Rendtel, Ulrich; Buechel, Felix

Tests for non-ignorable panel attrition and their application on wage estimates from the German Socio-Economic Panel (GSOEP)

DIW Discussion Papers, No. 89

Provided in Cooperation with:
German Institute for Economic Research (DIW Berlin)

Suggested Citation: Rendtel, Ulrich; Buechel, Felix (1994) : Tests for non-ignorable panel attrition and their application on wage estimates from the German Socio-Economic Panel (GSOEP), DIW Discussion Papers, No. 89, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:
https://hdl.handle.net/10419/95751

DIW Diskussionspapiere
Discussion Papers

Discussion Paper No. 89

Tests for non-ignorable panel attrition and their application on wage estimates from the German Socio-Economic Panel (GSOEP)

by
Ulrich Rendtel and Felix Buechel

Deutsches Institut für Wirtschaftsforschung, Berlin
German Institute for Economic Research, Berlin

Deutsches Institut für Wirtschaftsforschung

Discussion Paper No. 89

Tests for non-ignorable panel attrition and their application on wage estimates from the German Socio-Economic Panel (GSOEP)

by
Ulrich Rendtel and Felix Buechel

Berlin, February 1994

Deutsches Institut für Wirtschaftsforschung, Berlin
Königin-Luise-Str. 5, 14191 Berlin
Telefon: 49-30 - 82 991-0
Telefax: 49-30 - 82 991-200

Tests for non–ignorable panel attrition and their application on wage estimates from the German Socio–Economic Panel (GSOEP)*

by
Ulrich Rendtel†
and
Felix Buechel‡

February 1994

1 Introduction

Panel surveys are plagued by the successive drop–out of persons, which refuse to continue to participate or which are lost due to the problem to recontact them in the next wave of the panel. Such losses not only reduce the sample size, they may also bias estimates on the basis of the persons that remain in the sample. This may occur if the drop-out is related with $f(Y \mid X)$, the distribution of interest. The panel attrition is ignorable if conditioning on the participation does not affect $f(Y \mid X)$; i. e. we have $f(Y \mid X, S = 1) = f(Y \mid X, S = 0)$, where $S = 1$ indicates a participation and $S = 0$ indicates a drop–out.

It is easy to pose the hypothesis of non–ignorable panel attrition: For example, if one tries to explain the log–income Y by some X-covariates, one may assume that a move from one town to another will happen if the earning at the new place is more attractive than at the old place. But movers are at a higher risk to drop–out off the panel than persons who don't move. So under this hypothesis one would assume that the income estimation on the basis of the observed incomes gives at least an underestimate of the model's constant.

There is also a hypothesis which results in an opposite attrition effect. This hypothesis assumes that low–income persons tend to show a reduced interest in surveys and hence have a higher drop–out rate. As a result of such a hypothesis

*We would like to thank Johannes Schwarze for helpful comments.

†German Institute for Economic Research, Berlin
‡Technical University Berlin, Dept. of Economics

an income estimate of the observed incomes would result in an over–estimate at least of the model's constant.

It may turn out that both hypotheses contain some true aspects which in the end will compensate each other in their effect on the estimation of the regression model. So in the end the standard regression estimates may still give consistent estimates of the regression parameters.

While it is easy to pose the question of effects of the panel attrition on the distributional properties of some model estimate, it is hard to answer such questions. The key difficulty is the knowledge of the model distribution $f(Y \mid X, S = 0)$ for those who dropped out of the panel. To be honest, one has to state that without any knowledge about $f(Y \mid X, S = 0)$ the problem is not solvable on the basis of the observed data alone.

In this case there is need for some a–priori knowledge. The standard cross-sectional selection models, which link a regression equation and probit selection equation by a correlation of the error terms (cf. for example Amemiya 1984, Heckman 1979, Lee 1982 and Lee/Maddala 1985), use model assumptions to estimate the effect of the selection rule on the estimation of the regression model. The basic tools are: (a) Distributional assumptions (usually a bivariate normal distribution for the error terms) and (b) a-priori zero coefficients for variables that appear in the probit but not in the regression equation. Usually the importance of such assumption — which cannot be verified from the observed data — is widely ignored, although the warning of Little (1982, p. 246) is apparent: "If selection is non–ignorable one can eliminate bias only by constructing a model that correctly represents the response mechanism." One might add that the detection of a bias depends also on a correct specification of the selection rule.

In a panel however there exists a lot of information about non–respondents, i.e. people that attrit. The information arises from the characteristics observed in the previous panel waves. Also the field work of the present panel wave produces relevant information; namely whether the person or the household has moved since the last interview or whether the household has split up into two separate households. There is also information about temporal dependencies: Participation in a panel turns out to be a sequential decision, where drop-out is in most — but not all — cases an absorbing state. In the case of a household panel there is also a nearly perfect dependency of participation behaviour within households. Such informations may be used to estimate drop–out probabilities for panel members. These drop-out probabilities condense our knowledge about attrition on the basis of attriters and non–attriters.

In the case of the German Socio-Economic Panel (GSOEP), which is an ongoing household panel started in 1984 (cf. Wagner et al. 1993), it turned out, that field–related characteristics are the most relevant for drop–out during the panel (cf. Rendtel (1990, 1993). For a panel study which is based on personal interviews, this is a very plausible result.

In order to answer the question whether panel attrition affects the estima-

tion of the model of interest, it is crucial to exploit the relation of the model characteristics with the field work characteristics. Under this aspect we discuss the selection model of Ridder (1990) and Verbeek (1991), which generalize the standard cross–sectional selection model to multi–wave panels.

The estimation of a joint regression–selection model is one possibility to deal with panel attrition. In many cases it would be wise to use some kind of a test to check whether there is any attrition bias before one uses a selection model which depends basically on untestable model assumptions. Here we propose a routine that uses efficiently the non–respondent information condensed in the probabilities of drop-out. The basic idea of the routine is to simulate the panel attrition on the sample $\mathcal{S}$. Such a simulation results in a smaller sample $\mathcal{S}^*$. If the distribution of $\hat{\beta}_{\mathcal{S}^*}$ is not concentrated around $\hat{\beta}_{\mathcal{S}}$, the model estimate based on $\mathcal{S}$, this may be taken as an indication that the panel attrition is non–ignorable. The paper discusses also situations, where the test does not detect a non–ignorable selection rule.

The paper also describes an alternative routine which compares the model estimate on the subsample of persons without panel attrition and the estimate on the subsample of persons with panel attrition. This routine is similar to the strategy of Becketti et al. (1988), which test wage estimates from the PSID sample for affects of panel attrition. The test routine is related with various variable addition tests which try to check whether the expectation of the error terms in the regression equation for the non–attrited observations is different from zero.

All methods are compared in two empirical examples which use data from the first 8 waves of the GSOEP. Both examples deal with wage estimation. In the first example we test an extended Mincer–type human capital model, where income is explained by schooling, experience and tenure. The data set of the second consists of males who experienced a period of unemployment of at least 12 month. We want to know whether a joint analysis of incomes before and after the unemployment phase is affected by panel attrition.

2 A joint regression–selection model for multi–wave panel data

Cross–sectional selection models link a regression equation and a probit selection equation by assumptions about the joint distribution of the error terms, cf. for example Ameniya (1984), Heckman (1979) and Lee (1982).

This basic model was extended by Ridder (1990) to treat attrition in multi–wave panels. The panel character of observations is treated by a random effects model for the error terms of the regression model and the probit selection model.

The regression model is given by:

$$Y_{i,t} = X'_{i,t}\beta + \mu_i + \nu_{i,t} \quad t \in \mathcal{T}_i \tag{1}$$

where $Y_{i,t}$ is the dependent variable for person i at panel wave t, $X_{i,t}$ is a set of regressors, μ_i is the person related variance component and $\nu_{i,t}$ is a shock component which is independent from μ_i. The set $\mathcal{T}_i$ consists of those wave numbers, for which the regression relation is defined for person i. In the original formulation Ridder used $\mathcal{T}_i = \{1, \ldots, T\}$, but there are apparent cases where such a relation is not valid for all sample members. For example, if $Y_{i,t}$ is the earned income, then $Y_{i,t}$ is undefined, if a person is not in the labor force at wave t.

The attrition model is given by:

$$S^*_{i,t} = Z'_{it}\gamma + \xi_i + \eta_{i,t} \tag{2}$$

and

$$S_{i,t} = \begin{cases} 1 & \text{if} \quad S^*_{i,t} \geq 0 \\ 0 & \text{else} \end{cases} \tag{3}$$

where $S_{i,t}$ is an indicator that person i is at wave t in the sample ($S_{i,t} = 1$) or not ($S_{i,t} = 0$). $Z_{i,t}$ is a set of covariates for the latent variable $S^*_{i,t}$. The random effect ξ_i describes a person-related propensity towards participation and $\eta_{i,t}$ is a shock component.

These two equation are linked by the assumption that there may be correlations between μ_i and ξ_i and/or between $\nu_{i,t}$ and $\eta_{i,t}$. The joint distribution of $\nu_i = (\ldots, \nu_{i,t}, \ldots)'$, $\eta_i = (\ldots, \eta_{i,t}, \ldots)'$, μ_i and ξ_i is taken to be a multivariate normal distribution with:

$$\begin{pmatrix} \nu_i \\ \eta_i \\ \mu_i \\ \xi_i \end{pmatrix} \sim N\left(0, \begin{pmatrix} \sigma^2_\nu I_i & \sigma_{\nu\eta} I_i & 0 & 0 \\ \sigma_{\nu\eta} I_i & \sigma^2_\eta I_i & 0 & 0 \\ 0 & 0 & \sigma^2_\mu & \sigma_{\mu\xi} \\ 0 & 0 & \sigma_{\mu\xi} & \sigma^2_\xi \end{pmatrix}\right) \tag{4}$$

where I_i is the identity matrix with the dimension equal to the number of elements in $\mathcal{T}_i$.

If the covariate sets $X_{i,t}$ and $Z_{i,t}$ are equal, it is easy to show that the selection rule is ignorable, if $\sigma_{\nu,\eta} = 0$ and $\sigma_{\mu,\xi} = 0$. In this case the likelihood of the parameter $\vartheta = (\beta, \sigma^2_\nu, \sigma^2_\mu, \gamma, \sigma^2_\eta, \sigma^2_\xi)$ factors into a product where the first factor involves only the parameter values $\vartheta_1 = (\beta, \sigma^2_\nu, \sigma^2_\mu)$ and observed values of $Y_{i,t}, X_{i,t}$ ($t \in \mathcal{T}_i$), and the second factor involves only the parameter values $\vartheta_2 = (\gamma, \sigma^2_\eta, \sigma^2_\xi)$ and the observations $S_{i,t}, X_{i,t}$ ($t \in \mathcal{T}_i$). Therefore the ML-estimation for the regression model and the selection model may be carried out independently[1].

[1]We do not regard here the case that there are parameter restrictions between ϑ_1 and ϑ_2.

As a consequence of the likelihood factorization $Y_{i,t}$ and $S_{i,t}$ are conditionally independent given $X_{i,t}$, cf. Dawid (1979). Hence conditioning on the $S_{i,t}$ ($t \in \mathcal{T}_i$) is irrelevant for the distribution of $f(Y_{i,t} \mid X_{i,t})$.

If $X_{i,t}$ is different from $Z_{i,t}$ — as suggested by Ridder — the previous situation applies if the coefficient of the variable that was omitted is zero. For example Ridder suggests to add variables like $S_{i,t-1}$ or $L_{i,t} = \sum_{s=1}^{t-1} \prod_{r=1}^{s} S_{i,t-r}$, the length of the present participation period, to $Z_{i,t}$ but not to $X_{i,t}$. As long as the corresponding β–coefficient in the regression equation is zero, the previous situation applies.

If $S_{i,t-1}$ or $L_{i,t}$ are not conditionally independent from $Y_{i,t}$ given $X_{i,t}$, the ML-estimation of the joint regression selection model will give biased estimates[2]. For the cross–sectional case, which is a special case of this model, Rendtel (1992) has shown in a simulation study that in such cases the ML–estimates of β inhibit a bias. In some cases the bias is higher than the bias of the OLS estimate that was induced by the non–ignorable selection. Such features contrast remarkably with the original intention of models for non–ignorable selection: the consistent estimation of β, σ_ν^2 and σ_μ^2.

There is no way to test the β-coefficient of such extra Z–variables from the observed data alone. Therefore the inclusion of extra Z-variables not contained in $X_{i,t}$ is equivalent to a–priori knowledge[3] about the β-coefficient of such variables. If the extra variable is $S_{i,t-1}$, this a–priori knowledge is equivalent to assuming that $\sigma_{\mu,\xi}$, the correlation between the random effects, is 0. So we cannot test such a hypothesis with a model where $S_{i,t-1}$ is included in $Z_{i,t}$ but not in $X_{i,t}$.[4]

The same arguments are also true for other field related variables that are known to be important for the participation in a panel. Unless it is known that for example the change of the interviewer is due to the field institute there is no way to get the information whether such a change is correlated with $Y_{i,t}$ from the observed data alone. So the use of non–respondent information which exists in a panel is rather restricted in such selection models. If there are no extra Z-variables the identification of $\sigma_{\nu,\eta}$ and $\sigma_{\mu,\xi}$ depends solely on the assumption of the multivariate normal distribution.

The attrition model treated here has also problems to deal with time dependent covariates in the regression equation, since these covariates have missing values in the selection equation for waves with $S_{i,t} = 0$ and $t \in \mathcal{T}_i$. So we have to omit these variables from the selection equation. If the corresponding γ-coefficient is zero, no problems arise. If the γ-coefficient is different from zero, again biased ML-estimates of β, γ and the covariance terms may arise, cf. Rendtel (1992) for the cross–sectional case. Like in the case of additional Z-variables there is no

[2]See also Little/Rubin (1987, p.230), who discuss the behaviour of the two step estimation of the cross-sectional model.

[3]In most empirical studies the "knowledge" turns out to be a mere assumption.

[4]Indeed one can use $S_{i,t-1}$ as an additional variable in the regression equations to test such a hypothesis, see section 5.

possibility to test on the basis of the observed data whether the γ-coefficient is zero.

Again a–priori knowledge is needed. Some authors (for example Verbeek 1991) try to circumvent this problem by imputing the unobserved time–dependent X value by the last observed X value. Although this rule of thumb seems reasonable in some cases, there is no theoretical justification for such a rule nor it is clear what it does imply for the estimation of the model. By the same strategy one could "solve" the whole attrition problem simply by imputing the missing $Y_{i,t}$ by the last observed value.

If drop–out is an absorbing state we get statistical information only for the first wave of drop–out. In this case the attrition model assures the validity of parameter estimates only for that additional wave, but not for all $t \in \mathcal{T}_i$ as it might be intended. The attrition model also does not care about $\mathcal{T}_i$, for example the persons participation in the labour force. It is implicitly assumed that also for $S_{i,t} = 0$ there exists a $Y_{i,t}$. While this may be a plausible assumption in some cases, there may be situations where this is a crucial assumption. This would advocate for a much more complex model, which consists of three equations: A regression equation, a participation equation for $\mathcal{T}_i$, and an attrition equation. As shown above, the "simple" two equations model is hard to specify and its estimation is numerically tedious (see below). A three equations model would worsen the situation. It is not treated here.

There is a last topic to be mentioned in the general discussion of the attrition model: It assumes constant coefficients for the regression equation and the attrition model. While this seems reasonable for the regression equation, the attrition behavior seems to be more unstable especially at the initial panel waves. Although it is easy to formulate a model with time dependent γ this may drastically increase the number of model parameters.

The computation of the likelihood of the joint regression selection model involves two–fold integration. This is seen as follows, cf. Verbeek (1991).

Let Y_i^{obs} the vector of the observed $Y_{i,t}$ values and S_i^{sel} the vector of all selection indicators $S_{i,t}$ $t \in \mathcal{T}_i$. Let T_i the number of elements of S_i^{sel} and O_i the number of elements of Y_i^{obs}.

Then the log–likelihood l_i factors into:

$$\begin{aligned} l_i &= lnf(Y_i^{obs}, S_i^{sel}) \\ &= lnf(S_i^{sel} \mid Y_i^{obs}) + lnf(Y_i^{obs}) \end{aligned} \tag{5}$$

$lnf(Y_i^{obs})$ may be computed according to Hsiao (1986, p. 38). In order to compute the likelihood contribution of $lnf(S_i^{sel} \mid Y_i^{obs})$ one has to determine the conditional distribution $\xi_i 1_i + \eta_i \mid \mu_i 1_{O_i} + \nu_i^{obs}$, where 1_i is the vector with T_i ones, 1_{O_i} is the vector of O_i ones and ν_i^{obs} is the vector of shocks that correspond to Y_i^{obs}. One may show (cf. Verbeek 1991, p. 104 ff.) that the distribution of $\xi_i + \eta_{i,t} \mid \mu_i 1_{O_i} + \nu_i^{obs}$

can be represented by the sum $u_{i,t} + \delta_{i,1} + S_{i,t}\delta_{i,2}$ where the summands are normal distributed with means:

$$E(\delta_{1,i}) = E(\delta_{2,i}) = 0 \tag{6}$$

$$\begin{aligned} E(u_{i,t}) &= S_{i,t}\frac{\sigma_{\nu\eta}}{\sigma_\nu^2}\left(\epsilon_{i,t} - \frac{\sigma_\mu^2}{\sigma_\nu^2 + 0_i\sigma_\mu^2}\sum_{\tau\in\mathcal{T}_i} S_{i,\tau}\epsilon_{i,\tau}\right) + \frac{\sigma_{\mu\xi}}{\sigma_\nu^2 + 0_i\sigma_\mu^2}\sum_{\tau\in\mathcal{T}_i} S_{i,\tau}\epsilon_{i,\tau} \\ &=: c_{i,t} \end{aligned} \tag{7}$$

Where $\epsilon_{i,t} = \mu_i + \nu_{i,t}$. $S_{i,t}$ is treated here as a constant. For the variances of $u_{i,t}, \delta_{1,i}$ und $\delta_{2,i}$ we get:

$$\begin{aligned} V(u_{i,t}) &= \sigma_\eta^2 - S_{i,t}\frac{\sigma_{\nu\eta}^2}{\sigma_\nu^2} \\ &=: s_{i,t}^2 \end{aligned} \tag{8}$$

$$\begin{aligned} V(\delta_{1,i}) &= \sigma_\xi^2 - \frac{\sigma_{\mu\xi}^2 0_i}{\sigma_\mu^2 0_i + \sigma_\nu^2} \\ &=: \omega_{i,1} \end{aligned} \tag{9}$$

$$\begin{aligned} V(\delta_{2,i}) &= \frac{\sigma_{\nu\eta}^2\sigma_\mu^2}{\sigma_\nu^2(\sigma_\mu^2 0_i + \sigma_\nu^2)} \\ &=: \omega_{i,2} \end{aligned} \tag{10}$$

The $u_{i,t}$ are mutually independent and also independent from $\delta_{1,i}$ and $\delta_{2,i}$. The covariance of $\delta_{1,i}$ and $\delta_{2,i}$ is given by:

$$\begin{aligned} Cov(\delta_{1,i}, \delta_{2,i}) &= \frac{\sigma_{\mu\xi}\sigma_{\nu\eta}}{\sigma_\mu^2 0_i + \sigma_\nu^2} \\ &=: \omega_{i,1,2} \end{aligned} \tag{11}$$

Hence we get for $f(S_i^{sel} \mid Y_i^{obs})$:

$$\begin{aligned} &f(S_i^{sel} \mid Y_i^{obs}) = \\ &\int\int \prod_{\tau\in\mathcal{T}_i} \Phi\left[(2S_{i,t} - 1)\frac{Z_{i,t}'\gamma + c_{i,t} + \delta_{1,i} + S_{i,t}\delta_{2,i}}{s_{i,t}}\right] f(\delta_{1,i}, \delta_{2,i}) d\delta_{1,i}, d\delta_{2,i} \end{aligned} \tag{12}$$

Where $f(\delta_{1,i}, \delta_{2,i})$ is the bivariate density of $\delta_{1,i}$ and $\delta_{2,i}$. For the computation of $c_{i,t}$, $\epsilon_{i,t}$ is replaced by $Y_{i,t} - X'_{i,t}\beta$.

One may transform the above 2-fold integral such that the integration is with respect to the product of two standard normal distributions, cf. Rendtel (1994). For each integration one may use the Hermite approximation:

$$\int H(x)e^{-x^2}dx \approx \sum_{j=1}^{J} H(x_j)g_j \tag{13}$$

where x_j are preassigned evaluation points and g_j are known weights, cf. Stroud/-Secrest (1966). The approximation is exact for polynomials up to order $2J-1$. The empirical results (see below) indicate that the estimation results are quite stable with respect to the approximation order. This extends the findings of Buttler/Moffitt (1982) for random effects probit models.

3 The simulation of panel attrition as a tool for the detection of non–ignorable drop–out

In this section we present a routine that may serve as a test for the detection of non–ignorable panel attrition. It uses efficiently all information about the non–respondents and indicates which coefficients are biased in a certain direction.

The basic idea is as follows. Let $\mathcal{P}$ the panel sample if no panel attrition is present. By assumption $\hat{\beta}_{\mathcal{P}}$, the estimate of β on the basis of $\mathcal{P}$, is a consistent estimate. Let $\mathcal{S}$ be the actual panel sample, which is affected by attrition, and $\hat{\beta}_{\mathcal{S}}$ the estimate of β on the basis of $\mathcal{S}$. If attrition yields an inconsistent estimation of β then $\hat{\beta}_{\mathcal{P}}$ and $\hat{\beta}_{\mathcal{S}}$ will differ with high probability.

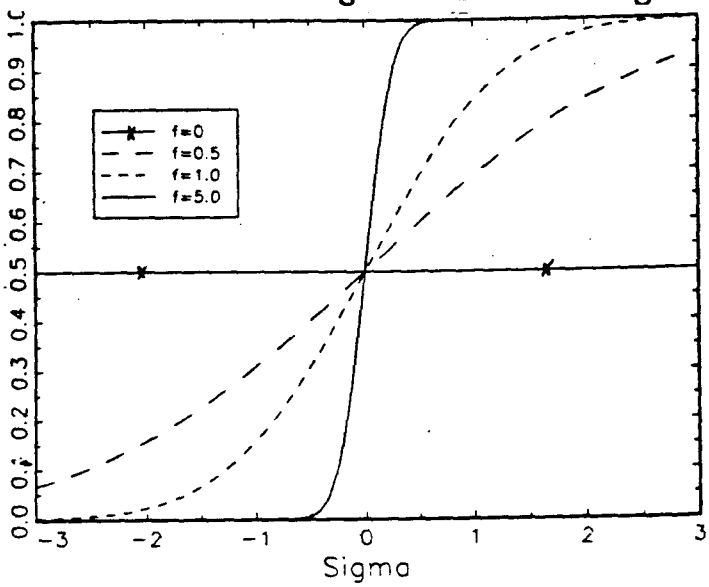
In many instances attrition is regarded as a stochastic censoring with affects the shape of a distribution in a "smooth" way. For example in the attrition model of the previous section one gets for $\sigma^2_\mu = \sigma^2_\xi = 0$ and the normalization $\sigma^2_\eta = 1$:

$$\begin{aligned} Pr(S_{i,t} = 1 \mid \nu_{it}) &= Pr(Z'_{i,t}\gamma + \eta_{i,t} \geq 0 \mid \nu_{i,t}) \\ &= \Phi\left(\frac{Z'_{it}\gamma + \rho_{\nu\eta}\frac{\nu_{i,t}}{\sigma_\nu}}{\sqrt{1-\rho^2_{\nu\eta}}}\right) \end{aligned} \tag{14}$$

This is the stochastic censoring model. The tightness of the stochastic censoring depends on the ratio $f = \rho_{\nu\eta}/\sqrt{1-\rho^2_{\nu\eta}}$. Figure 1 shows $Pr(S_{i,t} = 1 \mid \nu_{i,t})$ for $f = 0, 0.5, 1$ and 5 and $Z'_{i,t}\gamma = 0$. The corresponding values for $\rho_{\nu,\eta}$ are 0, 0.44, 0.71 and 0.98.

Now suppose we perform an independent second attrition experiment with those units that survived the first attrition. This would result in a sample $\mathcal{S}^*$.

Figure 1: **Stochastic censoring with different degrees of tightness**



This is indicated by the selection dummies $S^*_{i,t}$. The probability that unit i is in $\mathcal{S}^*$ is therefore:

$$Pr(S^*_{i,t} = 1 \mid \nu_{i,t}) = \Phi\left(\frac{Z'_{i,t}\gamma + \rho_{\nu\eta}\frac{\nu_{i,t}}{\sigma_\nu}}{\sqrt{1-\rho^2_{\nu\eta}}}\right)^2 \tag{15}$$

Figure 2 compares $Pr(S_{i,t} = 1 \mid \nu_{i,t})$ and $Pr(S^*_{i,t} = 1 \mid \nu_{i,t})$ for $f = 0.5, 1$ and 5. It gives the impression that $Pr(S^*_{i,t} = 1 \mid \nu_{i,t})$ equals approximately $Pr(S_{i,t} = 1 \mid \nu_{i,t})$ shifted to the right, where the size of the shift depends on f. For $f = 5$ which is a nearly deterministic censoring the shift is smallest.

Now we may regard the selected distribution of $\nu_{i,t}$:

$$f(\nu_{i,t} \mid S_{i,t} = 1) \;=\; \frac{Pr(S_{i,t} = 1 \mid \nu_{i,t}) f(\nu_{i,t})}{\int Pr(S_{i,t} = 1 \mid \nu_{i,t}) f(\nu_{i,t}) d\nu_{i,t}} \tag{16}$$

where $f(\nu_{i,t})$ is the distribution of $\nu_{i,t}$ in the unattrited Panel $\mathcal{P}$. Figure 3 compares $f(\nu_{i,t})$, $f(\nu_{i,t} \mid S_{i,t} = 1)$ and $f(\nu_{i,t} \mid S^*_{i,t} = 1)$ for the censoring factors f=0.5, 1 and 5. The figure clearly display a shift of $f(\nu_{i,t} \mid S_{i,t} = 1)$ vs. $f(\nu_{i,t})$ and also a shift of $f(\nu_{i,t} \mid S^*_{i,t} = 1)$ vs. $f(\nu_{i,t} \mid S_{i,t} = 1)$.

Since the size of the shift indicates the approximate bias of the parameter estimates, an estimate on the basis of $\mathcal{S}^*$ is biased with respect to $\hat{\beta}_{\mathcal{S}}$, the estimate of β on the basis of S. We may take such a bias as an indication of the bias of $\hat{\beta}_{\mathcal{S}}$ with respect to $\hat{\beta}_{\mathcal{P}}$. As figure 3 immediately shows such conclusion fails in the case of deterministic censoring. Also the size of t[illegible] t diminishes if we compare the effect of the original attrition with [illegible] the simulated

Figure 2: **Comparision of $Pr(S_{i,t} = 1 \mid \nu_{i,t}) = \Phi(f * \nu_{i,t})$ and $Pr(S^*_{i,t} = 1 \mid \nu_{i,t}) = \Phi(f * \nu_{i,t})^2$ for f=0.5, 1 and 5.**

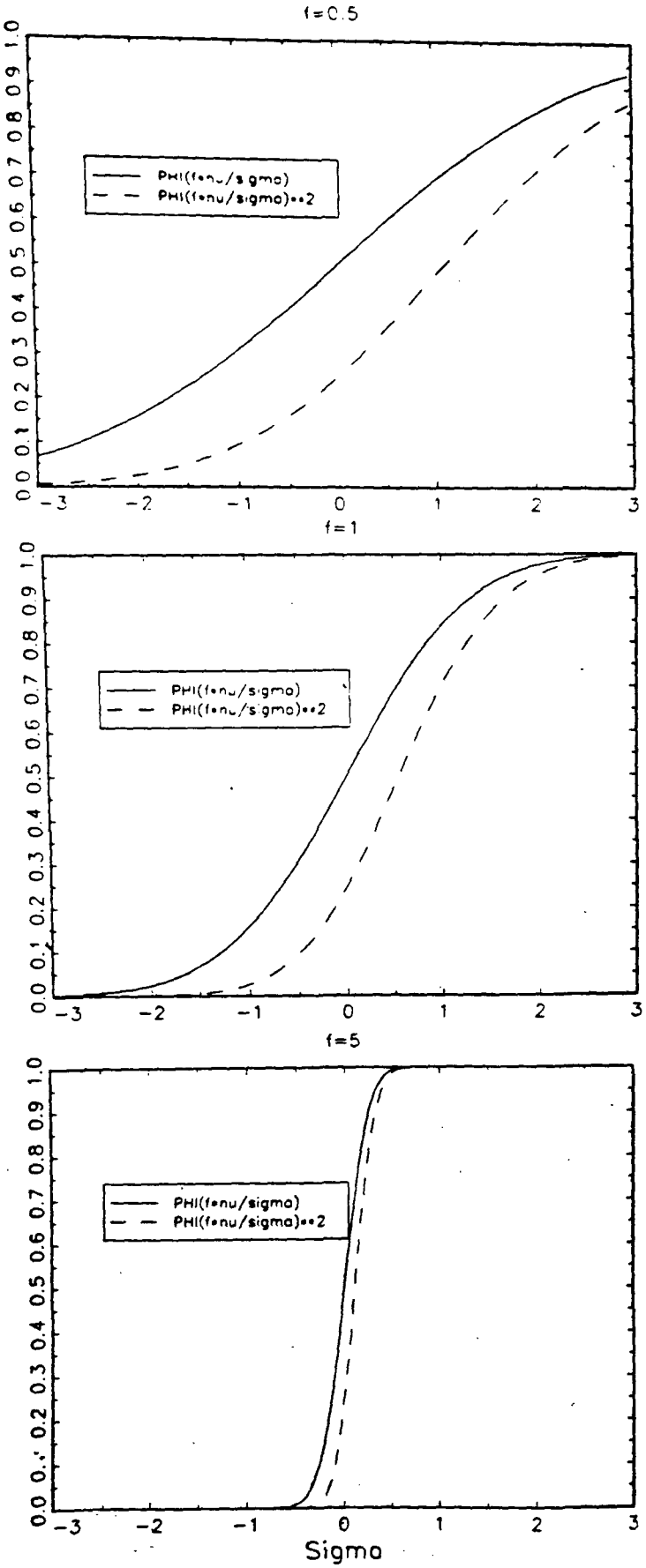
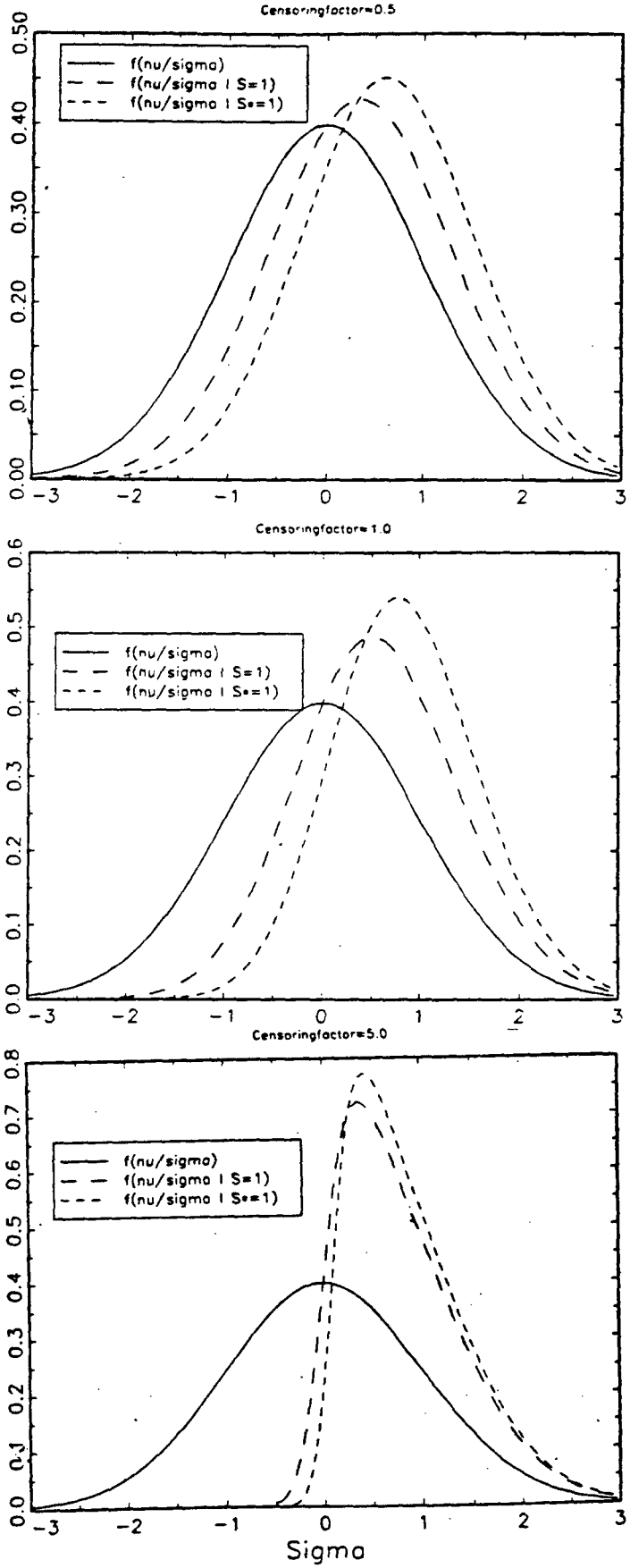


Figure 3: **Comparison of the unselected distribution of $\nu_{i,t}$ with the censored distributions $\nu_{i,t} \mid S_{i,t}$ and $\nu_{i,t} \mid S^*_{i,t}$ for f=0.5, 1 und 5.**



attrition. Nevertheless the correlation $\rho_{\nu,\eta}$ has to become very high until the simulated attrition looses all its power.

One may show similar effects also for discrete characteristics, cf. Rendtel (1994). The essential point is that the first attrition leaves enough observations in S that can be dropped out by the artificial attrition.

In order to perform a artificial attrition experiment one has to know the probabilities of attrition. Here we have to use our information about the non-respondents. The analysis of drop-out may be carried out on the basis of personal characteristics in the previous panel wave and field work characteristics of the present panel wave. Of course, one has to admit that the personal characteristics of the present panel wave are unknown for the non-respondents.

Since the value in the previous wave is known, there may be only a potential influence of a change of a characteristic on the participation behavior that remains undetected. In many cases such changes are more or less correlated with fieldwork characteristics. For example, a change of the employer may result in a move of a household which is recognized in the present panel wave. There are good reasons to believe that the causation of drop–out is in most cases a matter of fieldwork. In the example: It is not the change of the employer but the change of the interviewer that causes the drop–out.

Therefore one should model drop–out by known field–related variables. The relation of the field–related variables to the present wave variables (which are unknown for the attriters) is established by the simulation of the attrition on the respondents. In the example: If people with a change of the interviewer are more likely to attrit and attrition is simulated according to the probabilities estimated on the basis of this characteristic this will lead to an increased simulated drop–out rate of movers with a change of the employer.

The simulation should also reproduce the most relevant dependencies in the panel survey. Here the almost perfect sequential character of the survey should be reproduced. This is achieved by simulating attrition between the single panel waves[5] and the rule that a simulated drop–out is an absorbing state. In the case of the GSOEP, where all adult household members are interviewed, there is also a nearly perfect interaction in participation behavior, i.e. the participaton decision takes place at the household level. If the analysis is on the personal level one should not forget this dependency. This is achieved by modelling and simulating drop–out on the household level.

There are cases where the absorbing state strategy and the household decision strategy conflict with each other. This happens if "new" persons move in existing panel households. In this case a new attrition simulation should be carried out, if the "old" panel members have been "dropped–out" by simulation. Thus

[5]The estimated probabilities of drop–out between waves are part of the GSOEP data base. So there are no efforts necessary to estimate these probabilities in order to use the simulation strategy discussed here.

observations after a move–in of new persons have a higher probability to be in the simulated sample. Such a rule clearly reflects the design effects of household panels on inclusion probabilities.

Let $\hat{\beta}_{S^\bullet}$ be the estimate of β on the basis of the simulated sample. Under the hypothesis that attrition is ignorable

$$\Delta(S^*) = \hat{\beta}_S - \hat{\beta}_{S^\bullet}$$

should be concentrated around 0. In order to make statistical inferences, we have to know the distribution of $\Delta(S^*)$ under the artificial drop-out experiment. This may be done by the bootstrap technique, where the simulation experiment is repeated independently.

The replications give a univariate distribution of $\Delta_p(S^*)$ for each component of β $(p = 1, \ldots, P)$. If the distribution of $\Delta_p(S^*)$ is approximately unimodal, one would reject the hypothesis of ignorable attrition if the mean of $\Delta_p(S^*)$ departs more than two standard deviations from zero for one component.

If we assume a multivariate normal distribution for $\Delta(S^*)$ we may simultaneously test deviations of $\Delta(S^*)$ from the zero vector. For that purpose we compute

$$\hat{\mu}_\Delta = \frac{1}{R}\sum_{r=1}^{R}\Delta(S_r^*) \tag{17}$$

and

$$\hat{\Sigma}_\Delta = \frac{1}{R-1}\sum_{r=1}^{R}(\Delta(S_r^*) - \hat{\mu}_\Delta)'(\Delta(S_r^*) - \hat{\mu}_\Delta) \tag{18}$$

where $\Delta(S_r^*)$ is replication no. r $(r = 1, \ldots, R)$ of $\Delta(S^*)$. Thus one would reject the hypothesis of ignorable panel attrition if

$$T_{SIM} = \hat{\mu}_\Delta' \hat{\Sigma}_\Delta \hat{\mu}_\Delta \tag{19}$$

exceeds the critical value of a χ^2–distribution with P degrees of freedom.

The detection routine described in this section may be used not only to check effects of panel attrition on the estimation of the regression coefficients. It may be applied for the estimation of every statistical model. It uses efficiently the information about characteristics of non–respondents and dependencies of drop–out events.

The method has two drawbacks: It depends basically on the assumption that the simulated probabilities of drop–out are at least a good approximation for the true drop–out probabilities. This criticism may be met by a thorough analysis of drop–out. There are also some persons re–entering a panel. Such cases may

be used to check the assumptions about the influence of characteristics that are unknown for non–respondents on the participation behavior.

The second drawback is vanishing power of the routine if attrition acts like a non–stochastic censoring. But there are hints for presence of non–stochastic censoring: A small subgroup of persons with high estimated drop–out rate, that is completely removed by the simulation, may be taken as an indication of a nearly perfect non–stochastic censoring process.

4 The comparison of model estimates between groups with different participation behaviour

Under the assumption of ignorable drop–out the distribution of $Y_{i,t} \mid X_{i,t}$ is not affected by any conditioning on a participation scheme. Hence under the regression model and the assumption of ignorable drop–out regression estimates on different groups of participation behaviour should not differ systematically.

Let for example $\mathcal{S}_I$ the set of all panel members, whose participation behaviour is not affected by field–related drop–out[6]. Denote its sample complement by $\mathcal{S}_{II}$. The 2 subsets give the estimates $\hat{\beta}_I$ and $\hat{\beta}_{II}$. If the observations in $\mathcal{S}_I$ and $\mathcal{S}_{II}$ are independent from each other, we have:

$$V(\hat{\beta}_I - \hat{\beta}_{II}) = V(\hat{\beta}_I) + V(\hat{\beta}_{II}) \tag{20}$$

Thus one would reject the hypothesis that there is no systematic difference between $\hat{\beta}_I$ and $\hat{\beta}_{II}$ if

$$T_{ATTRIT} = (\hat{\beta}_I - \hat{\beta}_{II})'(V(\hat{\beta}_I) + V(\hat{\beta}_{II}))^{-1}(\hat{\beta}_I - \hat{\beta}_{II}) \tag{21}$$

exceeds a preassigned critical value. Under the assumption of asymptotic normality of $\hat{\beta}_I$ and $\hat{\beta}_{II}$ we may use the χ^2–distribution for T_{ATTRIT} where the degrees of freedom are given by the number of regression parameters.

This routine was used by Löwenbein/Rendtel (1991). There is some similarity with the routine of Becketti et al. (1988), which compares regression estimates for the first wave of the PSID on different subsets. While the first estimate is done on the basis of all male heads of households in wave 1, the second estimate is restricted to those persons who are in the sample at wave 8. A third estimate bases on those persons which remain in the sample until wave 14. Apparently their strategy confounds demographic and field–related drop–out. Also the subsamples are nested, which implies that the corresponding estimates are dependent.

[6]The set does not exclude person with demographic drop–out during the panel.

The major drawback of the routine discussed here is its incapability to detect certain non–ignorable selection rules. If we denote the indicator for no field–related drop–out by S_i, i.e. $S_i = \prod_{t \in \mathcal{T}_i} S_{i,t}$, the routine checks the following equality:

$$E(\epsilon_{i,t} \mid X_{i,t}, S_i = 1) = E(\epsilon_{i,t} \mid X_{i,t}, S_i = 0, S_{i,t} = 1) \tag{22}$$

where $\epsilon_{i,t} = \mu_i + \nu_{i,t}$. The routine does not check whether the distribution of the error terms of the observed and the unobserved part of a person's longitudinal profile is identical, i.e. whether

$$E(\epsilon_{i,t} \mid X_{i,t}, S_i = 0, S_{i,t} = 1) = E(\epsilon_{i,t} \mid X_{i,t}, S_i = 0, S_{i,t} = 0) \tag{23}$$

holds. In the situation of Ridder's attrition model, the above equality holds only for $\sigma_{\nu,\eta} = 0$, i.e. the shock components of the regression equation and the selection equation must not be correlated.

5 Additional variables in the regression equation

For the expectation of the observed $Y_{i,t}$ given $X_{i,t}$ and the observed participation profile $S_{i,s}$ $(s \in \mathcal{T}_i)$ we get:

$$E(Y_{i,t} \mid X_{i,t}, S_{i,s}(s \in \mathcal{T}_i)) = X'_{i,t}\beta + E(\epsilon_{i,t} \mid X_{i,t}, S_{i,s}(s \in \mathcal{T}_i)) \tag{24}$$

Under the assumption of the joint regression-selection model $E(\epsilon_{i,t} \mid X_{i,t}, S_{i,s}(s \in \mathcal{T}_i))$ has also a variance component structure:

$$E(\epsilon_{i,t} \mid X_{i,t}, S_{i,s}(s \in \mathcal{T}_i)) = \sigma_{\mu,\xi} A_{1,i} + \sigma_{\nu,\eta} A_{2,i,t} \tag{25}$$

where $A_{1,i}$ and $A_{2,i,t}$ are given by:

$$A_{1,i} = \frac{1}{\sigma_\eta^2 + T_i \sigma_\xi^2} \sum_{t \in \mathcal{T}_i} E(\xi_i + \eta_{i,t} \mid X_{i,t}, S_{i,s}(s \in \mathcal{T}_i)) \tag{26}$$

$$A_{2,i,t} = \frac{1}{\sigma_\eta^2}(E(\xi_i + \eta_{i,t} \mid X_{i,t}, S_{i,s}(s \in \mathcal{T}_i)) - \frac{\sigma_\xi^2}{\sigma_\eta^2 + T_i \sigma_\xi^2} \sum_{s \in \mathcal{T}_i} E(\xi_i + \eta_{i,s} \mid X_{i,s}, S_{i,r}(r \in \mathcal{T}_i)) \tag{27}$$

In the panel case however the computation of the additional variables is much more complicated than in the cross–sectional case. In order to determine $A_{1,i}$ and $A_{2,i,t}$ one has to know $E(\xi_i + \eta_{i,t} \mid X_{i,t}, S_{i,s}(s \in \mathcal{T}_i))$. Because of

$$E(\xi_i + \eta_{i,t} \mid X_{i,t}, S_{i,s}(s \in \mathcal{T}_i)) = \tag{28}$$
$$E_\xi(\xi \mid X_{i,t}, S_{i,s}(s \in \mathcal{T}_i)) + E_\xi(E_{\eta|\xi}(\eta_{i,t} \mid \xi_i, X_{i,t}, S_{i,s}(s \in \mathcal{T}_i)))$$

one gets:

$$E(\xi_i + \eta_{it} \mid X_{i,t}, S_{i,s}(s \in \mathcal{T}_i)) = \tag{29}$$
$$\int (\xi_i + E_\eta(\eta_{it} \mid \xi_i, X_{i,t}, S_{i,s}(s \in \mathcal{T}_i)))f(\xi_i \mid X_{i,t}, S_{i,s}(s \in \mathcal{T}_i))d\xi_i$$

Standard calculation give (cf. for example Appendix C of Rendtel (1994)):

$$E_\eta(\eta_{i,t} \mid \xi_i, X_{i,t}, S_{i,s}(s \in \mathcal{T}_i)) = (2S_{i,t} - 1)\frac{\sigma_\eta \phi(\frac{X'_{i,t}\gamma + \xi_i}{\sigma_\eta})}{\Phi((2S_{i,t} - 1)\frac{X'_{i,t}\gamma + \xi_i}{\sigma_\eta})} \tag{30}$$

and:

$$f(\xi_i \mid X_{i,t}, S_{i,s}(s \in \mathcal{T}_i)) = \frac{\prod_{t \in \mathcal{T}_i} \Phi((2S_{i,t} - 1)\frac{X'_{i,t}\gamma + \xi_i}{\sigma_\eta})\frac{1}{\sigma_\xi}\phi(\frac{\xi_i}{\sigma_\xi})}{\int_{-\infty}^{\infty} \prod_{t \in \mathcal{T}_i} \Phi((2S_{i,t} - 1)\frac{X'_{i,t}\gamma + \xi_i}{\sigma_\eta})\frac{1}{\sigma_\xi}\phi(\frac{\xi_i}{\sigma_\xi})d\xi_i} \tag{31}$$

In order to determine $A_{1,i}$ and $A_{2,i,t}$ one has to estimate a random effects probit model, which gives estimates $\hat{\gamma}, \hat{\sigma}_\epsilon$ and $\hat{\sigma}_\eta$. For each person and each wave $t \in \mathcal{T}_i$ one has to compute $E(\xi_i + \eta_{i,t} \mid X_{i,t}, S_{i,s}(s \in \mathcal{T}_i))$ by the use of equations (29), (30) and (31).

If one has overcome this computational burden, one may estimate the regression estimation augmented by $A_{1,i}$ and $A_{2,i,t}$. The result may serve as a test for non-ignorable panel attrition, where the estimates of $\sigma_{\mu\xi}$ and $\sigma_{\nu\eta}$ are tested to be different from zero. In case of $\hat{\sigma}_{\mu\xi} \neq 0$ and/or $\sigma_{\nu\eta} \neq 0$ the resulting estimate of β may serve as a consistent estimate under the joint attrition-selection model. The procedure, which was described by Verbeek (1991), is the panel extention of Heckmans (1979) two-step procedure in the cross-sectional case. Note that this procedure bases essentially on the model assumptions that were discussed in section 2.

In the special case, where $\sigma_{\mu\xi} = 0$ and $\sigma_\xi = 0$, $A_{2,i,t}$ reduces to

$$A_{2,i,t} = \frac{1}{\sigma_\eta^2} E(\xi_i + \eta_{it} \mid X_{i,t}, S_{i,s}(s \in \mathcal{T}_i)) \tag{32}$$

$$= \frac{1}{\sigma_\eta^2} E(\eta_{i,t} \mid \xi_i = 0, T_{i,t}, S_{i,s}(s \in \mathcal{T}_i))$$

$$= \frac{(2S_{i,t}-1)}{\sigma_\eta} \frac{\phi\left(\frac{X'_{i,t}\gamma}{\sigma_\eta}\right)}{\Phi\left((2S_{i,t}-1)\frac{X'_{i,t}\gamma}{\sigma_\eta}\right)}$$

which is the cross-sectional "selection correction"-variable. In general it holds that:

$$E(\epsilon_{i,t} \mid X_{i,t}, S_{i,s}(s \in \mathcal{T}_i)) \neq E(\epsilon_{i,t} \mid X_{i,t}, S_{i,t}) \tag{33}$$

Thus a cross-sectional selectivity check for each wave will in general not detect whether panel attrition is ignorable or not.

Because of the complex estimation of $E(\epsilon_{i,t} \mid X_{i,t}, S_{i,s}(s \in \mathcal{T}_i))$ one may try to find some simple ad-hoc solution, which approximates this conditional expectation. Verbeek (1991, p. 144) makes 3 proposals:

(1) $A_i = \sum_{t \in \mathcal{T}_i} S_{i,t}$, the total number of observations from person i.

(2) $A_i = \prod_{t \in \mathcal{T}_i} S_{i,t}$, the indicator for a drop-out during the panel.

(3) $A_{i,t} = S_{i,t}$, the participation indicator of the previous wave.

Note that the last variable is not informative, if the drop-out is an absorbing state.

The additional variable test rejects the hypothesis of ignorable drop-out if the corresponding estimated β coefficient differs significantly from zero. The power of such a routine depends on the fact how good $E(\epsilon_{i,t} \mid X_{i,t}, Si, s(s \in \mathcal{T}_i))$ is approximated by $A_{i,t}$. The power may be enhanced if we replace $A_{i,t}$ by some low degree polynomial in $A_{i,t}$.

The test routine which compares the estimates of β between the subsample of persons with no panel attrition and the persons with drop-out (cf. section 4) may be seen as a special case of this additional variables routine. The added variables are in this case $A_i = \prod_{t \in \mathcal{T}_i} S_{i,t}$ and all interactions $A_i * X_{i,t}$ with the regression covariates[7].

6 Wage estimates from the GSOEP

Wages and labor force participation are the central topics of the German Socio-Economic Panel (GSOEP). This ongoing household panel, which is similar to the Panel Study of Income Dynamics (PSID), was started in 1984 with a sample of

[7]There remains a slight difference: The routine in secton 4 does not restrict the variances to be equal between the two groups, while the additional variable routine does so.

about 6000 households and 12000 interviewed persons. A short description of the data base is given in Wagner et al. (1993a). Detailed information can be found in Hanefeld (1984), Wagner et al. (1991,1993b) and Witte (1989).

All household members who are older than 16 years are interviewed. The regular interviewing method is the personal interview or the self–filling in of the questionaire in the presence of the interviewer. All persons that have given an interview are followed up if they move within Germany[8].

The GSOEP was at its 10^{th} wave in 1993. Until that time it has lost 37% of its wave–1–members by panel attrition.

Because wages are the basic variables of this survey they have been choosen to be "tested" for effects of panel attrition. But it is clear from the preceding sections that such a test is valid only for a special model estimation. A simple change of the covariate set may result in a different conclusion with respect to the ignorability of the panel attrition.

6.1 Example 1: Earned income of male German employees

The model used in this section is the basic model of human capital theory, which explains the log of the earned monthly gross income by the duration of the education (schooling), the duration of the participation in the labor force (experience) and the firm specific human capital expressed by the length of the job at the present employer (tenure). Such a model was also used by Becketti et al. (1988) to evaluate the PSID[9].

During the first 8 waves of the GSOEP there are 3523 male German full-time employees in the sample which give 15670 valid income observations.

Figure 4 displays the participation behavior of these persons within each of the 8 panel waves. There is a remarkable exchange of persons with valid income data and without valid income data, which figure 4 does not reveal. There are only 831 (=23.4%) men with valid income data in all 8 waves.

A "balanced" panel analysis would only use this fraction of the sample, which excludes systematically all persons that enter the labor force or leave it during the reference period. By this way one would "loose" about 58% of the valid income observations. We do not think that this is a careful exploitation of the data.

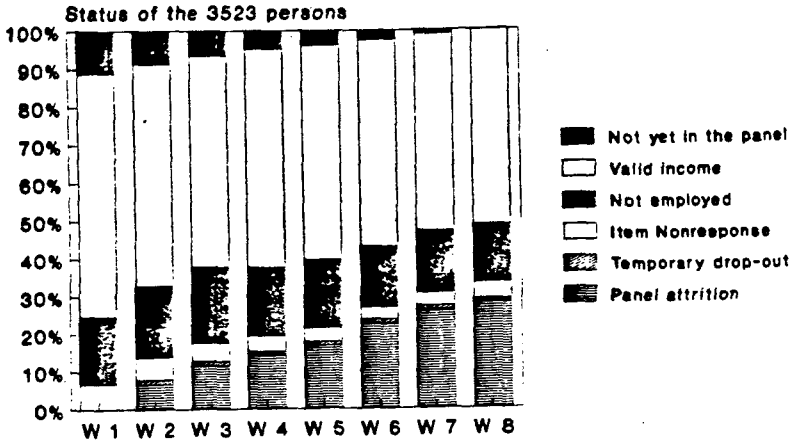
Table 1 compares the results of 3 estimates:

- The pooled regression, which assumes independent error terms.
- The feasible generalized least squares estimate (FGLS), which assumes a variance component structure for the error terms, cf. Hsiao (1986, p. 34).

[8]Until wave 5 a different follow–up rule was used, which excluded non–wave–1–persons from being followed. This is still the PSID follow–up rule.

[9]Becketti et al. did not use the tenure. Instead they used some race dummies.

Figure 4: **Participation behavior during the panel: German male employees in waves 1 to 8 of the GSOEP.**



- The ML–estimate, which assumes a normal distribution of the variance components, cf. Hsiao (1986, p. 38).

The FGLS–estimate and the ML–estimate give nearly identical results. Since the incomes have not been deflated, there are also time dummies in the regression equation, that measure time dependent income effects. If attrition is non–ignorable one might expect that at least the estimation of the time effects is biased. In table 1 all estimated coefficients of the model appear to be of reasonable size.

Table 2 compares the estimates[10] of persons without and with attrition (Column ATTRIT). The hypothesis that the two estimates are equal is to be rejected at the 2 percent level. This gives some evidence that non–ignorable drop out is present. If we compare the coefficients, we see that the returns of schooling and experience are somewhat higher for the attriters, while their constant is 10 % points lower than the constant for the non–attriters. Such differences are consistent with the joint presence of two drop–out reasons: low paid, young males with little schooling leave the panel with an above average propensity, and also persons with high returns of schooling and experience. With respect to time there is no clear trend but in most waves the attriters have higher time effects than the non–attriters. This diminuishes somehow the differences with respect to the constant.

In table 2 there are also the results of 3 tests with additional variables, which

[10]For the estimates of table 2 the FGLS estimator was used.

Table 1: **Estimation of income of male German employees.** Dependent variable: ln(monthly gross income). Source: Waves 1 to 8 of the GSOEP. Number of persons: 3523. Number of valid income observations: 15670. t-values in parenthesis.

Characteristic	Pooled Regression	Feasible Generalized Least Squares	ML-Estimate
Constant	6.675	6.717	6.717
	(428.4)	(254.9)	(256.7)
Schooling	0.075	0.070	0.070
	(79.4)	(36.3)	(36.6)
Experience	0.042	0.042	0.042
	(46.4)	(38.1)	(38.1)
Exp. * Exp.	-7.2×10^{-4}	-6.9×10^{-4}	-6.9×10^{-4}
	(-39.1)	(-30.9)	(-30.9)
Tenure	0.003	0.001	0.001
	(8.5)	(1.9)	(2.0)
1985	0.016	0.012	0.012
	(1.9)	(2.6)	(2.6)
1986	0.047	0.045	0.045
	(5.4)	(9.6)	(9.6)
1987	0.062	0.070	0.070
	(7.1)	(14.7)	(14.7)
1988	0.099	0.106	0.107
	(11.2)	(21.6)	(21.6)
1989	0.154	0.155	0.156
	(17.4)	(30.5)	(30.6)
1990	0.188	0.197	19.7
	(20.9)	(37.2)	(37.4)
1991	0.227)	0.239	0.240
	(25.1)	(43.7)	(43.8)
R^2	0.412	-	-
σ_ϵ^2	0.0813	-	-
σ_μ^2	-	0.0742	0.0721
σ_ν^2	-	0.0203	0.0204

Table 2: **Comparison of the results of some test for non–ignorable panel attrition. ATTRIT**: Comparison of model estimates on persons without panel attrition and with panel attrition. **ADDVAR**: Addition of a variable that is correlated with panel attrition. t–values in parenthesis .

	ATTRIT		**ADDVAR**		
Charac–teristic	without attrition	with attrition	W_1	S_{t-1}	$\prod_t S_t$
Constant	6.739	6.635	6.697	6.710	6.708
	(218.7)	(129.7)	(245.2)	(244.9)	(252.0)
Schooling	0.068	0.075	0.069	0.070	0.070
	(30.5)	(20.0)	(35.8)	(36.3)	(36.4)
Experience	0.041	0.047	0.042	0.042	0.042
	(32.8)	(18.9)	(36.3)	(37.9)	(38.0)
Exp.* Exp.	-6.6×10^{-4}	-7.8×10^{-4}	-6.8×10^{-4}	-6.9×10^{-4}	-6.9×10^{-4}
	(-27.2)	(-14.6)	(-30.4)	(-30.8)	(-30.9)
Tenure	0.001	-0.001	0.001	0.001	0.001
	(2.5)	(-0.8)	(1.8)	(1.8)	(1.9)
1985	0.010	0.018	0.012	0.012	0.012
	(1.8)	(2.1)	(2.7)	(2.6)	(2.7)
1986	0.046	0.041	0.046	0.045	0.045
	(8.5)	(4.4)	(9.8)	(9.6)	(9.7)
1987	0.068	0.082	0.072	0.070	0.071
	(12.3)	(8.3)	(14.9)	(14.7)	(14.8)
1988	0.104	0.122	0.108	0.107	0.107
	(18.5)	(10.9)	(21.7)	(21.6)	(21.7)
1989	0.153	0.171	0.158	0.156	0.156
	(26.8)	(13.5)	(30.5)	(30.5)	(30.6)
1990	0.195	0.217	0.200	0.197	0.198
	(33.1)	(14.2)	(37.0)	(37.3)	(37.4)
1991	0.240	0.232	0.244	0.240	0.241
	(39.3)	(12.9)	(42.9)	(43.7)	(43.8)
σ_μ^2	0.0747	0.0720			
σ_ν^2	0.0204	0.0197			
W_1	-	-	0.043	-	-
			(2.8)		
S_{t-1}	-	-	-	0.009	-
				(1.0)	
$\prod_t S_t$	-	-	-	-	0.026
					(2.5)
N	12665	3005	15670	15670	15670
	$T_{ATTRIT} = 23.75$ p-value=0.02				

are displayed under the column ADDVAR. The first additional variable W_1 indicates an interview in the first panel wave, the second variable S_{t-1} indicates an interview at the previous panel wave, while the third variable, $\prod_t S_t$ indicates panel attrition. Since the additional variables are dummies, there is no gain in using a polynomial of these variables. The indicator W_1 for wave 1 participation was used instead of the length of the participation, since it appeared to be more powerful to predict attrition behavior, cf. Rendtel 1993.

As one expects from the comparison of attriters and non–attriters the coefficient of the panel attrition indicator $\prod_t S_t$ is significant. Its size is somewhat smaller than the difference of the constants for attriters and non–attriters. This is due to the fact that the "ATTRIT"–routine regarded as an additional variable routine is not so restrictive as the addition of a mere interaction of $\prod_t S_t$ with the constant.

Also the coefficient for the wave 1 participation is significant and also its size is remarkable (4.3% points). Both results indicate that drop–out is related to some characteristics that are stable over time.

A note about the sample size is necessary at this place. It is well known that the significance of estimated parameters depends on the sample size. If the sample size is big enough, almost every estimated coefficient becomes "significant". Also this example "suffers" from an extraordinary high sample size, which forces almost every coefficient in this example to be significant. As a consequence tests for non-ignorability tend to reject the hypothesis of ignorable drop–out, if the sample size increases. This is an unpleasant feature of the ATTRIT and the ADDVAR routines.

Next we use the strategy of simulating the panel attrition according to the estimated probabilities of drop–out[11]. Table 3 displays the means and the standard deviations of $\hat{\beta}_S - \hat{\beta}_{S_*}$ for 100 replications of the simulation. For both estimates the FGLS estimate was used.

With the exception of schooling, for all other characteristics the differences between $\hat{\beta}_S$ and $\hat{\beta}_{S_*}$ may be ignored. Figure 5 displays the distribution[12] of $\hat{\beta}_S - \hat{\beta}_{S_*}$ for each characteristic[13]. All distributions are approximately bell–shaped. So the 2–σ–rule which was used to justify the ignorability of the drop–out appears to be reasonable. The distribution for the schooling coefficient is an apparent exception.

If we use the global test statistic T_{SIM}, we get $T_{SIM} = 8.9$, which yields the p–value $p = 0.71$ for 12 degrees of freedom. Therefore there is not much evidence from the simulation routine for a serious attrition bias. Also the size of the potential bias of the schooling coefficient, indicated by the mean of $\hat{\beta}_S - \hat{\beta}_{S_*}$,

[11]A documentation of the models that were used to calculate these probabilities for the GSOEP is given in Rendtel (1994).

[12]The distribution is displayed here by a kernel density estimate. For the kernel function a normal density was used.

[13]The distributions were standardized to unit variance.

Table 3: **Results of 100 replications of the artificial panel attrition.** $\hat{\beta}_S$: Estimate of β on the basis of the original sample. $\hat{\beta}_{S^*}$: Estimate of β on the basis of the artificially attrited sample S^*.

Characteristic	Mean of $\hat{\beta}_S - \hat{\beta}_{S^*}$	Std. Deviation of $\hat{\beta}_S - \hat{\beta}_{S^*}$
Constant	0.0082	0.0122
Schooling	-0.0013	0.0008
Experience	-0.0004	0.0007
Exp. * Exp.	-3.1×10^{-6}	13.3×10^{-6}
Tenure	-0.0001	0.0003
1985	-0.0009	0.0014
1986	-0.0002	0.0021
1987	0.0001	0.0023
1988	-0.0014	0.0028
1989	-0.0025	0.0029
1990	-0.0003	0.0030
1991	-0.0019	0.0033

appears to be small.

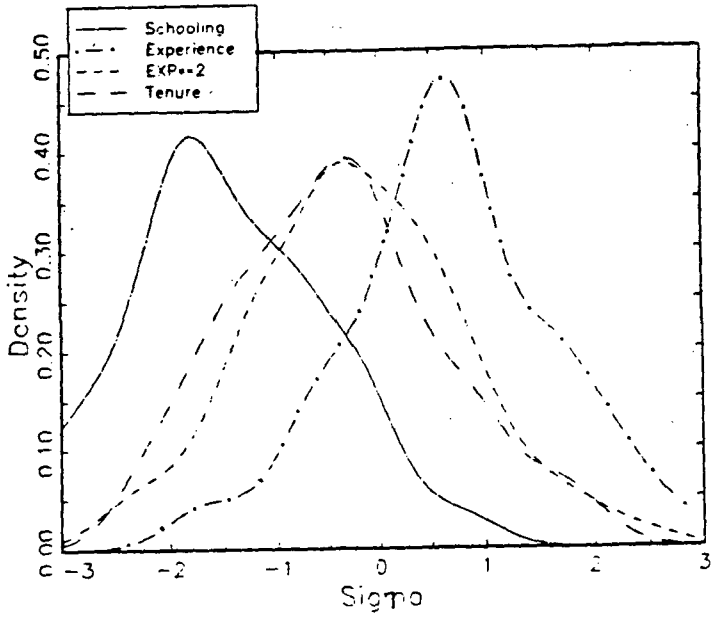
Note however that the procedure indicates a negative bias for the schooling coefficient. This implies that the estimated coefficient is an over–estimation of the true value. The implications of the ATTRIT routine go into the other direction: Here one would conclude that the schooling coefficient gets under–estimated.

One may ask for the reasons of this discrepancy between these two routines. One simple answer woud be: The estimated probabilities of drop out ignore relevant aspects of panel attrition. Note however that the decrease of the schooling coefficient and the increase of the constant — as indicated by the simulated attrition — is consistent with an "under–class" bias.

But it may be that the simulation strategy gives the correct results. One may show that the FGLS estimate remains consistent under the condition $E(\mu_i + \nu_{i,t} \mid X_{i,t}, S_{i,s}(s \in \mathcal{T}_i)) = 0$, cf. Verbeek (1991). This assumption is fulfilled if $\sigma_{\nu\eta} + \sigma_{\mu\xi} = 0$ and $E(\xi_i + \eta_{i,t} \mid X_{i,t}, S_{i,s}(s \in \mathcal{T}_i))$ does not vary over time. Since the ATTRIT routine does not check any effect with respect to $\sigma_{\nu\eta}$, the covariance of the shock components, it may well be that the above restriction holds. In this case the FGLS estimate is consistent under panel attrition as indicated by the global test statistic T_{SIM}.

Now we want to compare these results with an estimate of the joint regression–selection model. In order to do this, there are several difficulties to overcome.

Figure 5: **The distribution of $\hat{\beta}_S - \hat{\beta}_{S^*}$ for the coefficients: Schooling, Experience, Exp.* Exp. and Tenure.** Kernel density estimate on the basis of 100 replications of the artificial attrition experiment.



First we realize that we have also to deal with item nonresponse. The treatment of item nonresponse did not cause any complications for the above used test routines. In the case of the joint regression–selection model however, we find situations were persons participate but give no valid income information. There are 3 possibilities to deal with such situations:

(a) One may confound item and unit nonresponse. As a consequence the above used routines test a somewhat different drop–out bias than the estimates of the joint regression–selection model.

(b) One may "skip" observations with item nonresponse but participation at the unit level. This is technically achieved by dropping such waves from the set $\mathcal{T}_i$ of relevant time indices. This strategy requires that such dropping is ignorable.

(c) One could try to formulate a more complex model with an extra equation for item nonresponse. Because of the high computational burden of the "simple" regression–selection model and its untestable model assumptions this alternative was not used here.

Next we have to check, whether we have to drop some X–covariates from the selection equation because they are unknown in the case of panel attrition.

Since schooling is a very stable characteristic, it seems reasonable to use the last observed value in the case of panel attrition. The experience was calculated here[14] by age - 6 - schooling, so it is known also in the case of panel attrition. Apparently the tenure is unknown in the case of drop–out. But persons with a change of the employer have a higher propensity to drop out, if they move. So there is no a–priori justification that tenure has no impact on the participation behavior. If the coefficient of tenure is as small as indicated in table 1 the omission of the tenure from the participation equation should not have severe consequences. Note however that the results of table 1 are to be checked for a possible bias. Thus it might happen that the true coefficient of tenure is different from the small estimated value of table 1.

Finally we should look for a variable that explains drop–out but is independent from Y given X. Since in the GSOEP 80% of all interviewer changes happened in households that did not move[15], while in about 35% of households with a move there was no change of the interviewer, one might try to use this variable as a Z–variable that is not contained in X. Of course, such a use ignores the possibility that persons with a change of the employer that caused a move and increased the income also change the interviewer. As a consequence there would be a positive correlation between income and a change of the interviewer. We may estimate the regression coefficient of the change variable simply by adding this variable to the set of the income regressors. This yields the estimate $\hat{\beta}_{Change} = 0.007$ with a t–value of 2.9. This does indicate a small - although significant — influence of the change of the interviewer on the income, which might justify the use of this variable as an extra Z–variable. But the basis of such a reasoning is the assumption that such estimates are unbiased.

Table 4 displays the results for the separate and the simultaneous ML–estimation of the regression–selection model. The table compares the results of 4 model variants:

Model I: Item nonresponse confounded with unit nonresponse; No extra Z–variables.

Model II: Item nonresponse confounded with unit nonresponse; Change of the interviewer used as an extra Z–variable.

Model III: Item nonresponse observations skipped; No extra Z–variables.

Model IV: Item nonresponse observations skipped; Change of the interviewer used as an extra Z–variable.

For the computation of the twofold integrals in the likelihood function the Hermite approximation was used. The order of the Hermite approximation was 6.

[14] In the example below we used the actual time in the labor force

[15] This was partly due to the policy of the field institute that reduced the number of interviewers working for the GSOEP.

Table 4: **Results of separate and simultaneous ML-estimation of the joint regression–selection model. Income of German male employees.** Source: Waves 1 to 8 of the GSOEP. Number of persons: 3523. Number of valid income observations: 15670. t-values in parenthesis .

Characteristic	Separate estimation	Simultaneous estimation			
		model I	model II	model III	model IV
Regression equation					
Constant	6.717	6.715	6.715	6.716	6.716
	(256.7)	(255.9)	(254.8)	(255.0)	(255.2)
Schooling	0.070	0.070	0.070	0.070	0.070
	(36.7)	(36.7)	(36.7)	(36.6)	(36.6)
Experience	0.042	0.042	0.042	0.042	0.042
	(38.1)	(38.1)	(38.0)	(37.9)	(37.9)
Exp. * Exp.	-6.9×10^{-4}	-6.9×10^{-4}	-6.8×10^{-4}	-6.8×10^{-4}	-6.8×10^{-4}
	(-39.0)	(-31.0)	(-31.0)	(-31.0)	(-30.9)
Tenure	0.001	0.001	0.001	0.001	0.001
	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)
1985	0.012	0.012	0.012	0.012	0.012
	(2.6)	(2.6)	(2.6)	(2.5)	(2.5)
1986	0.045	0.045	0.045	0.045	0.045
	(9.6)	(9.6)	(9.5)	(9.5)	(9.5)
1987	0.070	0.070	0.070	0.070	0.070
	(14.7)	(14.7)	(14.7)	(14.6)	(14.6)
1988	0.106	0.107	0.106	0.106	0.107
	(21.6)	(21.6)	(21.5)	(21.5)	(21.5)
1989	0.156	0.156	0.156	0.156	0.156
	(30.6)	(30.5)	(30.5)	(30.5)	(30.5)
1990	0.197	0.197	0.197	0.197	0.197
	(37.4)	(37.3)	(37.3)	(37.3)	(37.2)
1991	0.240	0.240	0.240	0.240	0.240
	(43.8)	(43.8)	(43.7)	(43.7)	(43.7)
σ_μ^2	0.072	0.072	0.072	0.072	0.072
	(36.4)	(36.4)	(36.4)	(36.4)	(36.4)
σ_ν^2	0.020	0.020	0.020	0.020	0.020
	(78.0)	(78.0)	(78.0)	(78.0)	(77.9)
$\rho_{\mu\xi}$	-	-0.005	0.023	-0.001	-0.022
	-	(-0.2)	(0.5)	(-0.1)	(-0.5)
$\rho_{\nu\eta}$	-	0.000	0.004	-0.011	0.032
	-	(0.0)	(0.1)	(-0.1)	(0.6)

(continued)

Table 4: **Continuation: Estimated coefficients of the selection equation**

Characteristic	Separate estimation			
	model I	model II	model III	model IV
Constant	0.419	0.504	0.060	0.713
	(4.6)	(5.5)	(5.9)	(6.7)
Schooling	0.018	0.020	0.016	0.018
	(2.9)	(3.1)	(2.3)	(2.7)
Experience	0.037	0.033	0.049	0.049
	(6.9)	(6.4)	(7.9)	(7.5)
Exp. * Exp.	-0.0007	-0.0007	-0.0009	-0.0009
	(-6.8)	(-6.3)	(-7.4)	(-7.0)
Change of	-	-0.404	-	-0.555
Interviewer		(-14.8)		(-10.8)
$\sigma_\eta^2/(\sigma_\xi^2+\sigma_\eta^2)$	0.560	0.567	0.770	0.806
	(17.0)	(17.1)	(7.0)	(6.3)
	Simultaneous estimation			
	model I	model II	model III	model IV
Constant	0.400	0.487	0.600	0.714
	(4.3)	(5.2)	(5.9)	(6.7)
Schooling	0.018	0.020	0.016	0.018
	(2.8)	(3.0)	(2.3)	(2.7)
Experience	0.034	0.032	0.049	0.049
	(6.7)	(6.2)	(7.9)	(7.5)
Exp. * Exp.	-0.0007	-0.0007	-0.0009	-0.0009
	(-6.6)	(-6.1)	(-7.4)	(-7.0)
Change of	-	-0.394	-	-0.555
Interviewer	-	(-14.6)		(-10.9)
$\sigma_\eta^2/(\sigma_\xi^2+\sigma_\eta^2)$	0.547	0.550	0.770	0.806
	(15.5)	(15.7)	(7.0)	(6.4)

For the example in the next section which has much less observations also higher orders up to 12 were used. There was almost no difference in the estimation results, so we considered the order 6 to be sufficient also for this example.

If we look at the estimated values of $\rho_{\mu\xi}$ and $\rho_{\nu\eta}$, we find that all models confirm the hypothesis that the panel attrition is ignorable. Hence the estimated values of the separate and the simultaneous estimation are almost identical. There is some variation for the estimates of the selection model with respect to models I to IV. All models state that young males with low schooling have a higher propensity to leave the panel. The negative effect of a change of the interviewer is diminuished if item- and unit nonresponse are confounded. In this case the variance due to the shock component η amounts half of the total variance in the selection equation. In the models that skip item nonresponse observations this ratio increases up to 80%. This indicates that item nonresponse tends to occur for several times in an income profile.

The results for the joint regression-selection model state in accordance with the test routine T_{SIM} that the panel attrition is ignorable. Yet there remain some doubts whether the model fits the situation as indicated by the comparison of attriters and non-attriters. This analysis suggested that there are typically two groups of drop-outs: Low skilled, young males with c. p. low wages, and high educated, "experienced" employees with c. p. high wages. Such a two-sided censoring process cannot be fitted by the selection model used here. As a consequence the test of the inadequate parameters for non-ignorability may lead to inadequate conclusions.

6.2 Example 2: Income of male employees before and after a period of long-term unemployment

The choice of this example was motivated by the following considerations:

- Choice of a group with a high potential risk of drop-out.
- The group should be censored in "one direction" in order to fit the requirements of the selection model.
- The aim of the analysis should be panel specific, for example a before/after event comparison.

We have chosen therefore the incomes of male employees before and after long-term unemployment. The aim of the analysis is to assess effects of unemployment on earned income, if we control for basic variables like schooling, experience and tenure. The aim is here to investigate the effect of unemployment on wages for those, who are in the labor force, but not for those, where no before or after income exists; i. e. those, who enter or leave the labor force.

It is immediately clear that the selection of this sample from the population of male employees is non–ignorable, which means that the conditional distribution of incomes with respect to the control variables differs for the employees with long–term unemployment and other employees. Hence the results of our analysis are only valid for employees with such a restriction.

The emphasis in this example is again on the selective effect of panel attrition on the estimation of a regression model. Here we use the regression model of the previous example augmented by interaction terms with the indicator LTU, for observations after the long–term unemployment. Hence the coefficient of LTU*schooling measures the depreciation of school specific human capital, while LTU*experience and LTU*(exp.*exp.) describe the depreciation of occupationally achieved human capital. We did not include an interaction term with the tenure since by the nature of the problem the firm–specific human capital is usually completely devaluated after long–term unemployment[16].

In order to reduce the number of parameters we deflated the gross income. Therefore we used only 1 time dummy which measures real wage increases after 1987, the year that marks the end of the previous economic recession.

The sample consists of 224 male employees with 761 valid income observations. This sample size seems to be more typical for model estimations than the data set of the previous example, which is about 20 times larger[17]. From tables 9 and 10 (see Appendix) we report here that the gross incomes are about 30% lower than in the previous sample. The sample members are less educated (1.5 years in the mean) and have less work experience (about 5 years). 68% of the income observations are recorded after the long–term unemployment period.

Figure 6 shows the sample status of these persons during the first 8 waves of the GSOEP. If we compare this figure with the corresponding figure 4 of the previous example, we will notice that there appear almost no losses due to panel attrition during the first 3 waves of the panel. This contradicts with the knowledge that the panel attrition is highest at the beginning of the panel. The reason for this discrepancy arises from the fact that it is in most cases necessary to observe a person in two subsequent waves to assess a long–term unemployment period. Those persons who are unemployed for less than a year and attrit are not in the sample. One has to admit that this is the source of a potential attrition effect that is ignored in the subsequent analysis. The situation like the one described here is typical for a panel analysis that investigates measurements before and after an event. Again, one possible way to tackle the problem is a 3 equation model that consists from an income equation, an equation for the occurrence of long–term unemployment spells and finally one equation for the participation during the panel.

[16]We assume that employees do not return to their old firms: The usual pattern of re–entering labor force after an unemployment spell in Germany.

[17]With respect to the number of income measurements

Figure 6: **Participation behavior during the panel: Male employees with long–term unemployment.** Waves 1 to 8 of the GSOEP.

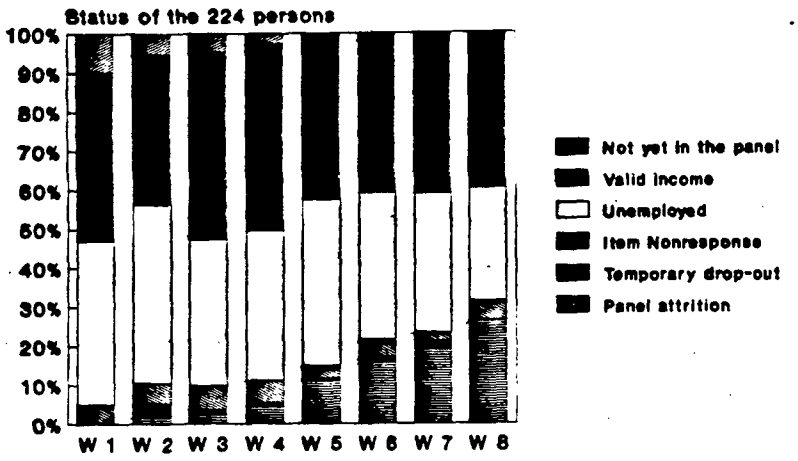


Table 5 compares the estimates of the model by pooled regression, FGLS– and ML–estimation. Again there are no differences between the FGLS– and the ML–estimate. All interactions of LTU with the other covariates are not significant. Thus a simple answer to the initial question of income effects after long–term unemployment would be: There is no further depreciation of human capital. But one has to state, that there are permanent effects by at least one year of missing experience.

A more careful answer would take the low sample size into account. If we look a the estimated coefficients for the LTU interactions we notice a tendency that higher skilled and more experienced (— and hence older —) persons are more probable to have negative income effects that result from the unemployment phase, which seems to be a plausible result. The positive coefficient for the after LTU–phase looks somewhat strange, but if one evaluates the estimated incomes for reasonable values of schooling and experience and subtracts the losses of one year of missing experience one results always in a drop of incomes after the unemployment phase.

If we compare in table 6 the results of separate estimates for attriters (with 194 income observations) and non–attriters (with 567 income observations) there appears no significant difference. The estimation for the attriters obviously suffers from the small sample size, which results in an estimate where the constant remains the only significant effect. Also the ADDVAR routine does not exhibit a systematic effect of drop–out, cf. table 6.

Table 5: **Estimation of the income of male employees with an observed period of long–term unemployment (LTU) of at least 12 month.** Dependent variable: ln(monthly gross income). Source: Waves 1 to 8 of the GSOEP. Number of persons: 224. Number of valid income measurements: 761. t-values in parenthesis .

Characteristic	Pooled Regression	Feasible Generalized Least Squares	ML-Estimate
Constant	7.241	7.261	7.264
	(59.5)	(50.9)	(51.8)
After 1987	0.051	0.047	0.047
	(1.9	(2.2)	(2.2)
Schooling	0.044	0.031	0.031
	(4.2)	(2.5)	(2.6)
Experience	0.009	0.020	0.020
	(1.1)	(2.4)	(2.4)
Exp. * Exp.	-0.6×10^{-4}	-3.6×10^{-4}	-3.6×10^{-4}
	(-0.3)	(-1.6)	(-1.6)
Tenure	0.002	0.006	0.006
	(0.8)	(2.1)	(2.1)
After LTU	0.178	0.132	0.130
	(1.2)	(0.9)	(0.9)
After LTU*Schooling	-0.017	-0 007	-0.007
	(-1.4)	(-0.6)	(-0.6)
After LTU*Experience	0.005	-0.004	-0.004
	(0.6)	(-0.5)	(-0.5)
After LTU*(Exp. * Exp.)	-2.6×10^{-4}	0.3×10^{-4}	0.3×10^{-4}
	(-1.1)	(0.1)	(0.1)
R^2	0.08	-	-
σ_ϵ^2	0.0911	-	-
σ_μ^2	-	0.0528	0.0495
σ_ν^2	-	0.0405	0.0412

Table 6: **Comparison of the results of some tests for non–ignorable panel attrition. ATTRIT**: Comparison of model estimates on persons without panel attrition and with panel attrition. **ADDVAR**: Addition of a variable that is correlated with panel attrition. t–values in parenthesis.

	ATTRIT		**ADDVAR**		
Charac–teristic	without attrition	with attrition	$\prod_t S_t$	S_{t-1}	W_1
Constant	7.277	7.340	7.271	7.238	7.31
	(45.2)	(17.1)	(50.1)	(47.5)	(49.1)
After 1987	0.051	0.041	0.046	0.046	0.043
	(2.1)	(1.0)	(2.2)	(2.2)	(2.1)
Schooling	0.031	0.021	0.031	0.031	0.032
	(2.4)	(0.5)	(2.5)	(2.5)	(2.5)
Experience	0.021	0.025	0.020	0.020	0.021
	(2.0)	(1.1)	(2.4)	(2.4)	(2.6)
Exp. * Exp.	-3.8×10^{-4}	-6.5×10^{-4}	-3.7×10^{-4}	-3.6×10^{-4}	-3.8×10^{-4}
	(-1.4)	(-0.8)	(-1.6)	(-1.6)	(-1.7)
Tenure	0.007	0.003	0.006	0.006	0.006
	(2.2)	(0.4)	(2.1)	(2.1)	(2.1)
After LTU	0.151	-0.127	0.133	0.128	0.136
	(0.9)	(-0.3)	(0.9)	(0.9)	(1.0)
After LTU*	-0.010	0.024	-0.007	-0.007	-0.007
Schooling	(-0.8)	(0.5)	(-0.6)	(-0.6)	(-0.6)
After LTU*	-0.005	-0.009	-0.005	-0.004	-0.005
Experience	(-0.5)	(-0.4)	(-0.5)	(-0.5)	(-0.6)
After LTU*	0.4×10^{-4}	3.0×10^{-4}	0.3×10^{-4}	0.2×10^{-4}	0.3×10^{-4}
Exp. * Exp.	(0.2)	(0.3)	(0.1)	(0.1)	(0.1)
$\prod_\tau S_\tau$	-	-	-0.015	-	-
			(-0.4)		
S_{t-1}	-	-	-	0.042	-
				(0.6)	
W_1	-	-	-	-	-0.076
					(-1.1)
N	567	194	761	761	761
	$T_{ATTRIT} = 3.32$ p–value=0.97				

Table 7: **Results of 100 replications of the artificial panel attrition.** $\hat{\beta}_S$: Estimate of β on the basis of the original sample. $\hat{\beta}_{S^*}$: Estimate of β on the basis of the artificially attrited sample S^*.

Characteristic	Mean of $\hat{\beta}_S - \hat{\beta}_{S^*}$	Std. Deviation of $\hat{\beta}_S - \hat{\beta}_{S^*}$
Constant	0.0437	0.0650
After 1987	-0.0090	0.0126
Schooling	-0.0025	0.0057
Experience	-0.0021	0.0049
Exp. * Exp.	0.4×10^{-4}	1.1×10^{-4}
Tenure	-0.0004	0.0018
After LTU	-0.0577	0.1173
After LTU*Schooling	0.0051	0.0113
After LTU*Experience	0.0019	0.0058
After LTU*(Exp. * Exp.)	-0.5×10^{-4}	1.4×10^{-4}

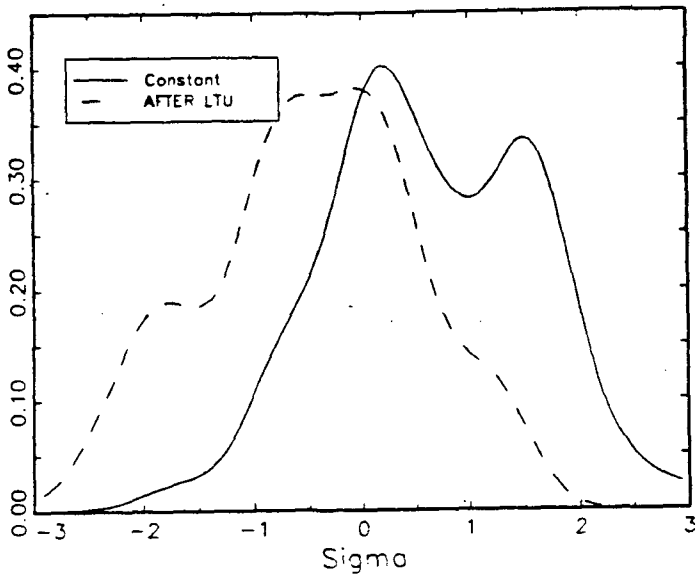
If we switch to the results of the simulated artificial panel attrition in table 7, again we will find no indication of a systematic bias. The highest bias is indicated for the constant and for the interaction of LTU with the constant. The distribution of $\hat{\beta}_S - \hat{\beta}_{S^*}$ for these two coefficients is displayed in figure 7 . There is an apparent bimodality for the simulated distribution of both coefficients and it is obvious that the minor modal values of each distribution correspond with each other[18]. Such a plausible correspondence indicates that the estimated trade–off of both coefficients is sensible to the panel attrition. The global test statistic $T_{SIM} = 2.78$ however does not indicate a systematic in the deviations: For 10 degrees of freedom one gets a p–value of p=0.98.

It is interesting to see that in this example there is a good concordance of the direction of the estimated bias in table 7 and the deviation of the estimate of the attriters form the corresponding estimate of the non–attriters. Both routines suggest, that the depreciation of school specific human capital by long–term unemployment is somewhat overstated by the model estimate.

Next we want to estimate the joint regression–selection model. First we have to fix the selection model. In the example the experience was computed from the actual time in the labor force, because this seems to be more accurate for the problem analyzed here. Since we do not know the employment status of the persons that drop out we have to remove from the selection equation all variables

[18]For the distribution of the other coefficients, which is not documented here, there is no such bimodality.

Figure 7: **The distribution of $\hat{\beta}_S - \hat{\beta}_{S^\bullet}$ for the constant and the coefficient LTU.** Kernel density estimate on the basis of 100 replications of the artificial attrition experiment.



that use the experience. Like in the first example we have to remove also the tenure. Note that in the first example the experience turned out to have an effect on the participation behavior. Hence the omission of this variable from the selection equation may result in biased estimates.

Like in the previous example we will use two selection models, where one model does exclude the variable "change of the interviewer" while the alternative model uses this variable. Under the assumption, that drop–out is ignorable we can get a consistent estimate of the coefficient of the "change of the interviewer" in the income equation. The corresponding ML–estimate results in $\hat{\beta}_{CHANGE} = 0.069$ with a t–value of 3.2. This coefficient is 10 times higher than in the previous example. This difference appears to be plausible, since in the case of long–term unemployment the percentage of persons that move in order to get a new job will be higher than in the case of all male employees.

Thus, unless the estimation of β_{CHANGE} is severely biased by non–ignorable drop–out, the change of the interviewer is positively correlated with the gross income. On the other hand it is known — and also confirmed by the estimation — that this variable has also a massive impact on the attrition behaviour. Thus one has to expect a non–ignorable effect of the panel attrition.

The estimation results displayed in table 8 are somewhat puzzling. The most striking results are the differences in the estimation of $\rho_{\nu\eta}$, the correlation of the

shock components. Both models I and III do not use the change variable in the selection equation. In model III there is a detection of a non–ignorable drop–out behavior while in model I there is no indication of such an effect. On the basis of our a–priori knowledge, we know that the implications of the model I estimate are wrong. In the case of models I and III the identification of the attrition effects depends solely on the distributional assumptions about the joint distribution of the error terms of the regression equation and the selection equation. If these untestable assumptions are not fulfilled one may end up in wrong conclusions like in the case of model I. The models II and IV that include the variable change in the selection equation result both in a significant value of $\rho_{\nu\eta}$ and are therefore in accordance with what we expect.

If we take a look at the differences in the estimated β–coefficients, we notice (comparing the estimate of model II with the separate estimation) that the major differences appear for the constant, the time after LTU and the depreciation of the education. The latter two differences are in the same direction as the preceding routines suggest, cf. tables 6 and 7. In the case of the constant the selectivity correction according to model II is just opposite to the corrections that may be derived from tables 6 and 7, which suggest an increase of the constant.

This may be a hint that there is some bias in the estimation of model II. As was mentioned in section 2, the ML–estimation of the joint regression–selection model may be biased if a Z–variable is omitted from the X–variables that has a non–zero regression coefficient. Apparently this case applies here[19].

[19]A bias may also result from the omission of the experience variables in the selection equation.

Table 8: **Results of separate and simultaneous ML-estimation of the joint regression–selection model. Income of male employees with an observerd period of long–term unemployment (LTU).** Source: Waves 1 to 8 of the GSOEP. Number of persons: 224. Number of valid income measurements: 761. t–values in parenthesis.

Charac-teristic	Separate estimation	Simultaneous estimation			
		model I	model II	model III	model IV
Regression equation					
Constant	7.263	7.263	7.234	7.257	7.245
	(51.8)	(51.8)	(50.3)	(51.1)	(50.1)
After 1987	0.047	0.047	0.055	0.040	0.038
	(2.2)	(2.2)	(2.5)	(1.8)	(1.8)
SCHOOLING	0.031	0.031	0.030	0.032	0.032
	(2.6)	(2.6)	(2.3)	(2.6)	(2.6)
EXPERIENCE	0.020	0.020	0.022	0.020	0.020
	(2.4)	(2.4)	(2.6)	(2.4)	(2.5)
EXP. * EXP.	-3.6×10^{-4}	-3.6×10^{-4}	-3.6×10^{-4}	-3.6×10^{-4}	-3.6×10^{-4}
	(-1.6)	(1.6)	(-1.8)	(1.6)	(1.6)
TENURE	0.006	0.006	0.006	0.006	0.006
	(2.1)	(2.1)	(2.0)	(2.0)	(2.0)
After LTU	0.130	0.131	0.074	0.094	0.092
	(0.9)	(0.9)	(0.5)	(0.7)	(0.6)
After LTU*	-0.007	-0.007	-0.003	-0.005	-0.005
SCHOOLING	(-0.6)	(-0.6)	(-0.2)	(-0.4)	(-0.4)
After LTU*	-0.004	-0.004	-0.006	-0.004	-0.005
EXPERIENCE	(-0.5)	(-0.5)	(-0.7)	-(0.5)	(-0.6)
After LTU*	0.3×10^{-4}	0.3×10^{-4}	-3.6×10^{-4}	0.2×10^{-4}	0.3×10^{-4}
EXP.*EXP.	(0.1)	(0.1)	(0.2)	(0.1)	(0.1)
σ_μ^2	0.050	0.050	0.052	0.050	0.051
	(7.8)	(7.8)	(7.7)	(7.8)	(7.8)
σ_ν^2	0.041	0.041	0.048	0.044	0.045
	(16.4)	(16.4)	(10.4)	(10.9)	(13.3)
$\rho_{\mu\xi}$	-	-0.002	0.113	0.309	0.377
	-	(-0.1)	(0.8)	(0.8)	(0.7)
$\rho_{\nu\eta}$	-	-0.009	0.683	0.461	0.608
	-	(-0.1)	(5.0)	(1.7)	(4.4)

(continued)

Table 8: **Continuation: Estimated coefficients of the selection equation**

Charac-teristic	Separate estimation			
	model I	model II	model III	model IV
Constant	1.623	1.655	1.591	1.928
	(2.6)	(2.6)	(1.1)	(1.3)
After 1987	0.100	0.053	-0.283	-0.397
	(0.9)	(0.5)	(-1.2)	(-1.5)
SCHOOLING	-0.047	-0.037	0.026	0.015
	(-0.8)	(-0.7)	(0.3)	(0.2)
After LTU	-1.294	-1.159	-0.881	-1.065
	(-1.9)	(-1.7)	(-0.8)	(-0.9)
After LTU*	0.081	0.073	0.042	0.066
SCHOOLING	(1.3)	(1.2)	(0.4)	(0.6)
CHANGE	-	-0.682	-	-0.869
of interv.		(-4.3)		(-1.5)
$\sigma_\eta^2/(\sigma_\xi^2+\sigma_\eta^2)$	0.816	0.807	0.927	0.914
	(2.6)	(2.7)	(0.7)	(0.7)
	Simultaneous estimation			
	model I	model II	model III	model IV
Constant	1.623	1.526	1.333	1.666
	(2.6)	(2.5)	(0.9)	(0.7)
After 1987	0.100	0.036	-0.303	-0.428
	(0.9)	(0.2)	(-1.0)	(-0.8)
SCHOOLING	-0.047	-0.028	0.051	0.039
	(-0.8)	(-0.5)	(0.5)	(0.4)
After LTU	-1.294	-1.070	-0.638	-0.813
	(-1.9)	(-1.6)	(-0.5)	(-0.6)
After LTU*	0.081	0.072	0.020	0.046
SCHOOLING	(1.3)	(1.2)	(0.2)	(0.4)
CHANGE	-	-0.727	-	-0.929
of interv.		(-4.3)		(-0.7)
$\sigma_\eta^2/(\sigma_\xi^2+\sigma_\eta^2)$	0.816	0.830	0.943	0.957
	(2.6)	(2.5)	(0.5)	(0.4)

7 Concluding remarks

If we resume the results of the empirical examples for the GSOEP, we conclude that there are small if any attrition effects on the estimation of income equations. If one wants to see a selection effect of non–ignoable size one should compare the estimated coefficients for schooling, experience and tenure between example 1 and example 2[20].

It turned out, that the detection of non–ignorable attrition effects is a cumbersome and hazardous task. All methods presented here have their pro's an con's. Two empirical examples suffice to demonstrate the pitfalls of all routines.

We recommend to distrust the use of a single method. If the results of all methods are concordant this is a good reason to accept the hypothesis of non–ignorable drop–out.

The tests for non–ignorable panel attrition that were discussed here and the estimation of a full joint regression–selection model differ in situations where an attrition bias is indicated. While the test routines give only hints, which coefficient is biased in what direction, the estimates of the full model promise to give consistent estimates under non–ignorable attrition, if the model is correctly specified. As shown in the examples the correctness of the specification is a crucial problem here.

The simulation of the panel attrition according to estimated probabilities of drop–out is the "inverse" of a weighted regression analysis which uses the reciprocal of the probabilities of drop–out as weights. Hence a regression analysis weighted in this way would be a natural canditate for a consistent estimator in the case of non–ignorabble drop–out. We do not advocate for a comparison of the unweighted and the weighted estimation as a test for non-ignorable drop out, since both estimates are dependent. So the estimated variances of the estimates are not valid to judge differences between these estimates.

The use of survey weights in cross–sectional regression analysis has been discussed for stratified samples by DuMouchel/Duncan (1983), Little (1991) and Nathan/Holt (1980). However it is still an open question under what kind of selection model such a weighted analysis turns out to be consistent or ML.

[20]Of course we did not intend to use the coefficients of example 2 for all male employees.

Appendix: Means and standard deviations of variables used in examples 1 and 2

Table 9: **Example 1: Mean and standard deviation of characteristics in the income equation of German male employees.** Source: Waves 1 to 8 of the GSOEP. Number of persons: 3523. Number of valid income observations: 15670.

Characteristic	Mean	Std. Deviation
ln(Income)	8.16	0.37
Schooling	11.67	2.46
Experience	22.35	11.38
Exp. * Exp.	628.91	543.81
Tenure	12.77	10.06
1985	0.13	0.34
1986	0.13	0.33
1987	0.13	0.33
1988	0.12	0.33
1989	0.12	0.33
1990	0.12	0.32
1991	0.11	0.32

Table 10: **Example 2: Mean and standard deviation of characteristics of male employees with an observed period of long-term unemployment (LTU) of at least 12 months.** Source: Waves 1 to 8 of the GSOEP. Number of persons: 224. Number of valid income measurements: 761.

Characteristic	Mean	Std. deviation
ln(Income)	7.85	0.31
Schooling	10.26	1.77
Experience	17.27	10.61
Exp. * Exp.	410.83	393.19
Tenure	4.77	5.17
After 1987	0.62	0.49
After LTU	0.68	0.47
After LTU * Schooling	6.84	4.93
After LTU * Experience	11.02	11.69
After LTU * (Exp. * Exp.)	258.08	374.71

Literature

Amemiya, Takeshi 1984: Tobit Models: A Survey. Journal of Econometrics, 24, 3–61.

Becketti, Sean; Gould, William; Lillard, Lee; Welch, Finis 1988: The Panel Study of Income Dynamics After 14 Years: An Evaluation. Journal of Labor Economics, 6, 472–492.

Buttler, James; Moffitt, Robert 1982: A Computationally Efficient Quadrature Procedure for One-Factor Multinomial Probit Models. Econometrica, 50, 761–764.

Dawid, A. P. 1979: Conditional Independence in Statistical Theory. Journal of the Royal Statistical Society, B, 41, 1–31.

DuMouchel, William; Duncan, Greg 1983: Using Sample Survey Weights in Multiple Regression Analysis of Stratified Samples. Journal of the American Statistical Association, 78, 535–543.

Gottschalk, Peter; Moffitt, Robert 1992: A Study of Sample Attrition in the Michigan Panel Study of Income Dynamics. Manuscript, Institute for Social Research, University of Michigan, Ann Arbor MI.

Hanefeld, Ute 1984: The German Socio–Economic Panel. In: American Statistical Association (Ed.): 1984 Proceedings of the Social Statistics Section, Washington D.C., 117–124.

Heckman, James 1979: Sample Selection Bias as a Specification Error, Econometrica, 47, 153–161.

Hsiao, Cheng 1986: Analysis of Panel Data, Cambridge University Press, Cambridge.

Lee, L. 1982: Some Approaches to the Correction of Selectivity Bias. Review of Economic Studies, 49, 352–372.

Lee, L.; Maddala, G. 1985: The Common Structure of Tests for Selectivity Bias, Serial Correlation, Heteroskedasticity and Non–Normality in the Tobit Model. International Economic Review, 26, 1–20.

Little, Roderick 1982: Models for Nonresponse in Sample Surveys. Journal of the American Statistical Association, 77, 237–250.

Little, Roderick 1991: Inference with Survey Weights. Journal of Official Statistics, 7, 405–424.

Little, Roderick; Rubin, Donald 1987: Statistical Analysis with Missing Data. Wiley, New York.

Löwenbein, Oded; Rendtel, Ulrich 1991: Selektivität und Panelanalyse. In: Rendtel, Ulrich; Wagner, Gert (Hrsg.): Lebenslagen im Wandel – Zur Einkommensdynamik in Deutschland seit 1984, Campus, Frankfurt/M. – New York, 156–187.

Nathan, G. ;Holt, D. 1980: The Effect of Survey Design on Regression Analysis. Journal of the Royal Statistical Society, B, 42, 377–386.

Rendtel, Ulrich 1990: Teilnahmeentscheidung in Panelstudien: Zwischen Beeinflussung, Vertrauen und sozialer Selektion. Über die Entwicklung der Antwortbereitschaft im Sozio–ökonomischen Panel. Kölner Zeitschrift für Soziologie und Sozialpsychologie, 42, 280–299.

Rendtel, Ulrich 1992: On the Choice of a Selection–Model when Estimating Regession Models with Selectivity. DIW–Discussion Paper, 53, Berlin.

Rendtel, Ulrich 1993: Über die Repräsentativität von Panelstichproben — Eine Analyse der feldbedingten Ausfälle im Sozio–oekonomischen Panel (SOEP). DIW–Discussion Paper, 70, Berlin.

Rendtel, Ulrich 1994: Die Analyse von Paneldaten unter Berücksichtigung von Panelmortalität. Mimeo, Berlin.

Ridder, Gert 1990: Attrition in Multi–Wave Panel Data. In: Hartog, G.; Ridder, G.; Theeuwes, J. (eds): Panel Data and Labor Market Studies, North–Holland, Amsterdam, 45–68.

Stroud, A.; Secrest, D. 1966: Gaussian Quadrature Formulas. Prentice–Hall, Englewood Cliffs.

Verbeek, Marno 1991: The Design of Panel Surveys and the Treatment of Missing Observations. Dissertation at the University of Brabant, Tilburg, Netherlands.

Wagner, Gert; Burkhauser, Richard; Behringer, Friederike 1993: The English Language Public Use File of the German Socio–Economic Panel. Journal of Human Resources, 28, 429–433.

Wagner, Gert; Schupp, Jürgen; Rendtel, Ulrich 1991: The Socio–economic Panel for Germany, Methods of Production and Management of Longitudinal Data. DIW–Discussion Paper, 31a, Berlin.

Wagner, Gert; Schupp, Jürgen; Rendtel, Ulrich 1993: Das Sozio–oekonomische Panel (SOEP) — Methoden der Datenproduktion und –aufbereitung im Längsschnitt. In: Hauser, R.; Ott, N.; Wagner, G. (Hrsg.): Mikroanalytische Grundlagen der Gesellschaftspolitik — Band 2: Erhebungsverfahren, Analysemethoden und Mikrosimulation, Akademie Verlag, Berlin, 70–112.

Witte, James 1989: The Potential for Comparative Panel Research Using Data from the Survey of Income and Program Participation and the German Socio–economic Panel. SIPP Working Paper Series No. 8926, Washington D.C., US Bureau of the Census.