Panel Data Econometrics: Modelling and Estimation

Olaf Hübler

Diskussionspapier Nr. 319
August 2005
ISSN 0949-9962

Address:
University of Hannover
Institute of Quantitative Economic Research
Königsworther Platz 1
D-30167 Hannover
huebler@mbox.iqw.uni-hannover.de

Abstract:
This paper presents a survey on panel data methods in which I emphasize new developments. In particular, linear multilevel models with a new variant are discussed. Furthermore, nonlinar, nonparametric and semiparametric models are analyzed. In contrast to linear models there do not exist unified methods for nonlinear approaches. In this case FEM are dominated by CML methods. Under REM assumptions it is often possible to use the ML method directly. As alternatives GMM and simulated estimators exist. If the nonlinear function is not exactly known, nonparametric or semiparametric methods should be preferred.

Zusammenfassung: Der Beitrag gibt einen Überblick über Methoden zur Paneldatenanalyse, wobei neuere Methoden im Vordergrund stehen. Insbesondere werden lineare Mehrebenenmodelle unter Berücksichtigung einer neuen Variante diskutiert. Außerdem ist die Analyse auf nichtlineare, nicht- und semiparametrische Verfahren ausgerichtet. Im Gegensatz zu linearen Modellen existiert bei nichtlinearen Ansätzen keine einheitliche Schätzstrategie. Im Falle von Fixed-Effects-Modellen dominiert die bedingte ML-Methode. Unter den Annahmen eines Random-Effects-Ansatzes ist es oft möglich, die ML-Methode direkt zu nutzen. Alternativen bilden GMM-Schätzer und simulierte Schätzer. Wenn die nichtlineare Funktion nicht genau bekannt ist, sind nicht- oder semiparametrische Schätzer zu präferieren.

JEL-Classification: C14, C23, C24, C25
Keywords: panel data, linear multilevel, nonlinear, non- and semiparametric models

1 Introduction

The use of panel data regression methods has become popular as the availability of longitudinal data sets has increased. Panel data usually contain a large number of cross sectional units (individuals, households, companies, regions, countries), which are repeated observed over time. The advantages of panel data compared with cross sectional data on the one one hand and time series data on the other hand are the following: Firstly, the large number of observations give more informative data, less multicollinearity, more degrees of freedom and a higher efficiency of econometric estimates. Secondly, it is possible to separate between cohort, period and age effects. Thirdly, the analysis can determine intra- and interindividual effects. Fourthly, panel data allow researchers to control for unobserved heterogeneity. Pure time series and pure cross sectional data are usually contaminated by these effects. Fifthly, longitudinal observations improve the possibilities of evaluating the effects of policy interventions and it is possible to determine under which conditions the effects can be interpreted as causal effects. The availability of panel data allows us to estimate treatment effects consistently without assuming ignorability of treatment and without an instrumental variable, provided the treatment varies over time and is uncorrelated with time-varying unobservables that affect the response. Problems with panel data are possible by attrition, time-varying sample size and structural changes. Furthermore, it must be considered that a panel data survey is very cost-intensive.

A wide range of methods was developed in the last 40 years. The major contributions to this topic have been in four volumes edited by MADDALA (1993) and BALTAGI (2002). Furthermore, recent textbooks and surveys of panel econometrics exist (ARELLANO 2003, ARELLANO/HONORE 2001, BALTAGI 2005, CAMERON/TRIVEDI 2005, HSIAO 2004, HÜBLER 2003, LECHNER 2002, LEE 2002, MATYAS/SEVESTRE 1996, WOOLDRIDGE 2002). The modelling of panel data approaches distinguishes in the time dependence, in the assumptions of the error term and in the measurement of dependent variables. Due to the specific assumption, consequences for the estimation methods follow. Beside the classical methods in panel data econometrics like least squares and maximum likelihood estimators we find conditional and quasi ML estimators, GEE (generalized estimating equations), GMM (generalized methods of moments), simulated, non- and semiparametric estimators. For linear panel data models with predetermined regressors we can apply conventional techniques. The main objective is to eliminate and determine unobserved heterogeneity. Two situations are distinguished: regressors and unobserved heterogeneity are independent or interact. Much less is known about nonlinear models. Simple first differences methods but also conditional likelihood approaches fail for many models to eliminate unobserved heterogeneity. As the specification of nonlinearity is often unknown non- and semiparametric methods are preferred.

We distinguish between several types of panel data models and proceed from general to more specific models. The endogenous variable y_{it} can be determined by observed exogenous time invariant ($\tilde{x}_i$) and time-varying (x_{it}) variables, unobserved time invariant regressors ($\alpha_{i.}$) and a time-varying error term (u_{it}). The term $m(\cdot)$ tells us that the functional relation is unknown, i.e. nonparametric approaches are formulated, which may vary between the periods (m_t). If y_{it} is not directly determined by $\tilde{x}_i$, x_{it}, $\alpha_{i.}$ and u_{it} but across an unobservable variable we call this a latent model, expressed by $g(\cdot)$. This relation may also be time-varying ($g_t(\cdot)$). Furthermore, we assume that besides the individual and the time dimension

$(i = 1, ..., N; t = 1, ..., T)$ L further levels such like firms, sectoral or regional level are observable which can induce specific effects:

- Generalized time-varying multilevel nonparametric latent model

$$y_{iLt} = g_t[m_t(\tilde{x}_i L, x_{it}, \alpha_{iL}, u_{iLt})]$$

  The major problem to handle this model is the high nonlinear dimensionality. Therefore, it is helpful to consider more specific models:

- Generalized additive time-varying nonparametric latent model

$$y_{it} = g[m_{1t}(x_{it}) + m_{2t}(\alpha_{iL}) + u_{it}]$$

  The determinants are additively separated in observed variables, unobserved individual terms and the error term. Additionally, observed time invariant determinants may be incorporated ($\sum m_{3t}(\tilde{x})$). Nevertheless, also this type is still to general for practical estimations.

- Additive nonparametric multilevel model

$$y_{it} = \sum_{k=1}^{K} m_1(x_{it}) + \sum_{l=1}^{L} \alpha_{il} + u_{it}$$

  In contrast to the latter approach the effects are additively split within one variable group. Now, the curse of dimensionality is solved.

- Partial linear model

$$y_{it} = m_1(x_{i1t}) + z'_{it}\gamma + \alpha_{i1} + u_{it}$$

  As an alternative to the additive formulation of observed variables now only one or few regressors are described by an unspecific nonparametric form while the remainder is defined by a linear combination.

- Linear model with individual effects and time varying coefficients

$$y_{it} = x'_{it}\beta_t + \alpha_{i1} + u_{it}$$

  This is the conventional panel data model defined by an unobserved individual effect and time-varying coefficients.

- Linear model

$$y_{it} = x'_{it}\beta + u_{it}.$$

  This represents the most simple panel data model without any specific panel character.

The last three types can be found in applications, whereas the other types have, until now, only been used in more specific formulations.

2 Parametric Models

2.1 Linear Models

Standard panel data analysis starts with linear models

$$y_{it} = x'_{it}\beta + u_{it} \qquad i = 1, ..., N \quad t = 1, ..., T \tag{1}$$

where y is the dependent variable, x is a K×1 regressor vector, β is a K×1 vector of coefficients and u is the error term. The number of cross sectional observations is N and these units are repeatedly measured. When the cross sectional data are only pooled over T periods, the coefficients can be estimated by OLS under classical assumptions about the error term. The advantage of panel data in this case should be improved efficiency. But let us note that in very large samples too many effects of regressors are significant. This "curse of large number" should be considered in significance comparisons with cross sectional estimates.

Some modifications follow if cross sectional heteroscedastic or cross sectional correlated and time-wise autoregressive models are analyzed with panel data. Conventional EGLS estimates are the consequence. But the main objective of all panel data analyzes is the consideration of unobserved heterogeneity and its estimation. The methods which are developed for this purpose depend on the assumptions of the error term, the regressand, the regressors and the coefficients of the model. Furthermore, it is important to recognize whether all cross sectional units are observed in the same periods. Sometimes there will be gaps in the data. Some individuals have dropped out and can no longer be included in the panel and new entrants have emerged over the sample period. Some panel data sets cannot be collected every period due to lack of resources or cut in funding. This missing value problem leads to unbalanced panels. E.g. WANDSBEEK and KAPTEYN (1989) or DAVIS (2002) study various methods for this unbalanced model. In many cases, methods developed for a balanced panel can also be employed for an unbalanced panel with only slight modifications as long as the incompleteness is due to randomly missing observations.

2.1.1 Individual effects under alternative specifications

Conventional panel data analysis starts with a simple linear model and an unobserved time invariant individual term α_i

$$y_{it} = x'_{it}\beta + \alpha_i + \epsilon_{it} =: x'_{it}\beta + u_{it}. \tag{2}$$

The basic task is to estimate the unobserved heterogeneity. If we assume under classical conditions of ϵ_{it} that α_i is uncorrelated with ϵ_{it} and the observed regressors x_{it}, we call this a "random effects model - REM" or "error component model". Standard procedures exist to separate the estimation of β and α_i. In a random effects approach the distribution of α_i is parameterized. This makes the model fully parametric. BALESTRA and NERLOVE (1966) were amongst the first who solved this problem.

The coefficient vector β can be determined by OLS of the transformed model

$$y_{it} - \hat{\delta}\bar{y}_i = (x_{it} - \hat{\delta}\bar{x}_i)'\beta^* + u_{it} - \hat{\delta}\bar{u}_i, \tag{3}$$

where

$$\hat{\delta} = 1 - \frac{\hat{\sigma}_\epsilon^2}{\hat{\sigma}_\epsilon^2 + T\hat{\sigma}_\alpha^2}.$$

The variance of $\hat{\epsilon}$ can be estimated by the residuals of the within estimator, the OLS estimator of (4), and the estimated variance of $\hat{\alpha}$ follows by

$$\hat{\sigma}_\alpha^2 = \frac{1}{N-K}\sum_{i=1}^{N}\hat{\bar{u}}^2 - \frac{1}{T}\hat{\sigma}_\epsilon^2,$$

where the average residuals $\hat{\bar{u}}$ are determined by the between estimator, i.e. the OLS estimator of

$$\bar{y}_i = \sum_{k=1}^{K}\bar{x}_{ik}\beta_k + \bar{u}_i.$$

BREUSCH and PAGAN (1980) have developed a Lagrange multiplier test based on OLS residuals to determine whether random effects exist. WOOLDRIDGE (2002, 265) prefers a statistic which is distributed asymptotically as standard normal, also based on residuals, but without an assumption of the error term distribution.

If α_i and the regressors are correlated we call the approach "fixed effects model - FEM". In this case one attempts to find ways to estimate the parameters making only minimal assumptions on α_i which can be eliminated by a first difference estimator or a within estimator. In the latter case the weighted LS estimator

$$\beta_W = [\sum_{i=1}^{N} X_i'QX_i]^{-1}\sum_{i=1}^{N} X_i'Qy_i$$

follows where $Q = I_T - \frac{1}{T}\iota_T\iota_T'$. This coincides with the OLS estimator of the transformed model

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)'\beta^* + u_{it} - \bar{u}_i, \tag{4}$$

where β^* is the coefficient vector without a constant term. The estimated individual effect is

$$\hat{\alpha}_i = (y_{it} - \bar{y}_i) - (x_{it} - \bar{x}_i)'\hat{\beta}^*. \tag{5}$$

The significance of individual effects can be tested by a conventional F test where the residuals of the pooled model are compared with those of the model that incorporates the individual effect. Whether the REM or the FEM is preferred can be decided by a Hausman test (HAUSMAN 1978). In practice, we can integrate into the transformed REM (3) the transformed FEM

$$y_{it} - \hat{\delta}\bar{y}_i = (x_{it} - \hat{\delta}\bar{x}_i)'\beta + (x_{it} - \bar{x}_i)'\tilde{\beta} + \omega_{it} \tag{6}$$

and test whether $\tilde{\beta}$ is significantly different from zero.

The model can be extended by fixed time effects α_t. An elimination by a within estimator is usually not necessary because the number of periods is small and therefore the time effects can directly be estimated jointly with the coefficient vector β

$$y_{it} = x_{it}'\beta + \iota_t'\xi + \alpha_i + \epsilon_{it},$$

where ι is a unit vector and $\xi = (\xi_1, \xi_2, ..., \xi_t)'$. The next step to more advanced panel data models is to have fewer restricted assumptions of the error term. First order serial correlation and heteroscedasticity can be allowed. For example, a four step procedure solves the problem of determining the autocorrelation coefficient, ρ, and β where two transformations are necessary (HÜBLER 1990). If lagged dependent or jointly dependent variables or errors in the exogenous variables exist a GMM estimator suggested by ARRELANO and BOND (1991) is preferred (HÜBLER 1990, ARRELANO 2003). Alternatives are usually less efficient or even inconsistent. If, for example, the simple lagged dependent model exists

$$y_{it} = \gamma y_{i,t-1} + \alpha_i + \varepsilon_{it} = \gamma y_{i,t-1} + u_{it},$$

conventional first differences eliminate the individual term

$$y_{it} - y_{i,t-1} = \gamma(y_{i,t-1} - y_{i,t-2}) + \varepsilon_{it} - \varepsilon_{i,t-1},$$

however, an OLS estimator of the first differences is biased and even inconsistent under autocorrelation. Instrumental variables solve the problem. GMM estimators, which incorporate all valid instruments, are preferred compared to simple IV estimators. Valid instruments are those which are uncorrelated with the error term. This means, if observations from T waves are available, valid instruments Z fulfil the following orthogonality conditions

$$E\begin{bmatrix} du_{i3} \cdot dy_{i1} \\ du_{i4} \cdot dy_{i1} \\ du_{i4} \cdot dy_{i2} \\ \vdots \\ du_{iT} \cdot dy_{i,1} \\ \vdots \\ du_{iT} \cdot dy_{i,T-2} \end{bmatrix}_{m\times 1} = E(Z_i' du_i) = E(\xi_i) = 0$$

$$\begin{aligned} Z_i &= \text{diag}[dy_{i1}, \ldots, dy_{is}] \qquad\qquad s = 1, \ldots, T-2 \\ &= \begin{bmatrix} dy_{i1} & & & \\ & dy_{i1} dy_{i2} & & \\ & & \ddots & \\ & & & dy_{i1} \ldots dy_{i,T-2} \end{bmatrix}_{(T-2)\times m} \end{aligned}$$

$$du_i = [du_{i3}, \ldots, du_{iT}]',$$

where $du_{it} = u_{it} - u_{i,t-1}, dy_{i1} = y_{i1}, dy_{it} = y_{it} - y_{i,t-1}$ for $t = 2, \ldots, T-2$.

Coefficients are determined by a minimum distance estimator

$$\arg\min_{\gamma}(\xi - \mathrm{E}(\xi_i))' W (\xi - \mathrm{E}(\xi_i)) = \arg\min_{\gamma} \xi' W \xi, \tag{7}$$

where W is a weighting matrix and estimation of γ is based on the empirical moments

$$\xi = \widehat{\mathrm{E}(\xi_i)} = (1/N) \sum_{i=1}^{N} Z_i' du_i = (1/N) Z' du.$$

$Z = (Z_1', \ldots, Z_N')'$ is a $N(T-2) \times m$ matrix and $du = (du_1, \ldots, du_N)'$ is a $N(T-2) \times 1$ vector. The GMM estimator is

$$\arg\min_{\gamma} \xi' W \xi = \hat{\gamma}_{GMM} = \frac{dy_{-1}' Z V_N^{-1} Z' dy}{dy_{-1}' Z V_N^{-1} Z' dy_{-1}}, \tag{8}$$

where the optimal weighting matrix W is the inverse of the covariance matrix of ξ_i

$$\hat{V}_N^{-1} = \left[\frac{1}{N} \sum_{i=1}^{N} \hat{\xi}_i \hat{\xi}_i' \right]^{-1} = \left[\frac{1}{N} \sum_{i=1}^{N} Z_i' d\hat{u}_i d\hat{u}_i' Z_i \right]^{-1}.$$

For practical applications we need preliminary consistent estimates of the error term.

This procedure can be extended to models with additional regressors

$$\begin{aligned} y_{it} &= \gamma y_{i,t-1} + x_{it}'\beta + \alpha_i + \varepsilon_{it} \\ &= \tilde{x}_{it}'\delta + u_{it}. \end{aligned} \tag{9}$$

Under the assumptions

$$E(\tilde{x}_{kit}\varepsilon_{it'}) \begin{cases} = 0 & \text{for } t' \geq t \\ \neq 0 & \text{otherwise} \end{cases} \quad k = 1, \ldots, K,$$

which means that all regressors and error terms of the same period or later are uncorrelated, we obtain as optimal instrumental variable matrix

$$Z_i = \operatorname{diag}\left[dy_{i1}, \ldots, dy_{is}, dx_{i1}', \ldots, dx_{i,s+1}' \right] \qquad \begin{array}{l} i = 1, \ldots, N \\ s = 1, \ldots, T-2 \end{array}. \tag{10}$$

Then the GMM estimator is

$$\hat{\delta}_{GMM} = \left[(D\tilde{X})' Z \hat{V}_N^{-1} Z' (D\tilde{X}) \right]^{-1} (D\tilde{X})' Z \hat{V}_N^{-1} Z' dy, \tag{11}$$

where matrix $D\tilde{X}$ is composed of $(T-2)N \times K$ elements of $d\tilde{x}_{it}$.

Thus far, we have assumed that all coefficients including the individual effect are time invariant. The following modification is less restrictive

$$y_{it} = \gamma_t y_{i,t-1} + x_{it}'\beta_t + \psi_t \mu_i + \varepsilon_{it}. \tag{12}$$

Time-varying individual effects $\psi_t \mu_i$ seem undesirable, because in conventional models the individual effects α_i is defined in contrast to the idiosyncratic error term ϵ_{it} as time invariant. Nevertheless, it can be argued that the effect varies e.g. with cyclical ups and downs, although individual characteristics stay the same. Conventional differences or within estimators do not eliminate the individual effect. CHAMBERLAIN (1984, p.1263) suggests a solution to determine the coefficients in (12). This equation in period t-1 is multiplied by $r_t = \psi_t / \psi_{t-1}$ and this expression is subtracted from (12), i.e. $y_{it} - r_t y_{i,t-1}$. So therefore the individual effect vanishes. A simultaneous equation system results

$$\begin{aligned} y_{i3} &= (\gamma_3 + r_3) y_{i2} - r_3 \gamma_2 y_{i1} + x_{i3}'\beta_3 - r_3 x_{i2}'\beta_2 + (\varepsilon_{i3} - r_3 \varepsilon_{i2}) \\ y_{i4} &= (\gamma_4 + r_4) y_{i3} - r_4 \gamma_3 y_{i2} + x_{i4}'\beta_4 - r_4 x_{i3}'\beta_3 + (\varepsilon_{i4} - r_4 \varepsilon_{i3}) \\ &\vdots \end{aligned}$$

$$y_{iT} \quad = (\gamma_T + r_T)y_{i,T-1} - r_T\gamma_{T-1}y_{i,T-2} + x'_{iT}\beta_T$$
$$-r_T x'_{i,T-1}\beta_{T-1} + (\varepsilon_{iT} - r_T\varepsilon_{i,T-1}) \qquad (13)$$

Equation y_{i1} and y_{i2} are suppressed as we have no information about y_{i0} and $y_{i,-1}$. As the error terms and the lagged dependent variables are correlated instrumental variables are used. Therefore, y_{i3} is also eliminated, because not enough instruments can be constructed. Altogether, the equations of period 4 to T can only be used to determine the coefficients. This residual system is expressed by

$$Y = WB + \varepsilon, \qquad (14)$$

where Y is a $(T-3)N\times 1$ vector, B is a $(T-3)(3+2(K-1))\times 1$ vector, $W = \text{diag}(W'_4, \ldots, W'_T)$ is a $(T-3)N \times (T-3)[3+2(K-1)]$ block diagonal matrix, $W_t = (y_{t-1}, y_{t-2}, 1, x'_{t-1}, x'_{t-2}) = (Y^*_t, X^*_t)$ is a $N \times (3+2(K-1))$ matrix. HOLTZ-EAKIN/NEWEY/ROSEN (1988) develop a twostep procedure, which combines the GMM estimator and the twostep instrumental variables estimator (2SIV) suggested by WHITE (1982). In the first step a conventional 2SLS estimator can be applied as an IV estimator. Based on these residuals an estimated covariance matrix $(\hat{V})$ is calculated and used in the second step. The instrumental variable model $Z'y = Z'X\beta + Z'u$ is estimated by EGLS. The White estimator is

$$\hat{\beta}_{2SIV} = (X'Z\hat{V}^{-1}Z'X)^{-1}X'Z\hat{V}^{-1}Z'y.$$

When the regressors are also correlated with the error term, the instrumental variable matrix is

$$Z_t = (y_{t-2}, \quad y_{t-1}, \quad 1, \quad X_{t-2}, \quad X_{t-1}) \sim N \times (3+2(K-1)).$$

Then a two-step approach follows:
Step 1:
Equation-wise the 2SLS estimator of the instrumental approach is determined:

$$Z'_t y_t = Z'_t W_t B_t + Z'_t \varepsilon_t \qquad t = 4, \ldots, T.$$

$$\hat{B}_t = (W'_t Z_t (Z'_t Z_t)^{-1} Z'_t W_t)^{-1} W'_t Z_t (Z'_t Z_t)^{-1} Z'_t y_t$$

Step 2:
The simultaneous EGLS estimator of the instrumental variables approach of all equations is

$$Z'Y = Z'WB + Z'\varepsilon$$

where $Z = \text{diag}(Z_4, \ldots, Z_T)$. The covariance estimator of the error term in $V(Z'\varepsilon) = \text{E}(Z'\varepsilon\varepsilon'Z)$ is

$$\hat{V}(Z'\varepsilon) = \begin{pmatrix} \hat{V}_{44} & \hat{V}_{45} & \ldots & \hat{V}_{4T} \\ \vdots & \vdots & & \vdots \\ \hat{V}_{T4} & \hat{V}_{T5} & \ldots & \hat{V}_{TT} \end{pmatrix}$$

where

$$\hat{V}_{tt'} = \frac{1}{N} \begin{pmatrix} (\sum_i z_{i,t,1}\hat{\varepsilon}_{it})(\sum_i z_{i,t',1}\hat{\varepsilon}_{it'}) & \ldots & (\sum_i z_{i,t,1}\hat{\varepsilon}_{it})(\sum_i z_{i,t',j}\hat{\varepsilon}_{it'}) \\ \vdots & \vdots & \vdots \\ (\sum_i z_{i,t,j}\hat{\varepsilon}_{it})(\sum_i z_{i,t',1}\hat{\varepsilon}_{it'}) & \ldots & (\sum_i z_{i,t,j}\hat{\varepsilon}_{it})(\sum_i z_{i,t',j}\hat{\varepsilon}_{it'}) \end{pmatrix}$$

$$= \frac{1}{N} \sum_{i=1}^{N} \hat{\varepsilon}_{it} \hat{\varepsilon}_{it'} z'_{it} z_{it'}$$

$$t = 4, \dots, T \qquad t' = 4, \dots, T$$

$j = 3 + 2(K-1)$, $z_{it} = (z_{it1}, \dots, z_{itj})$ die i-te Zeile von z_t. The final estimator is:

$$\hat{B}_{2SIVGMM} = \left(W'Z(\hat{V}(Z'\varepsilon))^{-1}Z'W\right)^{-1} W'Z(\hat{V}(Z'\varepsilon))^{-1}Z'Y.$$

2.1.2 Multilevel models

A critical assumption of the previous models with the exception of the last one is the general constancy of individual effects. E.g. we cannot expect that unobserved individual abilities have the same effects in different situations. One possibility, the time dependence, is described in (12). If the effects depend on exogenous variables, interaction variables can be modelled or separated regressions for each individual can be estimated. Then the differences in the estimated coefficients are due to individual characteristics. As usually only data from few periods are available, this procedure does not seem successful. But a natural extension is obtained when units are not only separated by one criterion, i=1,...,N, and then repeated observed over time (t=1,...,T), but further levels are considered, e.g. establishments and industries. We call these "multilevel models"(GOLDSTEIN 1995, LONGFORD 1993, RAUDENBUSCH, BRYK 2002). Typically, these models are formulated as "random effects models". A three-level model may be expressed by

$$y_{ijlt} = x'_{ijlt}\beta + \alpha_i + \alpha_{ij} + \alpha_{ijl} + \epsilon_{ijlt}, \tag{15}$$

where level 1 are the individuals (i=1,...,N), level 2 are the firms (j=1,...,J) and level 3 are the industries (l=1,...,L). One individual is assigned to one establishment and one establishment is assigned to one industry. If the latter two levels are neglected, but if they are effective on the dependent variable, the influence is incorporated by the former. Then the individual effect is biased. Other examples of three levels are establishments, districts and countries or individuals, households and cities. Recently, especially two-level models, i.e. linked employer-employee panel data (LEEP) models, have been introduced in the literature (ABOWD, CREEZY, KRAMARZ 2002, ABOWD, KRAMARZ 1999, 1999a, ABOWD, KRAMARZ, MARGOLIS 1999, GOUX, MAURIN 1999). In this context some methodological issues are of special interest. The basic model is described by

$$y_{it} = x'_{it}\beta + \alpha_i + \psi_{J(it)} + \epsilon_{it}$$

under the assumption $E[\epsilon_{it}|i, J(it), t, x_{it}] = 0$. The firm effect is expressed by $\psi_{J(it)}$. This means individual i in period t is assigned to one of the j=1,...,J establishments. The LEEP approach follows a "fixed effects model". In matrix notation we have

$$y = X\beta + D\alpha + F\psi + \epsilon, \tag{16}$$

where y is a N·T×1 vector, X is a N·T×K matrix. The design matrix for the individual effects D has the order N·T×N containing unit vectors. Analogously, the design matrix for the firm effects F is a N·T×J matrix. This means the number of columns is equal the number of firms.

The objective is to estimate the coefficients of β, α and ψ. The conventional technique to estimate the partitioned regression (16) does not work. The usual way is to "sweep out" the D matrix and then to determine the firm effects cannot be used in practice. The F matrix is too large a non-patterned matrix due to the large number of firms. Identification of the individual and firm effects, in order to estimate using the exact least squares estimator, requires finding the conditions under which equation (16) can be solved for some subset of the person and firm effects. ABOWD, CREECY and KRAMARZ (2002) present a procedure which involves applying methods from graph theory to determine groups of connected individuals and firms. Within a connected group identification can be determined using conventional methods from the analysis of covariance. A group contains all workers who ever worked for any of the firms in the group and all the firms at which any of the workers were ever employed. The algorithm constructs G mutually-exclusive groups of connected observations from the N workers in J firms observed over the sample period. However, usually approximate solutions to (16) are employed (ABOWD, KRAMARZ, MARGOLIS 1999). For this purpose an extended system is formulated which is easier managable under specific restrictions

$$y = X\beta + D\alpha + Z\lambda + M_Z F\psi + \epsilon, \tag{17}$$

where $\lambda = (Z'Z)^{-1}Z'F\psi$ is an auxiliary parameter, $M_Z = I - Z(Z'Z)^{-1}Z'$. The new matrix Z contains information or more directly specific columns of X, D and F. The intention behind creating Z is to incorporate all the relevant variables which determine interaction effects between X, F and D so that under the condition of Z, orthogonality conditions can be formulated. The selection of this information is a similar problem as the choice of instrumental variables. The following restrictions are imposed: (i) X and D are orthogonal, given Z, (ii) D and F are orthogonal, given Z. A four-step procedure can then be applied:

(i) Estimate

$$\begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} \begin{bmatrix} X'M_D X & X'M_D Z \\ Z'M_D X & Z'M_D Z \end{bmatrix}^{-1} \begin{bmatrix} X'M_D y \\ Z'M_D y \end{bmatrix}.$$

(ii) Using $\hat{\beta}$ and $\hat{\lambda}$ one can determine

$$\hat{\alpha} = (D'D)^{-1}D'(y - X\hat{\beta} - Z\hat{\lambda}).$$

(iii) Formulate the partitioned auxiliary regression $y = F\psi + Z\pi + v$ and determine

$$\hat{\pi} = (Z'M_F Z)^{-1}Z'M_F y.$$

(iv) With the help of $\hat{\pi}$ one can calculate

$$\hat{\psi} = (F'F)^{-1}F'(y - Z\hat{\pi}).$$

If the model is extended by a further level e.g. by industry effects we have to consider that each establishment is assigned to only one industry. The gross firm effect $F\psi$ has to be separated into the net firm effect $(F\psi - FA\kappa)$ and an industry effect $(FA\kappa)$ Therefore, model (16) passes into

$$y = X\beta + D\alpha + FA\kappa + (F\psi - FA\kappa) + \epsilon. \tag{18}$$

Matrix A assigns firms to a specific industry ($a_{jl} = 1$, if firm j belongs to industry l ; $a_{jl} = 0$ otherwise). The firm effect is estimated in dependence of industry effects, i.e. using

$$F\psi = FA\kappa + \omega,$$

we obtain

$$\hat{\kappa} = (A'F'FA)^{-1}A'F'F\psi.$$

Therefore,

$$\begin{aligned} F\psi - FA\hat{\kappa} &= F\psi - FA(A'F'FA)^{-1}A'F'F\psi \\ &= (I - FA(A'F'FA)^{-1}A'F')F\psi =: M_{FA}F\psi \end{aligned}$$

follows. If firm effects are suppressed, but effective, the industry effect is biased except FA and $M_{FA}F$ are orthogonal. Furthermore, individual effects are neglected, we obtain a raw estimated industry effect from

$$y = X\beta + FA\kappa + \epsilon,$$

which is a weighted average of individual and firm effects

$$\kappa^* = (A'F'M_X FA)^{-1}A'F'M_X D\alpha + (A'F'M_X FA)^{-1}A'F'M_X F\psi.$$

Analogously to the industry effects, it is possible to analyze occupation effects which are often interpreted as an interaction between the individual and firm effect (GROSHEN 1996). HILDRETH and PUDNEY (1999) discuss issues of non-random missing values and GOUX/MAURIN consider instrumental variables estimators which are necessary if jointly dependent variables or errors in variables exist.

While Hausman tests usually reject the random effects model (REM) the fixed effects model (FEM) has the problem that the within transformation of a model wipes out time invariant regressors as well as the individual effect, so that it is not possible to estimate the effects of those regressors on the dependent variable. One way to solve this problems is to replace random individual effects by estimated FE. We call this a "selectivity fixed-random effects model" and use the acronym SELFREM. The basic idea follows the sample selection approach. HECKMAN (1979) has substituted the conditional expected error term for an estimate, which is employed as an artificial regressor. Let us deal with this issue in a two-level model

$$y_{ijt} = x'_{ijt}\beta + \alpha_i + \alpha_j + \alpha_{ij} + \epsilon_{ijt}. \tag{19}$$

In the first step the general individual effects α_i, the general firm effects α_j and the firm specific individual effects α_{ij} is estimated by the within estimator of a FEM.

$$\hat{\alpha}_{1i} = (\bar{y}_i - \bar{y}) - (\bar{x}_i - \bar{x})'\beta^* = \hat{\beta}_{0i}^{(1)} - \hat{\beta}_0^{(1)}$$

$$\hat{\alpha}_{1j} = (\bar{y}_j - \bar{y}) - (\bar{x}_j - \bar{x})'\beta^* = \hat{\beta}_{0j}^{(1)} - \hat{\beta}_0^{(1)}$$

$$\hat{\alpha}_{1ij} = (\bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}) - (\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})'\beta^* = \hat{\beta}_{0ij}^{(1)} - \hat{\beta}_{0i}^{(1)} - \hat{\beta}_{0j}^{(1)} + \hat{\beta}_0^{(1)}, \tag{20}$$

where β^* is the coefficient vector without the constant term. The conventional RE estimator is not adequate if individual and firm effects, respectively, and regressors are correlated. In the second step the individual und firm effects are substituted by the estimates of (20). We incorporate these effects as linear combinations ($a_1\hat{\alpha}_{1i}$, $b_1\hat{\alpha}_{1j}$ and $c_1\hat{\alpha}_{1ij}$) and expect $\hat{a}_1 = 1, \hat{b}_1 = 1, \hat{c}_1 = 1$. OLS estimation of

$$y_{ijt} = x'_{ijt}\beta + a_1\hat{\alpha}_{1i} + b_1\hat{\alpha}_{1j} + c_1\hat{\alpha}_{1ij} + \epsilon_{ijt} \tag{21}$$

leads to new estimates of the individual and firm effects: $\hat{\alpha}_{2i}$, $\hat{\alpha}_{2j}$ and $\hat{\alpha}_{2ij}$. If a_1, b_1 and c_1 are not equal to one, the unobserved but estimated effects in (21) are substituted by new linear combinations and further OLS estimates are employed. The iterations (1,...,s) of this procedure are continued until a_s, b_s and c_s converge to one. The estimator is consistent if $\hat{a}_s \to 1$, $\hat{b}_s \to 1$ and $\hat{c}_s \to 1$ where $N \to \infty$. If $\hat{a}_s$, $\hat{b}_s$ and $\hat{c}_s$ do not converge to 1, this is evidence of misspecification. An alternative to (21) is a nonparametric specification

$$y_{ijt} = x'_{ijt}\beta + m(\hat{\alpha}_{1i}) + m(\hat{\alpha}_{1j}) + m(\hat{\alpha}_{1ij}) + \epsilon_{ijt}. \tag{22}$$

A simplified version of this approach is analyzed in section 3. It should be mentioned that the pure individual effects can only be separated from the firm specific individual effects if the individuals quit the firm and accept a job within the considered period. The general firm effects can only be determined if we have more than one employee in the firms of our sample.

If a test for poolability (Baltagi 2005) does not reject the null hypothesis, the unobserved individual and firms effects can be neglected. In this case another way might be useful to determine the relationship between individual and firms effects: a model with varying coefficients. In the first stage, individual observations are only considered and the model

$$y_{ijt} = x'_{ijt}\beta_j + \epsilon_{ijt}.$$

is estimated, where y_{ijt} is the endogeous variable with i = 1,...,N individuals, j = 1,...,M firms and t = 1,...,T periods. This variable is determined by a K×1-vector of individual characteristics x_{ijt}. The individual disturbance term is e $\epsilon_{ijt} \sim N(0, \sigma^2)$. We assume that the K×1 parameter vector β_j varies between the firms.

In the second stage the individual coefficients β_j are estimated by a 1×L vector w_j' including a constant term and L-1 firms characteristics

$$\beta_j = W_j\gamma_j + u_j,$$

where $W_j = I_K \otimes w_j$' is a K×K·L·T matrix, $w'_j = (w_{j11}, ..., w_{jL1}, ..., w_{j1T}, ..., w_{jLT})$ and $\gamma'_j = (\gamma_{j111}, ..., \gamma_{jL11}, ..., \gamma_{j1T1}, ..., \gamma_{jLT1}, ..., \gamma_{j11K}, ..., \gamma_{jL1K}, ..., \gamma_{j1TK}, ..., \gamma_{jLTK}$. Alternatively, w'_j can be substituted by $\bar{w}'_j = (\bar{w}_{j1}, ..., \bar{w}_{jL})$ and analogously the γ_j has to be adjusted. This means we assume that the firms effects on β_j are time invariant.

The K·L×1 parameter vector γ_j describes systematical firms influences on the individual y effects. We assume that the K×1 disturbance term vector on the firms level is $u_j \sim N(0, T)$. Additional assumptions are: Cov$[\epsilon_{ijt}, u_{kjt}] = 0$, Cov$[x_{kijt}, \epsilon_{ijt}] = 0$, Cov$[w_{ljt}, u_{kjt}] = 0$, Cov$[w_{ljt}, \epsilon_{ijt}] = 0$ and Cov$[x_{kijt}, u_{kjt}] = 0$ for all k, k' und l, k = 1...K, l = 1...L. The true parameter vector β_j is substituted by a firm specific OLS estimator ($\hat{\beta}_j$).

$$\hat{\beta}_j = W_j\gamma_j + u_j + e_j,$$

where the disturbance term is $e_j \sim N(0, V_j)$. A GLS estimation follows

$$\gamma_g = [\sum_{j=1}^{M}(W_j' \Delta_j^{-1} W_j)]^{-1}(W_j' \Delta_j^{-1} \hat{\beta}_j),$$

where $\Delta_j = Var(\hat{\beta}_j) = Var(u_j + e_j) = T + V_j$. The covariance matrix may iteratively estimated by the ML method. The variance of ϵ_{ijt} is modelled by

$$\sigma_{ij}^2 = exp(c_{ij}' \alpha),$$

where the vector c_{ij} incorporates heteroskedasticity inducing variables.

2.2 Nonlinear models

Many methods for nonlinear panel data models with fixed effects rely on the method of conditional ML, where a sufficient statistic of α_i is conditioned on to remove the unobserved individual effect. But in general nonlinear models it is not always easy and sometimes it is impossible to find a minimum sufficient statistic for α_i that is independent of β. Under random effects one can attempt to employ conventional ML methods. However, the combination of typical nonlinear models with panel structures often yields too complex models to use this procedure. Two alternatives exist. On the one hand, simulated ML estimation is manageable. Or, on the other hand, the GMM approach is a good alternative.

The most popular examples in microeconometric nonlinear models are logit, probit, count data, censored and selection models. A general nonlinear model may be characterized by

$$y_{it} = m(x_{it}; \theta) + u_{it}. \tag{23}$$

The conditional mean function is

$$E(y_{it}|x_{it}) = m(x_{it}; \theta).$$

A simplification follows by linear link functions (McCULLOUGH/NELDER 1983): $m(x_{it}; \theta) = g(x_{it}' \beta)$. The major problem of nonlinear panel data models is the removal of the individual effect. First differences or within estimates do not solve the elimination problem, β and the individual effect are not independent. The main limitation of much of the literature on nonlinear panel data methods is that the explanatory variables are assumed strictly exogenous. The discussion is focussed on models, in which the parameter that is usually interpreted as an intercept is allowed to be individual specific. Unfortunately, the features of the model that do not depend on α_i tend to be different for the different functional forms for $m(\cdot)$ in (23) and do not always exist, as for example in the case of a fixed effects probit model. The resulting estimation procedures are therefore different for different models. One then ends up estimating, say, a logit model in a way that is fundamentally different from the way one would estimate, say, a censored model. This is somewhat unsatisfactory.

2.2.1 Logit models

If the dependent variable is a dummy variable, most popular approaches are logit and probit models. Conditional maximum likelihood estimators can be applied to logit models with fixed effects (HSIAO 2004). The basic model is

$$y_{it}^* = x_{it}'\beta + \alpha_i + \epsilon_{it} = x_{it}'\beta + u_{it}$$

$$y_{it} = \begin{cases} 1 & \text{, if} \quad y_{it}^* \geq 0 \\ 0 & \text{, otherwise} \end{cases} \tag{24}$$

and the probability is

$$P(y_{it} = 1) = \frac{\exp\left(x_{it}'\beta + \alpha_i\right)}{1 + \exp\left(x_{it}'\beta + \alpha_i\right))},$$

where $i = 1, ..., N; t = 1, ..., T$. A simple ML estimator is inconsistent as FEM's allow a correlation between x and α. But in the literature on the estimation of nonlinear fixed effects panel data models have been developed alternative estimation strategies. The general idea is that, although the model does not have features that are linear in the α_i's, it is nonetheless sometimes possible to find features of the model that do not depend on α_i. One way is to use the conditional maximum likelihood estimation (CML)

$$L^c = \prod_{i=1}^{N} P(Y_{i1} = y_{i1}, \ldots, Y_{iT} = y_{iT} | \sum_t y_{it})$$

$$P(y_i | \sum_{t=1}^{T} y_{it}) = \exp\left((\sum_{i=1}^{T} x_{it} y_{it})'\beta\right) / \sum_{d \epsilon \tilde{B}_i} \exp\left((\sum_{t=1}^{T} x_{it} d_{it})'\beta\right),$$

where $\tilde{B}_i = (d_{i1}, \ldots, d_{iT}) | d_{it}$ are 0 or 1 and $\sum_t d_{it} = \sum_t y_{it}$. Under this condition the conditional ML estimator of β does not depend on α_i.

$\tilde{B}_i$ describe alternative data sets in relation to y_{it}, which fulfil the condition $\sum_t y_{it}$. When T observations exist, the number of elements with 1 can assume the following sum

$$\sum_{t=1}^{T} y_{it} = 0, 1, \ldots, T.$$

$\sum y_{it} = 0$ and $\sum y_{it} = T$ do not contribute to the likelihood function, as in this case only one combination exists. Therefore, the corresponding probability is equal to one and only the sum from 1 to T-1 is relevant. Alternatives exist if $1 \leq \sum y_{it} < T$. In other words, relevant for the estimation are only those individuals who change the status of the regressand once or more. We demonstrate this for the simple case $T = 2$. The conditional probability then follows

$$\begin{aligned} P(\omega_i = 1 | y_{i1} + y_{i2} = 1) &= \frac{P(\omega_i = 1)}{P(\omega_i = 1) + P(\omega_i = 0)} \\ &= \frac{exp[(x_{i2} - x_{i1})'\beta]}{1 + exp[(x_{i2} - x_{i1})'\beta]} = F[(x_{i2} - x_{i1})'\beta], \end{aligned}$$

where $\omega_i = 1$, if $(y_{i1}; y_{i2}) = (0; 1)$, and $\omega_i = 0$, if $(y_{i1}; y_{i2}) = (1; 0)$. This result demonstrates that for an individual for whom y changes, the probability that it changes from 1 to 0 or vice versa is a logit with explanatory variables $(x_{i2}-x_{i1})$. Since this probability does not depend on α_i, one can estimate β without making assumptions on α_i by considering only those individuals for whom $y_{i1}+y_{i2}=1$ and then estimating the logit model. The conditional likelihood function

$$\ln L^c = \sum_{i=1}^{N}\{\omega_i \cdot \ln F[(x_{i2}-x_{i1})'\beta] + (1-\omega_i)\cdot \ln[1-F[(x_{i2}-x_{i1})'\beta]]\}$$

can be maximized by conventional methods of a simple logit model. The problem becomes more complicated if lagged dependent variables representing state dependence also appear in specification (24). In that case we need $T \geq 4$ waves for the identification of a logit model.

A generalization of the standard logit panel data model is presented by REVELT and TRAIN (1998). They analyze a multinomial panel model and allow that the parameters associated with each observed variable vary randomly across individuals. Conditional on β_i, the probability that person i chooses alternative l in period t is

$$P(y_{it}=l) = \frac{\exp\,(x'_{ilt}\beta_i)}{\sum_{l=1}^{L}\exp\,(x'_{ilt}\beta_i)} =: L_{ilt}(\beta_i). \tag{25}$$

The unconditional probability over all values of β_i depends on the parameters of the distribution of β_i. For ML estimates we need the probability of each sampled person's sequence of observed choices. Conditional on β_i, the probability of person's i observed sequence of choices is the product of standard logits $(S_i(\beta_i)=\Pi_t P_{it}(\theta^*))$. The unconditional probability for the sequence of choices is

$$P_i(\theta^*) = \int S_i(\beta_i) f(\beta_i|\theta^*)d\beta_i, \tag{26}$$

where θ^* are the parameters of the density $f(\theta^*)$. The objective is to estimate θ^*, the population parameters that describe the distribution of the individual parameters. Exact maximum likelihood estimation is not possible since the integral in (26) cannot be calculated analytically. Instead, it is possible to approximate the probability by simulation and maximize the simulated log likelihood. For a given value of θ, a value of β_i is drawn from its distribution. Then $S_i(\beta_i)$ is calculated. This process is repeated. The average of the replicated results is taken as the estimated choice probability: $SP_i(\theta)=1/R\sum_{r=1}^{R}S_i(\beta_{i;r|\theta})$ where R is the number of replications. This is an unbiased estimator of $P_i(\theta)$. The estimated parameters of the simulated log likelihood function $SLL(\theta)=\sum_i ln(SP_i(\theta))$ are consistent and asymptotically normal under regularity conditions. The simulated score for each person is

$$SS_i(\theta) =: \frac{\partial lnSP_i(\theta)}{\partial\theta}$$

$$= [\frac{1}{SP_i(\theta)}]\frac{1}{R}\sum_{r=1}^{R}S_i(\beta_{i;r|\theta})[\sum_l\sum_t(d_{ilt}-L_{ilt;r|\theta})\frac{\partial\beta'_{i;r|\theta}x_{ilt}}{\partial\theta}],$$

where $d_{ilt}=1$, if individual i choose alternative l in period t and zero otherwise. These mixed logit approaches do not require the independence of irrelevant alternatives and general patterns of correlation over alternatives and time are allowed.

2.2.2 Probit models

As not so much is known about how to deal with fixed effects, it is often appealing to make assumptions on the distribution of individual effects. We cannot find simple functions for the parameters of interest that are independent of the nuisance parameter α_i for probit models. Starting point is the model (24), but the individual effect have to be uncorrelated with x_{it}, and α_i is a random variable where a normal distribution is assumed. Analogously to the linear model the variance of the error term is

$$V(u_{it}) \quad = \quad V(\alpha_i + \epsilon_{it}) = \sigma_\alpha^2 + \sigma_\epsilon^2$$

and square of correlation coefficient between the two error terms follows

$$\rho^2 \quad = \quad Corr^2(u_{it}; u_{it'}) = \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_\epsilon^2).$$

When the individual effects are treated as random variables we may use the FE estimator. However, this procedure implies a loss of efficiency. It can be even worse. In some cases the FE estimator is inconsistent. To obtain the ML estimator we must evaluate T-dimensional integrals

$$P(Y_{i1} = y_{i1}, ..., Y_{iT} = y_{iT}) \quad = \quad \int_{-x'_{i1}\beta}^{\infty} ... \int_{-x'_{iT}\beta}^{\infty} f(u_{i1}, ..., u_{iT}) du_{iT}, ..., du_{i1}.$$

Butler and Moffitt (1982) who follow the Gaussian quadrature procedure simplify the computation

$$P(Y_{i1} = y_{i1}, ..., Y_{iT} = y_{iT}) \quad = \quad \int_{-\infty}^{\infty} f(\alpha_i) \prod_{t=1}^{T} [F(\infty|\alpha_i) - F(-x'_{it}\beta|\alpha_i)] \cdot d\alpha_i.$$

Based on this probability the log likelihood $lnL = \sum lnL_i$ is calculated. The score vector can be used to employ the BHHH estimator where the covariance matrix is determined by the OPG of the score vector. The combination of random effects and probit approaches suggests in itself since both assume a normal distribution.

A more efficient alternative suggested by Chamberlain (1984) yields the minimum distance estimator which avoids numerical integration. This approach starts with a log likelihood function

$$\begin{aligned} lnL &= \sum_{i=1}^{N} lnL_i \\ L_i &= L_i(y_{it}^*) = L_i(\epsilon_{it}; \alpha_i) = L_i(y_{it}^*|\alpha_i) \cdot L(\alpha_i) \\ &= \int_{-\infty}^{+\infty} \prod_{t=1}^{T} F(x'_{it}\beta + \alpha_i)^{y_{it}} \cdot (1 - F(x'_{it}\beta + \alpha_i))^{1-y_{it}} \cdot dG(\alpha_i), \end{aligned}$$

where $G(\alpha_i)$ is a univariate distribution function. Following Chamberlain, who assumes

$$\alpha_i = \sum_{t=1}^{T} a'_t x_{it} + \eta_i = a' x_i + \eta_i,$$

we can write

$$L_i = \int_{-\infty}^{\infty} \prod_{t=1}^{T} F(x'_{it}\beta + a'x_i + \eta_i)^{y_{it}} \cdot (1 - F(x'_{it}\beta + a'x_i + \eta_i)^{1-y_{it}} \cdot dG^*(\eta_i),$$

where $a = (a_1', \cdots, a_T')' = (a_{11}, \cdots, a_{K1}, \cdots, a_{1T}, \cdots, a_{KT})' \sim TK \times 1$, $x_i \sim TK \times 1$, $F(\cdot)$ - standard normal distribution and $G^*(\eta)$ - normal distribution where $\eta \sim N(0, \sigma_\eta^2)$. Consequence is a multivariate probit model, where

$$y_{it} = 1, \text{if} \quad x_{it}'\beta + a'x_i + \eta_i + \epsilon_{it} > 0.$$

The components of the error term are independent normally distributed

$$\begin{pmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{iT} \end{pmatrix} + \begin{pmatrix} \eta_i \\ \vdots \\ \eta_i \end{pmatrix} = \epsilon_i + \eta_i \iota \sim N\left[0, (I_T + \sigma_\eta^2 \iota\iota')\right].$$

The conditional distribution of y_{it} due to x_i is modelled by

$$P(y_{it} = 1|x_i) = F[(x_{it}'\beta + a'x_i)/(1 + \sigma_\eta^2)^{1/2}] = F(x_i'\pi).$$

This is a marginal distribution in respect of α. Separated for each period $t = 1, \cdots T$ univariate probit specification can be determined as ML estimates which jointly converge to

$$\Pi = diag\left[(1 + \sigma_\eta^2)^{-1/2}\right](\underset{T \times TK}{I_T \otimes \beta'} + \underset{T \times 1}{\iota} \cdot \underset{1 \times TK}{a'}),$$

as $N \to \infty$. The minimum distance estimator follows if we derive

$$(\hat{\pi} - f(\theta))'\hat{\Omega}^{-1}(\hat{\pi} - f(\theta))$$

with respect to β, α and σ_η^2, where $\hat{\pi} = vec(\hat{\Pi}) \sim KT^2$ are stacked, unrestricted, univariate estimates, which are based on the ML method of T separate probit models and $\pi = vec(\Pi) = f(\theta) = f((\beta' \vdots \alpha' \vdots \sigma_\eta^2)')$. $\hat{\Omega}$ is a consistent estimator of the asymptotic covariance matrix of $\hat{\Pi}$, i.e. $\Omega = J^{-1}\Delta J^{-1}$ where

$$\underset{KT^2 \times T^2K}{J} = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_T \end{bmatrix} \quad ; \quad \underset{KT \times TK}{J_t} = E\left[\frac{f_{it}^2}{F_{it}(1 - F_{it})} x_i x_i'\right]$$

J-information matrix, $\Delta = E[\psi_i \otimes x_i x_i']$; $\psi_i \sim (T \times T)$ matrix, where a typical element is $\tilde{u}_{it} \cdot \tilde{u}_{it'}$, $\tilde{u}_{it} = \frac{y_{it} - F_{it}}{F_{it}(1 - F_{it})} f_{it}$ are generalized residuals and $F_{it} = F(x_i'\pi)$, $f_{it} = f(x_i'\pi)$. As an estimator of π_i we write $\hat{\pi}_i$. Expected values are substituted with means and π by $\hat{\pi}$ in f_{it} or F_{it}.

In the pure random effects model, one can also estimate the model by a pseudo-maximum likelihood method that ignores the panel structure altogether. The basic idea can be described as follows: the time correlation structure is only assumed as “nuisance“ with subordinated interest. Due to possible misspecification of this correlation structure, the application of the ML method is not completely valid. Therefore, in the literature this approach is called a quasi- or pseudo-ML estimation (QML). The procedure is the following. We define

$V(y_{it}) = F(x'_{it}\beta)[1 - F(x'_{it}\beta)]$

$V_i = \text{diag}\,[F(x'_{i1}\beta)(1 - F(x'_{i1}\beta))\ldots F(x'_{iT}\beta)(1 - F(x'_{iT}\beta))]$

$\Omega_i = V_i R(\delta) V_i$ covariance matrix, which yields the smallest KLIC under misspecification.

$R(\delta)$ is time correlation structure which depends from the unknown parameter vector δ - “nuisance“.

The objective is to minimize

$$S = \sum_{i=1}^{N} (y_i - F_i(x_{i\cdot}\beta))'\Omega_i^{-1}(y_i - F_i(x'_{i\cdot}\beta))$$

where

$$y_i = (y_{i1}, \cdots, y_{iT})' \quad , \quad F_i(\cdot) = (F(x'_{i1}\beta), \ldots, F(x'_{iT}\beta))'.$$

If Ω is known, the LS estimator fulfils the equation

$$\sum_{i=1}^{N} \frac{\partial F_i(\cdot)}{\partial \beta}\Omega_i^{-1}(y_i - F_i(\cdot)) = 0.$$

If Ω is unknown, a “working correlation matrix-$\tilde{\Omega}$“ is employed, which is usually misspecified, e. g. $R(\delta) = I$ or it is assumed that the correlations outside the main diagonals are equal. The equations are then called “generalized estimating equations - GEE“ and the solution corresponds to a QML estimation. This result is not identical with the minimization of S, because $\frac{\partial \Omega}{\partial \beta}$ is not considered. If the specification of $F(x'_{it}\beta)$ is correct, the QML estimator is consistent and asymptotically normally distributed, provided that the estimation of the covariance matrix $\hat{\beta}$ is robust.

If we consider a multinomial probit panel data model, the CML method fails. The Butler-Moffitt approach is restricted because of the underlying multidimensional integral. As mentioned in the introduction of nonlinear models two alternatives exist: simulated estimation methods and GMM approaches.

GMM estimators are based on the orthogonality conditions implied by the single equation conditional mean functions

$$E(y_{it} - F(x'_{it}\beta | X_i)) = 0,$$

where F$(\cdot)$ denotes the CDF of the univariate normal distribution. The orthogonality conditions are

$$E[A(X_i)\begin{pmatrix} (y_{i1} - F(x'_{i1}\beta)) \\ (y_{i2} - F(x'_{i2}\beta)) \\ \vdots \\ (y_{iT} - F(x'_{iT}\beta)) \end{pmatrix}] = 0, \tag{27}$$

where $A(X_i)$ is a $P \times T$ matrix of instrumental variables constructed from the exogenous data for individual i. The empirical counterpart to the left hand side of (27) substitutes the expected value with the sample mean so that

$$g_N(\beta) = \frac{1}{N}\sum_{i=1}^{N}[A(X_i)\begin{pmatrix} (y_{i1} - F(x'_{i1}\beta)) \\ (y_{i2} - F(x'_{i2}\beta)) \\ \vdots \\ (y_{iT} - F(x'_{iT}\beta)) \end{pmatrix}]$$

and the solution of the various estimators are

$$\hat{\beta}_{GMM} = argmin(g_N(\beta))'W(g_N(\beta)). \tag{28}$$

The GMM estimators differ by the choice of the instrument matrix A and the weighting matrix W.

By combining classical estimation methods and simulators, several approaches were developed. For example, simulated maximum likelihood methods (SMLM) including the GHK estimator can be used (GEWEKE, KEANE, RUNKLE 1997). Keane (1994) derived a computationally practical simulator for the panel probit model. Simulation methods replace the intractable integrals by unbiased Monte Carlo probability simulators. Further possibilities are the method of simulated moments, simulated scores and Markov chain Monte Carlo including the Gibbs and Metropolis-Hastings algorithm. Geweke, Keane and Runkle find that Gibbs sampling, simulated moments and maximum likelihood method using the GHK estimator all perform reasonably well in point estimation of parameters in a three alternative 10-period probit model. Monte Carlo studies of nonlinear panel data models (BERTSCHEK, LECHNER 1998, BREITUNG, LECHNER 1999) show that among different GMM estimators, the ranking is not so obvious while MLE performs best followed by the GMM estimator based on the optimal instruments derived from the conditional mean restrictions. GREENE (2004) also finds that the GMM estimator performs fairly well compared to the ML estimation.

2.2.3 Count data models

The most popular approach of panel data models, where the dependent variable is a count variable, is the poisson formulation.

$$P(y = j) = \frac{exp(-\lambda)\lambda^j}{j!} \qquad j = 0, 1, 2, ...$$

In contrast to linear models the individual effect is modelled as a multiplicative factor (CAMERON/TRIVEDI 1998, pp.275). The basic model can be written by

$$y_{it} = \alpha_i \cdot \mu_{it} + u_{it},$$

where

$$E[y_{it}|x_{it}, \alpha_i] = \lambda_{it} = \alpha_i \cdot exp(x'_{it}\beta) = exp(\tilde{\alpha}_i + x'_{it}\beta),$$

$\tilde{\alpha}_i = ln\alpha_i$ and $i = 1, \cdots, N; t = 1, \cdots, T$. Strict exogeneity of regressors x is assumed. Under fixed effects we can use the CML method analogous to the logit model. The conditional likelihood function

$$lnL^c = \sum_{i=1}^{N}[ln(\sum_{t=1}^{T} y_{it})! - \sum_{t=1}^{T} ln(y_{it}!) + \sum_{t=1}^{T} y_{it} ln[\frac{exp(x'_{it}\beta)}{\sum_{\tau=1}^{T} exp(x'_{i\tau}\beta)}]]$$

is maximized, where c in lnL^c means "conditional". The condition is given by $\sum y_{it} = T\bar{y}_{it}$. HAUSMAN, HALL and GRILICHES (1984) follow this procedure. However, they also estimate a negative binomial model with fixed effects. BLUNDELL, GRIFFITH and WINDMEIJER (2002) show that under strict exogeneity of the regressors the CML estimator of the poisson model corresponds to a moments estimator of an mean scale model

$$y_{it} = exp(x'_{it}\beta)\frac{\bar{y}_i}{\bar{\mu}_i} + u^*_{it},$$

where $\bar{y}_i = T^{-1}\sum_t y_{it}$, $\bar{\mu}_i = T^{-1}\sum_t \mu_{it}$ and $u^*_{it} = u_{it} - (\mu_{it}/\bar{\mu}_i)\bar{u}_i$. The ratio of the two means measures the individual effect. The method of moments is given by

$$\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}u^*_{it} = \sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}(y_{it} - exp(x'_{it}\beta)\frac{\bar{y}_i}{\bar{\mu}_i}) = 0.$$

Under weak exogeneity this condition is not consistent as x_{it} and u^*_{it} are correlated via $\bar{u}_{it}$. An alternative presents quasi differences of the condition of moments. These weighted differences eliminate the individual effect

$$w_{it} = y_{it}\frac{\mu_{i,t-1}}{\mu_{it}} - y_{i,t-1} = u_{it}\frac{\mu_{i,t-1}}{\mu_{it}} - u_{i,t-1}.$$

Using instrumental variables which fulfil the condition

$$E[u_{it}|\alpha_i, z_{i1}, \cdots, z_{i,t-1}] = 0$$

so that

$$E[w_{it}|z_{i1}, \cdots, z_{i,t-1}] = E_{\alpha|z}[E(w_{it}|z_{i1}, \cdots, z_{i,t-1})] = 0,$$

the GMM estimator can be employed, which minimizes

$$(\frac{1}{N}\sum_{i=1}^{N} w'_i Z_i)W_N^{-1}(\frac{1}{N}\sum_{i=1}^{N} Z'_i w_i).$$

The optimal weighting matrix is

$$W_N = \frac{1}{N}\sum_{i=1}^{N} Z'_i w_i(\hat{\beta}_0)w_i(\hat{\beta}_0)'Z_i,$$

where $w_i(\hat{\beta}_0)$ is based on a consistent initial estimator $\hat{\beta}_0$. If the instruments are exactly identified and $z_{it} = x_{it}$, the sample moment condition corresponds to the structure of moment conditions at strict exogeneity

$$\sum_{i=1}^{N}\sum_{t=2}^{T} z_{i,t-1}w_{it} = \sum_{i=1}^{N}\sum_{t=2}^{T} z_{i,t-1}(y_{i,t-1} - exp(x'_{i,t-1}\beta)\frac{y_{it}}{\mu_{it}}) = 0.$$

Only if the variance of the individual effects is relatively large, is this estimator satisfactory, as Monte Carlo studies demonstrate. Estimators with pre-sample information instead of prospective values in the weights improve the estimates. Then the bias is only small. Analytical expressions of the unconditional density do not exist for Gaussian RE poisson models. An alternative is once again simulation estimators. Therefore, Chip/Greenberg/Winkelmann (1998) employ the MCMC method.

2.2.4 Sample selection models

HECKMAN's (1979) seminal work and further studies have demonstrated that OLS estimates restricted to subgroups are usually biased. This basic result derived for cross section data models was generalized for panel data models by HAUSMAN/WISE (1979), VERBEEK (1991), VERBEEK and NIJMAN (1992) using random effects models. VERBEEK (1991), NIJMAN/VERBEEK (1992) and ZABEL (1992) consider analogous fixed effects approaches. A systematic discussion of fixed effects selectivity models is presented by WOOLDRIDGE (1995). Starting point is the formulation of the output function

$$y_{it} = x'_{it}\beta + \alpha_i + u_{it} \tag{29}$$

and the selectivity function

$$d^*_{it} = z'_{it}\gamma + \eta_i + \epsilon_{it}; \quad d_{it} = I[d^*_{it} \geq 0], \tag{30}$$

where $I[\cdot]$ is an indicator function which is equals one if the argument in $[\cdot]$ is fulfilled and otherwise zero. Both unobserved individual effects α_i and η_i may be correlated with the observed regressors x_{it} and z_{it}. A special case follows if $z_{it} = x_{it}$. The dependent variable y_{it} is only observable if $d_{it} = 1$. The following restrictions are imposed

A1 : η_i is linearly dependent on z_i

$$\eta_i = z_{i1}\gamma_1 + \cdots + z_{iT}\gamma_T + c_i$$

A2 : $\nu_{it} = \epsilon_{it} + c_i$ is independent of x_i and z_i; $\nu_{it} \sim N(0, \sigma_t^2)$.

A3 : α_i is linearly dependent on x_i and ν_{it}

$$(\alpha_i | x_i, \nu_{it}) = \psi_{t0} + x'_{i1}\psi_{t1} + \cdots + x'_{iT}\psi_{tT} + \phi_t\nu_{it}$$

A4 : u_{it} is in mean independent on x_i and z_i and the conditional expected value depends linear on ν_{it}

$$E(u_{it}|x_i, z_i, \nu_{it}) = E(u_{it}|\nu_{it}) = \rho_t\nu_{it}.$$

Under the assumptions A1-A4, β is identified

$$E(y_{it}|x_i, z_i, \nu_{it}) = \psi_{t0} + x'_{i1}\psi_{t1} + \cdots + x'_{iT}\psi_{tT} + x'_{it}\beta + \delta_t\lambda(\frac{z'_i\gamma_t}{\sigma_{z'_i\gamma_t}}),$$

where $\delta_t = \rho_t + \phi_t$. The estimation proceeds in the following steps:

(i) In each period t the selection function is estimated by a probit approach and from this result the selection variable $\hat{\lambda}(\cdot)$ is determined.

(ii) The pooled OLS estimator is applied to the function extended by the selection term $\delta_t \lambda(\frac{z_i'\gamma_t}{\sigma_{z_i'\gamma}})$

$$y_{it} = \hat{w}_{it}'\theta + e_{it},$$

where $\hat{w}_{it}' = (1, x_{i1}', ..., x_{iT}', x_{it}, 0, ..., 0, \hat{\lambda}_{it}, 0, ..., 0)$. The estimator is

$$\hat{\theta} = (\hat{\psi}', \hat{\beta}', \hat{\delta}')' = (\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}\hat{w}_{it}\hat{w}_{it}')^{-1}(\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}\hat{w}_{it}y_{it}).$$

(iii) The asymptotic covariance matrix of $\hat{\theta}$ is defined by

$$V(\hat{\theta}) = \frac{1}{N}\hat{A}^{-1}\hat{B}\hat{A}^{-1},$$

where

$$\hat{A} = N^{-1}(\sum_{t=1}^{T} d_{it}\hat{w}_{it}'\hat{w}_{it}); \quad \hat{B} = N^{-1}(\hat{p}_i\hat{p}_i'); \quad \hat{p}_i = \hat{q}_i - \hat{D}\hat{r}_i;$$

$$\hat{q}_i = \sum_{t=1}^{T} d_{it}\hat{w}_{it}'\hat{e}_{it}; \quad \hat{D} = N^{-1}\sum_{t=1}^{T} d_{it}\hat{w}_{it}'\hat{\theta}'\nabla_\delta w_{it}(\hat{\delta})';$$

$$\hat{e}_{it} = y_{it} - \hat{w}_{it}'\hat{\theta}; \quad \hat{r}_i = r_i(\hat{\delta}).$$

The gradient matrix $\nabla_\delta w_{it}(\hat{\delta}))'$ of $w_{it}(\delta)'$ is evaluated at $\hat{\delta}$.

ROCHINA-BARRACHINA (1997) develops an approach with first differences to eliminate the individual term and assumes a trivariate normal distribution. CARRASCO (2001) and VELLA/ VERBEEK (1999) have presented some further modified approaches.

2.2.5 Censored models

A pooled estimator of a censored model

$$y_{it}^* = x_{it}'\beta + u_{it}$$

$$y_{it} = \begin{cases} 1 & \text{if} \quad y_{it}^* \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

is not different from the cross section estimator and under random effects, where $u_{it} = \alpha_i + \epsilon_{it}$, $\alpha_i \sim N(0, \sigma_\alpha^2)$,$\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$, we have only to incorporate the additional variable α_i. This means the log likelihood function is written by

$$\begin{aligned} lnL &= \sum_{i=1}^{N} ln \int_{-\infty}^{+\infty} \prod_{t=1}^{T} F(\frac{x_{it}'\beta}{\sigma_\epsilon} + \frac{\alpha_i}{\sigma_\epsilon})^{1-y_{it}} \\ &\cdot \quad [\frac{1}{\sigma_\epsilon} f(\frac{y_{it}}{\sigma_\epsilon} - \frac{x_{it}'\beta}{\sigma_\epsilon} - \frac{\alpha_i}{\sigma_\epsilon})]^{y_{it}} f(\frac{\alpha_i}{\sigma_\alpha})d(\frac{\alpha_i}{\sigma_\alpha})) \end{aligned}$$

and is partially derived with respect to θ, where $\theta = (\beta', \sigma_\epsilon, \sigma_\alpha)'$. The first order conditions are highly nonlinear. An iterative procedure is necessary to solve the maximization problem. Gradients have to be determined. Analogously to the RE probit model, the Gauss Hermite quadrature procedure introduced by Butler and Moffitt can be employed. Chamberlain's idea of a correlated RE probit model, where the individual effect α_i is $a'x_i + \eta_i$, can be transferred to the tobit model detailed presented in Jacubson (1988). FE tobit models are suggested by Heckman and McCurdy (1980).

3 Non- and semiparametric models

The discussion starts in section 3.1 with pooled models and nonparametric terms. But usually, individual effects are incorporated in panel data models. This problem is managed in the next two sections. Furthermore, attrition of observations is a common phenomenon over a panel. Attrition in one period may be followed by reappearance in future periods. Section 3.4. discusses this issue.

3.1 Pooled models

In a general formulation the causal dependence of the dependent variable y_{it} on independent variables and the error term is typically described by

$$y_{it} = g[m(x_{it}) + u_{it}], \tag{31}$$

where $g(\cdot)$ calls a mapping which induces the variable y_{it}. Possibly, y_{it} depends on an unobserved endogenous variable $y^*_{it} = m(x_{it} + u_{it})$. If y_{it} is directly created by x_{it} and u_{it}, the relation can be simplified by $y_{it} = m(x_{it}) + u_{it}$ and the linear model results if $m(x_{it}) = x'_{it}\beta$. While parametric models assume a known structural relation under unknown parameters and an error term a simple nonparametric panel data model formulates a mean regression

$$y_{it} = E(y_{it}|x_{it}) + u_{it} = m(x_{it}) + u_{it}. \tag{32}$$

The incentive to use nonparametric instead of parametric methods is the higher degree of flexibility. The basic problem is the enormous amount of calculations, especially if the number of regressors increase. Furthermore, it is difficult to interpret the estimation. This is called "the curse of dimension". Two possibilities exist to solve this problem. Either additive or partial linear models are assumed. The former are discussed in HASTIE/TIBSHIRANI (1997). We focus on the presentation about partial linear models.

If a pooled estimation of panel data are employed the same procedures as with cross section data can be used. It is necessary to test whether the pooled procedure is convenient. Parametric procedures are described by BALTAGI (2001). A nonparametric test present BALTAGI, HIDALGO and LI (1996). LI and HSIAO (1996) test whether individual effects exist. When the test does not reject the null hypothesis of poolability, the individual effects can be neglected. Starting point is a partial linear approach of panel data

$$y_{it} = z'_{it}\gamma + m(x_{it}) + u_{it}. \tag{33}$$

The fundamental idea is to eliminate the nonparametric part. Then the linear term can be estimated separately following ROBINSON (1988). In other words, the new regressand is the difference between y and the conditional expected value which is induced by the nonparametric regressors. Due to the identity $E[m(x)|x] = m(x)$ the nonparametric term vanishes by first differences

$$y - E(y|x) = (z - E(z|x))'\gamma + u. \tag{34}$$

Before we can estimate γ the conditional expected value has to be determined by a nonparametric procedure (PAGAN/ULLAH 1999,199). The problem is that the denominator of a kernel estimator is a random variable

$$\hat{y}_{it} = \hat{E}(y_{it}|x_{it}) = \frac{1}{NT \cdot h} \sum_{j=1}^{N} \sum_{t=1}^{T} \frac{K\frac{x_{it}-x_{jt}}{h}}{\hat{f}_{it}} \cdot y_{jt},$$

where $\hat{f}_{it} = \frac{1}{NT \cdot h} \sum_{j=1}^{N} \sum_{t=1}^{T} K(\frac{x_{it}-x_{jt}}{h})$ is the kernel density estimator and h is the bandwidth. If we do not only consider a univariate nonparametric term, a multivariate kernel is necessary. A simplified form can be assumed in this case, namely the product of the univariate kernels, i.e. $K(x_{it}) = \Pi_{d=1}^{D} K(x_{dit})$. The denominator can be neglected under a density weighted estimation (POWELL/STOCK/STOKER 1989, LI/STENGOS 1996), where the density function follows from a kernel estimation ($\hat{f}$). The differences model (34) has to be weighted

$$\hat{f}_{it}(y_{it} - \hat{y}_{it}) = \hat{f}_{it}(z_{it} - \hat{z}_{it})'\gamma + \hat{f}_{it} u_{it}.$$

The least squares estimator of γ follows

$$\hat{\gamma} = (\sum_{i=1}^{N} \sum_{t=1}^{T} (z_{it} - \hat{z}_{it})(z_{it} - \hat{z}_{it})' \hat{f}_{it}^2)^{-1} (\sum_{i=1}^{N} \sum_{t=1}^{T} (z_{it} - \hat{z}_{it})(y_{it} - \hat{y}_{it}) \hat{f}_{it}^2). \tag{35}$$

This OLS estimator is consistent. But a GLS estimator presented by LI and ULLAH (1998) achieves the semiparametric efficiency border and is therefore superior. In the second step we obtain the multivariate nonparametric term $m(x)$ under a Taylor series approximation

$$\begin{aligned} y_{it} - z_{it}'\hat{\gamma} &= m(x_{it}) + \beta(x)'(x_{it} - x) + R(x_{it}, x) + z_{it}'(\gamma - \hat{\gamma}) + u_{it} \\ &= m(x_{it}) + \beta(x)'(x_{it} - x) + \tilde{u}_{x,it} \end{aligned}$$

using a local LS estimator. The parameter vector $\tilde{\beta}(x)' = (m(x), \beta(x)')$ results from

$$\hat{\tilde{\beta}}(x) = argmin \frac{1}{NT \cdot h} \sum_{i=1}^{N} \sum_{t=1}^{T} K_{it}(y_{it} - z_{it}'\hat{\gamma} - m(x) - (x_{it} - x)'\beta(x))^2,$$

where $K_{it} = K(\frac{x_{it}-x}{h})$ and h is the bandwidth. The local linear least squares estimator is

$$\begin{aligned} \hat{\tilde{\beta}}(x) &= [\sum_{i=1}^{N} \sum_{t=1}^{T} \bar{K}(\frac{x_{it}-x}{h}) \begin{pmatrix} 1 & x_{it} \\ x_{it} - x & (x_{it} - x)(x_{it} - x)'. \end{pmatrix}]^{-1} \\ &\cdot \sum_{i=1}^{N} \sum_{t=1}^{T} \bar{K}(\frac{x_{it}-x}{h}) \begin{pmatrix} 1 \\ x_{it} - x \end{pmatrix} (y_{it} - z_{it}\hat{\gamma}), \end{aligned}$$

where $\bar{K}(\cdot)$ is the product of kernel functions of second order.

3.2 Random effects models

If the following simple nonparametric models with individual effects α_i exist

$$y_{it} = m_t(x_{it}) + (\alpha_i + \epsilon_{it}) = E(y_{it}|x_{it}) + u_{it}, \tag{36}$$

where $i = 1, ..., N$ and $t = 1, ..., T$, all conditional moment procedures can be used to estimate the nonparametric term $m_t(x_{it})$, if a random effects model is convenient. A known distribution of the individual effects is assumed and α_i are identically independently distributed. Furthermore, the individual effects are independent of the regressors. Under unspecific time dependence of $E(y_{it}|x_{it})$ it is possible to estimate the parameters separately for each wave. If time invariance is assumed, i.e. $m_t(x_{it}) = m(x_{it})$, the pooled procedure can be employed to determine the nonparametric term. A local linear approach is possible

$$y_{it} \approx m(x) + (x_{it} - x)'\beta(x) + \alpha_i + \epsilon_{it} =: \tilde{x}'_{it}\tilde{\beta} + \alpha_i + \epsilon_{it},$$

where $\tilde{x}' = (1, (x_{it} - x)')$ and $\tilde{\beta}' = (m(x), \beta(x)')$. In order to determine the nonparametric terms $m(x)$ and $\beta(x)$ we can choose ULLAH and ROY's (1998) GLS approach. This corresponds to the conventional within transformation,where α is eliminated. Under semiparametric partial linear panel data models

$$y_{it} = m(x_{it}) + z'_{it}\gamma + \alpha_i + \epsilon_{it} \tag{37}$$

we can follow LI and STENGOS (1996) analogously to pooled models.

The estimator of the time invariant nonparametric term of a balanced panel can be assigned to an unbalanced panel (KÖNIG 1997). An extension to models with time variable nonparametric models is also possible (KÖNIG 2002). In this case a wave specific procedure is suggested.

An alternative to LI and STENGOS is developed by KÖNIG (2002, 176ff). The advantage of this method is that the peculiarity of the panel structure can be better considered. It is also possible to model a time variable nonparametric term. Simple estimators are supplied. But in this case it is necessary to consider some restrictions. Conventional first differences and within estimators, well-known from pure linear models can be applied. This approach is independent from the sensitivity of bandwidth. We can expect better behavior in small samples. Disadvantages may be possible under asymptotic consideration. But for applications infinite large samples are less relevant.

The partial linear model may be interpreted as a simple linear model with fixed effects if nonparametric regressors ($x_{it} = x_i$) are time invariant

$$y_{it} = z'_{it}\gamma + c_i + \epsilon_{it},$$

where $c_i = m(x_i) + \alpha_i$. The parameter vector γ is determined by first differences or within estimators if the linear regressors are strictly exogenous. Biased estimators result at a direct application to time variable regressors. The bias can significantly be reduced, if we follow a suggestion by KÖNIG (2002, 182). The more the values of two periods differ, the stronger is the bias. Therefore, only small weights are assigned to those values. Such a weighting function yields the kernel function. The corresponding estimator is

$$\hat{\gamma} = (\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{\tau=1}^{T} K_{(x_{it}-x_{i\tau})}(z_{it} - z_{i\tau})(z_{it} - z_{i\tau})')^{-1}$$

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{\tau=1}^{T} K_{(x_{it}-x_{i\tau})}(z_{it}-z_{i\tau})(y_{it}-y_{i\tau}), \tag{38}$$

which is consistent and asymptotically normally distributed. HOROWITZ and MARKATOU (1996) suggest a kernel estimator to determine the CDF of a random effects model without nonparametric term. The empirical distributions of ϵ_{it} and $u_{it} = \alpha_i + \epsilon_{it}$ are separately determined. The isolation of ϵ_{it} usually takes place by creation of first differences.

$$\hat{\epsilon}_{it} = (y_{it} - y_{i1}) - (z_{it} - z_{i1})'\hat{\gamma}.$$

Under weak restrictions the estimator of the density function of $\hat{u}_{it}$ and $\hat{\epsilon}_{it}$ is consistent.

3.3 Fixed effects models

In contrast to random effects models there exist an additional problem in nonparametric panel data models with fixed effects. Due to the allowed correlation between α_i and x_{it} the conditional expected value of y_{it} differs from the nonparametric term. Instead we obtain

$$E(y_{it}|x_{it}) = m(x_{it}) + E(\alpha_i|x_{it}).$$

Therefore, it is not possible to determine the nonparametric part by the conditional moment approach. The conventional solution by first differences or within estimators breaks down. The individual effect is eliminated, but not identified by this procedure. ULLAH and ROY (1998) suggest a Taylor series of the nonparametric expression as a starting point

$$y_{it} = m(x) + (x_{it} - x)'\frac{\partial m(x_{it})}{\partial x_{it}}|_{x_{it}=x} + \frac{1}{2}(x_{it} - x)'\frac{\partial^2 m(\tilde{x})}{\partial x \partial x'}(x_{it} - x) + \alpha_i + \epsilon_{it}$$

$$=: m(x) + (x_{it} - x)'\beta(x) + R_2(x_{it}, x) + \alpha_i + \epsilon_{it}$$

$$=: m(x) + (x_{it} - x)'\beta(x) + \alpha_i + \tilde{\epsilon}_{x,it}, \tag{39}$$

where $\tilde{x}$ is assumed within the range x and x_{it}. It is intended to estimate $\beta(x)$ of this local linear model. A local within estimator with simple kernel function weights gives biased and inconsistent estimates due to residual terms ($E(\bar{R}_2(x_{i,x}|x_{it} = x) \neq 0$). The same problem follows under analogous first differences estimators. But a double weighting of first differences eliminates the bias (KÖNIG 2002,61ff). We define the product kernel from period t and t-1

$$K(\frac{x_{it} - x}{h}, \frac{x_{i,t-1} - x}{h}) = K(\frac{x_{it} - x}{h}) \cdot K(\frac{x_{i,t-1} - x}{h}) =: K_{it} \cdot K_{i,t-1},$$

where h is again the bandwidth. Instead of weighting with $K(\frac{x_{it}-x}{h})$ as in a conventional differences estimator, the weight is the product of the local kernels

$$\hat{\beta}(x)_D = \{\sum_{i=1}^{N}\sum_{t=2}^{T} K_{it}K_{i,t-1}\Delta x_{it}\Delta x'_{it}\}^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T} K_{it}K_{i,t-1}\Delta x_{it}\Delta y_{it}, \tag{40}$$

where $\Delta x_{it} = x_{it} - x_{i,t-1}$ and $\Delta y_{it} = y_{it} - y_{i,t-1}$. This estimator is not only consistent, but also asymptotically normally distributed with a null vector as the expected value vector and an asymptotic sandwich covariance matrix. A similar weighting is possible in the within model. For this case KÖNIG(2002, 68) derives the following consistent estimator

$$\hat{\beta}(x)_W = \{\sum_{i=1}^{N}\sum_{t=1}^{T} K_{it}\tilde{x}_{it}\tilde{x}_{it}'\}^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T} K_{it}\tilde{x}_{it}\tilde{y}_{it}, \tag{41}$$

where

$$\tilde{x}_{it} = x_{it} - \sum_{t=1}^{T}\frac{K_{it}}{\sum_{t=1}^{T} K_{it} + I[\sum_{t=1}^{T} K_{it} = 0]}x_{it}.$$

Analogously $\tilde{y}_{it}$ is defined. $I[\cdot]$ describes an indicator function where the value is 1, if the condition in $[\cdot]$ is fulfilled. Otherwise, the value is 0. Some similarities can be observed with BLUNDELL, GRIFFITH and WINDMEIJER's (2002) suggestion in count data models with non-strict exogenous variables. It seems plausible to note $\tilde{x}_{it}$ as quasi within variable. The asymptotic variances of the within estimator, which stem from a sandwich covariance matrix, are not larger than the corresponding variances of the differences estimator if only two periods exist. If $T \geq 3$, the difference of the asymptotic covariances $(V_{DIFF} - V_{WITHIN})$ is positive definite (KÖNIG 2002, 72).

Semiparametric partial linear models with fixed individual effects can be described by

$$y_{it} = \tilde{m}(x) + x_{it}'\beta(x) + z_{it}'\gamma + \alpha_i + \tilde{\epsilon}_{it},$$

where $\tilde{\epsilon}_{it} = \epsilon_{it} + R(x_{it}, x)$, $\tilde{m}(x) = m(x) - x'\beta(x)$. The individual term α_i may be correlated with x_{it} and z_{it}. The nonparametric term $m(x_{it})$ is developed by a Taylor series. In this case the problem of the parameter estimation (γ) persists also in the conditional expected value of y_{it}

$$E(y_{it}|x_{it}) = \tilde{m}(x) + x_{it}'\beta(x) + E(z_{it}|x_{it})'\gamma + E(\alpha_i|x_{it}).$$

Differences, i.e. $y_{it} - E(y_{it}|x_{it})$, eliminate the nonparametric term, but not the individual term. Therefore, it is necessary to remove α_i in the first step. LI and STENGOS (1996) employ a differences estimator of γ, where the estimator is weighted by the Nadaraya-Watson kernel estimator

$$\hat{\gamma}_D = \{\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta\hat{\tilde{z}}_{it}\Delta\hat{\tilde{z}}_{it}'\hat{\tilde{f}}(x_{it}, x_{i,t-1})^2\}^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta\hat{\tilde{z}}_{it}\Delta\hat{\tilde{y}}_{it}\hat{\tilde{f}}(x_{it}, x_{i,t-1})^2, \tag{42}$$

where $\hat{\tilde{f}}(x_{it}, x_{i,t-1})$ is the kernel density estimator and

$$\Delta\hat{\tilde{z}}_{it} = z_{it} - z_{i,t-1} - [\hat{E}(z_{it}|x_{it}, x_{i,t-1}) - \hat{E}(z_{i,t-1}|x_{it}, x_{i,t-1})].$$

Analogously, $\Delta\hat{\tilde{y}}_{it}$ is defined. By first differences α_i disappears but not the nonparametric term. In order to remove the difference $\tilde{m}(x_{it}) - \tilde{m}(x_{i,t-1})$ we additionally have to subtract the difference of the expected values. If $\Delta\hat{\tilde{z}}_{it} = z_{it} - z_{i,t-1} - [\hat{E}(z_{it}|x_{it}, x_{i,t-1}) - \hat{E}(z_{i,t-1}|x_{it}, x_{i,t-1})]$ and the error term is correlated, $\Delta\tilde{z}_{it}$ has to be instrumented. KÖNIG (2002, 215) suggests an alternative estimator without kernel weights

$$\hat{\gamma}_{D1} = \{\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta\hat{\tilde{z}}_{it}\Delta\hat{\tilde{z}}_{it}'\hat{I}_{it}\}^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta\hat{\tilde{z}}_{it}\Delta\hat{\tilde{y}}_{it}\hat{I}_{it}, \tag{43}$$

where $\hat{I}_{it} = I[\hat{\hat{m}}(x_{it}, x_{i,t-1}) > b_N]$ is a trimmed function and b_N is a threshold value where $b_N \to 0$ follows under $N \to \infty$. The idea of a trimmed function can be found in LI/LU/ULLAH (1996) and LI/ULLAH (1998). Again a within estimator can be formulated

$$\hat{\gamma}_W = \{\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{\tau=1}^{T}(z_{it} - z_{i\tau} - \hat{E}_{z_{it}-z_{i\tau}})(z_{it} - z_{i\tau} - \hat{E}_{z_{it}-z_{i\tau}})'\hat{I}_{it}\}^{-1} \tag{44}$$

$$\cdot \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{\tau=1}^{T}(z_{it} - z_{i\tau} - \hat{E}_{z_{it}-z_{i\tau}})(y_{it} - y_{i\tau} - \hat{E}_{y_{it}-y_{i\tau}})\hat{I}_{it}$$

where $\hat{E}_{y_{it}-y_{i\tau}} = \hat{E}(y_{it} - y_{i\tau} \mid x_{it}, x_{i\tau})$. A further estimator goes back to BERG, LI and ULLAH (2000). But this estimator is inconsistent, if $x_{i\tau}$ and z_{it} are not independent, given x_{it}.

MANSKI(1975, 1985, 1987) has developed nonparametric maximum score estimators for panel data models with fixed effects and dichotomous endogenous variables. Further models and an estimator are presented by LEE (1999a) and HONORE/LEWBEL (2002). A survey on tobit panel data models with nonparametric components which include the standard case of censored endogenous variables, selection models and censored multivariate models can be found in KYRIAZIDOU (1995,1997)and HONORE/KYRIAZIDOU (2000). They also develop some new variants, which do not require the parametrization of the distribution of the unobservables. However, it is necessary that the explanatory variables are strictly exogenous. Therefore, lagged dependent variables as regressors are excluded. KYRIAZIDOU (1997) obtains values near zero by differences between pairs of observations, because pairs with a large difference obtain small weights. HONORE (1992) suggests trimmed least absolute deviation and trimmed least squares estimators for truncated and censored regression models with fixed effects. He exploits the symmetry in the distribution of the latent variables and finds that when the true values of the parameters are known, trimming can transmit the same symmetry in distribution to observed variables. One can define pairs of residuals that depend on the individual effect in exactly the same way, so that differencing the residuals eliminates the fixed effects.

3.4 Models with sample attrition

DAS (2004) presents a two-step nonparametric random effects panel data model with sample attrition where the linear part is omitted. The model is

$$y_{it} = m(x_{it}) + u_{it} \qquad u_{it} = \alpha_i + \epsilon_{it}$$

$$d_{it} = I(v^t(x_{i1}, ..., x_{it}, w_{it}) - \eta_{it} > 0) =: I(v^t(z_{it} - \eta_{it}) > 0)$$

$$\eta_{it} = \delta_i + \tilde{\eta}_{it}, \tag{45}$$

where $z_{it} = (x_{i1}, ..., x_{it}, w_{it})$, x and w are vectors of covariates. The latter may contain elements of x. The attrition indicator is d_{it} where d_{it}=1, if there is no attrition. The outcome variable can only be observed if d=1. It is assumed that d_{it}, x_{it} and w_{it} are completely observable. The individual effects in the outcome and in the attrition function (α_i, δ_i), are i.i.d. The error terms ϵ_{it} and $\tilde{\eta}_{it}$ are time varying and may be autocorrelated. Furthermore, it is assumed that $E(u_{it}|x_{i1}, ..., x_{it}) = 0$. Under some weak restrictions the nonparametric term is identified, up to a constant, as an additive component of each $h_t(x_t, p_t)$ where $h_t(\cdot)$ is defined by

$$\begin{aligned} E(y_{it}|z_{it}, d_{it} = 1) &= m(x_{it}) + \kappa_t(p_{it}) \\ &= m(x_{it}) + g_{i1}\kappa_{i1}(p_{i1}) + ... + g_{iT}\kappa_{iT}(p_{iT}) =: h_t(x_{it}, p_{it}) \end{aligned}$$

and

$$p_{it} =: p_t(z_{it}) = E(d_{it}|z_{it}) = P(d_{it} = 1|z_{it})$$

$$E(u_{it}|z_{it}, d_{it} = 1) =: \kappa_t(p_{it}).$$

This is a generalization of HECKMAN's (1979) sample selection term.

The first step consists of obtaining estimates of probabilities p_{it} by a nonparametric procedure. Simple LS estimates of $d_i' = (d_{i1}, ..., d_{iT})$ are given by

$$\hat{\beta} = (\frac{1}{N}\sum_{i=1}^{N} r^L(z_i)r^L(z_i)')^{-1}(\frac{1}{N}\sum_{i=1}^{N} r^L(z_i)d_i) \tag{46}$$

and

$$\hat{p}_i = r^L(z_i)'\hat{\beta}. \tag{47}$$

Let

$$r^{L_t}(z_{it}) = (r_{1L_t}(z_{it}), ..., r_{L_tL_t}(z_{it}))'$$

represent a vector of approximating functions for $p_t(z_{it})$.

The second step is a nonparametric estimation of y_{it} on $\hat{v} = (x_{it}', \hat{p})'$. First

$$\hat{\gamma} = \frac{1}{N}\sum_{i=1}^{N} \hat{b}_T^K(\hat{v}_i)\hat{b}_T^K(\hat{v}_i)')^{-1}(\frac{1}{N}\sum_{i=1}^{N} \hat{b}_T^K(\hat{v}_i)y_i \tag{48}$$

is determined and

$$\hat{h}(\hat{v}_{it}) = \hat{b}^K(\hat{v}_{it})'\hat{\gamma} \tag{49}$$

follows, where $\hat{b}_T^K(\hat{v}_i) = (\hat{b}^K(\hat{v}_{i1}), ..., \hat{b}^K(\hat{v}_{iT}))$, $b^K(\hat{v}_{i1}) = d_{it}\tau(\hat{v}_{it})b^K(\hat{v}_{it})$, $\tau(\hat{v}) = \prod_{j=1}^{J} \iota(\tau_j^l \leq v_j \leq \tau_j^u)$, $0 < \tau_j^l < \tau_j^u < 1$, τ_j^l and τ_j^u are prespecified constants and $J = dim(v)$. The dependence of $\hat{b}^K(\hat{v}_{it})$ on d_{it} implies that possibly different subsets on the N observations contribute to the regression for each t.

Second, the estimation of the nonparametric term yields

$$\hat{m}(x_{it}) = \sum_{k=1}^{K} \hat{\gamma}_k b_{kK}(x_{it}). \tag{50}$$

A sandwich estimator is suggested to determine the covariance matrix.

4 Concluding remarks

Many new methods to estimate panel data models were developed in the past. The focus in this paper was directed on multilevel and nonlinear models. As the functional form of nonlinearity is usually unknown nonparametric estimates are corollary. Nowadays several methods are implemented in conventional packages such as STATA or S-PLUS, but others still require programming. In contrast to linear fixed effects panel data models, it is more difficult to manage the individual term in combination with a nonparametric term. Conventional differences and within estimators do not help to eliminate the latter. There do not exist uniform methods of nonlinear models. We have only specific estimation methods for several forms of nonlinearity and the results depend on the assumptions. While estimation of random effects panel data models is based on a fully specified model in which one can determine all the quantities of interest, fixed effects panel data models typically result in the estimation of some finite dimensional parameter from which one cannot calculate all functions of the distribution. Nevertheless, progress can also be observed in the estimation of fixed effects panel data models. Estimates of random effects models are usually more efficient. But very often the violation of the distributional assumptions yields inconsistent estimates. Fixed effects models make fewer assumptions and they react less sensitively to violations of the assumptions. Random effects models are usually preferable for prediction.

In future we have to analyze more completely the dynamic character of the panel data models. Almost nothing is known about nonlinear models with lagged dependent variables. Furthermore, non- and semiparametric methods should also be applied to multilevel models. In many situations it seems advantageous to start with nonparametric estimates. However, the next step would be to derive more fully specified parametric models based on the results of the first step.

References

Abowd, J.M., Creecy, R.H., Kramarz, F.(2002): Computing person and firm effects using linked longitudinal employer-employee data, Cornell university working paper.

Abowd, J.M., Kramarz, F. (1999): The analysis of labor markets using matched employer-employee data, Handbook of labor economics, O. Ashenfelter, D. Card, eds., Vol. 3B, Amsterdam, Elsevier, 2629-2710.

Abowd, J.M., Kramarz, F. (1999a): Econometric analyses of linked employer-employee data, Labour economics 6, 53-74.

Abowd, J.M., Kramarz, F., Margolis, D.N. (1999): High wage workers and high wage firms, Econometrica 67, 251-333.

Arellano, M. (2003): Panel data econometrics, University press: Oxford.

Arellano, M., Bond, S. (1991): Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, Review of economic studies 58, 277–297.

Balestro, P., Nerlove, M. (1966): Pooling cross-section and time series data in the estimation of a dynamic model: The demand for natural gas, Econometrica 34, 585–612.

Baltagi, B. (2002): Recent developments in the econometrics of panel data, Volume I and II, Elgar reference collection: Cheltenham.

Baltagi, B. (2005): Econometric analysis of panel data, 3rd ed., John Wiley & Sons: Chichester.

Berg, M.D., Li, Q., Ullah, A. (2000): Instrumental variable estimation of semiparametric dynamic panel data models: Monte Carlo results on several new and existing estimators, Advances in econometrics: nonstationary panels, panel cointegration, and dynamic panels, B. Baltagi, ed., J.A.I. Press: New York, 297–315.

Bertschek, I., Lechner, M. (1998): Convenient estimators for the panel probit model, Journal of econometrics 87, 329–372.

Blundell, R., Griffith, R., Windmeijer, F. (2002): Individual effects and dynamics in count data models, Journal of econometrics 108, 113-131.

Breitung, J., Lechner, M. (1999): Alternative GMM methods for nonlinear panel data models, Generalized method of moments estimation, L. Matyas, ed., 248–274.

Breusch, T.S., Pagan, A.R. (1980): The Lagrange multiplier test and its applications to model specification in econometrics, Review of economic studies 47, 239–253.

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

Cameron, A.C., Trivedi, P.K. (1998): Regression analysis of count data, Cambridge university press: Cambridge.

Cameron, A.C., Trivedi, P.K. (2005): Microeconometrics - Methods and Applications, Cambridge university press: Cambridge.

Carrasco, R. (2001): Binary choice with binary endogenous regressors in panel data: estimating the effect of fertility on female labor participation, Journal of business & economic statistics 19, 385-394.

Chamberlain, G. (1984): Panel data, Handbook of econometrics, Z. Griliches and M. Intriligator, eds., 1247–1318, North-Holland, Amsterdam.

Chip, S., Greenberg, E., Winkelmann, R. (1998): Posterior simulation and Bayes factors in panel count data models, Journal of econometrics 86, 33 - 54.

Das, M. (2004): Simple estimators for nonparametric panel data model with sample attrition, Journal of econometrics 120, 159–180.

Davis, P. (2002): Estimating multi-way error components models with unbalanced data structure, Journal of econometrics 106(1), 67-95.

Geweke, J., Keane, M., Runkle, D. (1997): Statistical inference in the multinomial multiperiod probit model, Journal of econometrics 80, 125-165.

Goldstein, H. (1987): Multilevel statistical models, 2nd ed., Edward Arnold, London.

Goux, D., Maurin, E. (1999): Persistence of interindustry wage differentials: a reexamination using matched worker-firm panel data, Journal of labor economics 17, 492-533.

Greene, W. (2004): Convenient estimators for the panel probit model: further results, Empirical economics 29, 21-47.

Groshen, E. (1996): American employer salary surveys and labor economics research: issues and contributions, Annales d'Economie et de Statistique 41/42, 413-442.

Hamerle, A., Ronning, G. (1995): Panel analysis for qualitative variables, Handbook of statistical modeling for the social and behavioral sciences, G. Arminger, C.C. Clogg, M.E. Sobel, eds., 401-451, Plenum press: New York.

Hastie, T.J. Tibshirani, R.J. (1997): Generalized additive models, Chapman and Hall: London.

Hausman, J.A. (1978): Specification tests in econometrics, Econometrica 46, 1251-1272.

Hausman, J.A., Hall, B.H., Griliches, Z. (1984): Econometric models for count data with an application to the patents-R and D relationship, Econometrica 52, 909-938.

Hausman, J.A., Wise, D.A. (1979): Attrition bias in experimental and panel data: the Gary income maintenance experiment, Econometrica 47, 455-473.

Heckman, J.J. (1979): Sample selection bias as a specification error, Econometrica 47, 153-161.

Heckman, J.J., McCurdy, T.E. (1980): A life cycle model of female labour supply, Review of economic studies 47, 47-74.

Hildreth, A.K., Pudney, S. (1999): Econometric issues in the analysis of Linked cross-section employer-worker surveys, The creation and analysis of employer-employee matched data, J. Haltiwanger et al., eds., North-Holland: Amsterdam, 461-488.

Holtz-Eakin, D., Newey, W., Rosen, H. (1988): Estimating vector autoregressions with panel data, Econometrica 56, 1371-1395.

Honore, B.E. (2002): Non-linear models with panel data, The Institute for Fiscal Studies, working paper CWP13/02.

Honore, B.E., Kyriazidou, E. (2000): Estimation of tobit-type models with individual specific effects, Econometric reviews 19, 341-366.

Honore, B.E., Kyriazidou, E., Udry, C. (1997): Estimation of type-3 tobit models using symmetric trimming and pairwise comparisons, Journal of econometrics 76, 107-128.

Honore, B.E., Lewbel, A. (2002): Semiparametric binary choice panel data models without strictly exogenous regressors, Econometrica 70, 2053-2063.

Horowitz, J.L., Markatou, M. (1996): Semiparametric estimation of regression models for panel data, Review of economic studies 63, 145-168.

Hsiao, C. (2004): Analysis of panel data, 2nd ed., Cambridge university press: Cambridge.

Hübler, O. (1990): Lineare Paneldatenmodelle mit alternativer Störgrößenstruktur, Neuere Entwicklungen in der Angewandten Ökonometrie, G. Nakhaeizadeh, K.-H. Vollmer, Physica-Verlag: Heidelberg, 65-99.

Hübler, O. (2003): Neuere Entwicklungen in der Mikroökonometrie, Empirische Wirtschaftsforschung - Methoden und Anwendungen, W. Franz, H.J. Ramser, M. Stadler (Hrsg.), Mohr Siebeck: Tübingen, 1-35.

Inkmann, J. (2001): Conditional moment estimation of nonlinear equation models, Springer: Berlin.

Jacubson, G. (988): The Sensitivity of labor-supply parameter estimates to unobserved individual effects: fixed- and random-effects estimates in a nonlinear model using panel data, Journal of labor economics 6, 302-329.

Keane, M. (1994): A computationally practical simulation estimator for panel data, Econometrica 62, 95-116.

König, A. (1997): Schätzen und Testen in semiparametrisch partiell linearen Modellen für die Paneldatenanalyse, Diskussionspapier Nr. 208, Universität Hannover.

König, A. (2002): Nichtparametrische und semiparametrische Schätzverfahren für die Paneldatenanalyse, Lit-Verlag: Münster.

Kyriazidou, E. (1995): Essays in estimation and testing of econometric models, Dissertation, Evanston (Illinois).

Kyriazidou, E. (1997): Estimation of a panel data sample selection model, Econometrica 65, 1335-1364.

Lechner, M. (2002): Eine Übersicht über gängige Modelle der Panelökonometrie und ihre kausale Interpretation, Allgemeines Statistisches Archiv 86(2), 125–143.

Lee, M.-J. (1999): Nonparametric estimation and test for quadrant correlation in multivariate binary response models, Econometric reviews 18, 387-415.

Lee, M.-J. (1999a): A root-N consistent semiparametric estimator for related-effect binary response panel data, Econometrica 67, 427-433.

Lee, M.-J. (2002): Panel data econometrics: Method-of-moments and limited dependent variables, Academic press: New York.

Li, D., Stengos, T.(2002): The Partially linear regression model: Monte Carlo evidence from the projection pursuit regression approach, Economics letters 75, 11-16.

Li, Q., Hsiao, C. (1998): Testing serial correlation in semiparametric panel data models, Journal of econometrics 87, 207-237.

Li, Q., Lu, X., Ullah, A. (2003): Estimating average derivative by local least squares method, Journal of nonparametric statistics.

Li, Q., Stengos, T. (1996): Semiparametric estimation of partially linear panel data models, Journal of econometrics 71, 389-397.

Li, Q., Ullah, A. (1998): Estimating partially linear panel data models with one-way error components, Econometric Review 17, 145-166.

Longford, N.T. (1993): Random coefficient models, Clarendon press: Oxford.

Maddala, G.S. (1993): The econometrics of panel data, Volume I and II, Elgar reference collection: Cambridge.

Manski, C.F. (1975): Maximum score estimation of the stochastic utility model of choice, Journal of Econometrics 3, 205-228.

Manski, C.F. (1985): Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator, Journal of econometrics 27, 313-333.

Manski, C.F. (1987): Semiparametric analysis of random effects linear models from binary panel data, Econometrica 55, 357-362.

Matyas, L., Sevestre, P.(ed.) (1996): The econometrics of panel data, Handbook of theory and applications, 2nd ed., Kluwer academic publishers, Dordrecht.

McCullagh, P., Nelder, J.A (1983): Generalized linear models, Chapman & Hall: London.

Nijman, T., Verbeek, M. (1992): Nonresponse in panel data: The impact on estimates of a life cycle consumption function, Journal of applied econometrics 7, 243-257.

Pagan, A., Ullah, A. (1999): Nonparametric analysis, Cambridge university press: Cambridge.

Powell, J.L., Stock, J.H., Stoker, T.M. (1989): Semiparametric estimation of index coefficients, Econometrica 57, 1403-1430.

Raudenbusch, S.W., Bryk, A.S. (2002): Hierarchical linear models: applications and data analysis methods, Sage publications.

Revelt, D., Train, K. (1998): Mixed logit with repeated choices of appliance efficiency levels, Review of economics and statistics 80, 647-657.

Rochina-Barrachina, M.E. (1997): A New estimator for panel data sample selection models, Thesis dissertation UCL.

Robinson, P.M. (1988): Root-N-consistent semiparametric regression, Econometrica 56, 931-954.

Ullah, A., Roy, N. (1998): Nonparametric and semiparametric econometrics of panel data, Handbook of applied economics, A. Ullah, D.E.A. Giles, eds., Marcel Dekker: New York, 579-604.

Vella, F., Verbeek, M. (1999): Estimating and interpreting models with endogenous treatment effects, Journal of business and economic statistics 17, 473-478.

Verbeek, M. (1991): The design of panel surveys and the treatment of missing observations, Dissertation, University of Brabant.

Verbeek, M., Nijman, T. (1992): Testing for selectivity bias in panel data models, International economic review 33, 681-703.

Wandsbeek, T.J., Kapteyn, A. (1989): Estimation of the error components model with incomplete panels, Journal of econometrics 41, 341-261.

White, H. (1982): Instrumental variables regression with independent observations, Econometrica 50, 483-499.

Wooldridge, J.M. (1995): Selection correction for panel data models under conditional mean independence assumptions, Journal of econometrics 68, 115-132.

Wooldridge, J.M. (2002): Econometric analysis of cross section and panel data, MIT Press: Cambridge (Mass.)

Zabel, J.E. (1992): Estimating fixed and random effects with selectivity, Economics letters 40, 269-272.