The Economic Journal, **114** (*July*), © Royal Economic Society 2004. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

Technical Appendix to

HOW IMPORTANT IS METHODOLOGY FOR THE ESTIMATES OF THE DETERMINANTS OF HAPPINESS?

Ada Ferrer-i-Carbonell and Paul Frijters

Economic Journal, vol. 114 (*July*), pp. 641–659

Appendix A: The GSOEP Sample

The German Socio-Economic Panel (GSOEP) is a representative panel of the German population that started in the Federal Republic of Germany in 1984. It currently tracks about 20,000 individuals and 12,000 households in both West Germany and the Former German Democratic Republic; see Wagner *et al.* (1993); Landua (1992); or Plug (1997) for a detailed description. We use the sample of 7,806 West-German workers, which forms around 75% of the West-German total sample. Because the transition to unemployment or work was low in this period (Hunt, 1999), this is a quite stable sample.

From this sample, we look at the six waves of the period 1992–7. The number of waves an individual is observed differs for various reasons. First, there are individuals who leave the panel for reasons such as death, immigration and (temporary or permanent) attrition. Second, there are new individuals included in the sample for reasons that include moving into a surveyed household, reaching the age of 16, or splitting-off from a surveyed household. Third, those who moved from working to non-working, East to West, or *vice versa*, also have fewer than six observations. All this means that we have 7,995 individuals and 30,569 observations in total, which can all be included in the cross-sectional models (models (1) and (3)). Of this total, 1,331 individuals only have one recorded wave in the period and hence they drop out in the fixed-effects OLS model (2), leaving 6,664 individuals and 21,104 observations. Of these remaining individuals, 863 have the same general satisfaction in all waves, meaning they cannot be used for the conditional estimator of the fixed-effect logit model presented in Section 2.5, leaving 5,801 individuals. Of these individuals, all the observations are used in estimation however, meaning that these 5,801 individual correspond to 25,442 observations for that model.

Regarding the variable definitions: age is calculated from the date of birth; income is net monthly household income in German Marks; the number of children is the number of dependent children younger than 16 who live in the household; whether the respondent lives in partnership is self-reported and does not only include marriage; health is the cardinal score on the answer 'how satisfied are you with your health situation' on a (0,10) scale.

Appendix B: Sensitivity Analyses of the Fixed-effect Ordered Logit Model

Our main worry is the endogeneity of GS and health since they are both subjectively evaluated. Therefore, we estimate the linear relation $\Delta Health_{it} = \Delta z_{it}\gamma + u_{it}$ and use $\Delta z_{it}\hat{\gamma}$ as an instrument for $\Delta Health_{it}$. The identifying variable in z_{it} is the number of days off from work because of illness.

Replacing health by predicted health greatly reduces the significance of the health coefficient and the overall likelihood but only qualitatively affects the age coefficient. Because age is, amongst others, a proxy for health, this was to be expected. Hence, although

the endogeneity of subjective health may indeed be responsible for the relatively high levels of R^2 found in Table 1, this endogeneity does not seriously affect most results.

Table 4 also presents the sensitivity analysis for the inclusion of time-dummies. These intercepts are, as expected, important to age and income results because their omission changes their coefficients. Nevertheless, because the time-period we look at here is shorter and more recent than those for Hamermesh (2001) and Winkelmann and Winkelmann (1998), this does not imply that the same change necessarily occurs in their papers.

Appendix C: The Das and van Soest Method

The Das and van Soest (1996; 1999) method first recodes each individual vector $\{GS_{i1},\ldots, GS_{iT}\}'$ into a set of K vectors $\{(GS_{i1} > k),\ldots, (GS_{iT} > k)\}'$ for $k = 0$ to $K - 1$, where $(K + 1)$ is the number of categories of the dependent variable and the lowest category equals 0. For each k, the parameter vector is estimated using the Chamberlain method. Because this yields a consistent estimator we have

$$\sqrt{n_k}(\beta_k - \beta) \rightarrow \mathrm{N}(0, \mathbf{\Sigma}_{kk}^{-1}), \quad k = 0, \ldots, K - 1, \tag{6}$$

whereby the data set for a particular k consists of all those individuals for whom $T > \sum_{t=1}^{t=T}(GS_{it} > k) > 0$. Asymptotically $\mathbf{\Sigma}_{kk} \rightarrow E[\mathbf{l}_k\mathbf{l}_k']$ where $\mathbf{l}_k$ is the score vector $\partial \ln L/\partial\beta_k$. To obtain the final estimator $\hat{\beta}$, Das and Van Soest use a minimum distance step:

$$\hat{\beta} = \arg\min_{\beta} \frac{1}{2}\left[\begin{pmatrix}\beta_0\\ \cdots\\ \beta_K\end{pmatrix} - \begin{pmatrix}\beta\\ \cdots\\ \beta\end{pmatrix}\right]' \mathbf{\Omega}^{-1}\left[\begin{pmatrix}\beta_0\\ \cdots\\ \beta_K\end{pmatrix} - \begin{pmatrix}\beta\\ \cdots\\ \beta\end{pmatrix}\right], \tag{7}$$

where the weighing matrix $\mathbf{\Omega} = [w_{a,b}]$ has entries $w_{a,b} = \mathbf{\Sigma}_{aa}^{-1}\mathbf{\Sigma}_{ab}^{-1}\mathbf{\Sigma}_{bb}^{-1}$ with $a, b = 0, \ldots, K - 1$. This estimator is made operational by replacing the unknown variance matrices with their sample analog. This for instance means

$$\hat{\mathbf{\Sigma}}_{ab}^{-1} = \left(\frac{1}{N_{ab}}\sum_i \begin{matrix}\dfrac{\partial \ln L_i}{\partial\beta_{a1}}\dfrac{\partial \ln L_i}{\partial\beta_{bM}} & \cdots & \dfrac{\partial \ln L_i}{\partial\beta_{a1}}\dfrac{\partial \ln L_i}{\partial\beta_{bM}}\\ \cdots & \cdots & \cdots\\ \dfrac{\partial \ln L_i}{\partial\beta_{aM}}\dfrac{\partial \ln L_i}{\partial\beta_{b1}} & \cdots & \dfrac{\partial \ln L_i}{\partial\beta_{aM}}\dfrac{\partial \ln L_i}{\partial\beta_{bM}}\end{matrix}\right)^{-1}, \tag{8}$$

Table 4

Sensitivity Analyses on the Fixed-effect Ordered Logit Model for the GSOEP

	Fixed-eff. ordered logit							
	(I)		(II)		(III)		(IV)	
	Estimate	t-val	Estimate	t-val	Estimate	t-val	Estimate	t-val
Age							−0.06	2.0
Age × age	−0.0002	0.7	−0.0006	1.8				
ln(household income)	0.19	3.8	0.18	3.7	0.11	2.3		
Number of children	0.002	0.1	0.01	0.5				
Steady partner (1 = yes)	0.08	1.4	0.11	2.0				
Subjective health	0.37	24.6						
Predicted health			0.46	4.5				
Time dummies	Yes		Yes		No		No	
Number of individuals	5,801		5,790		5,801		5,801	
−log(Likelihood)	9004.3		9286.3		9408.1		9332.1	
Number of cases	25,442		25,403		25,442		25,442	

where M is the number of parameters and N_{ab} is equal to the number of individuals that are both in the data set for $k = a$ and for $k = b$. Applying their method, we improve slightly on the Das and Van Soest estimator by using the sample hessian for $\boldsymbol{\Sigma}_{aa}^{-1}$ instead of $(1/n_a)\sum_i[\mathbf{l}_{ik}\mathbf{l}'_{ik}]$ because the sample hessian has better finite sample properties (Hayashi, 2000, p. 476).

When the sample sizes are very high and there is a lot of variation in the exogeneous variables, the Das and Van Soest estimator seems to make better use of all the available information than our estimator. In applying it to our data though, there were a number of limitations. For one, the estimation of β_k requires that there are sufficient individuals who have both some observations of GS_{it} higher than k and an observation equal or less than k. This in our case only held for sufficiently large k: the number of individuals reporting a 0 was for instance only 15. Even the number of individuals reporting anything lower than a 5 was less than 300. Also, for some variables, such as the number of children, there is not very much time-variation. This increases the number of individuals one needs per k to get sufficient variation for the estimator to have good properties. Additionally, the estimation of $\boldsymbol{\Sigma}_{ab}^{-1}$ requires individuals both in the data set for $k = a$ and for $k = b$. This involves fewer individuals than that are in either the data set for $k = a$ or $k = b$. For these reasons, we could only apply the Das and Van Soest method to four groups: $k = 5$, $k = 6$, $k = 7$, and $k = 8$. This implies a loss of data to the extent that the Das and Van Soest estimates in the text are based on 5,222 individuals which is about 11% less than the number of individuals for our own estimator.

Apart from these practical limitations appearing in our data, there is also a theoretical disadvantage to the estimator of Das and Van Soest: $\boldsymbol{\Sigma}_{ab}^{-1} = E[\mathbf{l}_a\mathbf{l}'_b]$ depends on the joint distribution of the sets $k = a$ and $k = b$ and hence on the distributions of λ_k^i and f_i. This creates regularity problems. For instance, if $\lambda_a^i = \lambda_{a+1}^i$ for some $a \neq \{0, K\}$ then category a is empty. This does not affect the validity of any of the estimators β_k or their asymptotic properties. Neither will this affect our new estimator. However, in this case $\boldsymbol{\Sigma}_{ab}^{-1} = \boldsymbol{\Sigma}_{a-1,b}^{-1}$ and $\boldsymbol{\Sigma}_{b,a}^{-1} = \boldsymbol{\Sigma}_{b,a-1}^{-1}$ for all b, which means $\boldsymbol{\Omega}$ is singular and the method breaks down. Another example: if $\lambda_a^i = -\infty$ for some individuals and $\lambda_b^i = +\infty$ for all other individual with $1 < a < b < K + 1$, then no category is empty. One still has consistent estimates for any β_k and for our new estimator, but there are no observations to estimate $\boldsymbol{\Sigma}_{a-1,b}^{-1}$.

Summarising, the Das and Van Soest method requires stricter regularity assumptions on the distributions of λ_k^i and f_i. For samples that are very big, with a lot of variation in x_{it}, combining the estimators β_k in a fashion suggested by Das and Van Soest would seem to work well though. In case of limited variation in some x_{it} and where little can be presumed about the distributions of λ_k^i and f_i, our method is more robust.

Appendix D: The Properties of the Estimator

Our discussion of the estimator and our strategy for efficiency takes the notionally convenient case that β is of dimension 1 and T is equal for all individuals, but carries over to the case that T is variable and β is multi-dimensional.

We first transform our data in a way that only preserves the information we can use. We introduce the notation C_i for the set of possible different conditioning events for individual i. For a vector $\{GS_{i1} = 5,\ GS_{i2} = 7,\ GS_{i3} = 4\}'$ this for instance means $\mathbf{C}_i = \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$ where the first element belongs to the conditioning event that $GS_{i1} > 4$ and the second element belongs to the conditioning event that $GS_{i1} > 5$. The vectors in $\mathbf{C}_i$ are denoted as

$\mathbf{C}_{ij}$, where j runs from 1 to n_i^C. Each vector $\mathbf{C}_{ij}$ is implicitly related to a k, termed k_{ij}. The time observations in each vector $\mathbf{C}_{ij}$ are denoted by C_{ijt}. The general problem is to find weights for the maximisation problem

$$\hat{\beta} = \arg\max_{\beta}[M = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{n_i^C} w_{ij} \ln L(\mathbf{C}_{ij}|\sum_{t=1}^{T} C_{ijt}, \beta, \mathbf{x}_i)] \tag{9}$$

$$s.t. \quad \sum_{i=1}^{N}\sum_{j=1}^{n_i^C}(w_{ij})^2 = N, \tag{10}$$

where M is the function to be maximised. Now, $\mathbf{C}_{ij}$ is independently distributed over individuals, but not identically distributed. We have:

$$\ln L_{ij} = \ln L(\mathbf{C}_{ij}|\sum_{t=1}^{T} C_{ijt}, \beta, x_i) = \ln \frac{e^{\sum_{t=1}^{T} C_{ijt}\mathbf{x}_{it}\beta}}{\sum_{S(\sum_{t=1}^{T} Cijt)} e^{\sum_{t=1}^{T} C_{ijt}\mathbf{x}_{it}\beta}},$$

$$\frac{\partial \ln L_{ij}}{\partial \beta} = \frac{1}{L_{ij}} \times \left\{ \frac{\sum_{t=1}^{T} C_{ijt}x_{it}e^{\sum_{t=1}^{T} C_{ijt}x_{it}\beta}}{\sum_{S(\sum_{t=1}^{T} C_{ijt})} e^{\sum_{t=1}^{T} C_{ijt}\mathbf{x}_{it}\beta}} - \frac{\left[\sum_{S(\sum_{t=1}^{T} C_{ijt})}(\sum_t C_{ijt}x_{it})e^{\sum_{t=1}^{T} C_{ijt}\mathbf{x}_{it}\beta}\right] e^{\sum_{t=1}^{T} \mathbf{C}_{ijt}\mathbf{x}_{it}\beta}}{\left[\sum_{S(\sum_{t=1}^{T} C_{ijt})} e^{\sum_{t=1}^{T} C_{ijt}\mathbf{x}_{it}\beta}\right]^2} \right\},$$

$$\mathrm{E}\frac{\partial \ln L_{ij}}{\partial \beta} = \sum_{C_{ijt}^* \in S(\sum_{t=1}^{T} C_{ijt})} L(C_{ijt}^*)\frac{1}{L_{ij}(C_{ijt}^*)} \times \left\{ \frac{\sum_{t=1}^{T} C_{ijt}^*\mathbf{x}_{it}e^{\sum_{t=1}^{T} C_{ijt}^*\mathbf{x}_{it}\beta}}{\sum_{S(k_i,\sum_{t=1}^{T} C_{ijt})} e^{\sum_{t=1}^{T} C_{ijt}\mathbf{x}_{it}\beta}} - \frac{\left[\sum_{S(\sum_{t=1}^{T} C_{ijt})}(\sum_t C_{ijt}\mathbf{x}_{it})e^{\sum_{t=1}^{T} C_{ijt}\mathbf{x}_{it}\beta}\right] e^{\sum_{t=1}^{T} C_{ijt}^*\mathbf{x}_{it}\beta}}{\left[\sum_{S(\sum_{t=1}^{T} C_{ijt})} e^{\sum_{t=1}^{T} C_{ijt}\mathbf{x}_{it}\beta}\right]^2} \right\} = 0,$$

$$\mathrm{E}\frac{\partial^2 \ln L}{\partial^2 \beta} = \mathrm{E}(\frac{\partial \ln L_{ij}}{\partial \beta})^2 = \sum_{C_{ijt}^* \in S(\sum_{t=1}^{T} C_{ijt})} \frac{1}{L_{ij}(C_{ijt}^*)} \times \left\{ \frac{\sum_{t=1}^{T} C_{ijt}^*\mathbf{x}_{it}e^{\sum_{t=1}^{T} C_{ijt}^*\mathbf{x}_{it}\beta}}{\sum_{S(k_i,\sum_{t=1}^{T} C_{ijt})} e^{\sum_{t=1}^{T} C_{ijt}\mathbf{x}_{it}\beta}} - \frac{\left[\sum_{S(\sum_{t=1}^{T} C_{ijt})}(\sum_t C_{ijt}\mathbf{x}_{it})e^{\sum_{t=1}^{T} C_{ijt}\mathbf{x}_{it}\beta}\right] e^{\sum_{t=1}^{T} C_{ijt}^*\mathbf{x}_{it}\beta}}{\left[\sum_{S(\sum_{t=1}^{T} C_{ijt})} e^{\sum_{t=1}^{T} C_{ijt}\mathbf{x}_{it}\beta}\right]^2} \right\}^2,$$

where $C_{.}$ denotes the random variable and $\mathbf{C}_{.}$ the realisation. Because C_i is independently distributed and $\mathrm{E}[L(\mathbf{C}_{ij}|\sum_{t=1}^{T} C_{ijt}, \beta, \mathbf{x}_i)]$ is shown above to be maximised at the true β for any conditioning set, this establishes that the estimator $\hat{\beta}$ follows the regularity conditions required for extremum estimators to be consistent and normally distributed under mild conditions on w_{ij} (see Hayashi, 2000, ch. 7). Most importantly, it implies that $\mathrm{E}(\partial \ln L_{ij}/\partial\beta) = 0$ for any k_{ij}. Our approach is to impose the restriction that $w_{ij} = 0$, 1 and that $\sum_{j=1}^{n_i^C} w_{ij} = 1$. One advantage of this is that we can interpret the ensuing estimator as a Maximum Likelihood estimator. Starting out with a consistent estimator of β which can be obtained by applying the standard Chamberlain method, we in a second step set $w_{ij} = 1$ for the j that minimises the analytically calculated $\mathrm{E}(\partial^2 \ln L_{ij}/\partial^2\hat{\beta})$ for each particular individual i. This weighting strategy is analogue to weighted least-squares analyses where the variance is a known function of the conditioning information and the parameters β.

What our method circumvents is estimating $P[w_{ij} = 1]$ because this would require estimating the joint probability of $\mathbf{C}_{i1}, \ldots, \mathbf{C}_{in_i^c}$ which involves the unknown nuisance parameters. For the same reason, we cannot construct a maximum likelihood estimator in which $w_{ij} > 0$ for more than 1 j per individual because the joint probability of any pair $\mathbf{C}_{ij}$ and $\mathbf{C}_{il}$ involves the unknown nuisance parameters. Hence our method produces the maximum likelihood estimator with minimal variance.

The Das and Van Soest method also circumvents the problem of estimating the joint probability of $\mathbf{C}_{i1}, \ldots, \mathbf{C}_{in_i^c}$ by weighing $M - 1$ separate consistent estimators (where each estimate for β_k is, by the way, not based on i.i.d. data because $\sum_{t=1}^{T} C_{ijt}$ and even T varies per individual within the same set for k). This uses more information but involves the disadvantages for finite samples discussed in the previous Appendix and the implicit reliance on stronger regularity conditions for the nuisance parameters.

An open question is whether we can do better than maximum likelihood. The essential problem we have in finding variance minimising $w_{ij}/\sum_j w_{ij}$ is that this theoretically involves for each individual estimating $\mathrm{E}\dfrac{\partial \ln L_{ij}}{\partial \beta}\dfrac{\partial \ln L_{il}}{\partial \beta}$. This expression can not be estimated empirically because we only have one observation of $\dfrac{\partial \ln L_{ij}}{\partial \hat{\beta}}\dfrac{\partial \ln L_{il}}{\partial \hat{\beta}}$ per individual. It also cannot be analytically calculated with some initial estimate of β because $\mathrm{E}\dfrac{\partial \ln L_{ij}}{\partial \beta}\dfrac{\partial \ln L_{il}}{\partial \beta} = \sum_{S(\sum_{t=1}^{T} C_{ijt})} \sum_{S(\sum_{t=1}^{T} \mathbf{C}_{ilt})} \dfrac{L(C_{ij}^*, C_{il})}{L(C_{ij}^*) * L(C_{il})} \dfrac{\partial L_{ij}}{\partial \beta}\dfrac{\partial L_{il}}{\partial \beta}$ which involves $L(C_{ij}, C_{il})$ which does not factor out because this joint probability depends on the nuisance parameters. Hence, there seems no analytical way to optimally choose $w_{ij}/\sum_j w_{ij}$. A second-best option is to order the data in such a way that we have groups of observations with the same T and the same $\sum_{t=1}^{T} C_{ijt}$ because these are i.i.d. If each of these groups is large enough, then the optimal weighting of these different groups can use sample estimates of the cross-variance, circumventing the issue raised above. There is a large number of groups in our actual data however because both T and $\sum_{t=1}^{T} C_{ijt}$ vary in our data. Therefore, this is not an appealing way forward in our case, but may be an option when data sets are very large and less heterogeneous.

Finally, we note that the implicit sampling out of $\mathbf{C}_i$ via the free parameter k_i does not affect the estimators. For one, $L(\mathbf{C}_i|\beta, \mathbf{x}_i, f_i, \lambda_i)$ is also maximised at the true β. Hence, even though we do not know the true f_i and λ_i, the conditional likelihood of each individual $\mathbf{C}_{ij}$ with an implicit k_{ij} will also maximise the unknown unconditional likelihood. Because we can consistently estimate $\mathrm{E}(\partial^2 \ln L_{ij}/\partial^2 \beta)$ for each of the n_i^C conditioning events $\sum_{t=1}^{T} C_{ijt}$ and base our weighing on it, means we consistently estimate the variance of our final conditional estimator by $(1/N)\sum_i \sum_j w_{ij}(\partial^2 \ln L_{ij}/\partial^2 \hat{\beta})$. It is the case that likelihoods with other conditioning information, such as $L(\mathbf{C}_i|\beta, \mathbf{x}_i, f_i, \lambda_i)$ and $L(GS_i|\beta, \mathbf{x}_i, f_i, \lambda_i)$, depend on the nuisance parameters: the asymptotic variance of $L(\mathbf{C}_i|\beta, \mathbf{x}_i, f_i, \lambda_i)$ for instance is related to the nuisance parameters f_i and λ_i^k because $\mathrm{P}[w_{ij} = 1]$ depends on them. The variance of the unconditional likelihoods is therefore unknown.