$$E\left[\tilde{\psi}_1\left(Z_t^3,\beta_0\right)\right]=E\left[\psi_1\left(Z_t^1,\beta_0\right)\right] \tag{15}$$

which is done in Appendix A in a similar way as in Fitzgerald, Gottschalk, and Moffitt (1998), Horowitz and Manski (1998) and Imbens (2000). However, the moment functions (14) are generally unavailable because the propensity score $P\left(W_t\right)$ is unknown. A feasible estimation strategy consists of estimating the propensity score in a first step and estimating GMM using the moment functions (14) in a second step after replacing $P\left(W_t\right)$ with its estimator. The propensity score can be estimated using parametric, or semi-/nonparametric methods (see Horowitz, 1993) whereby the latter techniques usually focus on the binary case with $J=1$. Therefore parametric methods are proposed in the next section. Regardless of the method being used it has to be taken into account that $P\left(W_t\right)$ is replaced by an estimator in order to obtain reliable inference as emphasized by Little (1988) and Little and Schenker (1995).[5] Using a parametric estimate of the propensity score this adjustment is particularly simple in the GMM framework as shown by Newey (1984) and Newey and McFadden (1994) and demonstrated now.

In parametric multinomial response models the conditional mean of the vector of dummy variables D_t indicating the respective category of A_t can be written in general form as $E\left[D_t \mid W_t\right]=\pi\left(W_t,\gamma_0\right)$ with elements $\pi_j\left(W_t,\gamma_0\right)$, $j=0,\ldots,J$. Hence, the conditional moment approach to an unconditional moment function can be employed again and yields

$$\psi_2\left(Z_t^2,\gamma\right)=C\left(W_t\right)\cdot\left(D_t-\pi\left(W_t,\gamma\right)\right) \tag{16}$$

where $C\left(W_t\right)$ is some $r_2\times\left(J+1\right)$ matrix of instruments to be specified below. A feasible version of the moment function (14) can be written now as

$$\psi_3\left(Z_t^3,\delta\right)=\psi_1\left(Z_t^1,\beta\right)\cdot\frac{D_t^0}{\pi_0\left(W_t,\gamma\right)} \tag{17}$$

where $\delta=\left(\beta',\gamma'\right)'$. The unknown parameters $\delta_0=\left(\beta_0',\gamma_0'\right)'$ of both the propensity score equations and the equations of original interest can be estimated jointly using (8) and moment functions of the type $\psi\left(Z_t,\theta\right)\equiv\left(\psi_2\left(Z_t^2,\gamma\right)',\psi_3\left(Z_t^3,\delta\right)'\right)'$ in (7). The formula for the variance-covariance matrix of the stabilizing transformation of the resulting GMM estimator $\hat{\delta}$ automatically adjusts for the estimated propensity score. The same idea has been employed before

5 This potential source of erroneous inference is usually ignored in the treatment effects literature.

in the context of estimating a conditional response probability (with $J=1$) by Abowd, Crepon and Kramarz (1997).

The next section proposes alternative specifications of $\pi_j(W_t,\gamma)$ and $C(W_t)$ in (16) for estimating the propensity score..

4. Propensity Score Estimation

Well known parametric models for the estimation of multiple choice models include the multinomial logit and the multinomial probit. This section shows that the first order conditions for the Maximum Likelihood estimators of these models can be written in termini of the moment functions (16).

Because of its computational simplicity a convenient starting point is the multinomial logit model defined by the following parameterization of the conditional choice probabilities:

$$\pi_j(W_t,\gamma)=\frac{\exp\left(W_t'\gamma^j\right)}{1+\sum_{i=1}^{J}\exp\left(W_t'\gamma^i\right)} \tag{18}$$

for $j=0,\ldots,J$ and $\gamma=\left(\gamma'^0,\ldots,\gamma'^J\right)'$. To ensure that the probabilities add up to one the parameter values of the first (arbitrarily chosen) category have been restricted to zero, i.e. $\gamma^0=0$. In addition, exclusion restrictions may be imposed on some elements of W_t in some categories when the corresponding explanatory variables should not affect the conditional mean functions of these categories.

The matrix of instruments $C(W_t)$ will be chosen optimally for the given set of conditional moment restrictions $E\left[D_t-\pi(W_t,\gamma_0)\,|\,W_t\right]=0$ in such a way that (16) reflects the score function of the Maximum Likelihood estimator of the multinomial logit model. Hence,

$$C(W_t)=(W_t,\ldots,W_t) \tag{19}$$

which is of dimension $k_W\times(J+1)$. If there are exclusion restriction then the corresponding elements of the matrix have to be set to zero again. This matrix of instruments leads to the most efficient combination of the conditional moment functions. Using these instruments, γ_0 is exactly identified from the moment functions of a single period and therefore overidentified whenever $T>1$ (or $T>2$ when there is no initial nonresponse).

While the multinomial logit model can be readily estimated almost regardless of the number of categories, the model suffers from the well known IIA (independence of irrelevant alternatives) property which states that the ratio of two conditional choice probabilities is independent on the remaining categories as can be seen from (18). Therefore the multinomial probit model is considered as an alternative in the following.

Estimation of the multinomial probit model is much more difficult than estimation of the multinomial logit model. Conventional ML estimation using numerical optimization routines becomes impossible whenever there are more than four choice categories (i.e. $J > 3$). Therefore the score function of the model is derived for $J = 3$ in a first step and the method of simulated moments (McFadden, 1989) is discussed in a second step for applications with $J > 3$.

The $J = 3$ case will be outlined in some detail because it will serve as a reference for the Monte Carlo experiments in the next section. For the sake of providing simple identification conditions a random utility model is employed although this model can be hardly interpreted from an economic point of view in relationship with nonresponse heterogeneity. The "utility" associated with nonresponse category $j = 1, \ldots, J$ is specified as $U_t^j = W_t' \alpha_0^j + \varepsilon_t^j$ while the utility of the response category is normalized to $U_t^0 = 0$. It is assumed that the error terms $\varepsilon_t = \left(\varepsilon_t^1, \varepsilon_t^2, \varepsilon_t^3\right)'$ follow a trivariate normal distribution with mean zero and a within period variance-covariance matrix

$$\Sigma_0 = \begin{pmatrix} 1 & & \\ \sigma_{21} & \sigma_{22} & \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \tag{20}$$

where the upper left element is set to one for identification reasons. Otherwise the intratemporal correlation structure is left unrestricted.[6] Then the conditional choice probabilities can be derived from a pairwise comparison of the random utilities as

$$\begin{aligned} \pi_0\left(W_t, \gamma\right) &= \Pr\left[U_0 > U_1, U_0 > U_2, U_0 > U_3\right] \\ &= \Phi^{(3)}\left[-W_t'\alpha^1, -W_t' \frac{\alpha^2}{\sigma_{22}^{0.5}}, -W_t' \frac{\alpha^3}{\sigma_{33}^{0.5}}; \Omega_0\right] \end{aligned} \tag{21}$$

where $\Phi^{(3)}$ denotes the c.d.f. of the trivariate standard normal distribution and Ω_0 denotes the correlation matrix corresponding to Σ_0. The choice probabilities $\pi_j\left(W_t, \gamma\right)$, for $j = 1, 2, 3$, can be derived accordingly and are given in Appendix B. The parameter vector γ contains both

[6] The same holds for the inter-temporal correlation structure which needs not to be specified because only the marginal moments of each period are used. Börsch-Supan, Hajivassiliou, Kotlikoff, and Morris (1993) and Ziegler and Eymann (2000) discuss specifications of the inter-temporal correlation structure of the model.

$\alpha = \left(\alpha'^1, \ldots, \alpha'^J\right)'$ and the unknown elements of the lower triangular matrix L which is chosen such that $LL' = \Sigma_0$ to ensure positive definite variance-covariance matrices. The upper left element of L is normalized again to one. For computational reasons, Keane (1992) strongly recommends to impose at least one exclusion restriction on the elements of W_t in each category such that the same set of regressors is not used twice in different categories.

In order to mimic the score function of a ML estimator of the multinomial probit model which is given e.g. by McFadden (1989, equ. 4) with the moment function (16) it is necessary to chose the matrix of instruments correctly. In this model the optimal instruments are

$$C(W_t) = \left(\frac{\partial \ln \pi_0 (W_t, \gamma_0)}{\partial \gamma}, \ldots, \frac{\partial \ln \pi_J (W_t, \gamma_0)}{\partial \gamma} \right) \tag{22}$$

and depend on the unknown parameter vector γ_0.[7] One strategy of obtaining feasible optimal instruments consists of replacing γ_0 with a first step estimator. In the Monte Carlo experiments and the application given below γ_0 is replaced with the ML estimator of the multinomial probit model which has the additional advantage of providing good starting values for the GMM optimization routine. Alternatively, one might employ suboptimal instruments, for example the optimal instruments (19) for the multinomial logit model, which circumvent the first estimation step but will lead to asymptotically less efficient parameter estimators. Note, however, that (19) is not sufficient to identify all parameters of the multinomial probit model from the information of a single period because of the additional variance-covariance parameters. This should be no problem as soon as there are more waves available.

When the number of categories exceeds four ($J > 3$) then the moment condition (16) can be still used but its components in (21) and (22) have to be simulated with the help of Monte Carlo integration techniques as proposed in the influential paper by McFadden (1989). In this case (16) will be a set of simulated moment functions. Both components, the (non-)response probability in (21) and the matrix of optimal instruments in (22), contain multivariate normal probabilities which suggest using the GHK (Geweke-Hajivassiliou-Keane) simulator which has proven to be the most successful simulator in this particular application in a variety of experiments carried out by Hajivassiliou, McFadden and Ruud (1996).[8] Appendix C derives

7 The exact analytical form of these derivatives is given in Lemma 10 in McFadden (1989).

8 The GHK simulator for the multinomial logit model is presented e.g. by Börsch-Supan and Hajivassiliou (1993) and Geweke, Keane and Runkle (1994) in the context of simulated ML estimation and applied in a large number of recent studies (e.g. Inkmann, 2000a), in particular by Lechner (1999b) and Gerfin and Lechner (2000) who estimate propensity scores for multiple program participation.

the GHK simulator of the response probability (21) for $J > 3$. The simulator of the optimal instruments (22) can be obtained accordingly using a different set of underlying random numbers as emphasized by McFadden (1989, p.1004). Alternatively, one might consider using suboptimal instruments to circumvent this additional simulation step. McFadden derives the asymptotic distribution of the method of simulated moments (MSM) estimator using arguments from Pakes and Pollard (1989). The asymptotic variance-covariance matrix of the stabilizing transformation of the MSM estimator can be shown to be $(1+1/R)$ times the inverse of the information matrix of the multinomial probit model where R is the number of Monte Carlo replications used for simulating the response probability (see Lee, 1996, p. 80). Hence, with an increasing R MSM becomes asymptotically as efficient as ML. Since the moment functions (16) mimic the score of ML, the GMM variance-covariance matrix Λ given in Section 3 will remain valid despite simulation when $R \to \infty$.

5. Monte Carlo Experiments

In order to shed some light on the small sample performance of the GMM estimators introduced in the proceeding sections, a number of Monte Carlo experiments is carried out. Because the basic estimation procedure is not affected by the number of periods involved, the experiments are carried out on a cross-section such that the index t can be omitted in this section. Both, linear and nonlinear equations of interest are considered. There are $J = 3$ nonresponse categories. The conditional (non-)response probabilities are estimated using the multinomial probit specification with optimal instruments (22) which are obtained from a first step ML estimation of this model.

The data generating process is specified as follows

$$\begin{aligned} Y &= \tau(\beta_1 + \beta_2 X + \omega) \cdot 1(A = 0) && \text{with } \beta = (\beta_1, \beta_2)' \\ U^1 &= \alpha_1^1 + \alpha_2^1 X + \alpha_3^1 W + \alpha_4^1 D_1 + \varepsilon^1 && \text{with } \alpha^1 = (\alpha_1^1, \alpha_2^1, \alpha_3^1, \alpha_4^1)' \\ U^2 &= \alpha_1^2 + \alpha_2^2 X + \alpha_3^2 W + \alpha_4^2 D_2 + \varepsilon^2 && \text{with } \alpha^2 = (\alpha_1^2, \alpha_2^2, \alpha_3^2, \alpha_4^2)' \\ U^3 &= \alpha_1^3 + \alpha_2^3 X + \alpha_3^3 W + \alpha_4^3 D_3 + \varepsilon^3 && \text{with } \alpha^3 = (\alpha_1^3, \alpha_2^3, \alpha_3^3, \alpha_4^3)' \end{aligned} \tag{23}$$

where $\tau(e)$ denotes an observability rule defined as $\tau(e) = e$ in the first experiment (E1) and as $\tau(e) = 1(e > 0)$ in the second experiment (E2). Hence, the equation of interest is a linear regression model in E1 while it is a binary choice model in E2. Correspondingly, the moment functions $\psi_1\left(Z^1,\beta\right)$ of original interest are specified as the scores of the ML estimators of these models. The nonresponse indicator is determined by a random utility model with the utility of the response category normalized to zero, i.e. $U^0 = 0$ as demonstrated in Section 4. The explanatory variables D_1, D_2, D_3 are used to integrate Keane's (1992) recommendation to apply exclusion restrictions to the variables of the multinomial probit model in order to ensure identification of the unknown parameters in practical work.[9] Each of these dummy variables has a mean of 0.5. The continuous explanatory variable is generated as $X \sim N\left(2,4\right)$. The true mean parameters are set to

$$\beta_0 = \left(-1,1\right)' \quad \text{and} \quad \alpha_0^1 = \alpha_0^2 = \alpha_0^3 = \left(-1,1,-1,1\right)' \tag{24}$$

and the elements of Σ_0 in (20) are set to one on the main diagonal and to 0.5 off diagonal. The data generating process implements the MAR assumption by assuming that[10]

$$\begin{pmatrix} \varpi \\ W \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & 0.75 \\ 0.75 & 1 \end{pmatrix}\right). \tag{25}$$

The Tables 1-3 contain summary statistics for the parameter estimates of the equation of interest using estimators under both, the correctly specified MAR assumption and the misspecified MCAR assumption. The Tables A1-A3 in Appendix D contain summary statistics for the unknown parameters of the propensity score equations under the MAR assumption. The sample sizes are N = 500 in E1 and N = 500 and N = 2000 in E2. The number of Monte Carlo replications is M = 1000 for the experiment with N = 500 and M = 500 for N = 2000.

As expected, the correctly specified MAR estimator exhibits much less bias in the linear regression model in Table 1 than the MCAR estimator ignoring the selection on observables mechanism. The same holds for the estimators of the slope coefficient in the binary choice model in Table 2. Surprisingly, the intercept is estimated with less bias by the misspecified MCAR estimator. However, this result is mitigated when the sample size increases from N = 500 to N = 2000 as Table 3 clearly shows and should reverse in a larger sample.

---

9 Of course, these restrictions are not required from a theoretical point of view. The specification of the multinomial probit model presented in Section 4 ensures identification. In a discussion of this paper, Richard Blundell has pointed out that these exclusion restrictions also serve as a main reason for considering nonresponse heterogeneity. In the absence of these restrictions the different nonresponse categories can be pooled for the estimation of the propensity score *provided* there is no original interest in the parameters of the selection process itself. This pooling issue will be investigated in future work in more detail.

10 Torsten Persson has suggested to evaluate the performance of the estimators which implement the selection on observables assumption under a data generating process which produces selection on unobservables in order to shed some light on the estimator's performance under misspecification. This idea remains an important task for future work as well.

Table 1: GMM estimation of the equation of interest under MCAR and MAR
Summary statistics from 1000 replications of Monte Carlo experiment E1 (N = 500)

	param.	true	mean	bias	se	rmse
MCAR	β_1	-1.0000	-0.8217	0.1783	0.0679	0.1908
MAR		-1.0000	-1.0058	-0.0058	0.1162	0.1164
MCAR	β_2	1.0000	1.1506	0.1506	0.0436	0.1568
MAR		1.0000	1.0752	0.0752	0.1002	0.1253

Table 2: GMM estimation of the equation of interest under MCAR and MAR
Summary statistics from 1000 replications of Monte Carlo experiment E2 (N = 500)

	param.	true	mean	bias	se	rmse
MCAR	β_1	-1.0000	-0.9543	0.0457	0.1722	0.1782
MAR		-1.0000	-1.1109	-0.1109	0.1944	0.2238
MCAR	β_2	1.0000	1.2795	0.2795	0.1649	0.3245
MAR		1.0000	1.1819	0.1819	0.2328	0.2954

Table 3: GMM estimation of the equation of interest under MCAR and MAR
Summary statistics from 500 replications of Monte Carlo experiment E2 (N = 2000)

	param.	true	mean	bias	se	rmse
MCAR	β_1	-1.0000	-0.9322	0.0678	0.0868	0.1101
MAR		-1.0000	-1.0721	-0.0721	0.1163	0.1368
MCAR	β_2	1.0000	1.2505	0.2505	0.0775	0.2622
MAR		1.0000	1.1034	0.1034	0.1694	0.1985

6. Application

In this section the proposed estimators are applied to a firm level data set to estimate R&D (research and development) equations for the amount of R&D investments devoted to the development of process and product innovations, process R&D and product R&D, henceforth. According to a hypothesis formulated by Cohen and Klepper (1996) and Klepper (1996) the proportion of process R&D on total R&D increases with increasing firm size. This hypothesis will be investigated in the following by estimating a SURE model of process and product R&D equations and comparing the estimated parameters of the firm size variables. If the Cohen/Klepper hypothesis holds then the firm size parameter in the process R&D equation should exceed the corresponding parameter in the product R&D equation. A similar estimation strategy has been employed by Fritsch and Meschede (2000) who find weak evidence for the Cohen/Klepper hypothesis in separate regressions for product and process R&D,

A data set which contains information on process and product R&D for German manufacturing firm is the Mannheim Innovation Panel (MIP) which is collected by the Centre for European Economic Research (ZEW) in Mannheim. This data set contains detailed information on the R&D and innovation activities of the firms (see Harhoff and Licht, 1994, for details). A distinction between process and product R&D is possible in the first two waves raised in 1993 and 1994 which are used in the following. Similar to other surveys included in the European Community Innovation Surveys (CIS) program, the MIP is affected by the problem that the R&D information is only collected for those firms who plan to introduce an innovation within the three years subsequent to the interview date (so called "innovation filter"). In the first wave the innovation filter is defined slightly different: firms which have introduced an innovation within the three years before the interview or expect to introduce an innovation within the three years following the date of the interview pass the innovation filter. Those firms who pass the innovation filter in the first wave are defined as the population of interest in the subsequent analysis. This assumption facilitates the estimation procedure because it rules out initial nonresponse. Looking at the response behavior in 1994 of the 1019 firms included in the 1993 sample (after deleting item nonresponse information), it turns out that more than 50% (525) of these firms do not respond in the second wave in 1994. However, 172 of the 1994 nonrespondents return in later waves (1995 – 1997). To distinguish these firms from permanent nonrespondents and to account for the innovation filter described above which absorbs 59 firms in the second wave, the (non-)response indicator is defined as follows:

$$A_2 = \begin{cases} 0 & \text{response in period 2 and innovation filter passed,} \\ 1 & \text{response in period 2 but innovation filter not passed,} \\ 2 & \text{nonresponse in period 2 but at least one response in periods 3-5,} \\ 3 & \text{nonresponse in periods 2-5.} \end{cases} \qquad (26)$$

The R&D equations of interest are linear in the log of process and product R&D. The underlying moment functions are specified as the scores of the ML estimators of the two separate equations. The main explanatory variable is the log of the number of employees as a firms size measure. To control for different R&D capabilities, an indicator for the presence of a R&D department is included. To control for industry heterogeneity a vector of 10 sector dummies is added. Finally, the list of explanatory variables is extended by an indicator for a firm located in the former East Germany. The list of attributes consists of (lagged) firm size, lagged total R&D expenditures, the R&D department indicator, the dummy variable for East German firms and indicators for innovation constraints and expectations regarding the future (3 ys.) development of the firm. Table 4 presents the GMM estimation results for the process and product R&D equations of interest under MCAR and MAR.[11] Table A4 in Appendix E contains summary statistics for the respondents. Table 5 displays the GMM estimation results for the (non-)response equations used for the estimation of the propensity score under MAR. Table A5 in Appendix E compares the means of the attributes over the (non-)response categories. From the list of attributes listed in Table A5 only those are usually maintained in the final regression in Table 5 which turned out significant.

The GMM estimator is based on a multinomial logit specification of the conditional response probabilities[12] and an optimal weight matrix using the usual two step estimation procedure. Apart from the J test (or Hansen-Sargan test) of overidentifying restrictions, a LM test developed by Imbens, Spady and Johnson (1998) is employed to test the model specification whereby the latter (based on the empirical likelihood approach) is expected to have better small sample size properties according to the Monte Carlo experiments carried out by these authors. Neither the MAR nor the MCAR specification is rejected by the LM test as Table 4 shows.

[11] A statistical test whether MCAR or MAR can be preferred is not available unless refreshment samples are exploited as shown by Hirano, Imbens, Ridder, and Rubin (1999). However, because the current paper focuses on estimation under nonresponse heterogeneity and not on hypotheses testing, this test is left to future research.

[12] The multinomial probit specification is left for future work.

Table 4: GMM estimation of the R&D equations (N = 1019)

	MCAR		MAR	
parameters	estimate	t-value	estimate	t-value
equation 1: product R&D				
Intercept	-5.1640	-25.1732	-5.2941	-25.9882
Log number of employees	0.7382	20.2409	0.7534	20.9891
East German firm	-0.2118	-2.4804	-0.2491	-2.8661
R&D department	1.2871	14.2021	1.2628	13.8153
equation 2: process R&D				
Intercept	-5.8346	-28.7268	-5.9593	-29.4326
Log number of employees	0.6611	17.8507	0.6772	18.0051
East German firm	-0.2748	-3.4299	-0.3442	-4.1998
R&D department	0.5908	7.1618	0.5324	6.2484
hypotheses tests:	$\chi^2(\text{dof})$	p-value	$\chi^2(\text{dof})$	p-value
LM test of overident. restrictions	32.02 (28)	0.2739	32.93 (28)	0.2385
Wald test: employ. 1 = employ. 2	6.6086 (1)	0.0101	5.2563 (1)	0.0219
Wald test: joint sign. equation 1	1723 (13)	0.0000	1775 (13)	0.0000
Wald test: joint sign. equation 2	770.1 (13)	0.0000	839.6 (13)	0.0000
Wald test: sector dummies equ. 1	154.9 (10)	0.0000	144.4 (10)	0.0000
Wald test: sector dummies equ. 2	33.89 (10)	0.0002	37.92 (10)	0.0000

Table 5: GMM estimation of the (non-)response equations (N = 1019)

	MAR	
parameters	estimate	t-value
equation 3: innovation filter		
Intercept	-4.7062	-10.3687
East German firm	-1.0349	-2.6016
Log number of employees	0.5510	7.9871
Lagged log R&D	-0.6424	-6.4956
Qualified staff missing	0.8203	2.5928
equation 4: transitory nonresponse		
Intercept	-2.0887	-4.3739
East German firm	-0.6000	-2.6222
Lagged log number of employees	0.2053	2.5344
Lagged log R&D	-0.1105	-1.6610
High risk of imitation	0.4234	2.3969
equation 5: permanent nonresponse		
Intercept	-1.0243	-2.7485
East German firm	-0.1495	-0.9013
Lagged log number of employees	0.1363	2.0985
Lagged log R&D	-0.1464	-2.9277
Decreasing demand expected	0.2895	1.9731
hypotheses tests:	$\chi^2(\text{dof})$	p-value
Wald test: joint significance equation 3	78.70 (4)	0.0000
Wald test: joint significance equation 4	21.87 (4)	0.0002
Wald test: joint significance equation 5	14.45 (4)	0.0060

Contrary to the results predicted by the Cohen/Klepper model, the impact of firm size on product R&D is larger than the impact on process R&D. Equality of the firms size coefficients in the product and process equations is clearly rejected by a Wald test. These results holds regardless on the specification (MCAR or MAR) being used. Actually, the magnitude of the parameter estimates is very similar in both specifications of the (non-) response mechanism. The only exception is the coefficient of the dummy variable for East Germany which increases in absolute terms if selection on observables is taken into account. This variable is also an important predictor of the nonresponse reasons "innovation filter" and "transitory nonresponse" but turns out insignificant in the "permanent nonresponse" category.

7. Conclusion

This paper has suggested a possible way of dealing with nonresponse heterogeneity in panel data. The proposed GMM estimator simultaneously exploits the propensity-score-weighted moment functions of interest and the score of the ML estimator of a multinomial choice model for the propensity score estimation.

It should be emphasized that the general estimation procedure is not restricted to the attrition problem but can be applied to any other selection problem as well which is characterized by multiple selection possibilities (e.g. programme heterogeneity in the treatment effect literature, see e.g. Lechner, 2001).

The basic prerequisite of the proposed method is the availability of rich data sets which include all relevant information for the selection on observables assumption. A very interesting route for future research is the analysis of the performance of those estimators which rely on the propensity score under circumstances which are characterized by missing information on the set of relevant conditioning variables or by selection on unobservables which has been the dominant assumption in the econometric sample selection literature since Heckman's early contributions (e.g. Heckman, 1979).

The results in this paper indicate that the proposed estimator works well in experimental setups but behaves very similar to an estimator ignoring the attrition process in an application to real data. The latter result might be explained by the lack of the relevant variables driving the selection on observables assumption.

Appendix A

Proposition: $E\left[\tilde{\psi}_1\left(Z_t^3,\beta_0\right)\right]=E\left[\psi_1\left(Z_t^1,\beta_0\right)\right]$

Proof:

$$E\left[\psi_1\left(Z_t^1,\beta_0\right)\cdot\frac{D_t^0}{P\left(W_t\right)}\right]=E\left[E\left[\psi_1\left(Z_t^1,\beta_0\right)\cdot\frac{D_t^0}{P\left(W_t\right)}\mid W_t\right]\right]$$

$$=E\left[E\left[\frac{\psi_1\left(Z_t^1,\beta_0\right)}{P\left(W_t\right)}\mid W_t,D_t^0=1\right]\cdot\Pr\left[D_t^0=1\mid W_t\right]\right]$$

$$=E\left[E\left[\frac{\psi_1\left(Z_t^1,\beta_0\right)}{P\left(W_t\right)}\mid W_t\right]\cdot P\left(W_t\right)\right]=E\left[\psi_1\left(Z_t^1,\beta_0\right)\right]$$

MAR is used to obtain the first expression in the last row.

Appendix B

$$\pi_1\left(W_t,\gamma\right)=\Phi^{(3)}\left[W_t'\alpha^1,W_t'\frac{\alpha^1-\alpha^2}{\sqrt{1+\sigma_{22}-2\sigma_{21}}},W_t'\frac{\alpha^1-\alpha^3}{\sqrt{1+\sigma_{33}-2\sigma_{31}}};\Omega_1\right]$$

$$\pi_2\left(W_t,\gamma\right)=\Phi^{(3)}\left[W_t'\frac{\alpha^2}{\sqrt{\sigma_{22}}},W_t'\frac{\alpha^2-\alpha^1}{\sqrt{1+\sigma_{22}-2\sigma_{21}}},W_t'\frac{\alpha^2-\alpha^3}{\sqrt{\sigma_{22}+\sigma_{33}-2\sigma_{31}}};\Omega_2\right]$$

$$\pi_3\left(W_t,\gamma\right)=\Phi^{(3)}\left[W_t'\frac{\alpha^3}{\sqrt{\sigma_{33}}},W_t'\frac{\alpha^3-\alpha^1}{\sqrt{1+\sigma_{33}-2\sigma_{31}}},W_t'\frac{\alpha^3-\alpha^2}{\sqrt{\sigma_{33}+\sigma_{22}-2\sigma_{32}}};\Omega_3\right]$$

Here Ω_i denotes the correlation matrix corresponding to Σ_i, $i=1,2,3$, where Σ_i can be expressed in termini of the elements of the matrix Σ_0 defined in (20) as

$$\Sigma_i=\Sigma_0^{ii}+\begin{pmatrix}0 & & \\ -\Sigma_0^{ij} & \Sigma_0^{jj}-2\Sigma_0^{ij} & \\ -\Sigma_0^{ik} & \Sigma_0^{jk}-\Sigma_0^{ij}-\Sigma_0^{ik} & \Sigma_0^{kk}-2\Sigma_0^{ik}\end{pmatrix}$$

where a term like Σ_0^{ij} denotes the element in the i-th row and j-th column of Σ_0 and where j and k are the remaining two numbers from the set $[1,2,3]$ after deleting i.

Appendix C

To simulate J-variate probabilities of the form given in Appendix B (for $J = 3$) for $J > 3$ write the conditional choice probabilities for $j = 0, \ldots, J$ in general form as

$$\pi_j\left(W_t, \gamma\right) = E\left[D_t \mid W_t\right] = \Pr\left[\tilde{\varepsilon}_t^j < B_t^j \mid W_t\right] = \Phi^{(J)}\left(B_t^j, \Omega_j\right)$$

where $\tilde{\varepsilon}_t^j \sim N\left(0, \Omega_j\right)$ is a $J \times 1$ vector of differences in the elements of the error term vector ε_t^j of the random utility model, where the appropriate differences are implied by the J pairwise comparisons of the utilities. B_t^j is a $J \times 1$ vector of upper bounds of the multiple integrals over the density function of the multivariate standard normal distribution (see Appendix B for the exact bounds in the trivariate case). The GHK simulator is based on $J \times 1$ vectors of standard normal distributed random variables η^r satisfying $L_j\eta^r \sim N\left(0, \Omega_j\right)$ where L_j is chosen such that $L_j L_j' = \Omega_j$. Denote the elements of L_j and B_t^j by l and b, respectively, and write

$$\Pr\left[L_j\eta^r < B_t^j \mid W_t\right] = \Pr\left[\begin{pmatrix} l_{11} & & & & \\ l_{21} & l_{22} & & 0 & \\ l_{31} & l_{32} & l_{33} & & \\ & & \cdots & & \\ l_{J1} & l_{J2} & l_{J3} & \cdots & l_{JJ} \end{pmatrix}\begin{pmatrix} \eta_1^r \\ \eta_2^r \\ \eta_3^r \\ \vdots \\ \eta_J^r \end{pmatrix} < \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_J \end{pmatrix} \mid W_t\right]$$

which implies a triangular structure of the form

$$\begin{aligned} &\eta_1^r < b_1 / l_{11} \\ &\quad\vdots \\ &\eta_j^r < (b_j - l_{j1}\eta_1^r - l_{j2}\eta_2^r - \ldots - l_{j,j-1}\eta_{j-1}^r) / l_{jj} \\ &\quad\vdots \\ &\eta_J^r < (b_J - l_{J1}\eta_1^r - l_{J2}\eta_2^r - \ldots - l_{J,J-1}\eta_{J-1}^r) / l_{JJ} \end{aligned}$$

which can be used to generate estimators of $\pi_j\left(W_t, \gamma\right)$ by calculating the product of the single probabilities that each η_j^r falls in the respective truncated interval. Replicating this procedure R times to reduce the sampling variance yields the unbiased simulated response probability

$$\begin{aligned} \hat{\pi}_j\left(W_t, \gamma\right) &= \tfrac{1}{R}\sum_{r=1}^{R}\left(\Pr\left[\eta_1^r < b_1 / l_{11} \mid W_t\right] \cdot \ldots \cdot \Pr\left[\eta_J < (b_J - l_{11}\eta_1^r - l_{21}\eta_2^r - \ldots - l_{J-1,1}\eta_{J-1}^r) / l_{JJ} \mid W_t\right]\right) \\ &= \tfrac{1}{R}\sum_{r=1}^{R}\left(\Phi\left(b_1 / l_{11}\right) \cdot \ldots \cdot \Phi\left((b_J - l_{11}\eta_1 - l_{21}\eta_2 - \ldots - l_{J-1,1}\eta_{J-1}) / l_{JJ}\right)\right) \end{aligned}$$

(cf. Börsch-Supan and Hajivassiliou, 1993, for more details) which is based on products of the c.d.f. Φ of the univariate standard normal distribution which can be readily calculated.

Appendix D

Table A1: GMM estimation of the propensity score equations under MAR
Summary statistics from 1000 replications of Monte Carlo experiment E1 (N = 500)

parameter	true	mean	bias	se	rmse
α_1^1	-1.0000	-1.0036	-0.0036	0.4046	0.4046
α_2^1	1.0000	1.0147	0.0147	0.1582	0.1589
α_3^1	-1.0000	-1.0065	-0.0065	0.1826	0.1827
α_4^1	1.0000	0.9860	-0.0140	0.2866	0.2870
α_1^2	-1.0000	-1.0263	-0.0263	0.5244	0.5250
α_2^2	1.0000	1.0206	0.0206	0.2066	0.2076
α_3^2	-1.0000	-1.0153	-0.0153	0.2242	0.2248
α_4^2	1.0000	1.0002	0.0002	0.3713	0.3713
α_1^3	-1.0000	-1.0391	-0.0391	0.5207	0.5222
α_2^3	1.0000	1.0242	0.0242	0.2056	0.2070
α_3^3	-1.0000	-1.0220	-0.0220	0.2190	0.2201
α_4^3	1.0000	1.0061	0.0061	0.3510	0.3511
ℓ_{21}	0.5000	0.4893	-0.0107	0.3298	0.3300
ℓ_{22}	0.8660	0.8248	-0.0413	0.2969	0.2997
ℓ_{31}	0.5000	0.4868	-0.0132	0.3541	0.3544
ℓ_{32}	0.2887	0.2580	-0.0307	0.3178	0.3192
ℓ_{33}	0.8165	0.7315	-0.0850	0.2825	0.2950
σ_{21}	0.5000	0.4893	-0.0107	0.3298	0.3300
σ_{22}	1.0000	1.1163	0.1163	0.6176	0.6285
σ_{31}	0.5000	0.4868	-0.0132	0.3541	0.3544
σ_{32}	0.5000	0.5228	0.0228	0.4509	0.4514
σ_{33}	1.0000	1.1445	0.1445	0.6374	0.6536

Note: An estimate of the variance-covariance matrix Σ_0 was obtained by optimizing with respect to the elements ℓ_{ij} of the lower triangular matrix L such that $LL' = \Sigma_0$. The elements σ_{ij} of Σ_0 were obtained afterwards by means of the delta method.

Table A2: GMM estimation of the propensity score equations under MAR
Summary statistics from 1000 replications of Monte Carlo experiment E2 (N = 500)

parameter	true	mean	bias	se	rmse
α_1^1	-1.0000	-0.9911	0.0089	0.4182	0.4183
α_2^1	1.0000	1.0208	0.0208	0.1579	0.1592
α_3^1	-1.0000	-1.0204	-0.0204	0.1778	0.1790
α_4^1	1.0000	0.9912	-0.0088	0.2943	0.2944
α_1^2	-1.0000	-1.0299	-0.0299	0.5314	0.5323
α_2^2	1.0000	1.0287	0.0287	0.2027	0.2047
α_3^2	-1.0000	-1.0275	-0.0275	0.2227	0.2244
α_4^2	1.0000	1.0076	0.0076	0.3657	0.3657
α_1^3	-1.0000	-1.0457	-0.0457	0.5523	0.5542
α_2^3	1.0000	1.0303	0.0303	0.2038	0.2060
α_3^3	-1.0000	-1.0251	-0.0251	0.2177	0.2192
α_4^3	1.0000	1.0077	0.0077	0.3728	0.3729
ℓ_{21}	0.5000	0.4987	-0.0013	0.3524	0.3524
ℓ_{22}	0.8660	0.8336	-0.0324	0.2896	0.2914
ℓ_{31}	0.5000	0.4967	-0.0033	0.3308	0.3309
ℓ_{32}	0.2887	0.2600	-0.0286	0.3153	0.3166
ℓ_{33}	0.8165	0.7262	-0.0903	0.3118	0.3246
σ_{21}	0.5000	0.4987	-0.0013	0.3524	0.3524
σ_{22}	1.0000	1.1515	0.1515	0.6384	0.6562
σ_{31}	0.5000	0.4967	-0.0033	0.3308	0.3309
σ_{32}	0.5000	0.5328	0.0328	0.4510	0.4522
σ_{33}	1.0000	1.1474	0.1474	0.6318	0.6488

Note: See Table A1.

Table A3: GMM estimation of the propensity score equations under MAR
Summary statistics from 500 replications of Monte Carlo experiment E2 (N = 2000)

parameter	true	mean	bias	se	rmse
α_1^1	-1.0000	-1.0081	-0.0081	0.1917	0.1919
α_2^1	1.0000	1.0000	0.0000	0.0703	0.0703
α_3^1	-1.0000	-0.9998	0.0002	0.0864	0.0864
α_4^1	1.0000	0.9996	-0.0004	0.1345	0.1345
α_1^2	-1.0000	-1.0041	-0.0041	0.2323	0.2323
α_2^2	1.0000	1.0006	0.0006	0.0905	0.0905
α_3^2	-1.0000	-1.0014	-0.0014	0.1037	0.1037
α_4^2	1.0000	0.9977	-0.0023	0.1636	0.1636
α_1^3	-1.0000	-1.0082	-0.0082	0.2332	0.2334
α_2^3	1.0000	1.0007	0.0007	0.0920	0.0920
α_3^3	-1.0000	-1.0013	-0.0013	0.1030	0.1030
α_4^3	1.0000	0.9981	-0.0019	0.1611	0.1611
ℓ_{21}	0.5000	0.4928	-0.0072	0.1531	0.1533
ℓ_{22}	0.8660	0.8545	-0.0115	0.1383	0.1388
ℓ_{31}	0.5000	0.4865	-0.0135	0.1647	0.1652
ℓ_{32}	0.2887	0.2808	-0.0079	0.1493	0.1495
ℓ_{33}	0.8165	0.8004	-0.0160	0.1346	0.1356
σ_{21}	0.5000	0.4928	-0.0072	0.1531	0.1533
σ_{22}	1.0000	1.0155	0.0155	0.2562	0.2567
σ_{31}	0.5000	0.4865	-0.0135	0.1647	0.1652
σ_{32}	0.5000	0.4965	-0.0035	0.1813	0.1813
σ_{33}	1.0000	1.0236	0.0236	0.2621	0.2632

Note: See Table A1.

Appendix E

Table A4: Descriptive statistics for respondents (A = 0)

	1993		1994	
	mean	std.dev.	mean	std.dev.
Log process R&D expenditures	-1.8422	1.9346	-1.7642	1.8734
Log product R&D expenditures	-0.7736	2.2660	-0.7584	2.2167
Log number of employees	5.4598	1.6325	5.3433	1.6384
East German firm	0.2905	0.4542	0.3356	0.4728
R&D department	0.5594	0.4967	0.5931	0.4918
Observations	1019		435	

Table A5: Means of attributes by (non-)response indicator

1994	A = 0	A = 1	A = 2	A = 3
East German firm	0.3356	0.2034	0.1977	0.2946
Log number of employees	5.3433	4.7357	-	-
Lagged log number of employees	5.4248	4.8047	5.7580	5.4664
Log process R&D expenditures	-1.7472	-2.5157	-1.6755	-1.9265
Log product R&D expenditures	-0.6227	-2.0830	-0.4628	-0.8918
Lagged R&D department	0.5885	0.2712	0.5988	0.5524
High risk of imitation	0.2805	0.3729	0.3953	0.3229
Qualified staff missing	0.2046	0.4068	0.2791	0.2578
External funds missing	0.4667	0.4237	0.3663	0.4334
Decreasing employment expected	0.4161	0.4915	0.4826	0.4448
Decreasing sales expected	0.2207	0.3220	0.3081	0.2550
Decreasing demand expected	0.2161	0.3390	0.3031	0.3031
Observations	435	59	172	353

References

Abowd, J. M., B. Crepon, and F. Kramarz (1997): "Moment Estimation with Attrition", NBER technical working paper no. 214.

Angrist, J. D. (1997): "Conditional Independence in Sample Selection Models", *Economics Letters*, 54, 103-112.

Börsch-Supan, A., and V. A. Hajivassiliou (1993): "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models", *Journal of Econometrics*, 58, 347-368.

Börsch-Supan, A., V. A. Hajivassiliou, L. J. Kotlikoff, and J. N. Morris (1993): "Health, Children, and Elderly Living Arrangements: A Multiperiod-Multinomial Probit Model with Unobserved Heterogeneity and Autocorrelated Errors", NBER working paper no. 3343.

Breitung, J., and M. Lechner (1999): "Alternative GMM Methods for Nonlinear Panel Data Models", in: L. Matyas (ed.), *Generalized Method of Moments Estimation*, Cambridge: Cambridge University Press.

Cassel, C.-M., C.-E. Särndal, and J. H. Wretman (1983): "Some Uses of Statistical Models in Connection with the Nonresponse Problem", in: W. G. Madow (ed.): *Incomplete Data in Sample Surveys*, New York: Academic Press, 143-160.

Cohen, W. M., and S. Klepper (1996): "Firm Size and the Nature of Innovation within Industries: The Case of Process and Product R&D", *Review of Economics and Statistics*, 78, 232-243.

Czaja, J. L., S. M. Hirabayashi, R. J. A. Little, and D. B. Rubin (1992): "Projecting from Advance Data Using Propensity Modeling: An Application to Income and Statistics", *Journal of Business & Economic Statistics*, 10 (2), 117-131.

Dehejia, R. H., and S. Wahba (1999): "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs", *Journal of the American Statistical Association*, 94 (448), 1053-1062.

Dawid, A. P. (1979): "Conditional Independence in Statistical Theory", *Journal of the Royal Statistical Society*, Series B, 41, 1-31.

Fitzgerald, J., P. Gottschalk, and R. Moffitt (1998): "An Analysis of Sample Attrition in Panel Data", *Journal of Human Resources*, 23 (2), 251-299.

Fritsch, M., and M. Meschede (2000): "Product Innovation, Process Innovation, and Size", working paper, Technical University Bergakademie Freiberg.

Gerfin, M., and M. Lechner (2000): "Microeconometric Evaluation of the Active Labour Market Policy in Switzerland", working paper, Universities of Bern and St. Gallen.

Geweke, J., M. Keane, and D. Runkle (1994): "Alternative Computational Approaches to Inference in the Multinomial Probit Model", *Review of Economics and Statistics*, 76 (4), 609-632.

Hajivassiliou, V. A., D. McFadden, and P. A. Ruud (1996): "Simulation of Multivariate Normal Rectangle Probabilities and their Derivatives: Theoretical and Computational Results", *Journal of Econometrics*, 72, 85-134.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 50 (4), 1029-1054.

Harhoff, D., and G. Licht (1994): "Das Mannheimer Innovationspanel", in: U. Hochmuth and J. Wagner (eds.), *Firmenpanelstudien in Deutschland, konzeptionelle Überlegungen und empirische Analysen*, 255-284, Tübinger Volkswirtschaftliche Schriften, Tübingen.

Harhoff, D., K. Stahl, and M. Woywode (1998): "Legal Form, Growth and Exit of West German Firms: Empirical Results for Manufacturing, Trade and Service Industries", *Journal of Industrial Economics*, 46 (4), 453-488.

Hausman, J. A., and D. A. Wise (1979): "Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment", *Econometrica*, 47 (2), 455-473.

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error", *Econometrica*, 47, 153-162.

Heckman, J. J., and V. J. Hotz (1989): "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs", *Journal of the American Statistical Association*, 84, 862-874.

Heckman, J. J., and R. Robb (1985): "Alternative Methods for Evaluating the Impact of Interventions", in: J. J. Heckman and B. Singer (eds.): *Longitudinal Analysis of Labor Market Data*, New York: Cambridge University Press, 156-245.

Hirano, K., G. W. Imbens, G. Ridder, and D. B. Rubin (1999): "Combining Panel Data Sets with Attrition and Refreshment Samples", NBER technical working paper no. 230, forthcoming in *Econometrica.*

Holland, P. W. (1986): "Statistics and Causal Inference", *Journal of the American Statistical Association*, 81, 945-970.

Horowitz, J. L. (1993): "Semiparametric and Nonparametric Estimation of Quantal Response Models", in: G. S. Maddala, C. R. Rao, and H. D. Vinod (eds.), *Handbook of Statistics*, vol. 11, Amsterdam: North-Holland.

Horowitz, J. L., and C. F. Manski (1998): "Censoring of Outcomes and Regressors Due to Survey Nonresponse: Identification and Estimation Using Weights and Imputation", *Journal of Econometrics*, 84, 37-58.

Horvitz, D., and D. Thompson (1952): "A Generalization of Sampling Without Replacement from a Finite Population", *Journal of the American Statistical Association*, 47, 663-685.

Imbens, G. W. (2000): "The Role of the Propensity Score in Estimating Dose-Response Functions", *Biometrika*, 87 (3), 706-710.

Imbens, G. W., R. H. Spady, and P. Johnson (1998): "Information-Theoretic Approaches to Inference in Moment Condition Models", *Econometrica*, 66 (2), 333-358.

Inkmann, J. (2000a): "Misspecified Heteroskedasticity in the Panel Probit Model: A Small Sample Comparison of GMM and SML Estimators", *Journal of Econometrics*, 97, 227-259.

Inkmann, J. (2000b): "Finite Sample Properties of One-step, Two-step, and Bootstrap Empirical Likelihood Approaches to Efficient GMM Estimation", CoFE discussion paper no. 00/03.

Inkmann, J. (2001): *Conditional Moment Estimation of Nonlinear Equation Systems - With an Application to an Oligopoly Model of Cooperative R&D*, Lecture Notes in Economics and Mathematical Systems, Vol. 497, Heidelberg: Springer.

Keane, M. P. (1992): "A Note on Identification in the Multinomial Probit Model", *Journal of Business & Economic Statistics*, 10 (2), 193-201.

Klepper, S. (1996): "Entry, Exit, Growth, and Innovation over the Product Life Cycle", *American Economic Review*, 86 (3), 562-583.

Lechner, M. (1995): *Ökonometrische Modelle für qualitative Paneldaten: Das Beispiel der selbständigen Erwerbstätigkeit in den neuen Bundesländern*, Frankfurt: Campus.

Lechner, M. (1999a): "Nonparametric Bounds on Employment and Income Effects of Continuous Vocational Training in East Germany", *Econometrics Journal*, 2, 1-28.

Lechner, M. (1999b): "Programme Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labour Market Policies, working paper, University of St. Gallen.

Lechner, M. (2001): "Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption", in: M. Lechner and F. Pfeiffer (eds.), *Econometric Evaluation of Labour Market Policies*, Heidelberg: Physika, 43-58.

Lechner, M., and J. Breitung (1996): "Some GMM Estimation Methods and Specification Tests for Nonlinear Models", in: L. Mátyás and P. Sevestre (eds.), *The Econometrics of Panel Data*, 2nd edition, Dordrecht: Kluwer, ch. 22.

Lee, M.-J. (1996): *Method of Moments and Semiparametric Econometrics for Limited Dependent Variable Models,* New York: Springer.

Little, R. J. A. (1986): "Survey Nonresponse Adjustments for Estimates of Means", *International Statistical Review*, 54 (2), 139-157.

Little, R. J. A. (1988): "Missing-Data Adjustments in Large Surveys", *Journal of Business & Economic Statistics*, 6 (3), 287-301.

Little, R. J. A., and D. B. Rubin (1987): *Statistical Analysis with Missing Data*, New York: Wiley.

Little, R. J. A., and N. Schenker (1995): "Missing Data", in: G. Arminger, C. C. Clogg, and M. E. Sobel (eds.), *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, New York: Plenum Press, 39-75.

McFadden, D. (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration", *Econometrica*, 57 (5), 995-1026.

Newey, W. K. (1984): "A Method of Moments Interpretation of Sequential Estimators", *Economics Letters*, 14, 201-206.

Newey, W. K., and D. L. McFadden (1994): "Large Sample Estimation and Hypothesis Testing", in: R. F. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, Amsterdam: North-Holland, 2111-2245.

Pakes, A., and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators", *Econometrica*, 57 (5), 1026-1057.

Rendtel, U. (1995): *Lebenslagen im Wandel: Panelausfälle und Panelrepräsentativität*, Frankfurt: Campus.

Ridder, G. (1990): "Attrition in Multi-Wave Panel Data", in: J. Hartog, G. Ridder, J. Theeuwes (eds.): *Panel Data and Labor Market Studies*, Amsterdam: North-Holland, 45-68.

Rosenbaum, P. R., and D. B. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika*, 70, 41-55.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1995): "Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data", *Journal of the American Statistical Association*, 90, 106-121.

Rubin, D. B. (1976): "Inference and Missing Data", *Biometrika*, 63, 581-592.

Rubin, D. B. (1977): "Assignment to Treatment Group on the Basis of a Covariate", *Journal of Educational Statistics*, 66, 688-701.

Schnell, R. (1997): *Nonresponse in Bevölkerungsumfragen*, Opladen: Leske + Budrich.

Ziegler, A., and A. Eymann (2000): "Zur Simulated Maximum-Likelihood-Schätzung von Mehrperioden-Mehralternativen-Probitmodellen", Beiträge zur angewandten Wirtschaftsforschung, No. 588-00, Universität Mannheim.