Zeitschrift für Wirtschafts- u. Sozialwissenschaften (ZWS) 116 (1996), S. 339 - 358
Duncker & Humblot, Berlin

Combining Panel Data and Macro Information for the Estimation of a Panel Probit Model*

By François Laisney and Michael Lechner**

1. Introduction

When studying particular subgroups of a population the econometrician typically has few observations at hand. We found ourselves in such a situation when attempting to estimate a participation model for lone mothers in West Germany on the basis of the only widely available data set containing relevant information, the Socio-Economic Panel (SOEP).[1] Despite many advantages, the SOEP is not ideal for studying such a special group, because the number of lone mothers in any single wave varies between 157 (1985) and 85 (1990). The obvious alternative would be to work with a much larger sample from the outset, such as the Microcensus, a 1% representative sample which serves as basis for the German Labour Force Survey. Unfortunately, neither the latter nor the original Microcensus were released by the Federal Office of Statistics (Statistisches Bundesamt) at the time when we started this study. If either had been, we could have tried combining the informations it contains with those contained in the SOEP, as proposed for instance by Arellano and Meghir (1992). On the other hand, the Federal Statistical Office does publish information on the basis of the Microcensus, and in our situation it appears vital to take advantage of any additional relevant information we can obtain from that source.

---

* Verantwortlicher Herausgeber / editor in charge: J. W.

** Support from the Deutsche Forschungsgemeinschaft (DFG) and from the European Commission under SPES grant CT910051 is gratefully acknowledged. Furthermore, we thank the DIW Berlin for providing us with the data of the Socio-economic panel (SOEP), and Susanne Salomon and Stefan Vögele for able research assistance. We benefited from discussions with Matthias Staat and Gerhard Wagenhals, our coauthors on the companion paper Laisney et al. (1993a), with Guido Imbens and Johannes Velling, and from the comments of participants in the conference of the Verein für Socialpolitik in Jena, 1994, especially Heinz Galler, Olaf Hübler and Gerd Ronning. Finally, the perceptive comments of two anonymous referees helped us improve significantly upon the previous version.

1 We use the wording "lone parents" rather than "single parents" because the former could also be widowed, divorced or separated.

The purpose of this paper is to document the relative benefits derived from using the panel structure of the data and from including macro information in the form of extra moments, as proposed for cross-section analysis by Imbens and Lancaster (1994). In our example these extra moments will be population proportions of lone mothers working, given some characteristics like age or number of children in different age groups. The compatibility between the micro and the macro information can be tested before estimation is carried out, and this step considerably enriches the data analysis that one should anyway perform before engaging in estimation. We see this as an important by-product of this approach, but will not emphasize it further here, as a companion paper, Laisney et al., 1993a, thoroughly documents the data analysis. We will instead focus on the econometric aspects of the study. As far as we know, the extension to panel data of the idea of Imbens and Lancaster is new, although straightforward.

Section 2 describes the main features of the approach combining Generalized Method of Moments (GMM) estimation of limited dependent variable models on panel data and the use of additional moments retracing macro information. We briefly review the data in Section 3 in order to allow the discussion of further aspects of the estimation strategy in Section 4. The results are presented in Section 5. An Appendix explains how we have taken account of the fact that the SOEP is a stratified sample.

2. Main Econometric Aspects

Our approach combines ideas concerning the GMM estimation of limited dependent variable models on panel data, outlined in Avery et al. (1983), with an approach combining micro and macro data sources in order to achieve better efficiency, suggested by Imbens and Lancaster (1994).

The type of model we consider here can be written as

$$y_{it} = \mathbf{1}[x_{it}\beta + \alpha_t + u_{it} > 0]\,, \qquad i = 1, \ldots, n, \quad t = 1, \ldots, T, \tag{1}$$

where y_{it} is a dichotomous variable, $\mathbf{1}[.]$ denotes the indicator function of the event in the bracket, x_{it} is a vector of explanatory variables assumed independent of the error term u_{it}, and α_t is a period-specific intercept. What this model says in terms of our example is that the probability for woman i to participate in the labour market at time t given her own characteristics at that time, x_{it}, and the term α_t subsuming the common residual effect of business-cycle related variables on participation behaviour, is

$$P_t(y_{it} = 1|x_{it}) = P_t(u_{it} > -x_{it}\beta - \alpha_t)\,. \tag{2}$$

The subscript t of P_t indicates that the distribution of u_{it} is allowed to depend on t. Indeed, the model specification is completed by the choice of a family of distributions for the vector $u_i = (u_{i1}, \ldots, u_{iT})'$. The choice here will be the family of centred multivariate normal distributions $u_i \sim N(0, \Sigma)$. Using the panel structure in this context means taking account of the fact that observations of the same individual over time may well be correlated, and thus that the $T \times T$ matrix Σ needs not be diagonal. However, we do restrict Σ to be constant across individuals.

Note that we have to use an "unbalanced" panel here, since otherwise (i.e. if we discarded individuals for whom we do not have an observation in each single time period, in order to obtain a so-called balanced panel) we would be left with too few observations to conduct any reasonable analysis: with our data, using a balanced panel means using observations on 20 individuals only.[2]

Efficiency is important given our small sample size. One way to try and attain it is to use full information maximum likelihood. Yet this is difficult for T larger than 3 ($T = 7$ is the time dimension of our example), unless it is combined with restrictive assumptions on the covariance structure of the error terms. The reason is that in the absence of such restrictions, the computation of the likelihood function requires the evaluation of integrals over several dimensions, and thus the use of simulation methods.

On the other hand, the problem with assumptions on the source of correlation in the error terms, like for instance in the popular random effects probit model, or its extension to random effects plus an autoregressive process in the error term, is that these will need to be tested; unless one is lucky some assumptions will be rejected and others not, leading to unpalatable pre-testing problems in the final estimator.

Therefore, a more appealing approach, in our opinion, is to use a GMM estimator necessitating no such assumption, with a first set of moments given by the scores of the cross-section likelihood functions, and with an additional set of moments in order to increase efficiency. Other extensions of the idea of Imbens and Lancaster are conceivable, like for instance improving the random effects panel probit estimator, but we have said above why we do not favour the random effect approach.

We now spell out the reasons that lead us to take the pooled probit estimator as a bench-mark instead. We consider the vector of random variables (y, x) with $y = (y_1, \ldots, y_t, \ldots, y_T)'$ and $x = (x_1, \ldots, x_t, \ldots, x_T)'$, and the vectors

[2] A word of caution may avoid a possible confusion: we use the terms "balanced" and "unbalanced" as in the econometric literature on panel data, not as in the statistical literature on experiment design.

$(y_i, x_i), i = 1, \ldots, n$, of realisations from n independent draws from their joint distribution. The conditional distribution of y given x is characterised by a parameter vector $\varphi = (\beta', \alpha', (\text{vec}\Sigma)')' = (\theta', (\text{vec}\Sigma)')'$, where θ denotes the parameter of interest, with true value θ_0, and the vector vecΣ contains the $T(T+1)/2$ free elements of Σ. Asymptotic arguments here and in all references quoted concern the case where the time dimension T is fixed but we can continue indefinitely to sample individuals, i.e. let n increase without limit.

We shall not attempt to add to the wealth of excellent and accessible expositions of the GMM estimation principle available in the recent literature (e.g. Newey and McFadden, 1994), but it will be useful to introduce some notation for later reference. Given a moment restriction $\mathrm{E}h(y, x; \theta_0) = 0$ based on the expectation of a vector-valued function h of the observations and of the parameter, and a possibly data-dependent weighting matrix C, the GMM estimator $\hat{\theta}_n$ associated with C is defined as:

$$(3) \qquad \hat{\theta}_n = \arg\min_{\theta \in \Theta} \left[\frac{1}{n} \sum_{i=1}^{n} h(y_i, x_i; \theta) \right]' C \left[\frac{1}{n} \sum_{i=1}^{n} h(y_i, x_i; \theta) \right] .$$

The *optimal* choice of the weighting matrix C is $\{\mathrm{E}h(y, x; \theta_0) h(y, x; \theta_0)']\}^{-1}$, or a sequence converging to that limit. Under regularity conditions, the asymptotic distribution of the optimal GMM estimator for the case of independent identically distributed observations is:

$$(4) \qquad \sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, V), \quad \text{with}$$

$$(5) \qquad V^{-1} = \mathrm{E} \frac{\partial h'}{\partial \theta}(y, x; \theta_0) \Big\{ \mathrm{E}\big[h(y, x; \theta_0) h(y, x; \theta_0)' \big] \Big\}^{-1} \mathrm{E} \frac{\partial h}{\partial \theta'}(y, x; \theta_0) .$$

The covariance matrix can be consistently estimated by replacing expectations with sample averages, and θ_0 with a consistent estimate.[3]

Before proceeding, we assume that the observability rule for the unbalanced panel, is *ignorable*, i.e. $\mathrm{E}\big[h_t(y_t, x_t; \theta_0) | r_t = 1\big] = \mathrm{E}\big[h_t(y_t, x_t; \theta_0) | r_t = 0\big] = 0$, where the dichotomous variable r_t equals one if the individual is observed in period t and zero otherwise.[4] This assumption and the use of modified moments of the form $r_t h_t(y_t, x_t; \theta_0)$ will allow us to estimate all the necessary moments from the complete population without the need of further

[3] For a complete list of assumptions and proofs of these properties, the reader is referred to Hansen (1982). The exact formulas used in our case of stratified sampling are detailed in the Appendix.

[4] This corresponds to the definition given in Verbeek and Nijman (1992): it is less restrictive than the definition given in Rubin (1976) or Little and Rubin (1987) that

corrections. For ease of exposition and for the sake of brevity of notation, we shall assume in this section that $r_t = 1$ for each period, i.e. that we have a balanced panel. However, in estimation we do use an unbalanced panel.

Let us now partition the vector of moments according to $h = (h_1^{1'}, \ldots, h_T^{1'}, h_1^{2'}, \ldots, h_T^{2'})'$ and let $h_t = (h_t^{1'}, h_t^{2'})'$. This partition will correspond to the distinction between information from the panel data set used, and information obtained from macro data. The latter gives rise to additional individual moments in the following way: suppose that we know the expectation g_t^* of some function $\tilde{g}(y_{it}, x_{it})$ with respect to the joint distribution of (y_{it}, x_{it}) over the population at time t. Using the law of iterated expectations, we have

$$g_t^* = \mathrm{E}\Big[\mathrm{E}\{\tilde{g}(y_{it}, x_{it})|x_{it}\}\Big] =: \mathrm{E}g(x_{it}; \theta_0) \,, \tag{6}$$

so that the macro information g_t^* yields the individual moment $h_t^2(x_{it}; \theta) = g(x_{it}; \theta) - g_t^*$. The precise nature of this type of information, alluded to in the introduction, is discussed in Subsection 3.2 and in Section 4. Here it will suffice to accept that macro data yield extra moments that can be used in the GMM procedure.

The type of panel data models considered here allows for an arbitrary correlation structure Σ over time, but requires independence between the error term and the regressors. Under these assumptions, each separate cross section estimation based on maximisation of the likelihood for y_t given x_t yields consistent parameter estimates of the components of θ_0 which can be identified from that single cross section, namely β/σ_t and α_t/σ_t, where σ_t^2 denotes the variance of u_t. This suggests using the corresponding scores as elements of h_t^1.

It has first been noted by Avery et al. (1983) that furthermore imposing – in the absence of macro information – that $h(x, y; \theta) = \sum_{t=1}^{T} h_t^1(y_t, x_t; \theta)$, and that the weighting matrix C is the identity matrix, yields the pooled estimator – i.e. the pseudo-maximum likelihood estimator obtained by ignoring the panel structure and treating all observations as if they were independent realisations. Hence the pooled estimator is a GMM estimator and therefore it is consistent, whatever the cross-period correlations between observations of the same individual may be.[5] However, when computing its covariance matrix, these correlations are taken into account by use of the

prevails in the statistical literature. It still remains perforce an untested assumption: to assess its validity would require much more information than that available here.

[5] Strictly speaking, this argument is only valid under homoscedasticity over time, i.e. if the diagonal terms of Σ are identical. But the generalization to the heteroscedastic case is straightforward.

appropriate GMM formula in place of the usual MLE formula. We will use this pooled estimator as the bench-mark from which to measure efficiency gains obtained either by drawing on macro information, or by using the panel structure optimally, or both.

Thus the panel estimator we shall consider uses the period scores $h_t^1(y_t, x_t; \theta)$ individually, rather than their sum over time. Since it uses more information than the pooling estimator does, it is thereby asymptotically more efficient.[6] However, these efficiency gains over the pooled estimator come at the cost of expanding the moment space: this may become a problem when estimating the optimal weighting matrix and the covariance matrix of the coefficient estimates.

The elements of h^2 take outside information into account, as suggested by Imbens and Lancaster (1994). A simple way to interpret the efficiency gains that can be expected from the use of such outside information – as indeed the above mentioned gains from using the panel structure – is to realize that extra moments can be viewed as parameter restrictions: the fact that valid parameter restrictions result in more precise estimation is well known. The interesting point here is that since they are drawn from population magnitudes, these restrictions will indeed be valid, provided the sample design is correctly taken into account. The details of the implementation (choice of moments, etc.) will be discussed below. Further technicalities arising from the fact that the SOEP is a stratified sample are discussed in the Appendix.

A simple specification test available for all these estimators, including the pooled probit estimator, is provided by the fact that under the null hypothesis of a correct specification, i.e. that all moments used in the estimation are valid (expectation equal to 0), the *distance* statistic

$$\frac{1}{n}\sum_{i=1}^{n} h(y_i, x_i; \hat{\theta}_n)' \left[\sum_{i=1}^{n} h(y_i, x_i; \hat{\theta}_n) h(y_i, x_i; \hat{\theta}_n)'\right]^{-1} \sum_{i=1}^{n} h(y_i, x_i; \hat{\theta}_n) \tag{7}$$

converges to a χ^2 distribution with a number of degrees of freedom (df) equal to the rank of the covariance matrix of the moments minus the number of unrestricted parameters.[7] Sargan Tests of overidentifying restrictions can also be performed, using the difference of the relevant distance statistics: this is asymptotically χ^2 with df equal to the difference of the df of the distances considered.

---

[6] For a discussion of the relative efficiency of the various estimators that can be constructed along these lines, see Lechner and Breitung (1996) and Bertschek and Lechner (1995).

[7] Note that this test statistic uses the individual moments for the pooled estimators also, not the summed moments used in estimation.

Table 1

Means of variables used in the estimation

Year	1984	1985	1986	1987	1988	1989	1990
dependent variable participation	0.55	0.58	0.62	0.65	0.61	0.63	0.69
schooling							
Realschule[1]	0.19	0.20	0.19	0.16	0.12	0.11	0.11
Abitur	0.05	0.06	0.07	0.09	0.11	0.14	0.19
number of children							
younger than 4	0.13	0.11	0.03	0.02	0.07	0.12	0.08
4-6 years old	0.25	0.15	0.19	0.14	0.10	0.07	0.09
younger than 7	0.37	0.25	0.22	0.16	0.15	0.19	0.16
7-14 years old	0.57	0.61	0.64	0.66	0.67	0.70	0.78
15-17 years old	0.35	0.32	0.34	0.31	0.35	0.32	0.35
density of child care relevant child dummy[2]* 10*							
0-3 years	0.03	0.04	0.00	0.00	0.02	0.03	0.01
4-6 years	1.60	0.99	1.22	0.98	0.64	0.45	0.63
7-10 years	0.05	0.06	0.07	0.07	0.07	0.10	0.09
age							
younger than 32	0.23	0.21	0.21	0.18	0.14	0.14	0.15
33-40	0.28	0.25	0.29	0.32	0.36	0.35	0.36
41-48	0.25	0.27	0.29	0.27	0.27	0.26	0.28
marital status							
single	0.11	0.11	0.12	0.13	0.08	0.11	0.16
divorced	0.45	0.47	0.43	0.45	0.45	0.47	0.58
widow	0.27	0.22	0.25	0.24	0.27	0.29	0.16
not German	0.23	0.19	0.19	0.19	0.22	0.23	0.19
regions[3]							
northern	0.17	0.17	0.18	0.15	0.14	0.18	0.14
southern	0.30	0.26	0.26	0.33	0.33	0.30	0.33
urbanisation							
< 20'000 inhab.	0.27	0.29	0.36	0.36	0.32	0.34	0.27
> 500'000 inhab.	0.68	0.62	0.52	0.54	0.58	0.55	0.61
net unearned income[4]							
Y^{NP} (in DM 10,000)	1.69	1.78	1.80	1.82	1.82	1.88	1.87
observations	150	157	129	119	110	102	85

Notes:

1. Realschule corresponds to successful completion of 10 years of schooling, Abitur to 13 years.

2. Density of child care is defined as the ratio of the number of day-care places available for a category of children in a Land to the number of children of that category in the Land. This varies both across Federal States (Länder) and across years. The variable used in estimation is 10 times the product of this variable with the relevant child dummy, and this varies across households. For a woman with children in all three age categories, the three variables are non zero. That the means are close to zero comes from the combined facts that (i) means are taken over zero and non zero observations and (ii) the variables are small even for non zero observations. Yet the size of the estimated coefficients (Table 2) shows that the scaling is appropriate.

3. The reference category consists of the Federal States of Saarland, Rhineland-Palatinate and Hesse.

4. Net unearned income is household income after transfers and taxes when the woman does not participate in the labour market.

3. Data

3.1 The data set

The sample we use is an unbalanced panel drawn from the first seven waves of the socio economic panel of West Germany (SOEP, see Hanefeld, 1987, or Wagner et al., 1993, for an extensive description of this data source). Our selection of observations of lone mothers is based on households classified in waves 1984 to 1990 as single parent households, with the lone mother being the head of the household. She is younger than 59, her oldest child living in the household is younger than 27 and her youngest child is younger than 21 years. After deleting cases with missing information, there are 296 individual observations left.

The means of the labour force states and of the explanatory variables are given in Table 1. Since our sample size does not allow the estimation of complicated models, we estimate only a simple reduced-form participation probit model, i.e. a model attempting to explain the probability of participation of a woman in the labour market by determinants of her preference for leisure *and* determinants of her potential wage rate. The rationale for inclusion of the different variables listed in Table 1 is as follows: the time dummies are supposed to subsume business-cycle effects; schooling, age, region, urbanisation grade and the "not German" dummy can be viewed as determinants of both preferences and potential wage; the marital status variable should control for some of the heterogeneity in preferences; children have an influence on preferences, through the increased needs associated with an increased consumption, and through a direct impact on a woman's time budget (this time budget effect should be refined by taking account of the variability of child care availability both geographically and over time); the effect of the net unearned income variable is purely a budget constraint effect: if leisure is a normal good, one expects a negative coefficient for that variable in a participation equation. More detailed information on the selection of the sample and a full set of descriptive statistics can be found in Laisney et al. (1993a).

3.2 Additional information

The outside information used here is based on a very large data set, the *Microcensus* (the published data used is taken from the "Fachserien" of the Statistisches Bundesamt). This is a 1% representative sample of the total population of the Federal Republic of Germany: with over 600,000 individuals, it is more than 40 times larger than the SOEP. This sheer difference in sizes allows us to ignore the sampling error in this information and to treat the latter as the knowledge of population parameters.

This information consists in participation rates of lone mothers by age of the youngest child (younger than 15 or 18) and marital status (single, separated, divorced, widowed). These are available for 1985 to 1990. Decreasing the age for the youngest child resulted in all too sparse cells (given marital status). In order to avoid an excessive number of moments, those corresponding to the marital categories in each period have been summed up, which results in only 12 additional moments instead of 48. An additional moment for 1990 was constructed as the sum over exclusive age groups (25-34, 35-44, 45-55) of moments based on participation rates by age group for the divorced with children below 18. Only the divorced have been used, because the other cells do not contain enough observations. Furthermore, four coarser groupings have also been used, resulting in 24 additional moments. These are participation rates for lone mothers with (i) children younger than 6 years, (ii) age between 25 and 34, (iii) age between 35 and 44, and (iv) age between 45-54. Altogether we thus have 37 additional moments (12+1+24).

We have tested the compatibility between macro and micro information for all individual components of the moments above, and thus performed 69 simple binomial tests of equality between each sample frequency $\hat{p}_{jt}$ and its population counterpart p_{jt} (macro data). For these tests we obtained a marginal rejection at the 5% level for only two moments. For the 37 "aggregated" moments used in estimation no single rejection appeared.

Note that we do not use moments concerning the explanatory variables alone, although these could in principle increase efficiency through their correlation with the other moments. There are two reasons for this. Firstly, as mentioned above, we are concerned with the numerical problems that arise when the number of moments becomes important. Secondly, Imbens and Lancaster (1994, 662) have shown that, given a partition of the observations space based on explanatory variables, using marginal probabilities on top of the corresponding conditional probabilities of participation given this partition is not informative, which justifies our disregard of such marginal probabilities.[8]

[8] Strictly speaking, this statement needs to be qualified in two ways: firstly, we have not always used all conditional probabilities corresponding to a partition; secondly, the proof is valid for a cross-section but may not extend to panel data, since more correlations are available there. In the same vein, whereas h_t^1 and h_t^2 are asymptotically uncorrelated, because the former has a zero conditional expectation given x_t, as a score, while the latter is a function of x_t, the same does not hold for h_t^1 and h_t^2 if $t \neq s$.

4. Further Considerations on Estimation

In Section 2 we have seen that the marginal likelihood function for a single period is the simple cross-section probit likelihood function, and the expectation of the scores of these marginal likelihood functions will be zero for the true parameter values in each period.[9] These individual scores will be the elements of h^1 and their expression is:

(8) $$h^1_{it}(y_{it}, x_{it}; \theta_t) = \left(r_{it} \frac{y_{it} - \Phi_{it}}{\Phi_{it}(1 - \Phi_{it})} \phi_{it} \right) \begin{bmatrix} D_t \\ x'_{it} \end{bmatrix} ,$$

where Φ_{it} and ϕ_{it} denote the cumulative density and the density of the standard normal evaluated in $(x_{it}\beta + \alpha_t)/\sigma_t$, and D_t denotes the relevant time dummy.

Furthermore, recall that we have denoted $\hat{p}_{jt}$ the population participation frequencies in each particular socio-demographic group $j = 1, \ldots, J$. Let j_{it} be one if individual i belongs to group j in period t , and zero otherwise. The following expressions are used as elements of h^2:

(9) $$h^{2j}_{it}(x_{it}; \theta_t) = r_{it} j_{it} \{p_{jt} - \Phi_{it}\} ,$$

whereby the homogeneity between the two sides of the equations as regards arguments comes from the fact that both x_{it} and θ_t appear in Φ_{it} and j_{it} is a function of x_{it}.[10] Identification in the probit model is only up to scale, so that some normalization is necessary for a meaningful comparison of coefficients. For the panel probit model (1) it turns out that one possible restriction would be to set one of the error term variances σ_t^2 to one. A further restriction of the specification would then be to set all such variances to one. However, we are mainly interested in the comparison of different estimators of coefficients β and α_t in (1); since the quotient of any two such coefficients is identified provided that the denominator is not zero, we have chosen the equivalent normalization of setting the intercept to 1 for the first period (in the pooled probit estimation, the first intercept was indeed significant) and restricted all variances to be equal. The other intercepts are left free and the correlations of the error terms over time are unrestricted.

---

[9] "Marginal" refers here to integrating out the dependent variables y_{is} for $s \neq t$ in the conditional density of y_i given x_i: this yields the "marginal" conditional density of y_{it} given x_i, and given the assumptions made for model (1) this is the density of y_{it} given x_{it}.

[10] To make all these dependencies explicit in the equation would result in a very clumsy expression.

5. Results

Table 2 shows the estimated coefficients and t-values for the participation model, for four GMM estimators differing in the choice of moments and in whether or not macro information was used.[11] The first pair of columns corresponds to the pooled probit estimator, at least as regards the estimated coefficients: the estimated variance of this estimator used here differs from the output of the maximum likelihood pooled probit estimator in that it does take account of the correlations implied by the panel structure of the data (see Avery et al., 1983). The second pair of columns is relative to a GMM estimator based on the moments of the pooled probit estimator supplemented with macro moments. The weighting matrix is chosen to be the identity matrix, as for the pooled probit. The panel GMM estimators of the last two pairs of columns are computed on the basis of a consistent estimate of the optimal weighting matrix, i.e. the inverse of $p\lim(1/n)\Sigma h(y_i, x_i; \theta_0)$ $h(y_i, x_i : \theta_0)'$ which is obtained from the corresponding pooled probit estimate (i.e. with or without macro information).[12]

Starting with the comparison between the pooled probit estimates with and without use of macro information (the first two pairs of columns), we see that the number of "well-determined" coefficients (p-value < 1%), leaving the intercepts aside, moves from 1 to 5, and that this is *not* due to an increase in the estimates, but to a decrease in their estimated variance. Considering only these coefficients, we see that while the coefficients of the two variables number of children younger than 4 and widow dummy have much smaller absolute values in the estimation using macro information, the estimates for the urbanisation grade variables and for the net unearned income hardly change. The sign of the net unearned income is in accordance with the assumption that leisure is a normal good, and the other signs do not contradict the intuition.

Moreover, both specifications appear to be rejected, as the distance statistics and the corresponding tests show. The fact that the use of macro information allows a clearer rejection results from the associated efficiency gains. Since the restrictions embodied in the macro information are valid, the Sargan test of overidentifying restrictions must also be interpreted as a specification test in this context. It does also reject without ambiguity. Still,

---

[11] Admittedly, there would have been scope also for comparison of different types of such information.

[12] There is an inescapable arbitrariness in this type of choice: even reporting results for all possible choices of the estimator of the optimal weighting matrix based on reported coefficient estimates would not exhaust the possibilities, since we could have considered a wealth of other estimators. Thus, we do not mean to imply that the choice of the weighting matrix made here have any optimality property other than asymptotic.

Table 2

Estimates for the reduced-form participation model

GMM estimator	pooled		pooled		panel		panel	
Macro data used	no		yes		no		yes	
Variable	coef	*t-val*	coef	*t-val*	coef	*t-val*	coef	*t-val*
time effects[1]								
1985	1.02	*0.3*	1.14	*3.2*	0.97	*0.4*	1.13	*6.1*
1986	1.10	*0.8*	1.14	*2.6*	1.05	*0.7*	1.13	*5.6*
1987	1.10	*0.8*	1.15	*2.9*	1.09	*1.2*	1.14	*5.7*
1988	1.05	*0.4*	1.13	*2.2*	0.99	*0.1*	1.12	*4.9*
1989	1.17	*1.1*	1.19	*2.9*	1.12	*1.4*	1.17	*6.2*
1990	1.23	*1.3*	1.20	*3.0*	1.17	*1.9*	1.18	*6.3*
schooling								
Realschule	0.43	*1.6*	0.12	*1.6*	**0.58**	*3.0*	**0.16**	*3.4*
Abitur	0.42	*1.3*	0.08	*0.7*	**0.83**	*2.6*	0.09	*1.4*
number of children								
younger than 4	–1.23	*–2.3*	**–0.47**	*–2.9*	**–1.50**	*–4.3*	**–0.48**	*–8.9*
4-6 years old	–0.17	*–0.3*	0.65	*1.6*	–0.49	*–0.9*	**0.61**	*3.9*
7-14 years old	–0.10	*–0.9*	–0.00	*–0.0*	–0.11	*–1.8*	–0.02	*–0.9*
15-17 years old	0.15	*1.5*	0.09	*2.3*	**0.14**	*2.7*	**0.09**	*4.8*
density of child care relevant child dummy* 10*								
0-3 years	0.72	*1.7*	0.31	*2.0*	**0.93**	*4.0*	**0.34**	*9.0*
4-6 years	–0.05	*–0.5*	–0.12	*–1.8*	0.01	*0.1*	**–0.11**	*–4.7*
age								
younger than 32	–0.13	*–0.7*	–0.16	*–2.0*	–0.16	*–1.4*	**–0.15**	*–3.9*
33-40	0.17	*0.9*	0.00	*0.1*	0.24	*2.0*	0.02	*0.6*
41-48	0.08	*0.5*	0.02	*0.3*	0.11	*1.1*	0.03	*1.1*
marital status								
single	0.03	*0.1*	–0.08	*–0.9*	0.09	*0.6*	–0.07	*–1.4*
divorced	–0.02	*–0.1*	–0.08	*–1.4*	0.08	*0.8*	**–0.08**	*–2.6*
widow	–0.67	*–2.3*	**–0.30**	*–3.3*	**–0.65**	*–3.7*	**–0.29**	*–6.0*
not German	0.42	*1.7*	0.06	*0.9*	**0.55**	*3.1*	0.04	*1.2*
regions								
northern	0.04	*0.2*	–0.07	*–1.0*	–0.02	*–0.2*	–0.09	*–2.3*
southern	0.47	*2.0*	0.13	*2.2*	**0.58**	*3.1*	**0.17**	*5.4*
urbanisation								
< 20'000 inhab.	–0.20	*–1.2*	**–0.20**	*–2.7*	–0.20	*–1.7*	**–0.21**	*–5.4*
> 500'000 inhab.	–0.27	*–1.6*	**–0.24**	*–3.5*	**–0.40**	*–3.7*	**–0.24**	*–6.8*
net unearned income								
Y^{NP} (in DM 10,000)	**–0.42**	*–2.9*	**–0.47**	*–7.8*	**–0.41**	*–4.2*	**–0.46**	*–14.4*
specification test	χ^2(df)	p-%	χ^2(df)	p-%	χ^2(df)	p-%	χ^2(df)	p-%
distance	168.3	0.24	256.5	0.00	151.5	2.73	256.1	0.00
df	120		157		120		157	

Notes:

1. The coefficient of the 1984 intercept has been normalized to 1, and the *t*-values reported for the time effects concern the test of equality to 1 of the other intercepts.

2. Boldface emphasizes "well-determined" slope coefficients (p-value < 1 %).

the fact that we estimate large variance matrices with few observations may give these tests a tendency to over-reject.

Consider next the comparison between the panel estimates and the pooled estimates, both with use of macro information (second and fourth pairs of columns). Here the estimates are *very* similar, even for not so well determined coefficients, while the number of well-determined coefficients now moves from 5 to 13. This similarity in the estimates is a reassuring feature of these results. It encourages us to comment on the ceteris paribus interpretation of a few more coefficients: schooling appears to increase the participation probability of lone mothers, Realschule more significantly than Abitur (which may come as a surprise considering the evolutions in Table 1). Children younger than 4 discourage participation, while older children encourage it (recall that we consider only *mothers* here: older children take less time *and* cost more). Child care availability encourages participation (although it is difficult to find a direct interpretation for the negative impact of child care availability for the 4-6 years category). The only age group that significantly differs from the 49+ group is the youngest group, an effect that is surely not independent of the "child age" variables. The separated women have the highest participation probability, the widows the lowest. Being non German has no impact, at least in the additive form postulated here. There appears to be an increase in participation in the south (whereas the reverse seems to be the case for married women, see Laisney et al., 1993b) and the highest participation probability is achieved in smaller cities. The rejection of the specification is confirmed.

The comparison of the panel and pooled estimators without macro information is more puzzling (first and third column pairs). While 10 coefficients appear well-determined in the panel estimation, this seems due more to an increase in the absolute value of the estimates than to a decrease in the corresponding variances: this is the case for the "Abitur" and "not German" dummies, which appear significant only in this estimation. Even more puzzling is the coefficient for the "large city" dummy (> 500,000 inhabitants) which is about twice its size in all other estimations – and this is *not* a misprint. Indeed, given these discrepancies the similarity between the two estimates with macro information is a real surprise.

Table 3 shows the efficiency gains achieved by each estimator over the simple pooled probit estimator in terms of the proportional increase in sample size that would be required to obtain the same precision with the simple pooled probit estimator as with the estimator in question.

Only using the macro information amounts to having more than 6 times as many observations (we refer here to the median multiplier). This is far from being as spectacular as the multiplier of 50 reported by Imbens and

Table 3

Inverse relative asymptotic efficiency
(with reference to the pooled probit estimator without macro information)

GMM estimator	pooled	panel	panel
Macro data used	yes	no	yes
time effects			
1985	4.03	2.50	17.2
1986	5.51	2.83	26.7
1987	6.23	3.02	27.2
1988	5.67	2.30	30.5
1989	5.56	3.39	31.4
1990	6.58	3.67	36.0
schooling			
Realschule	11.85	1.90	33.4
Abitur	7.89	*1.12*	28.3
number of children			
younger than 4	10.79	2.34	**98.0**
4-6 years old	*2.29*	*1.46*	*16.5*
7-14 years old	8.70	2.84	37.8
15-17 years old	6.81	**3.73**	32.9
density of child care relevant child dummy* 10*			
0-3 years	7.22	**3.36**	**126.9**
4-6 years	*2.15*	1.58	*15.7*
age			
younger than 32	6.29	2.62	25.2
33-40	7.74	2.42	40.1
41-48	6.62	2.90	30.5
marital status			
single	6.03	2.14	21.0
divorced	6.00	1.98	23.7
widow	10.47	2.86	38.1
not German	**15.75**	1.93	60.3
regions			
northern	6.70	2.27	21.7
southern	**14.45**	1.55	51.7
urbanisation			
< 20'000 inhabitants	5.47	1.99	19.7
> 500'000 inhabitants	5.84	2.33	22.6
net unearned income			
Y^{NP} (in DM 10,000)]	5.89	2.21	20.4
median multiplier	6.44	2.33	29.39

Notes:

1. Entries in this table are of the type $[V_{asy}(\sqrt{n}\hat{\theta}_n)/V_{asy}(\sqrt{n}\hat{\theta}_n^0)]^{-1}$, where $\hat{\theta}_n$ denotes the estimator for the coefficient of one of the variables and for one of the columns, and $\hat{\theta}_n^0$ denotes the pooled probit estimator without macro information for the same coefficient.
2. Each entry can be interpreted as the proportional increase in sample size that would be required to obtain the same precision with the naive pooled probit estimator as with the estimator in question.
3. The last line gives the column medians.
4. Emphasis is placed on the two highest (bold) and lowest (italics) gains for slope coefficients.

Lancaster for a very parsimonious participation equation for Dutch males, but it is still worth having. Part of the explanation of the difference between the efficiency gains in their study and ours is that the participation rate in their sample is above 90%, which means that the sample gives very little information about the parameters of interest. By contrast, the participation rates in our sample are between 55% and 69%, making the dichotomous variable much more informative in our case. The two largest gains, corresponding to multipliers around 15, are obtained for the "not German" and "southern" dummies, which still both remain insignificant. The lowest two gains still amount to more than doubling the number of observations and are obtained for variables connected with children between 4 and 6 (number and child care density). This comes as a surprise, since we use no macro information concerning regions or nationality, whereas we do use information on children: according to intuition, and to the simulation results of Imbens and Lancaster (1994, 675), we would have expected more efficiency gains for the latter than for the former.

The gains from using the panel structure alone appear much more modest, ranging between almost no increase (for the Abitur dummy and for the number of children aged 4-6) and a multiplier around 3.5 (number of 15-17 years old, and density of child care for children below 3).

In this light, the large gains obtained for the combined use of macro information and panel structure may appear somewhat over-optimistic, with a median multiplier of almost 30, and a lowest multiplier above 15. The lowest two gains are obtained for the same variables as with the use of macro information only, while the highest, 98 and 127, are obtained for variables connected with infants. A more conservative estimate of these gains could be obtained by multiplying the first two columns. This would amount to trusting those figures but denying the possibility of interaction between the two modes of efficiency improvement; it yields a median gain of 7.60, minimum gains of 3.34 and 3.40 for children aged 4-6 (number and density of child care), and maximum gains of 29.9 and 30.4 for the widow and not German dummies, and these are still substantial.

6. Conclusions

What have we learned by doing this exercise? First of all that it is feasible to use the approach proposed by Imbens and Lancaster (1994) with panel data. Compared with a straightforward analysis of micro data, there are some extra costs in terms of programming, computing, and data analysis time, but in our opinion, the efficiency gains obtained amply repay the effort. Moreover, drawing macro information into the data analysis has the positive effect of leading to a better documentation of the data than is usual in studies based on micro data only.

Of course, the whole approach is inferior to what could be achieved by getting hold of the micro data from which the aggregate data has been computed, but the latter is publicly available, and at a low cost, which is often not the case for the former.

For our example, the most significant efficiency gains are achieved by using the macro information. Once this was done, furthermore using the panel structure does not change the estimates much and only yields a better precision, giving us more confidence in the interpretation of the coefficients. In particular the variables pertaining to child care appear to play a significant role in the end, which was not the case for the naive pooled probit estimates. The improved efficiency also yields more stringent inference, leading to unambiguous rejection of the specification.

The approach followed here is both widely applicable and extremely flexible. In our opinion its potential has still to be fully discovered. For empirical demand analysis, for instance, it can be can be set against approaches that necessitate exact aggregation of some form – and thus strong restrictions on functional form – in order to obtain identification or efficiency gains in the estimation of price reactions from the combined use of cross-section and macro information. Such approaches put artificial, and mostly empirically rejected, restrictions on the class of admissible micro models, since aggregation must result in a macro model that in some ways mimics the micro model (see for instance Jorgenson et al., 1982, or Nichèle and Robin, 1995).

By contrast, once identification is achieved, the approach followed here places no constraint on the nature and number of extra moments used, provided that they are compatible with the micro information (under some usual regularity conditions). In particular, there is no necessity for the macro information to be available at all dates corresponding to the waves of the panel, and it is legitimate to discard some moments if their number makes computations problematic.

Appendix: Taking Account of a Stratified Sample

The SOEP is a stratified sample, and in particular there is a special file containing most non Germans, and these are over-sampled. For commodity of exposition, we will refer to the subsamples as subsamples A and B, and to their individuals as Germans and non Germans, respectively, even though some non Germans are included in subsample A. Table A1 reports statistics on individual sampling weights for both subsamples. These are the INFRA-Test weights, and we shall admit that they are inversely proportional to the sampling probabilities.[13]

Table A1

Statistics on sampling weights

Year	1984	1985	1986	1987	1988	1989	1990
Germans (A)							
mean weight	1151	1199	1150	1102	1110	1225	1252
std of weights	357	357	325	458	367	389	344
observations	118	127	103	95	81	79	68
non Germans (B)							
mean weight	263	260	265	250	263	220	246
std of weights	164	158	160	159	127	129	115
observations	32	30	26	24	29	23	17

Although there is substantial variability within each subsample, it is striking that the weights in subsample A are on average 4 times larger than those for subsample B. Since the weights themselves are estimates that correlate with some of the explanatory variables we use, we are reluctant to use them directly in estimation. On the other hand, ignoring them would make unweighted sample means non comparable with the corresponding population magnitudes. Thus we choose a middle path, assuming constancy of the sampling probabilities within each subsample, but estimating the various conditional probabilities we need by using the corresponding sampling weights.

Computation of macro moments

In detail, in the notations of Section 4, and leaving the time index aside for the moment, the macro moments that we use have the form $\mathrm{P}[y_i = 1 | j_i = 1]$, or, using the law of iterated expectations $\mathrm{E}[y_i | j_i = 1] = \mathrm{E}\{\mathrm{E}[y_i | x_i] | j_i = 1\} = \mathrm{E}[\Phi_i | j_i = 1]$. But

$$\text{(A.1)} \qquad \mathrm{E}[\Phi_i | j_i = 1] = \mathrm{E}[\Phi_i | j_i = 1, i \in A]\mathrm{P}[i \in A | j_i = 1] + \mathrm{E}[\Phi_i | j_i = 1, i \in B]\mathrm{P}[i \in B | j_i = 1]\,.$$

13 We have also tried the DIW weights and found only a marginal impact on our results. Therefore we report only one set of results. For details on the various sampling weights available with the SOEP, we refer the reader to Pischner (1994).

We estimate the quantities of type $P[i \in A|j_i = 1]$, period by period, by dividing the sum of the weights of individuals in A with $j_i = 1$ by the sum of the weights of all individuals with $j_i = 1$, and call the corresponding estimate $\hat{P}_{Ajt}$.

Furthermore, quantities of the type $E[\Phi_i|j_i = 1, i \in A]$ are consistently estimated by

$$\text{(A.2)} \qquad \frac{1}{n_{Aj}} \sum_{i=1}^{n} \mathbf{1}[i \in A, J_i = 1]\Phi_i = \frac{1}{n_j} \sum_{i=1}^{n} j_i \frac{n_j}{n_{Aj}} \mathbf{1}[i \in A]\Phi_i ,$$

where n_{Aj} denotes the number of observations of subsample A satisfying $j_i = 1$, and n_j the number satisfying the latter condition. Finally, the expression that replaces equation (9) in the text is

$$\text{(A.3)} \qquad h_{it}^{2j}(x_{it}; \theta_t) = r_{it} j_{it} n_j \left\{ p_{jt} - \Phi_{it} \left[\frac{\hat{P}_{Ajt}}{n_{Ajt}} \mathbf{1}[i \in A] + \frac{\hat{P}_{Bjt}}{n_{Bjt}} \mathbf{1}[i \in B] \right] \right\} .$$

This is analogue to what Imbens and Lancaster (1994, 667) report, except that we have chosen to estimate the sampling probabilities separately instead of including the corresponding moments in the estimation procedure, as they do. We thus incur an efficiency loss, but avoid doubling the number of macro moments, which is almost certainly advantageous in small samples.

Computation of the asymptotic variance of the estimator

The asymptotics considered here assume indefinitely repeated stratified sampling, and it is important to note that we are not interested in population moments, but in the parameter θ of the conditional distribution of y given x (in the notations of Section 2). Thus, although we do use reweighting in the computation of some of the moments in order to ensure that their expectation is indeed zero, we do not subsequently reweight the individual moments (i.e. moments indexed with i). The restriction that makes this procedure feasible is that the stratification is defined on the basis of the conditioning variable x. The independence of the individual moments and the fact that their expectation is zero ensure the applicability of a uniform law of large numbers and a central limit theorem (see Davidson and MacKinnon, 1993, Definition 4.16 p. 136). Given this, equation (4) of Section 2 remains valid, with the following equation replacing equation (5):

$$\text{(5')} \qquad V = (D'CD)^{-1} D'C \left[p \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} h(y_i, x_i; \theta_0) h(y_i, x_i; \theta_0)' \right]^{-1} CD(D'CD)^{-1} ,$$

with

$$D = p \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \frac{\partial h'}{\partial \theta} (y_i, x_i; \theta_0) . \tag{5''}$$

In both expressions the probability limit corresponds to perpetual sampling along the lines above. We show the full expression rather than the simplified analogue to expression (5) which obtains if the weighting matrix C is chosen equal to the probability limit in (5′) because (i) some of our results are based on other choices of C, and (ii) even when C is chosen as optimal we do use the full expression, as this leads to a more robust estimate of the asymptotic variance.

References

Arellano, M. / *Meghir,* C. (1992), "Female Labour Supply and On the Job Search: An Empirical Model Estimated Using Complementary Data Sources", Review of Economic Studies, 59, 537-557.

Avery, R. B. / *Hansen,* L. P. / *Hotz,* V. J. (1983), "Multiperiod Probit Models and Orthogonality Condition Estimation", International Economic Review, 24, 21-35.

Bertschek, I. / *Lechner,* M. (1995), "Convenient Estimators for the Panel Probit Model", Discussion Paper 528-95, Institut für Volkswirtschaftslehre und Statistik, Universität Mannheim.

Davidson, R. / *MacKinnon,* J. (1993), Estimation and Inference in Econometrics. Oxford University Press, Oxford.

Hanefeld, U. (1987), Das Sozioökonomische Panel. Campus Verlag, Berlin.

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

Imbens, G. W. / *Lancaster,* T. (1994), "Combining Micro and Macro Data in Microeconometric Models", Review of Economic Studies, 61, 655-680.

Jorgenson, D. W. / *Lau,* L. J. / *Stoker,* T. M. (1982), "The Transcendental Logarithmic Model of Aggregate Consumer Behavior", in Advances in Econometrics, R.Basman and G.Rhodes (eds.). JAI Press, Greenwich.

Laisney, F. / *Lechner,* M. / *Staat,* M. / *Wagenhals,* G. (1993a), "Labour Force and Welfare Participation of Lone Mothers in West Germany", ZEW Discussion paper 93-26.

Laisney, F. / *Lechner,* M. / *van Soest,* A. / *Wagenhals,* G. (1993b), "A Life-Cycle Labour Supply Model with Taxes Estimated on German Panel Data: The Case of Parallel Preferences", Economic and Social Review, 24, 335-368.

Lechner, M. / *Breitung,* J. (1996), "Some GMM Estimation Methods and Specification Tests for Nonlinear Models", Chapter 22 in , in The Econometrics of Panel Data: A Handbook of the Theory with Applications, L. Mátyás and P. Sevestre (eds.), second revised ed., 582-612. Kluwer, Amsterdam.

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

Newey, W. K. / *McFadden,* D. L. (1995), "Large Sample Estimation and Hypothesis Testing", Chapter 36 in Handbook of Econometrics, Vol. IV, R.F. Engle and D.L. McFadden (eds.), 2113-2241. North Holland, Amsterdam.

Nichèle, V. / *Robin,* J. M. (1995), "Simulation of Indirect Tax Reforms Using Pooled Micro and Macro French Data", Journal of Public Economics, 56(2), 225-244.

Pischner, R. (1994), "Quer- und Längsschnittgewichtung des Sozio-ökonomischen Panels", in Gewichtung in der Umfragepraxis, S. Gabler, J. Hoffmeyer-Zlotnick and D. Kreps (eds.), 166-187. Westdeutscher Verlag, Opladen.

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

Verbeek, M. / *Nijman,* T. (1992), "Incomplete Panels and Selection Bias", in The Econometrics of Panel Data: Handbook of Theory and Applications, L. Mátyás and P. Sevestre (eds.), 262-302. Kluwer, Amsterdam.

Wagner, G. / *Burkhauser,* R. / *Behringer,* F. (1993), "The English Language Public Use File of the German Socio Economic Panel Study", Journal of Human Resources, 28, 429-433.

Zusammenfassung

Zur gemeinsamen Nutzung von Paneldaten und makroökonomischer Information in der Schätzung eines Panel-Probit-Modells

Zur Schätzung von Modellen für besondere Gruppen einer Bevölkerung verfügt der Ökonometriker typischerweise über wenige Beobachtungen und sollte jede gültige, zusätzlich verfügbare Information berücksichtigen. Hier werden für die Schätzung eines Partizipationsmodells für alleinerziehende Mütter die relativen Effizienzgewinne untersucht, die im Rahmen einer GMM-Schätzung aus der Ausnutzung, einerseits der Panelstruktur, andererseits von Makroinformation in der Form von zusätzlichen Momenten, erzielt werden.

Abstract

When studying particular subgroups of a population the econometrician typically has few observations, and should draw upon any additional relevant information. We illustrate, for the estimation of a participation model for lone mothers, the relative benefits derived from using the panel structure of the data in a GMM framework and from including macro information in the form of extra moments.

JEL-Klassifikation: C23, C25, J22