Psychometrika (2025), **90**, 870–903
doi:10.1017/psy.2024.19

PSYCHOMETRIC SOCIETY

THEORY AND METHODS

Estimation of Linear Models from Coarsened Observations: A Method of Moments Approach

Bernard M. S. van Praag[1], J. Peter Hop[2] and William H. Greene[3]

[1]Tinbergen Institute, University of Amsterdam, The Netherlands; [2]Independent, The Netherlands; [3]University of South Florida, USA

Corresponding author: Bernard M. S. van Praag; Email: b.m.s.vanpraag@uva.nl

(Received 26 May 2023; accepted 31 October 2024; published online 10 March 2025)

J.P.H. deceased on August 2, 2024.

Abstract

In the last few decades, the study of ordinal data in which the variable of interest is not exactly observed but only known to be in a specific ordinal category has become important. To emphasize that the problem is not specific to a specific discipline we will use the neutral term coarsened observation. For single-equation models estimation of the latent linear model by Maximum Likelihood (ML) is routine. But, for higher-dimensional multivariate models it is computationally cumbersome as estimation requires the evaluation of multivariate normal distribution functions on a large scale. Our proposed alternative estimation method, based on the Generalized Method of Moments (GMM), circumvents this multivariate integration problem. It can be implemented by repeated application of standard techniques and provides a simpler and faster approach than the usual ML approach. It is applicable to multiple-equation models with K-dimensional error correlation matrices and J_k response categories for the kth equation. It also yields a simple method to estimate polyserial and polychoric correlations. Comparison of our method with the outcomes of the Stata ML procedure cmp yields estimates that are not statistically different, while estimation by our method requires only a fraction of the computing time.

Keywords: coarsened events; generalized method of moments; item response models; multivariate ordered probit; ordered qualitative data; ordinal data analysis; polychoric correlations

1. Introduction

The statistical tools of empirical Psychometrics, Econometrics, Political Science, and many other empirical sciences including marketing analysis, agriculture, health, and medical statistics, find their origin in the linear regression model.[1] The idea is that a random phenomenon Y can be *predicted* by

[1]Capitals will be used for random variables, vectors, and matrices. We denote the zero-vector by o = (0,. . .,0). Disturbance terms will be in Greek letters. Roman letters will be used for constants and realizations of random variables. Matrices will be denoted by capitals as well to conform to the traditional regression formulas. Indexes will be suppressed where the interpretation will be clear from the context. The ML-estimator of a parameter vector θ is denoted by $\widehat{\theta}$. The corresponding estimator from coarsened data is denoted by $\widehat{\widetilde{\theta}}$. An overbar above a symbol denotes a sample average.

variables X in the sense that $Y \approx f(X;\beta)$, where β is a parameter vector. Later, the same model is applied to *explain* the phenomenon Y, as caused by the variables X. In the econometrics literature since the 1960s this resulted in a host of different models, described in textbooks such as Greene (2018) and Cameron and Trivedi (2005). In psychometrics there was a similar development known as the Structural Equation Model (SEM) (Duncan, 1975; Hayduk, 1987; Bollen, 1989; Jöreskog, 1973). (For software, see, for example, Narayanan (2012).) The main idea behind the modeling approach is that the phenomenon Y to be studied has a conditional mean function that depends on other variables $X_1,\ldots,X_M$, say, $E\left(Y \middle| X = x\right) = f(x;\beta)$, where $f(.)$ is a continuous and differentiable function and β is a set of parameters. In practice the model is usually taken to be linear, that is, $E[Y|X = x] = \beta_0 + \beta_1 x_1 + \cdots + \beta_M x_M$, which produces the linear regression model. In the traditional approach, the variable Y is continuous and directly observable. In economics the model approach (the first influential introduction is by Hood and Koopmans (1953)) is used to describe dependencies between economic variables, e.g., purchase intentions as a function of income, prices, education, family size, and age. In marketing the modeling approach is used to develop and assess the effects of advertising, prices, promotions, etc. (see e.g., Fok (2017)). In medical statistics, a model is used to evaluate the response to diagnostic tests. These are a few examples to illustrate the pervasiveness of the regression model approach in empirical sciences.

The one-dimensional observation Y_i may be replaced by a K-dimensional vector, $Y_i = (Y_{i1},\ldots,Y_{iK})$ and the function $f(.)$ by a K-vector function, $f = (f_1,\ldots,f_K)$.

In practice, the parameters of interest in such a model are estimated from a set of N observations $(\{Y_i,X_i\})_{i=1}^N$. Randomness enters the observed outcomes through the difference between the individual responses, Y_i, and the conditional expectation for individual observation i, $E\left(Y_i \middle| X_i = x_i\right) = f(x_i)$. Hence, we introduce an "error," ε_i, or a K-vector of errors, $\varepsilon_i = (\varepsilon_{i1},\ldots,\varepsilon_{iK})$, which is defined for each response as $\varepsilon_i \stackrel{def}{=} Y_i - f(X_i;\beta)$. The error represents the aggregate of other possible, unobserved variables as well as the randomness of individual behavior. The random term is generated by a mean zero process that operates stochastically independently of X. If the model is linear, we denote the discrepancy as the residual $\varepsilon \stackrel{def}{=} Y - X'\beta$. This approach is applicable when the dependent variable(s) Y and the explanatory variable(s) X is (are) cardinal, i.e., are expressed in observable numerical values.

However, in many practical cases the dependent variable Y is coarsened; it is only observable in terms of ordinal categories on a preference scale, such as subjective health status, well-being, reported as "bad" or "good," or "like" or "dislike," or "poor, fair, good, very good, excellent," or some ordinal ranking from one to five where a cardinal interpretation becomes dubious. Although the observations take place in a coarsened mode, we must interpret the discrete answers as reflecting a latent variable Y, the range of which is a continuum. In that situation, we say the observations are "coarsened" or condensed into a set of J adjacent intervals $\{(v_{j-1},v_j]\}_{j=1}^J = \{S_j\}_{j=1}^J$ on the real axis. For an individual i, the latent observation is $Y_i \in S_{j_i}$, if the realized observation is j_i. Hence, we see that the estimation of the model is complicated by two factors. First, there is the statistical problem that there is always a random error term involved. Second, there is the additional observation problem that the continuum of observations is coarsened or condensed (Maris, 1995) and mapped on a discrete event space $\{j = 1,\ldots,J\}$. We will call such data Ordinally Coarsened (OC).

Since about 1934 in bioassay studies (Bliss, 1934; Finney, 1947, 1971), and in Economics, Sociology, and Political Science, much later in the 1960s and 1970s researchers realized that many variables of interest have an ordinal coarsened character. For instance, a question on self-assessed health status may be responded to with ordered labels varying from "very healthy," to "very unhealthy." In modern datasets, especially survey data, such verbal evaluations are abundant. The World Happiness Report, Oxford University (2024) is a notable example. Such coarsening is nearly always dictated by the fact that respondents are unable to quantify their answers directly on a numerical continuous scale but only in terms of a few ordered verbally described categories. These values are not expressed in numbers but in ordinal qualitative terms. In Psychometrics popular item response theory offers many examples (see

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

Wiberg et al. (2019)). To stress that the statistical problems are not specific to one discipline we will also speak of coarsened data.

In psychometric Structural Equations models (SEM) things may be further complicated by the fact that the explanatory X variables are sometimes coarsened as well (cf. Jöreskog (1973)). In this paper, we will assume that the explanatory X variables are either directly observable on a continuous scale, or coarsened in ordered classes labeled 1, 2,.., or by dichotomous dummy variables, taking the value zero or one.

A specific case of coarsened events is that of psychological testing (item-response theory IRT) and that of multiple-choice tests used in exams. Say, the exam or test consists of K items with each J *ordered* response categories. Then the test may be described by K item scores $Y_1, \ldots, Y_K$. The response to each separate item may be dichotomous (e.g. correct/false) or polychotomous (correct, not wholly correct,. . . , false).

The probability of the response on item k is $p_{ikj} = F(b_k; \theta_i)$, where b_k stands for the difficulty of the item k and θ_i for the ability of respondent i. From the joint probability of the K responses by individual i we try to estimate the latent ability θ_i (e.g., IQ) of the individual by maximizing per individual i the joint probability $\prod_{k=1}^{K} F(b_k; \theta_i)$, with respect to θ_i. The ability θ_i may also depend on (or be explained by) individual observable characteristics $X_i = (X_{i1}, \ldots, X_{iM})$, say $\theta_i = X_i'\beta + D_i\beta_{0i}$,where the dummy variable D_i equals one for individual i and zero for others and β_{0i} is the ability of individual i. The term $X_i'\beta$ gives then the part of ability or intelligence that can be explained by e.g., education, genetic factors, health, income, etc., while the individual parameter β_{0i} may be identified as the unexplainable truly individual ability component. We notice that the estimation of β_{0i} requires that $K > 1$ and preferably considerably larger than one.

The method developed below may be used to estimate these β's and θ_i's. The response probability $F(b_k; \theta_i)$ of a separate item score is frequently assumed to be described by a normal or logistic distribution function. Estimation of θ_i is then rather easy. However, intuition tells us that mostly, the item scores on different items by a respondent will be correlated. This is almost unavoidable if the response behavior depends on one θ_i. In that case, the ML- estimation is mostly difficult since estimation will require the evaluation of many multivariate integrals. The correlation means that the multivariate integrals cannot be reduced to products of one-dimensional integrals. The method proposed by us will avoid this problem.

For directly observed cardinal data, ordinary least squares (OLS) is usually the default estimator of choice.[2] There are many extensions of the OLS estimator that are used in nonstandard cases, such as nonzero covariances across observations. A familiar alternative to OLS is Generalized Least Squares (GLS), in which the disturbances of the K observations per observation unit i are heteroskedastic or correlated. Then an unknown error-covariance matrix has to be estimated as well. If this is feasible, we call the method Feasible Generalized Least Squares (FGLS). Another well-known example is Seemingly Unrelated Regressions (SUR), where K response variables are explained by K equations where the errors are correlated. We refer to well-known econometric textbooks such as Amemiya (1985), Cameron and Trivedi (2005), Greene (2018) and Verbeek (2017) for elaborate descriptions. We note that in modern work these estimations are usually based on the assumption of a known error distribution, usually multivariate normal, leading to Maximum Likelihood (ML) estimation. In the early received literature, least squares were not explicitly based on an underlying error distribution, but rather on the minimization of a Sum of Squared Residuals (SSR) that led to unbiased estimation of the regression coefficients. Later it was found that SSR minimization and ML-estimation led to the same estimator when the errors are normally distributed. (The normality assumption was also used to motivate certain inference procedures.) The common counterpart for the linear regression model in case of coarsening of Y is the Ordered Probit or Ordered Logit model (OP or OL). In the literature, it is frequently called Probit or Logit Regression. (See e.g., McKelvey and Zavoina (1975)).

[2]The original least squares method is due to Gauss (1809) and Legendre (1805). "Sur la Méthode des moindres quarrés". See, also, Stigler (1981) for a historical survey.

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

In this paper we will develop a novel approach to coarsened data, called Feasible Multivariate Ordered Probit (FMOP), where the errors are suspected of being correlated, as is the case in, e.g., item response models, in panel data, regional data or customer satisfaction data. It follows the analogy principle as formulated by Goldberger (1968, 1991) and Manski (1988), based on the method of moments (MoM).

The established way to treat such models with coarsened correlated observations in psychometrics or econometrics, and other empirical applications is by maximum likelihood (ML) estimation (see McFadden (1989) and Hajivassiliou and McFadden (1998)), where the likelihoods per observation include K-variate normal integrals (instead of densities). Those integrals are generally estimated by simulation such as by the Geweke-Hajivasilliou-Kean (GHK)-algorithm. Theoretically, this is a straightforward application of ML theory. However, the practical problem is that the evaluation of those integrals by simulation may be quite cumbersome and time-consuming.

Geweke (1989), McFadden (1989), Keane (1994), Hajivassiliou and McFadden (1998), Cappellari and Jenkins (2003) and Mullahy (2016), and others developed a multivariate probit estimator. Progress has also been made based on the simulated moments approach (McFadden, 1989), using the Gibbs sampler method (see Geman and Geman (1984) and Casella and George (1992)). An interesting historical survey on the (logit and) probit method is found in Cramer (2004). See also Hensher et al. (2015). Roodman (2007, 2020) developed a flexible working Stata estimation procedure (`cmp`) based on the GHK simulator.

Independently, scholars in Psychometrics expanded the vast literature on IRT models yielding different tools of analysis, inspired by the differences between the subject matters between disciplines (see Bollen (1989)). It is surprising that both psychometricians and econometricians are working on essentially the same methodological problems, but mostly without taking much notice of each other's literature. A rare exception is the econometrician Goldberger (1971) who explicitly recognized the commonalities between Econometrics and Psychometrics.

Our approach, based on sample moments, does not need the evaluation of likelihoods, i.e., multi-dimensional probability integrals or simulated moments. In this paper, we assume multivariate normal error distributions for ε in line with the established practice. Data on X are assumed to be generated by a random process that is statistically (and functionally) independent of that which generates ε. The important element is that the observed values of X_i convey no information about the errors ε_i, a property identified by econometricians as "strict exogeneity." It is the property that $E[\varepsilon|X] = 0$ for every X. It follows that $\mathrm{Cov}(X,\varepsilon) = \mathrm{Cov}(X,E[\varepsilon|X]) = 0$ as well. Econometricians call this condition strict exogeneity. It implies the same zero-covariance property. Data on X are assumed to be "well-behaved," meaning that in any random sample, the sample covariance matrix $(1/N)\sum_{i=1}^{N} X_i'X_i \overset{def}{=} \widehat{\Sigma}_X$ is always finite and positive definite. (Regularity conditions on X such as that the influence of any individual X_i in $(1/N)\sum_{i=1}^{N} X_i'X_i$ vanishes as N increases are also assumed.) Nothing further is assumed at this point about the distribution of X_i, e.g., normality, discreteness, etc.

The theoretical model, that is, the data-generating mechanism, mimics the classical linear statistical models. The substantive difference between exact and coarsened data is in the mode of observation of the dependent variable Y. In the classical framework the dependent variable Y is directly observed, while in the Ordered Coarsened (OC) -observation mode, the dependent variable Y is only observed to be in one of the J intervals $S_j = (v_{j-1}, v_j]$, where the cut points v_{j-1} and v_j are unknown parameters to be estimated. If Y is a K-vector, v_{j-1} and v_j are K-vectors and $S_j = (v_{j-1}, v_j]$ is a block in R^K. Since the cut points v_{j-1} and v_j are unknown, it follows that the unit of measurement of Y is unidentified. The usual identification is secured by setting the error variances equal to one; $\sigma_k^2 = 1 (k = 1, \ldots, K)$.[3]

The structure of this paper is as follows. In Section 2 we outline the basic probabilistic model in the presence of coarsening of the dependent variables. In Section 3 we develop the estimation method for a K-equations model based on the Ordered Coarsened data model. In IRT-models, this is equivalent to

[3]The important issue of whether the coarsened sample data $(\overline{Y}_i, X_i)$ contain sufficient information to identify estimators of β and the unknown cut-points in S_j is considered in Greene and Hensher (2010) among others.

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

K item responses per respondent. We call the method Feasible Multivariate Ordered Probit (FMOP). We call it Seemingly Unrelated Ordered Probit (SUOP) when we have a K-equation model with one observation $Y_i = (Y_{i1}, \ldots, Y_{iK})$ per observation unit i. Of course, mixtures of FMOP and SUOP are possible as well. Instead of differentiating high-dimensional log-likelihoods, with the likelihoods being multi-dimensional integrals, with respect to the unknown parameters $\theta = (\beta, v, \Sigma)$, we derive sample moment conditions $\widehat{\widehat{g}}(\theta) = 0$ from the coarsened data that are analogues of the likelihood equations, $\widehat{g}(\theta) = 0$ for direct observations. We then estimate θ from the equation set $\widehat{\widehat{g}}(\theta) = 0$. We find that $\widehat{\widehat{g}}(\theta)$ and $\widehat{g}(\theta)$ converge to the same probability limit for all values of θ. The estimation of θ from $\widehat{\widehat{g}}(\theta) = 0$ takes place by repeatedly applying the generalized method of moments (GMM), (Hansen, 1982). We note in passing that our approach employs elements of the EM algorithm (Dempster et al., 1977). In Section 4 we demonstrate how to estimate polychoric correlations from SUOP-estimates. Further, we generalize the concept of the coefficient of determination, R^2, to $K > 1$. In Section 5.1 we apply FMOP to an employment equation over a five-year panel dataset from the German SOEP dataset. In Section 5.2 we apply SUOP to a block of eight satisfaction questions extracted from the German SOEP data. In order to get more insight into the stability of the method, in Section 5.3. we do some experiments on a simulated data set. We use a recent update of the Stata procedure `cmp` as our benchmark to compare our alternative approach with the ML approach.

We find that the estimation results of regression coefficients and their standard errors do not differ substantially between the two methods. Where the established method may need hours, our method takes only minutes. In Section 6 we provide some concluding remarks. In the Appendix, we propose an easy intuitively appealing method to estimate the latent full error-covariance matrix, after the regression coefficients, β, have been estimated using the coarsened data.

2. Regression in the population space

Regression analysis often begins with assumptions about the distribution of the observed data. Each estimation procedure on a sample may be seen as a reflection of a similar procedure in the population. For convenience but without loss of generality we assume $E(X) = 0, E(Y) = 0$. The model of interest for equation k of K equations is

$$Y_k \mid x = \sum_{m=1}^{M} x_{m,k}\beta_{m,k} + \varepsilon_k = x_k'\beta_k + \varepsilon_k, k = 1, \ldots, K \tag{2.1}$$

Where the model design might call for differences across equations, they can be accommodated by suitable zero restrictions on the coefficients in β_k. For convenience, we ignore constant terms by setting $E(X) = 0$. The expectation of outcome or response Y_k is conditioned or co-determined by explanatory variables/stimuli X_m, $(m = 1, \ldots, M)$ assuming values x_m. There are M variables X_m, that are generated by a strictly exogenous process. The error vector ε derives from observation-specific variation around the theoretical conditional mean. We have $\varepsilon_k = Y_k - X'\beta_k$. Denoting the observations by $y_{i,k}$ the observed residual is defined as $e_{i,k} = y_{i,k} - x_i'\beta_k$. Deviations of observations $i = 1, \ldots, N$ of $Y_{i,k}$ from the conditional mean result from the presence of unobserved elements that enter the data-generating process, for example, variation across individuals in the self-assessment of health or well-being. Random elements are assumed to be generated by a zero mean, finite variance process; if $\varepsilon_k = Y_k - X'\beta_k$, it follows that $E[\varepsilon] = 0$ and $\mathrm{Var}[\varepsilon]$ is finite. We assume that the process that generates X is stochastically independent of that of ε. This implies $E[X\varepsilon] = 0$. Substitution yields the familiar regression equations

$$E\left[X_i'\left(Y_{i,k} - X_i'\beta_k\right)\right] = 0, k = 1, \ldots, K, \tag{2.2}$$

where X_i' is a $(M \times K)$-matrix and $X_i'\left(Y_{i,k} - X_i'\beta_k\right)$ a K-vector.

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

If this holds for all i, then we have

$$\sum_{i=1}^{N} E\left[X_i'\left(Y_{i,k} - X_i'\beta_k\right)\right] = 0 \forall k,m \tag{2.3}$$

from which the regression coefficients β can be solved.

We define the functions $g_{m,k}(\beta) = \sum_{i=1}^{N} E\left[\left(X'_{i,k,m}\right)\left(Y_{i,k} - X_i'\beta_k\right)\right]$. The equation system (2.3) in β is then shortly written as $g(\beta) = 0$.

Result (2.3) identifies the slopes β of the conditional mean function. The zero conditional mean result in (2.2) motivates least squares without reference to minimizing the mean squared error prediction of $Y|X$ or minimum variance linear unbiased estimation of β.

If the error covariance matrix $\Sigma = E\left(\varepsilon\varepsilon'\right)$ is not the identity matrix, we may want to correct for the unequal variances and correlations, and we weigh the observations by $\Sigma^{-1} = \left[E\left(Y - X'\beta\right)\left(Y - X'\beta\right)'\right]^{-1}$, producing

$$g_{m,k}(\beta) = \sum_{i=1}^{N} E\left[\left(X'_{i,m}\right)\Sigma^{-1}\left(Y_{i,k} - X_i'\beta_k\right)\right].$$

By weighting the K observations per individual i by Σ^{-1} variances are standardized, and the error correlations are accounted for. In that case, according to the Aitken theorem the covariance matrix of the estimator $\widehat{\beta}$ is minimized.

We have motivated least squares through the moment equations (2.2). We see that we can interpret these conditions (2.2) as first-order conditions for minimizing the expectation of the squared residuals $S = E\left(\varepsilon'\Sigma^{-1}\varepsilon\right) = E\left(Y - X'\beta\right)'\Sigma^{-1}\left(Y - X'\beta\right)$.

We do not need to specify the probability distribution of $X$and Y. We do assume well- behaved data generating processes, which will include a finite, positive variance of ε and a finite positive covariance matrix, $\mathrm{Var}[X, Y]$. If the marginal probability distribution of ε is multivariate normal, the regression estimator is Maximum Likelihood. We call $X'\beta$ the structural part of the model and ε the disturbance, where β is the $(M \times K)$ – matrix with columns ${\beta_1}', \ldots, {\beta_K}'$.

If the columns of the matrix β are identical, this is the typical setting for longitudinal and panel data. If the coefficients vary by response setting k, β_k we have the situation of K different model equations.

2.1. Regression for Coarsened observations

We call an observation $Y \in R$ coarsened if it is not observed directly, but only as belonging to one of the J intervals $\{(v_{j-1}, v_j]\}_{j=1}^{J} = \{S_j\}_{j=1}^{J}$ (The leftmost and rightmost terminals are infinity). These intervals constitute the class $\mathbb{C}$ of observable events. We will call $\mathbb{C}$ the observation grid. This is generalized to K-dimensional observations by replacing the J observed intervals $(v_{j-1}, v_j]$ by K-dimensional blocks.

$$(v_{j-1}, v_j] = \left[(v_{1,j_1-1}, v_{1,j_1}], \ldots, \left(v_{K,j_K-1}, v_{K,j_K}\right]\right] = S_j$$

where the one-dimensional random observations j_i are replaced by the index vectors $j_i = (j_{i1}, \ldots, j_{iK})$. $\mathbb{C}$ stands for a partition in R^K consisting of J^K adjacent blocks. (We take $J_1 = \cdots = J_K \stackrel{def}{=} J$ for convenience, but equality is not necessary.) We denote the $\mathbb{C}$-coarsened event space by $\Omega_{\mathbb{C}}$. We may then define the corresponding coarsened probability measure $P_{\mathbb{C}}$ on $\mathbb{C}$ by $P_{\mathbb{C}}$. We will call $\mathbb{C}$ the K-dimensional observation grid.

Coarsening of Yimplies that we do not directly observe $Y = y$, but the event $Y \in S_j$, and more explicitly for a K-vector Y that $v_{1,j_1-1} < Y_1 \leq v_{1,j_1}, \ldots, v_{K,j_K-1} < Y_K \leq v_{K,j_K}$. It follows that for given $X = x$ and

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

$j = (j_1, \ldots, j_K)$ there holds for the error vector

$$v_{1,j_1-1} - x_1'\beta_1 < \varepsilon_1 \leq v_{1,j_1} - x'\beta_1, \ldots, v_{K,j_K-1} - x_K'\beta_K < \varepsilon_K \leq v_{K,j_K} - x_K'\beta_K,$$

which we denote shortly as $\varepsilon \in (S_j - x\beta)$. We denote the marginal probability as $p_{k,j} = P(v_{1,j-1} - x\beta_k < \varepsilon_k \leq v_{1,j} - x\beta_k)$. We define the generalized residual as $\bar{\varepsilon} \stackrel{def}{=} E(\varepsilon | \varepsilon \in (S_j - x\beta))$. It is a random K-vector defined on the blocks $(v_{j-1} - x_i\beta, v_j - x_i\beta]$. These blocks constitute individual *i's* individualized observation grid $\mathbb{C}_i$. The grid $\mathbb{C}_i$ for observation unit i depends on $x_i'\beta$. However, for any given value x of X and any grid $\mathbb{C}$ $(x\beta)$ we have $\sum_{j=1}^{J} p_{k,j}\bar{\varepsilon}_{k,j} \stackrel{def}{=} E_{\mathbb{C}}(\bar{\varepsilon}_k) = E(\varepsilon) = 0$ according to the Law of Iterated Expectations(LIE). Then it follows that

$$x_{i,k}.E_{\mathbb{C}}(\bar{\varepsilon}_{i,k}) = 0, \forall i,k \tag{2.4}$$

and

$$(1/N)\sum_{i=1}^{N} E_{X_{i,k}}(X_{i,k}.\bar{\varepsilon}_{i,k}) = 0, k = 1, \ldots, K \tag{2.5}$$

This is the coarsened analogue of (2.2). We define the function $\bar{g}_k(\beta) = (1/N)\sum_{i=1}^{N} E(X_i(\bar{\varepsilon}_{i,k}))$. The equation system (2.4) is then shortly written as $\bar{g}(\beta) = 0$. If the error covariance matrix is not the identity matrix, we may want to correct for the unequal variances and correlations, and we write $\bar{g}(\beta) = \sum_{i=1}^{N} X_i\bar{\Sigma}^{-1}E(\bar{\varepsilon}_{i,k})$, where $\bar{\Sigma} = \mathrm{Var}(\bar{\varepsilon})$.

Finally, we have for the two functions

$$g(\beta) \equiv \bar{g}(\beta) \tag{2.6}$$

This holds for all values of β, not only for the zero roots of (2.2) and (2.4). (2.5) holds since for any x- value $xE(\varepsilon) = xE(\bar{\varepsilon})$.

3. Large sample results for regression

The Law of Large Numbers states that under standard regularity conditions, sample moments converge in probability to their population counterparts as the number N of observations grows large. Slutsky's theorem says that continuous and differentiable functions of random sample moments converge in probability to those functions of the population counterparts as N grows large, implying that the population functions $g(\beta, \Sigma)$ are consistently estimated by filling in the corresponding sample moments.

When we want to get its (large-)sample estimator $\widehat{g}(\beta, \Sigma)$ we replace the expectations in (2.2) with the corresponding sample moment conditions and we get

$$\frac{1}{N}\sum_{i=1}^{N} X_i'\widehat{\Sigma}^{-1}Y_i - \left(\frac{1}{N}\sum_{i=1}^{N} X_i'\widehat{\Sigma}^{-1}X_i'\right)\widehat{\beta} \stackrel{def}{=} \widehat{g}(\widehat{\beta}, \widehat{\Sigma}) = 0, \tag{3.1}$$

where $\widehat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \beta'X_i)(Y_i - \beta'X_i)'$. Solution of (3.1) with respect to the regression coefficients β yields

$$\widehat{\beta} = \left[\frac{1}{N}\sum_{i=1}^{N}(X_i'\widehat{\Sigma}^{-1}X_i)\right]^{-1}\frac{1}{N}\sum_{i=1}^{N}(X_i'\widehat{\Sigma}^{-1}Y_i). \tag{3.2}$$

This is the well-known OLS- estimator.

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

Estimation of the asymptotic covariance matrix Asy.Var$\left[\widehat{\beta}\right]$ is usually understood to mean under the condition that X equals the sample data x. Then the well-known template is

$$\text{Est.Asy.Var}\left[\widehat{\beta}\right] = \frac{1}{N}\left[\frac{1}{N}\sum_{i=1}^{N}\left(X_i'\widehat{\Sigma}^{-1}X_i\right)\right]^{-1} \tag{3.3}$$

3.1. Estimation from ordered coarsened data

Let us now consider the coarsened analogue. The events are elements of the observation grid $\mathbb{C}$. The corresponding coarsened probability measure is $P_{\mathbb{C}}$. If we would follow the conditional ML strategy, the information to be maximized is $E_{\mathbb{C}}\ (\ln P_{\mathbb{C}}) = \sum_{i=1}^{N}\sum_{j=1}^{J}P_{\mathbb{C}}(Y_i \in S_j|x_i)\ln P_{\mathbb{C}}(Y_i \in S_j|x_i)$.

The problem is here the evaluation of the $P_{\mathbb{C}}(Y_i \in S_j|x_i)$, being multivariate integrals over rectangular blocks in R^K. We can evaluate $P_{\mathbb{C}}(Y_i \in S_j|x_i)$ by its sample analogue, but this entails the evaluation of a multitude of K-dimensional integrals, making this procedure very cumbersome, albeit not impossible (see Roodman (2011)).

A much easier way is by making a detour and evaluating the coarsened analogue of the condition (3.1). Let j_i be the K-dimensional response by individual i. We notice that (the K-dimensional) $Y_i \in S_{j_i} \mid x_i$ implies $\varepsilon_i \in \left(v_{i,j_i-1} - x_i'B, v_{i,j_i} - x_i'B\right]$. In this way to each observation unit i is assigned its own observation grid $\mathbb{C}_i$ depending on $x_i'B$. We define the K-vector of the generalized residuals $\bar{\varepsilon}_i = E\left[\varepsilon_i \middle| \varepsilon_i \in \left(v_{i,j_i-1} - x_i'B, v_{i,j_i} - x_i'B\right], X_i = x_i\right]$.

The grid $\mathbb{C}_i$ over which the expectation is taken at the LHS in (3.1) is now a grid in the space R^{M+K} where the first X-coordinates are directly observed while the $K\varepsilon_i$-coordinates are coarsened by $\mathbb{C}_i$. Summing over the observations we get.

$$\frac{1}{N}\sum_{i=1}^{N}E\left[x_i'\Sigma^{-1}\varepsilon_i|x\right] = \frac{1}{N}\sum_{i=1}^{N}E_{C_i}\left[x_i'\bar{\Sigma}^{-1}\bar{\varepsilon}_i|x\right] \tag{3.4}$$

Then (3.4) may be summarized as the identity

$$g(\beta,\Sigma|x) \equiv \bar{g}\left(\beta,\bar{\Sigma}|x\right) \tag{3.5}$$

This implies that the equations $\bar{g}\left(\beta,\bar{\Sigma}|x\right) = 0$ and $g(\beta,\Sigma|x) = 0$ have the same roots β. The vector function $\bar{g}(\beta,\Sigma)$ may be interpreted as the vector of derivatives of a criterion function like a log-likelihood or the sum of squared residuals with respect to β. The simplest criterion function with $\bar{g}(\beta,\Sigma)$ as gradient vector[4] is

$$\begin{aligned}\bar{S} &= \frac{1}{N}\sum_{i=1}^{N}E\left(\bar{\varepsilon}'\,\bar{\Sigma}^{-1}\bar{\varepsilon}|X = x_i\right) = \\ &= \frac{1}{N}\sum_{i=1}^{N}E\left[\varepsilon_i \middle| \varepsilon_i \in \left(v_{i,j-1} - x_i'\beta, v_{i,j} - x_i'\beta\right], X = x_i\right]'\bar{\Sigma}^{-1}E\left[\varepsilon_i \middle| \varepsilon_i \in \left(v_{i,j-1} - x_i'\beta, v_{i,j} - x_i'\beta\right], X = x_i\right]\end{aligned}$$

This is the sum of Squared Generalized Residuals. The identity (3.5) implies that $\bar{S} = E\left(\bar{\varepsilon}'\bar{\varepsilon}|X\right)$ and $S = E\left(\varepsilon'\varepsilon|X\right)$ have the same derivatives with respect to β; consequently, they are identical except for a constant. When we decompose the residual variance into the sum of between- and within-variance $E\left(\varepsilon'\varepsilon|X\right) = E\left(\bar{\varepsilon}'\,\bar{\varepsilon}|X\right) + E\left(\ddot{\varepsilon}'\,\ddot{\varepsilon}|X\right)$, it is obvious that this constant difference is just the within-variance $S - \bar{S} = E\left(\ddot{\varepsilon}'\,\ddot{\varepsilon}|X\right)$, which appears not to depend on β. Things are complicated since each individual i has its own observation grid $\mathbb{C}_i$.

[4]Notice that $\frac{\partial}{\partial\beta}\bar{\varepsilon} = \frac{\partial}{\partial\beta}E\left(\varepsilon\middle|v_{j-1} - x\beta < \varepsilon \le v_j - x\beta, X = x\right) = \frac{\partial}{\partial\beta}\left[E\left(y\middle|v_{j-1} < y \le v_j\right) - x\beta, X = x\right] = x.$

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

The solution for β is found as the root vector of the K-equation system $\overline{g}(\beta,\Sigma|X)=0$.

We have now to construct the sample analogue of $\overline{g}(\beta,\Sigma|X)$. The $\overline{\varepsilon}_{i,j_i}$ have not drawn much attention in the empirical literature. One notable exception is Heckman (1976) who appears to be the first author in econometric literature to recognize the importance of this expected residual, later in the econometric literature sometimes called the Heckman-term (see also Van de Ven and Van Praag (1981)). They have been called by Gouriéroux et al. (1987) the generalized residuals. They used them in the analysis of residuals. If the exact errors ε are standard normally distributed, then we have for the coarsened errors the well-known formula

$$\begin{aligned}\overline{\varepsilon}_{i,j_i} &\overset{def}{=} E\left(\varepsilon_i \,|\, v_{j_i-1}-x_i\beta<\varepsilon_i\le v_{j_i}-x_i\beta,X\right)= \\ &= \frac{\varphi\left(v_{j_i-1}-x_i\beta\right)-\varphi\left(v_{j_i}-x_i\beta\right)}{\Phi\left(v_{j_i}-x_i\beta\right)-\Phi\left(v_{j_i-1}-x_i\beta\right)}\end{aligned} \tag{3.6}$$

If the errors are not normally distributed but logistically, the formulas for the truncated marginal distribution can be found for example in Johnson et al. (1994) or in Maddala (1983, p. 369). We shall restrict ourselves to the assumption of normally distributed errors.

Since there are no natural units observed, we can only estimate the β's in (2.1) up to their ratios. A way to make them identifiable is to assume $\sigma_k=1$ for $k=1,\ldots,K$, which is the traditional assumption in Probit and item response analysis.

The sample moment analogue of (3.1) is

$$\frac{1}{N}\sum_{i=1}^{N}\left[x_i\Sigma^{-1}\overline{e}_{i,j_i}\right]=\frac{1}{N}\sum_{i=1}^{N}\left[x_i\Sigma^{-1}\frac{\varphi\left(v_{j_i-1}-x_i\beta\right)-\varphi\left(v_{j_i}-x_i\beta\right)}{\Phi\left(v_{j_i}-x_i\beta\right)-\Phi\left(v_{j_i-1}-x_i\beta\right)}\right]=\widehat{\overline{g}}(\beta)=0 \tag{3.7}$$

where j_i is the index of the interval/block observed for the observation unit i.

Notice that (3.7) is a concise presentation of a system of K blocks of M equations, each corresponding with one of the elements of the coefficient matrix β, where we assume that each of the K blocks contains M different coefficients β_k.

The cut-points v remain to be estimated. There are $K\times(J-1)$ of them. Therefore, we derive another additional set of v-identifying equations. The cut-points v can be easily estimated one by one by applying the following strategy (called binarization). We define for each equation the $J-1$ auxiliary binary variables $\overline{\varepsilon}^{b}_{i,j}$ which may assume the lower value $E\left(\varepsilon_i|\varepsilon_i\le v_j-x_i\beta\right)$ or the upper value$E\left(\varepsilon_i|\varepsilon_i>v_j-x_i\beta\right)$.[5] We have

$$P\left(\varepsilon_i\le v_j-x_i\beta\right).E\left(\varepsilon_i|\varepsilon_i\le v_j-x_i\beta\right)+P\left(\varepsilon_i>v_j-x_i\beta\right).E\left(\varepsilon_i|\varepsilon_i>v_j-x_i\beta\right)=0 \tag{3.8}$$

Again, there holds $E\left(\overline{\varepsilon}^{b}_{i,j}\right)=0$, due to LIE. For the sample counterparts, this implies

$$\operatorname{plim}\left[\frac{1}{N}\sum\overline{e}^{b}_{i,j}\right]=0, j=1,2,\ldots,J-1.$$

The sample moment analogues are

$$\frac{1}{N}\left[\sum_{i(j_i\le j)}\frac{\varphi\left(v_{k,j}-x_{i,k}\beta_k\right)}{\Phi\left(v_{k,j}-x_{i,k}\beta\right)}-\sum_{i(j_i>j)}\frac{\varphi\left(v_{k,j}-x_{i,k}\beta\right)}{1-\Phi\left(v_{k,j}-x_{i,k}\beta\right)}\right]=0,k=1,\ldots,K,j=1,\ldots,J-1 \tag{3.8a}$$

from which the cut-points $v_{k,j}$ can be easily estimated, as both sums at the left increase in $v_{k,j}$. We notice that these observations are not yet weighted by an error-covariance matrix.

[5]This trick, called "binarization" is suggested by, e.g., Chris Muris, 2017. "Estimation in the Fixed-Effects Ordered Logit Model," *The Review of Economics and Statistics*, vol. 99(3), pages 465–477, July. The term has been used before in computer science.

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

Summing up we are ending with two-equation systems (3.7) and (3.8) from which the parameters β and ν are estimated. This can be done by applying the Generalized Method of Moments (GMM) (Hansen, 1982). We refer to well-known textbooks such as Cameron and Trivedi (2005), Greene (2018) and Verbeek (2017) for elaborate descriptions. The software can be found, e.g., in Stata. We use an iterative calculation scheme. Starting with assuming $\beta = 0$, a first-round yields $\beta^{(1)}, \nu^{(1)}$. Taking these values as a starting point we repeat this iterative procedure until convergence, which is rather rapidly reached. The GMM method provides us with an estimate of the covariance matrix of $\widehat{\widehat{\beta}}, \widehat{\widehat{\nu}}$ as well, using the well-known "sandwich" formula.

4. Polychoric correlations and coefficients of determination

Suppose we have two test items $Y_1, Y_2 (K = 2)$ available by which we may, for example, examine an individual or test the effect of a specific therapy or a response to a satisfaction question. For simplicity, we assume both items are yes/no questions. Then we are of course interested to know how correlated the two test items Y_1, Y_2 and therewith the responses on the two items are. The latent correlation between items is known in the psychometric literature as the polychoric correlation. The literature on polychoric correlations is massive. We refer out of the host of excellent contributions to the seminal (Olsson, 1979) and the more recent (Liu et al., 2021; Moss and Grønneberg, 2023) for more analysis. The problem is clearly how to estimate correlations between latent variables Y_1, Y_2, if we only have a 2×2 coincidence table at our disposition. We propose the following method.

The latent variables are modeled like (2.1). The latent model is

$$Y_{i,k} = X_{i,k,1}\beta_1 + \cdots + X_{i,k,M}\beta_M + \beta_0 + \varepsilon_{i,k} \quad i = 1, \ldots, N, k = 1, \ldots, K \tag{4.1}$$

where in this case $k = 1,2$. In this case, we have

$$\mathrm{Cov}\left(Y_1, Y_2\right) = \beta_1' E\left(X_1 X_2'\right)\beta_2 + \sigma_{1,2}\left(\varepsilon\right) \tag{4.2}$$

and more generally for a $K \times K$ coincidence table we find the $K \times K$-covariance matrix

$$\mathrm{Cov}(Y) = B'\Sigma_{XX}B + \Sigma_{\varepsilon\varepsilon} \tag{4.3}$$

where B' stands for the $K \times M$ matrix of structural effects and $\Sigma_{\varepsilon\varepsilon}$ for the latent error covariance matrix. Now, we derive the polychoric correlation from $\mathrm{Cov}(Y)$ in the usual way, that is, $\rho\left(Y_1 Y_2\right) = \mathrm{Cov}\left(Y_1 Y_2\right) / \sqrt{\sigma\left(Y_1\right).\sigma\left(Y_2\right)}$. The covariance matrix $\Sigma_{YY} = \mathrm{Cov}(Y)$ is estimated as

$$\widehat{\Sigma}_{YY} = \widehat{B}'\widehat{\Sigma}_{XX}\widehat{B} + \widehat{\Sigma}_{\varepsilon\varepsilon} \tag{4.4}$$

where $\widehat{B}$ is the estimated matrix of regression coefficients, $\widehat{\Sigma}_{XX} = \frac{1}{N}\sum_{i=1}^{N} X_i X_i'$, and $\widehat{\Sigma}_{\varepsilon\varepsilon}$the estimated full error-covariance matrix, as estimated in the Appendix.

We notice that in the case that there are no structural effects found, i.e., $B = o$, we still may have non-zero polychoric correlations due to correlated errors. The corresponding correlations are found from the covariance matrix $\widehat{\Sigma}_{YY}$ in the usual way.

The matrices B, Σ_{XX} are already consistently estimated. The latent (full) error-covariance matrix $\Sigma_{\varepsilon\varepsilon}$ is yet unknown. In the Appendix, we demonstrate how $\Sigma_{\varepsilon\varepsilon}$ is consistently estimated.

We note that this method does not assume the normality of the random vectors X or ε. We may also assume, for example, ε to be logistic. In those cases the formula (3.3) for the generalized residual has to be replaced by the corresponding formula for the logistic, or in fact, any distribution, provided that the covariance matrix is finite.

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

In the second application below, where we are estimating satisfaction, we present the estimated 8×8 polyserial correlations between satisfactions as the off-diagonal elements in Tables 5–7 and 8a–8e. For the first application in Section 5 we might estimate the polyserial correlations as well but given the panel nature of the data, it is not very interesting.

The relative explanatory power of the equation estimates depends on the question of how volatile the outcomes are due to random errors. Consider (5.1). We have

$$\mathrm{Var}\left(Y_1\right) = \beta_1' E\left(X_1 X_1'\right)\beta_1 + \sigma_{1,1}\left(\varepsilon\right) \tag{4.5}$$

An attractive measure of fit, that is explanatory power is the traditional coefficient of determination

$$R^2 = \frac{\beta_1' E\left(X_1 X_1'\right)\beta_1}{\beta_1' E\left(X_1 X_1'\right)\beta_1 + \sigma_{1,1}\left(\varepsilon\right)} = 1 - \frac{1}{\beta_1' E\left(X_1 X_1'\right)\beta_1 + \sigma_{1,1}\left(\varepsilon\right)} \tag{4.6}$$

The sample analogue for the first equation is

$$\widehat{R}_k^2 = \frac{\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{m=1}^{M}\widehat{\beta}_{k,m}x_{i,k,m}\right]^2}{\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{m=1}^{M}\widehat{\beta}_{k,m}x_{i,k,m}\right]^2 + 1} \tag{4.7}$$

where $\sigma_{1,1}\left(\varepsilon\right) = 1$, as postulated in Section 3. This is the same magnitude as proposed by McKelvey and Zavoina (1975). We notice that these numbers may be interpreted as coefficients of determination of the regression equations (5.1) for $k = 1, k = 2, \ldots, k = 5$, respectively. Of course, the regression is not performed as $Y_{i,k}$ is not observable per individual. However, the regression correlation coefficient can be estimated by a detour using (4.6) and (4.7). We call these satisfaction correlation coefficients. They measure the part of the satisfaction variation, which can be structurally explained by observable traits X. If there are K equations, we get $\widehat{R}_1^2, \ldots, \widehat{R}_K^2$. For curiosity we present in Table 2 our R^2 and the McFadden (1974) R^2 for Ordered Probit side- by -side.

5. Two empirical examples and one simulation experiment

In order to evaluate our method empirically we considered two data sets, both part of a 2009–2013 panel data-sequence from the German SOEP-panel data set and a block of eight satisfaction questions in wave 2013 of the SOEP data. The model in Section 5.1 is a set of five time-panel Ordered Probit equations where the errors are correlated. We call this estimation method Feasible Multivariate Ordered Probit (FMOP). It can be generalized to an arbitrary number of panel waves. The second data set consists of eight seemingly unrelated cross-section satisfaction questions, where errors are correlated. It is estimated in Section 5.2. We call this a Seemingly Unrelated Ordered Probit model (SUOP). In addition we present estimations on a simulated data set on request of one of the referees to this paper. The program code may be requested from the first author.

Given the results of the method, it becomes possible to estimate the latent full covariance matrix Σ as well. We defer the description of how to estimate the full latent covariance matrix to the Appendix.

5.1. Employment status evaluation on a German five-year panel data set

Now we apply the FMOP method to a specific data set. We choose the employment situation of German workers, where we do not pretend to make a study of German employment but merely test the feasibility of the method, using these employment data. Following the lines above, we try to estimate the employment equation and the error covariance matrix using FMOP.

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

The data are derived from the German, Socio-Economic Panel (SOEP) data set. Households are followed for a period of five successive years (2009–2013). We assume an unstructured error covariance matrix. All explanatory variables are measured as deviations from their averages.

We use the variable employment (variable e11103 in the German SOEP data set) in three self-reported categories "not working," "part-time working," and "full-time working." This implies that the five grids for the years 2009–2013 consist of three intervals each. We assume the explanatory variables "age (18–75 years of age)," "age-squared," dummy variables for "gender (female = 1)," "marital status: married (reference)," "marital status: single," "marital status: separated," "ln(household income minus individual labor income)," "number of children at home," "years of education," and dummy variables for "years of education unknown" and "living in East-Germany." The latent variable is assumed to be generated by the linear equation

$$\begin{aligned}\text{Employment} = {} & \beta_1 \text{Age} + \beta_2 \text{Age}^2 + \beta_3 \text{Female} + \beta_4 \text{Single} + \beta_5 \text{Separated} + \\ & + \beta_6 \text{Ln}\left(\text{HH.LabourInc}\right) + \beta_7 \text{Children} + \beta_8 \text{Years\_educ} + \\ & + \beta_9 \text{Years\_educ\_unknown} + \beta_{10} \text{East} + \varepsilon \end{aligned} \tag{5.1}$$

As already said, we assume an "unstructured" 5×5 covariance matrix where $\overline{\Sigma} = \left(\overline{\rho}_{k,k'}\right)$. The results are presented in Table 1. In the left-hand panel, we show the in-between error covariance matrix $\widehat{\overline{\Sigma}}$, in the right-hand panel the latent full covariance matrix $\widehat{\Sigma}$ as estimated by the simulation method described in the Appendix. The error correlation over time appears from the right panel to be quite considerable (1.0, 0.8775, 0.7800,...). When we look at the coarsened data the correlation is mitigated by the coarsened observation but still considerable.

Table 1. In-between and full error covariance matrices for FMOP.

	In-between covariance matrix (FMOP)					Full covariance matrix (FMOP)				
	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
2009	0.6303					1.0000				
2010	0.4760	0.6156				0.8775	1.0000			
2011	0.4049	0.4746	0.6149			0.7800	0.8856	1.0000		
2012	0.3695	0.4125	0.4690	0.6131		0.7343	0.7998	0.8837	1.0000	
2013	0.3336	0.3642	0.4059	0.4833	0.6168	0.6867	0.7280	0.7950	0.9079	1.0000

The regression estimates according to FMOP and Ordered Probit (errors independent) are presented in Table 2.

As expected, the regression estimates are of the same order, because both estimators are consistent. The difference is clearly in the calculated standard deviations. All FMOP standard errors are a factor of 1,5 to 2,0 larger than the OP estimates. This is caused because the assumption of error independence by OP instead of the observed strong error correlations is tantamount to a gross exaggeration of the reliability of the data material when we ignore the non-zero error correlations. The difference in standard deviations is a warning signal.

For curiosity we also look at the question of what standard deviations we would have found when we would have had the non-coarsened, that is exact, data material at our disposal. Those standard deviations are estimated by the roots of the diagonal elements of $\frac{1}{N}\left(X'\widehat{\Sigma}^{-1}X\right)^{-1}$ the elements of which are known. The latent error-covariance matrix Σ is estimated by $\widehat{\Sigma}$ according to the method described in the Appendix. We see from comparison that the FMOP-standard deviations (e.g., for the AGE-coefficient 0.0059) on the basis of the coarsened observations are much larger than the corresponding values found from Ordered Probit theoretical values (0.0032) or for GLS-estimation (0.0038) on the exact latent data.

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

Table 2. Regression estimates from FMOP and Ordered Probit.

	Feasible Multivariate Ordered Probit (FMOP)			Exact observ.	Ordered Probit (OP)		
	Number of obs. = 51760			FMOP	Number of obs. = 51760 McF pseudo R^2 = 0.2523 McK-Z R^2 = 0.5396		
EMPLOYMENT	Coeff.	Std.err.	z	Std.err.	Coeff.	Std.err.	z
AGE	0.3137	0.0059	53.50	0.0038	0.3236	0.0032	102.36
AGE square	−0.0037	0.00006	−58.34	0.00004	−0.0038	0.00003	−114.77
D. FEMALE	−0.7464	0.0223	−33.51	0.0180	−0.7727	0.0115	−67.01
D. SINGLE	0.0778	0.0309	2.52	0.0186	0.0541	0.0202	2.68
sD. SEPARATED	0.0538	0.0250	2.15	0.0164	0.1184	0.0175	6.77
Ln(HHLABOURINC)	0.0071	0.0015	4.65	0.0009	0.0150	0.0013	11.31
# of CHILDREN	−0.2051	0.0114	−17.97	0.0074	−0.2571	0.0077	−33.47
YEARS EDUCATION	0.0789	0.0039	20.03	0.0031	0.0742	0.0021	34.97
D. EDUC. unknown	−0.1172	0.0459	−2.55	0.0318	−0.1985	0.0294	−6.75
D. EAST Germany	−0.0542	0.0237	−2.28	0.0180	−0.0420	0.0128	−3.28
Cut point 1	−0.5892	0.0129	−45.79	–	−0.5984	0.0069	−86.68
Cut point 2	0.3569	0.0123	29.05	–	0.3753	0.0066	5.68

Next to our estimate of we also present the Probit R^2 as defined by McFadden (1974). We see that the latter is considerably smaller than 'our' R^2, which follows that of McKenzie-Zavoina (1975).

The computation time in total was 8 seconds. We used a laptop. The computation process can be split up into two parts: the first-stage OP estimation, taking 3 seconds and the second-stage estimation taking another 5 seconds.

We see that employment increases with age until age 42, after which employment decreases (we excluded respondents under 18 years of age and those over 75 years of age). Females are less often employed than males. In households with children, the respondents work less than in childless households. The more additional labor income in the household, the more the respondent works. The more education years one has, the more one works full-time, while respondents from East Germany are less employed than the West Germans.

5.2. Seemingly Unrelated Ordered Probit (SUOP) on a block of eight satisfaction questions

In the German panel questionnaire, we find a number of satisfaction questions referring to various life domains, like those presented in Figure 1. Here, we apply the SUOP method.

This type of questioning is abundantly used in marketing research and happiness research. Another very important instance, where the use of SUOP is at hand, is in the analysis of vignettes, also known as factorial surveys in sociological research or as conjoint analysis, now one of the major tools in psychology and marketing research (Green & Srinivasan, 1978; Atzmüller & Steiner 2010; Wallander, 2009; Van Beek et al., 1997).

The data set consists of about 15,000 observation units. Since the original formulation with 11 answer categories made the coarsened observations look very similar to continuous observations, we further coarsened the data into five response categories (0,1,2), (3,4),. . .,(9,10). In this paper, we apply SUOP analysis to the above-listed block of satisfaction questions with respect to life domains from the 2013

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

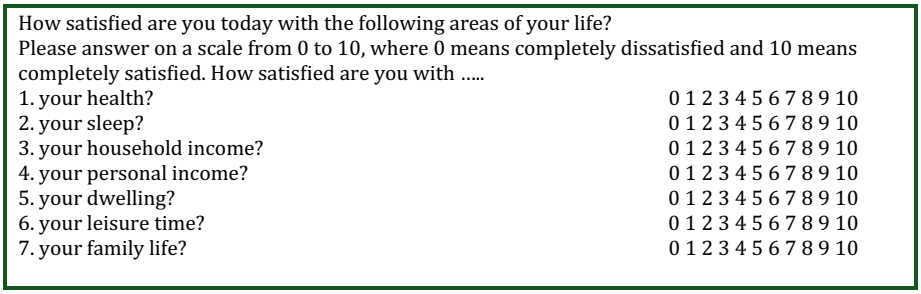

Fig. 1. A block of satisfaction questions with respect to various life domains.

wave of the GSOEP panel. We use the following explanatory variables: age and age-squared, dummies for being female, single and separated, ln(individual labor income), ln(household income *minus* individual labor income), the number of children, the number of years of education, living in East Germany, dummy disability status (0 (no), 1 (yes)), and health rating (1 (bad health),. . .,5 (very good health)). Our primary objective is to demonstrate the feasibility of SUOP. It stands to reason that for a substantive analysis of domain satisfactions, this model specification is probably too simplistic, however, for our objective, testing the feasibility of SUOP, this specific choice is no problem. In order to avoid that every dependent variable would be explained by the same set of explanatory variables we chose different subsets for each equation.

In Table 3 we present the estimates of the first two equations on Health and Sleep satisfaction. For the full table presenting all eight equation estimates we refer to the Appendix.

Table 3. Comparison of the parameter estimates and their s.e.'s for Ordered Probit, Method of Moments, and Maximum Likelihood.

McK-Z R^2	0.1489		0.1481		0.1154	
Health Satisfaction	β_{OP}	σ_{OP}	β_{SUOP}	σ_{SUOP}	β_{ML}	σ_{ML}
Health: AGE	−0.0695	0.0046	−0.0693	0.0048	−0.0651	0.0045
Health: AGE^2	0.0006	0.00005	0.0006	0.00005	0.0006	0.00005
Health: D_FEMALE	−0.0182	0.0175	−0.0181	0.0174	−0.0134	0.0166
Health: D_SINGLE	−0.0514	0.0291	−0.0510	0.0293	−0.0570	0.0278
Health: D_SEPARATED	−0.1241	0.0240	−0.1238	0.0243	−0.1108	0.0230
Health: Ln_LABOURINC	0.0289	0.0025	0.0288	0.0026	0.0227	0.0022
Health: CHILDREN	0.0357	0.0123	0.0356	0.0126	0.0308	0.0114
Health: D_EAST	−0.1678	0.0201	−0.1670	0.0196	−0.1402	0.0193
Health: DISABLE	−0.7991	0.0272	−0.7971	0.0268	−0.5825	0.0232
Health: Cut point 1	−1.8277	0.0184	−1.8240	0.0191	−1.8163	0.0181
Health: Cut point 2	−1.1161	0.0131	−1.1094	0.0134	−1.1122	0.0130
Health: Cut point 3	−0.3432	0.0109	−0.3415	0.0109	−0.3399	0.0107
Health: Cut point 4	0.9468	0.0123	0.9442	0.0124	0.9409	0.0122

(Continued)

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

Table 3. Continued

McK-Z R^2	0.0278		0.0278		0.0196	
Sleep Satisfaction	β_{OP}	σ_{OP}	β_{SUOP}	σ_{SUOP}	β_{ML}	σ_{ML}
Sleep: AGE	−0.0380	0.0041	−0.0381	0.0042	−0.0348	0.0041
Sleep: AGE^2	0.0004	0.00004	0.0004	0.00004	0.0003	0.00004
Sleep: D. FEMALE	−0.1213	0.0172	−0.1213	0.0172	−0.1204	0.0164
Sleep: D. SINGLE	0.0508	0.0302	0.0505	0.0307	0.0594	0.0284
Sleep: D. SEPARATED	−0.0762	0.0253	−0.0762	0.0260	−0.0658	0.0238
Sleep: Ln(HH.LABOURINC)	0.0067	0.0021	0.0066	0.0020	0.0062	0.0018
Sleep: # of CHILDREN	0.0219	0.0121	0.0220	0.0122	0.0164	0.0113
Sleep: YEARS of EDUCATION	0.0316	0.0032	0.0315	0.0031	0.0117	0.0028
Sleep: D. EAST GERMANY	−0.1063	0.0199	−0.1065	0.0196	−0.0763	0.0192
Sleep: Cut point 1	−1.6918	0.0175	−1.6912	0.0174	−1.7036	0.0174
Sleep: Cut point 2	−0.9791	0.0123	−0.9800	0.0123	−0.9831	0.0123
Sleep: Cut point 3	−0.9251	0.0106	−0.2970	0.0106	−0.2940	0.0104
Sleep: Cut point 4	0.7286	0.0114	0.7269	0.0115	0.7150	0.0113

In the first two columns we present the initial Probit estimates and their s.e.'s. In columns 3, 4 we present the corresponding SUOP-estimates and their s.e's. The two right-hand columns 5, 6 give the corresponding estimates by means of the ML-method. We take the `cmp` results as the touchstone of our comparison.

Our first conclusion is that the three methods OP, SUOP, ML yield estimates which do not differ significantly in most cases. This is not surprising as the three estimators are consistent. The standard deviations of the SUOP-estimators seem to be slightly larger than the ML-estimators, but the differences are mostly negligible.

In Table 4 we present the full correlation matrices as estimated by SUOP (estimation according to Appendix) and ML (according to Stata), respectively.

Table 4. Full error correlation matrices compared for SUOP and ML.

Residual								Stand.
Corr. SUOP	Health	Sleep	HH inc.	Ind inc.	Dwelling	Leisure	Family life	living
Health	1.0000							
Sleep	0.5328	1.0000						
HH. inc.	0.3196	0.2758	1.0000					
Ind. inc.	0.2851	0.2367	0.8414	1.0000				
Dwelling	0.3103	0.2894	0.4521	0.3769	1.0000			
Leisure	0.3246	0.3359	0.3274	0.2812	0.4557	1.0000		
Family life	0.3585	0.3295	0.3327	0.2838	0.4526	0.4743	1.0000	
Stand. living	0.3964	0.2520	0.7173	0.6044	0.5551	0.4626	0.5930	1.0000

(Continued)

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

Table 4. Continued

Residual Corr. ML	Health	Sleep	HH inc.	Ind inc.	Dwell ing	Leisure	Family life	Stand. living
Health	1.0000							
Sleep	0.5487	1.0000						
HH. inc.	0.3280	0.2802	1.0000					
Ind. inc.	0.2919	0.2487	0.8266	1.0000				
Dwelling	0.3144	0.2990	0.4530	0.3869	1.0000			
Leisure	0.3301	0.3463	0.3323	0.2899	0.4635	1.0000		
Family life	0.3643	0.3405	0.3355	0.2849	0.4563	0.4800	1.0000	
Stand. living	0.3973	0.3492	0.7099	0.6118	0.5582	0.4610	0.5930	1.0000

We see that of the 75 SUOP-estimated regression coefficients 17 fall out of the ML-confidence intervals. For the estimates of the correlation matrix we find a similar result.[6] Three of the 28 SUOP estimated correlation coefficients are just outside the ML-confidence intervals.

The polychoric correlation matrix is presented in Table 5.

Table 5. The polychoric correlation matrix.

Residual Corr. SUOP	Health	Sleep	HH inc.	Ind inc.	Dwelling	Leisure	Family life	Stand. living
Health	1.0000							
Sleep	0.7372	1.0000						
HH. inc.	0.7626	0.6347	1.0000					
Ind. inc.	0.7389	0.6028	0.9418	1.0000				
Dwelling	0.6985	0.5998	0.7752	0.7338	1.0000			
Leisure	0.6962	0.6258	0.7111	0.6704	0.7287	1.0000		
Family life	0.7496	0.6372	0.7550	0.7101	0.7558	0.7521	1.0000	
Stand. living	0.8015	0.6728	0.9220	0.8705	0.8169	0.7680	0.8538	1.0000

A naïve approach is to assign the values 0,1,…,9,10 to the satisfaction values and to calculate the Pearson correlations on that basis. This assignment is conform to daily usage, where average satisfaction values in a sample are also based on this assignment practice.

The Pearson correlations are presented in Table 6. We see that there is a considerable difference between Tables 5 and 6. We prefer 5 to 6, as the cardinalization by 0,1..,10 is arbitrary and might be replaced by another one (0,2,3,…) yielding a different Table 6, while the polyserial correlations are based on endogenous cardinalization. We notice that all Pearson correlations in Table 6 are considerably smaller than the corresponding numbers in Table 5.

[6]The cmp-procedure provides confidence intervals for the correlation estimates.

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

Table 6. The Pearson correlation matrix. (scale 0–10).

Pearsoncorr. scale (0–10)	Health	Sleep	HH inc.	Ind inc.	Dwelling	Leisure	Family life	Stand. living
Health	1.0000							
Sleep	0.5185	1.0000						
HH. inc.	0.3331	0.2829	1.0000					
Ind. inc.	0.2899	0.2575	0.7761	1.0000				
Dwelling	0.2414	0.2635	0.4307	0.3580	1.0000			
Leisure	0.2121	0.2804	0.2921	0.2448	0.4047	1.0000		
Family life	0.3027	0.2965	0.3331	0.2599	0.4085	0.3985	1.0000	
Stand. living	0.3852	0.3325	0.7112	0.5872	0.5201	0.3806	0.5387	1.0000

The computation process can be split up into three parts: the first-stage OP estimation took 1.8 seconds on our ASUS VivoBook 15 laptop, and the second-stage SUOP estimation took another 111 seconds. The whole calculation requires less than two minutes. The ML estimations by `cmp` (default) took 7 hours. In the SUOP method, we use the in-between covariance matrix $\widehat{\widehat{\Sigma}}$ and not the full covariance estimate. The computation time depends on the sample size N, the size K of the error covariance matrix, and the capacity of our laptop. In this example $K = 8$. We see that for the ML method, the time increases non-linearly with N. The number K seems to be important as well. For $K = 2$ equations both methods are roughly equally fast, where SUOP takes 11 seconds and ML-cmp 10.5 seconds for $N = 15535$. For $K = 3$ the SUOP computation increases to 16 seconds, while the ML method requires already 1,241 seconds. This is caused, it seems, by the fact that ML has to evaluate a lot of K-dimensional integrals. A colleague of ours (an expert Stata user) observed, quoting the "options" in the Stata text, that `cmp` uses the GHK-simulation method for evaluating the needed integrals and that in the default option cmp uses $2\sqrt{N}$ draws per evaluated likelihood. In the present case, this is about 250 simulations per observation. Cappellari and Jenkins (2003) suggested that for a large number of observations, the number of draws can be considerably reduced without severe efficiency loss. According to our colleague by taking 5 draws per likelihood we would reduce the computation time from the reported 7 hours to 8 minutes with only a slight efficiency reduction. That is probably still significantly slower than the new method, but the revision would be material. We followed this suggestion and found indeed comparable estimates for the coefficients β. To our surprise the standard deviations for the five draws were not significantly different from the 250 draws version. This seems to indicate that in the assessment of variance the additional contribution caused by the simulation variance is not taken into account.

Clearly, if we would reduce the number of equations from eight to a more manageable four or two equations and/or reduce the number of observations, both the ML and the SUOP methods would perform much faster.

Our conclusion is that the SUOP method is faster than the ML method. We are unable to say whether the Stata procedure cmp is to blame and could be improved or whether this is a general feature of the ML-GHK procedure. It might also be that we could have reduced the ML computation time by choosing specific options instead of the default procedure. Choosing too severe tolerance levels for the iterations involved would have increased the computation time in exchange for more exact confidence intervals. However, given that the `cmp` outcomes have about the same confidence intervals as our SUOP outcomes we do not believe that the tolerance levels chosen in `cmp` were more severe than in our method (see Table 8f).

5.3. A simulated example

Finally, we apply the estimation method to a simulated data set. We simulated a hard data set of 10,000 observations. We generated the set as follows. We assumed a latent model

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

$$
\begin{aligned}
y_{i,1} &= x_{i,1} + \quad x_{i,2} + \quad x_{i,3} + x_{i,4} + \varepsilon_{i,1} \\
y_{i,2} &= x_{i,1} + 2x_{i,2} + 3x_{i,3} + 4x_{i,4} + \varepsilon_{i,2} \\
y_{i,3} &= -x_{i,1} + 2x_{i,2} - 3x_{i,3} + 4x_{i,4} + \varepsilon_{i,3} \\
y_{i,4} &= -x_{i,1} + 0.5x_{i,2} - \quad x_{i,3} + x_{i,4} + \varepsilon_{i,4},
\end{aligned}
$$

where, x_1 is normal $N(0,1)$, where $x_2 = 0.5x_1 + D1$ with $D1$ a dummy variable equal to $+1$ or -1 with 50% each, where $x_3 = N(0,2) + 0.5x_2$, and where $x_4 = -0.5x_3 + D2$ and $D2$ is drawn to equal $+1$, 0, or -1 with a chance of $1/3$ each.

The error vector ε is i.i.d. $\mathrm{N}(0,\Sigma)$ with

$$
\Sigma = \begin{pmatrix}
1.0 & & & \\
0.5 & 1.0 & & \\
-0.5 & 0.3 & 1.0 & \\
0.2 & 0.6 & -0.1 & 1.0
\end{pmatrix}.
$$

We notice that all four variables x and the error vector has an expectation equal to zero. In order to avoid that (the non-conditioned) Y_i is approximately normal, we restricted the explanatory variables to a small number of four and we chose those variables to be non-normal and correlated, such that the structural part $\beta' x$ does not tend to normality. Our first aim is to look for the distribution of the exact data. The expectation $E(Y) = 0$, the empirical mean equals 0.0160 and the variance $\mathrm{var}(Y)$ equals 2.603. The correlation matrix of the variables X, and Y is the 8×8 matrix in Table 1.

Cut points are defined as
$$
\begin{aligned}
&v_{1,1} = 0 \\
&v_{2,1} = -1, \quad v_{2,2} = 1.5 \\
&v_{3,1} = -1, \quad v_{3,2} = 0.5 \\
&v_{4,1} = -1.5, v_{4,2} = -0.5, v_{4,3} = 1
\end{aligned}
$$

We define the response indicators:
$$
\begin{aligned}
j_{i,1} &= 1,2 \\
j_{i,2} &= 1,2,3 \\
j_{i,3} &= 1,2,3 \\
j_{i,4} &= 1,2,3,4
\end{aligned}
$$

The model is iteratively estimated by the FMOP method. $j_{i,k}$ is the interval index by respondent i for equation k, corresponding with the four equations $k = 1,\ldots,4$. We start with iteration $t = 0$ for $\beta = 0$. We define the under- and upper residuals

$$
E\left(\varepsilon \middle| \varepsilon \le v_{j_{i,k}}^{(t)} - \beta_k^{(t)\prime} x_{i,k}\right) = \underset{\sim}{e}_{j_{i,k}}^{(t)} = \frac{-\varphi\left(v_{j_{i,k}}^{(t)} - \beta_k^{(t)\prime} x_{i,k}\right)}{\Phi\left(v_{j_{i,k}}^{(t)} - \beta_k^{(t)\prime} x_{i,k}\right)}
$$

$$
E\left(\varepsilon \middle| \varepsilon > v_{j_{i,k}}^{(t)} - \beta_k^{(t)\prime} x_{i,k}\right) = \tilde{e}_{j_{i,k}}^{(t)} = \frac{\varphi\left(v_{j_{i,k}}^{(t)} - \beta_k^{(t)\prime} x_{i,k}\right)}{1 - \Phi\left(v_{j_{i,k}}^{(t)} - \beta_k^{(t)\prime} x_{i,k}\right)}
$$

We define the sets of respondents $S_{k,j}^1, S_{k,j}^2$ ($k = 1,\ldots 4$; $j = 1,\ldots,J_k$) who are in the response categories $\le j$ or $> j$respectively. We solve the equations

$$
\sum_{i \in S_{k,j}^1} \underset{\sim}{e}_{j_{i,k}}^{(t)} + \sum_{i \in S_{k,j}^2} \tilde{e}_{j_{i,k}}^{(t)} = 0, k = 1,\ldots,4, j = 1,\ldots,J_k \tag{5c.1}
$$

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

for $v_{k,j}^{(t)}$ and find estimated cut-points $v_{k,j}^{(t)}$ in the t^{th} iteration. These cut-points $v_{k,j}^{(t)}$ are substituted to define the generalized residuals. The estimated generalized residuals $\overline{e}_{j_{i,k}}^{(t)}$ in the t^{th} iteration are

$$\overline{e}_{j_{i,k}}^{(t)} = \frac{\varphi\left(v_{j_{i,k}-1}^{(t)} - \beta_k{}' x_{i,k}\right) - \varphi\left(v_{jik}^{(t)} - \beta_k{}' x_{i,k}\right)}{\Phi\left(v_{j_{i,k}-1}^{(t)} - \beta_k{}' x_{i,k}\right) - \Phi\left(v_{j_{i,k}-1}^{(t)} - \beta_k{}' x_{i,k}\right)},$$

where $j_{i,k}$ is the interval index by respondent i for equation k. corresponding with the four equations $k = 1,\ldots,4$. We now define the $(K \times M)$ orthogonality conditions

$$\frac{1}{N}\sum_{i=1}^{N} x_{i,k,m}\overline{e}_{j_{i,k}}^{(t)} = 0, k = 1,\ldots,4, m = 1,\ldots,M \tag{5c.2}$$

We have now two equation systems (5c.1) and (5c.2), which are simultaneously solved. Then we calculate the in-between error covariance matrix

$$\overline{\Sigma}^{(t)} = \left(\overline{\sigma}_{k,k'}\right) = \frac{1}{N}\left(\sum_{i=1}^{N}\overline{e}_{j_{i,k}}^{(t)}\overline{e}_{j_{i,k'}}^{(t)}\right).$$

We repeat (5c.1) and (5c.2) with the new $\beta^{(t+1)}$ and $v_{k,j}^{(t+1)}$, and find new estimates. We repeat (5c.1) and (5c.2) after weighting with the inverse covariance matrix solving

$$\sum_{i=1}^{N}\overline{e}_{j_{i,k}}^{(t)}\left[\overline{\Sigma}^{(t)}\right]^{-1} x_{i,k} = 0.$$

In the end we estimate the corresponding covariance matrix of the estimators $\widehat{\widehat{\beta}},\widehat{\widehat{v}}$ by the well-known sandwich formula.

We estimate each non-diagonal element $\sigma_{k,k'}$ of the latent full covariance matrix from the corresponding element $\overline{\sigma}_{k,k'}$ of the in-between error covariance matrix. The method is described in detail in the Appendix.

We conclude that the method is stable in the number *N of observations* and it does not differ significantly from the `cmp`-estimates.

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

Table 7. Beta's, Standard errors and Error correlations (N = 10,000).

N = 10000	beta's MoM	s.e. MoM	beta's cmp	s.e. cmp
Eq.y1: var. x1	0.9794	0.0264	0.9799	0.0269
Eq.y1: var. x2	0.9910	0.0240	0.9927	0.0241
Eq.y1: var. x3	0.9715	0.0222	0.9715	0.0223
Eq.y1: var. x4	0.9694	0.0281	0.9703	0.0282
Eq.y1: cutp. 1	−0.0024	0.0193	−0.0072	0.0191
Eq.y2: var. x1	1.0329	0.0290	1.0557	0.0279
Eq.y2: var. x2	1.9859	0.0396	1.9991	0.0387
Eq.y2: var. x3	3.0286	0.0550	3.0558	0.0535
Eq.y2: var. x4	4.0158	0.0733	4.0526	0.0711
Eq.y2: cutp. 1	−1.0215	0.0321	−1.0154	0.0311
Eq.y2: cutp. 2	1.4972	0.0386	1.5026	0.0358
Eq.y3: var. x1	−1.0234	0.0444	−1.0065	0.0406
Eq.y3: var. x2	2.0183	0.0640	1.9981	0.0586
Eq.y3: var. x3	−3.0813	0.0847	−3.0441	0.0805
Eq.y3: var. x4	4.1387	0.1152	4.0610	0.1097
Eq.y3: cutp. 1	−0.9753	0.0447	−0.9675	0.0428
Eq.y3: cutp. 2	0.5115	0.0407	0.4993	0.0387
Eq.y4: var. x1	−1.0079	0.0212	−1.0078	0.0209
Eq.y4: var. x2	0.4957	0.0169	0.4931	0.0165
Eq.y4: var. x3	−1.0180	0.0173	−1.0143	0.0171
Eq.y4: var. x4	0.9727	0.0224	0.9696	0.0218
Eq.y4: cutp. 1	−1.5233	0.0282	−1.5198	0.0273
Eq.y4: cutp. 2	−0.5125	0.0233	−0.5165	0.0223
Eq.y4: cutp. 3	1.0335	0.0253	1.0289	0.0245

Estimated Full correlation matrix MoM				Estimated Full correlation matrix cmp			
1.0000				1.0000			
0.4496	1.0000			0.4703	1.0000		
−0.5450	0.2527	1.0000		−0.4841	0.2306	1.0000	
0.1934	0.5877	−0.0846	1.0000	0.2110	0.6266	−0.0857	1.0000

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

Table 8a. Beta's, Standard errors and Error correlations (N = 5,000) (Base dataset of 10000 cases, only every second case is used).

N = 5000	beta's MoM	s.e. MoM	beta's cmp	s.e. cmp
Eq.y1: var. x1	1.0192	0.0385	1.0161	0.0392
Eq.y1: var. x2	0.9952	0.0341	0.9949	0.0344
Eq.y1: var. x3	0.9598	0.0319	0.9573	0.0312
Eq.y1: var. x4	0.9484	0.0398	0.9460	0.0397
Eq.y1: cutp. 1	−0.0330	0.0273	−0.0361	0.0270
Eq.y2: var. x1	1.0832	0.0418	1.0969	0.0402
Eq.y2: var. x2	1.9317	0.0552	1.9383	0.0534
Eq.y2: var. x3	3.0125	0.0790	3.0381	0.0749
Eq.y2: var. x4	4.0011	0.1053	4.0333	0.0997
Eq.y2: cutp. 1	−1.0462	0.0462	−1.0508	0.0442
Eq.y2: cutp. 2	1.4679	0.0542	1.4743	0.0501
Eq.y3: var. x1	−1.0418	0.0636	−1.0452	0.0567
Eq.y3: var. x2	1.9988	0.0847	2.0070	0.0807
Eq.y3: var. x3	−3.0718	0.1113	−3.0620	0.1128
Eq.y3: var. x4	4.1625	0.1512	4.1297	0.1548
Eq.y3: cutp. 1	−0.9182	0.0619	−0.8988	0.0598
Eq.y3: cutp. 2	0.6152	0.0562	0.6031	0.0554
Eq.y4: var. x1	−0.9814	0.0300	−0.9829	0.0292
Eq.y4: var. x2	0.4882	0.0238	0.4860	0.0233
Eq.y4: var. x3	−1.0281	0.0246	−1.0261	0.0244
Eq.y4: var. x4	0.9506	0.0322	0.9492	0.0306
Eq.y4: cutp. 1	−1.5350	0.0398	−1.5408	0.0390
Eq.y4: cutp. 2	−0.5381	0.0327	−0.5369	0.0321
Eq.y4: cutp. 3	1.0050	0.0362	1.0132	0.0343

Estimated Full correlation matrix MoM				Estimated Full correlation matrix cmp			
1.0000				1.0000			
0.5288	1.0000			0.4930	1.0000		
−0.3185	0.3089	1.0000		−0.3905	0.3245	1.0000	
0.3014	0.6468	−0.1165	1.0000	0.2201	0.6138	−0.0981	1.0000

0.3014: 95% confidence interval cmp [0.1462: 0.2916]

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

Table 8b. Beta's, Standard errors and Error correlations (N = 2,000) (Base dataset of 10000 cases, only every fifth case is used).

N = 2000	beta's MoM	s.e. MoM	beta's cmp	s.e. cmp
Eq.y1: var. x1	0.9929	0.0574	0.9814	0.0590
Eq.y1: var. x2	0.9259	0.0507	0.9186	0.0509
Eq.y1: var. x3	0.9448	0.0512	0.9392	0.0483
Eq.y1: var. x4	0.9605	0.0628	0.9593	0.0609
Eq.y1: cutp. 1	−0.0165	0.0418	−0.0081	0.0414
Eq.y2: var. x1	0.9729	0.0643	0.9747	0.0588
Eq.y2: var. x2	1.9411	0.0822	1.9198	0.0827
Eq.y2: var. x3	2.9799	0.1168	2.9654	0.1138
Eq.y2: var. x4	3.9235	0.1548	3.9077	0.1502
Eq.y2: cutp. 1	−1.0068	0.0705	−0.9718	0.0677
Eq.y2: cutp. 2	1.4167	0.0828	1.4069	0.0746
Eq.y3: var. x1	−1.0846	0.0992	−1.0725	0.0990
Eq.y3: var. x2	2.1905	0.1320	2.2170	0.1474
Eq.y3: var. x3	−3.3765	0.1985	−3.4092	0.2109
Eq.y3: var. x4	4.4998	0.2566	4.5139	0.2813
Eq.y3: cutp. 1	−1.2270	0.1100	−1.2034	0.1120
Eq.y3: cutp. 2	0.5252	0.0976	0.5239	0.0939
Eq.y4: var. x1	−0.9667	0.0465	−0.9641	0.0458
Eq.y4: var. x2	0.4765	0.0372	0.4680	0.0362
Eq.y4: var. x3	−1.0109	0.0378	−0.9994	0.0386
Eq.y4: var. x4	0.9413	0.0511	0.9278	0.0472
Eq.y4: cutp. 1	−1.5391	0.0629	−1.5537	0.0605
Eq.y4: cutp. 2	−0.5099	0.0509	−0.5206	0.0493
Eq.y4: cutp. 3	0.9616	0.0541	0.9675	0.0527

Estimated Full correlation matrix MoM				Estimated Full correlation matrix cmp			
1.0000				1.0000			
0.5982	1.0000			0.5334	1.0000		
−0.2635	0.3972	1.0000		−0.3878	0.2402	1.0000	
0.3114	0.6242	0.0889	1.0000	0.2159	0.6098	−0.0263	1.0000

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

Table 8c. Beta's, Standard errors and Error correlations (N = 1,000) (Base dataset of 10000 cases, only every tenth case is used).

N = 1000	beta's MoM	s.e. MoM	beta's cmp	s.e. cmp
Eq.y1: var. x1	0.9058	0.0798	0.8950	0.0820
Eq.y1: var. x2	0.9667	0.0719	0.9531	0.0718
Eq.y1: var. x3	0.8916	0.0726	0.8826	0.0651
Eq.y1: var. x4	0.9127	0.0880	0.9065	0.0839
Eq.y1: cutp. 1	0.0109	0.0581	0.0193	0.0578
Eq.y2: var. x1	1.0706	0.0998	1.0462	0.0885
Eq.y2: var. x2	1.8804	0.1089	1.8831	0.1162
Eq.y2: var. x3	2.9807	0.1617	2.9809	0.1636
Eq.y2: var. x4	3.9341	0.2122	3.9386	0.2157
Eq.y2: cutp. 1	−1.0250	0.0946	−1.0196	0.0955
Eq.y2: cutp. 2	1.3482	0.1172	1.3420	0.1055
Eq.y3: var. x1	−1.1157	0.1235	−1.1258	0.1419
Eq.y3: var. x2	2.1102	0.1724	2.1565	0.1987
Eq.y3: var. x3	−3.2868	0.2639	−3.3482	0.2912
Eq.y3: var. x4	4.4391	0.3299	4.5149	0.3847
Eq.y3: cutp. 1	−1.1548	0.1573	−1.1448	0.1512
Eq.y3: cutp. 2	0.5539	0.1232	0.5435	0.1322
Eq.y4: var. x1	−1.0006	0.0658	−1.0014	0.0664
Eq.y4: var. x2	0.4593	0.0526	0.4538	0.0507
Eq.y4: var. x3	−1.0533	0.0529	−1.0494	0.0562
Eq.y4: var. x4	0.9190	0.0742	0.9095	0.0665
Eq.y4: cutp. 1	−1.6660	0.0938	−1.7127	0.0892
Eq.y4: cutp. 2	−0.5368	0.0729	−0.5796	0.0724
Eq.y4: cutp. 3	1.0163	0.0762	1.0410	0.0778

Estimated Full correlation matrix MoM				Estimated Full correlation matrix cmp			
1.0000				1.0000			
0.7347	1.0000			0.5360	1.0000		
−0.3041	0.0570	1.0000		−0.3291	0.2410	1.0000	
0.1500	0.6412	−0.1904	1.0000	0.2013	0.5261	−0.1288	1.0000

0.7347: 95% confidence interval cmp [0.3862: 0.6583]

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

Table 8d. Beta's, Standard errors and Error correlations (N = 1,000) (A newly created dataset of 1000 cases).

N = 1000	beta's MoM	s.e. MoM	beta's cmp	s.e. cmp
Eq.y1: var. x1	0.9490	0.0846	0.9536	0.0827
Eq.y1: var. x2	0.9997	0.0726	1.0142	0.0747
Eq.y1: var. x3	0.9368	0.0702	0.9446	0.0696
Eq.y1: var. x4	0.9916	0.0871	0.9944	0.0908
Eq.y1: cutp. 1	0.0282	0.0604	0.0380	0.0599
Eq.y2: var. x1	0.8773	0.0956	0.8964	0.0795
Eq.y2: var. x2	1.9966	0.1537	1.9868	0.1219
Eq.y2: var. x3	2.9308	0.2170	2.9092	0.1597
Eq.y2: var. x4	3.8298	0.2892	3.8065	0.2112
Eq.y2: cutp. 1	−0.9652	0.0989	−0.9571	0.0952
Eq.y2: cutp. 2	1.2944	0.1303	1.3508	0.1049
Eq.y3: var. x1	−1.0870	0.1472	−1.0583	0.1484
Eq.y3: var. x2	2.1040	0.2332	2.1304	0.2109
Eq.y3: var. x3	−3.2458	0.2952	−3.2709	0.2838
Eq.y3: var. x4	4.2512	0.4029	4.3088	0.3815
Eq.y3: cutp. 1	−1.0302	0.1601	−1.0904	0.1513
Eq.y3: cutp. 2	0.6065	0.1224	0.5906	0.1338
Eq.y4: var. x1	−0.9638	0.0650	−0.9645	0.0649
Eq.y4: var. x2	0.4585	0.0521	0.4732	0.0518
Eq.y4: var. x3	−0.8987	0.0531	−0.9015	0.0500
Eq.y4: var. x4	0.9384	0.0712	0.9204	0.0672
Eq.y4: cutp. 1	−1.4835	0.0854	−1.4945	0.0842
Eq.y4: cutp. 2	−0.5339	0.0690	−0.5255	0.0680
Eq.y4: cutp. 3	0.8221	0.0821	0.8277	0.0699

Estimated Full correlation matrix MoM				Estimated Full correlation matrix cmp			
1.0000				1.0000			
0.5582	1.0000			0.3511	1.0000		
−0.5354	0.4227	1.0000		−0.5507	0.1285	1.0000	
0.3158	0.7640	−0.3565	1.0000	0.1896	0.6688	−0.1589	1.0000

0.5582: 95% confidence interval cmp [0.1834: 0.4989]
0.4227: 95% confidence interval cmp [−0.1869: 0.4199]

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

Table 8e. Beta's, Standard errors and Error correlations (N = 1,000) (Again a newly created dataset of 1000 cases).

	beta's MoM	s.e. MoM	beta's cmp	s.e. cmp
Eq.y1: var. x1	0.9592	0.0809	0.9417	0.0800
Eq.y1: var. x2	0.9477	0.0738	0.9352	0.0708
Eq.y1: var. x3	0.9064	0.0648	0.8982	0.0652
Eq.y1: var. x4	0.9397	0.0831	0.9374	0.0833
Eq.y1: cutp. 1	0.0766	0.0588	0.0701	0.0577
Eq.y2: var. x1	0.9359	0.0837	0.9287	0.0860
Eq.y2: var. x2	2.0244	0.1159	2.0294	0.1236
Eq.y2: var. x3	2.8800	0.1441	2.8982	0.1618
Eq.y2: var. x4	3.8478	0.1961	3.8657	0.2161
Eq.y2: cutp. 1	−0.9785	0.0996	−0.9942	0.0962
Eq.y2: cutp. 2	1.4898	0.1128	1.4802	0.1161
Eq.y3: var. x1	−0.8380	0.1291	−0.8675	0.1328
Eq.y3: var. x2	1.8184	0.1703	1.8429	0.1721
Eq.y3: var. x3	−2.8861	0.2167	−2.8964	0.2404
Eq.y3: var. x4	3.8474	0.2721	3.8491	0.3136
Eq.y3: cutp. 1	−1.0368	0.1240	−1.0327	0.1390
Eq.y3: cutp. 2	0.3357	0.1239	0.3267	0.1197
Eq.y4: var. x1	−1.0838	0.0679	−1.0809	0.0705
Eq.y4: var. x2	0.5400	0.0531	0.5334	0.0530
Eq.y4: var. x3	−1.0304	0.0550	−1.0243	0.0558
Eq.y4: var. x4	1.0777	0.0685	1.0711	0.0698
Eq.y4: cutp. 1	−1.5268	0.0823	−1.5160	0.0856
Eq.y4: cutp. 2	−0.4549	0.0726	−0.4652	0.0695
Eq.y4: cutp. 3	1.0477	0.0764	1.0275	0.0767

Estimated Full correlation matrix MoM				Estimated Full correlation matrix cmp			
1.0000				1.0000			
0.6303	1.0000			0.6316	1.0000		
−0.1647	0.5655	1.0000		−0.1230	0.2274	1.0000	
0.1451	0.6184	0.0078	1.0000	0.1891	0.5982	0.0324	1.0000

(0.5655: 95% confidence interval cmp [−0.1068: 0.5155)

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

Table 8f. The correlation matrix of the variables (N = 10,000).

	x1	x2	x3	x4	y1	y2	y3	y4
x1	1.0000							
x2	0.8797	1.0000						
x3	−0.3812	−0.2516	1.0000					
x4	−0.4803	−0.3452	0.9240	1.0000				
y1	0.6278	0.4450	−0.1164	−0.3281	1.0000			
y2	0.7033	0.6220	−0.0943	−0.2136	0.4566	1.0000		
y3	0.5598	0.5484	−0.9312	−0.8929	0.1250	0.2659	1.0000	
y4	−0.2500	−0.0479	0.9196	0.8447	−0.0921	−0.2116	−0.7871	1.0000

6. Concluding remarks

In this paper, we suggest a new approach to the statistical analysis of ordinal data, where the errors are supposed to be correlated. The basic idea is that the ordinally observed dependent variables reflect latent continuously-valued random variables Y and that the observations are coarsened corresponding to an interval grid $\mathbb{C} = \{(v_{j-1}, v_j]\}_{j=1}^{J}$ of Y on the real axis, where the unknown cut-points v have to be estimated as well. Each observed category corresponds to one interval on the real axis. For cases where Y is more dimensional, say K, and errors are correlated there is mostly a formidable impediment. The usual ML-estimation procedure requires to evaluate likelihoods, which are multivariate normal integrals over K-dimensional blocks. If this has to be performed this is very cumbersome and time - consuming. In this paper we show that estimation of the latent generating model behind the coarsened observations of dependent variables can be done in a much simpler way than usual without the need for multi-dimensional integration or large-scale simulations.

In our approach, we depart from the requirement that the difference between the observation Y and its predictor $\beta' X$, that is $\left(Y - \beta' X\right)$, cannot be further explained by X. This is translated into the zero-covariance conditions (2.5) and gives those conditions a significance on their own. When the errors are normally distributed this coincides with the ML-conditions.

The identifying moment conditions are found by substituting the residuals in the regular zero covariance-conditions for exact data by the corresponding generalized residuals corresponding to the ordinal data.

The approach closely resembles the traditional GLS- and SUR-approaches used to estimate linear models on exactly observed dependent variables.

For this method, an assumption about the marginal distributions of the error vector is required. We choose for normality, which enables us to use (3.6). Although we restricted ourselves to assuming normal errors, it is not difficult to generalize this method for other error distributions as well, where the logistic and the lognormal are the foremost candidates (see in those cases, e.g., Maddala (1983, p. 369) for the formulae of the generalized residuals). The estimation method remains unchanged.

This approach seems to smoothly close the gap between the analysis of exactly observable data and qualitative ordinal data. We saw in the above examples that the effect on variances (confidence bands and intervals) caused by SUOP compared to OP is in some cases small and in other cases large. The regression coefficients are mostly similar, which is not surprising as both estimators are consistent estimators. This is also the case for the comparison between traditional OLS and SUR estimates in traditional econometrics. The advantage lies in the possibility to account for error correlations, caused by using the additional information supplied by the error correlations. Standard error deviations are assessed without assuming a specific structure of the covariance matrix before estimation. In the panel data example in Section 5.1 it appears that the standard deviations of the estimates are doubled or more

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

taking error correlation into account. Hence, in this case the reliability reduction when taking error correlations into account is huge.

In this paper, we focus on the qualitative versions of FMOP and SUOP of FGLS and SUR. However, this method seems generally appropriate for two broad types of model estimation situations characterized by Roodman (2011) as:

> "1) those in which a truly recursive data-generating process is posited and fully modeled, and
> 2) those in which there is simultaneity but instruments allow the construction of a recursive set of equations, as in two-stage least squares (2SLS)."

Our method may be compared with the methods, (based on the GHK algorithm), developed by Cappellari and Jenkins (2003), based on simulated moments (Hajivassiliou & McFadden, 1998; Roodman, 2007, 2020). Those methods aim at getting numerical estimates of the log-likelihood by simulation and finding, by variation of the unknown parameters to be estimated, which parameter values maximize the simulated log-likelihood of the sample. This requires the repeated evaluation of multiple normal integrals and makes the procedure time-consuming. In our approach, we do not need to evaluate multiple integrals or large-scale simulations, and therefore the method is not restricted with respect to the size K of the equation system. Moreover, we can handle an arbitrary number $J\,(>2)$ of outcome categories. We do not have to restrict ourselves to dichotomous (biprobit) data only. The method can be used for any number of equations K and any number of interval categories J_k. For instance, in our SUOP-example (Section 5.2) we estimated eight equations, 75 effects and 32 cut points simultaneously. It is obvious that direct observation is a limiting case of coarsening and consequently the methos may also be used when the data set consists of a mixture of directly observed and ordinal data. Classical least-squares based estimation methods on exactly observed data may be seen as a specific limiting case.

Our estimation method appears to require only a few minutes of computing time, which compares favorably with the traditional methods each of which requires much more time. The method may be interpreted as a generalization of classical least squares models that deal with exact observations to include the estimation of models on the basis of more-dimensional ordered probit-type observations.

In this paper, we restricted ourselves to the most straightforward OP observation mode. In a forthcoming study, we will generalize this approach to tackle the case where the sample consists of a mixture of categorical, censored, and exactly observed data.

Acknowledgments. We are grateful for helpful remarks by the referees.

References

Aitkin, M. A. (1964). Correlation in a singly truncated bivariate normal distribution. *Psychometrika*, *29*, 263–270.

Atzmüller, C., & Steiner, P.M. (2010). Experimental vignette studies in survey research. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *6*, 128–138.

Amemiya, T. (1985). *Advanced econometrics*. Harvard university Press.

Bliss, C. L. (1934). The method of probits. *Science*, *79*, 38–39.

Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley & Son.

Cameron, A. C., & Trivedi, P. K. (2005). *Micro econometrics, methods and applications*. Cambridge University Press.

Cappellari, L., & Jenkins, S. P. (2003). Multivariate probit regression using simulated maximum likelihood. *The Stata Journal*, *3*, 278–294.

Casella, G., & George, E. (1992). Explaining the Gibbs sampler *American Statistician*, *46*(3), 167–174.

Cramer, J. S. (2004). The early origins of the logit model. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *35*, 613–626.

Dempster, P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm *Journal of the Royal Statistical Society, Series B (Methodological)*, *39*(1), 1–38.

Duncan, O. D. (1975). *Introduction to structural equation models*. New York: Academic Press.

Finney, D. (1947). *Probit analysis. A statistical treatment of the sigmoid respond curve*. Macmillan.

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

Finney, D. (1971). *Probit analysis*. Cambridge University Press.
Fok, D. (2017). Advanced individual demand models. In P. S. H. Leeflang, J. E. Wieringa, T. H. A. Bijmolt, & K. H. Pauwels (Eds.), *Advanced methods for modeling markets*. Springer.
Gauss (1809). *Theoria Motus Corporum Coelestium in sectionibus conicis solem ambientium*.
Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*(6), 721–741, doi: 10.1109/TPAMI.1984.4767596.
German SOEP-Socio-Economic Panel (DIW). https://www.eui.eu/Research/Library/ResearchGuides/Economics/Statistics/DataPortal/GSOEP
Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, *57*, 1317–1339.
Goldberger, A. S. (1968). *Topics in regression analysis*. MacMillan.
Goldberger, A. S. (1971). Econometrics and psychometrics: A survey of communalities. *Psychometrika*, *36*, 83–107. https://doi.org/10.1007/BF02291392
Goldberger, A. S. (1991). *A course in econometrics*. Harvard University Press.
Gouriéroux, C., Monfort, A., Renault, E., & Trognon, A. (1987). Generalised residuals. *Journal of Econometrics*, *34*(1–2), 5–32.
Green, P., & Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research*, *5*, 103–123.
Greene, W. H., & Hensher, D. A. (2010). *Modeling ordered choices: A primer*. Cambridge University Press.
Greene, W. H. (2018). *Econometric analysis* (8th ed.). Pearson.
Hajivassiliou, V., & McFadden, D. (1998). The method of simulated scores for the estimation of LDV models. *Econometrica*, *66*, 863–896.
Hayduk, L. (1987). *Structural equation modeling with LISREL: Essentials and advances*. Baltimore: Johns Hopkins University Press.
Hansen, L. P. (1982). Large sample properties of generalized methods of moments estimators. *Econometrica*, *50*, 1029–1054.
Heckman, J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, *5*(4), 475–492.
Hensher, D. A., Rose, J. M., & Greene, W. H. (2015). *Applied choice analysis* (2nd ed.) Cambridge University Press.
Hood, W. C., & Koopmans, T. C. (Eds.) (1953). *Studies in econometric method*. Wiley and Son)
Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distributions* (vol. *1*). New York: Wiley.
Jöreskog, K. G. (1973). A general method for estimating a linear structural equation-system. In A.S. Goldberger, & O.D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85–112). New York: Academic Press.
Keane, M. (1994). A computationally practical simulation estimator for panel data. *Econometrica*, *62*(1), 95–116.
Legendre, A. M. (1805). *Nouvelles methodes pour la determination des orbites des cometes*. Paris: Firmin Didot.
Liu, D., Li, S., Yu, Y., & Moustaki, I. (2021). Assessing partial association between ordinal variables: quantification, visualization, and hypothesis testing. *Journal of the American Statistical Association*, *116*(534), 955–968. https://doi.org/10.1080/01621459.2020.1796394
Maddala, G. S. (1983). *Limited dependent and qualitative variables in econometrics*. Cambridge University Press.
Manski, C. F. (1988). *Analog estimation methods in econometrics*. Chapman and Hall.
Maris, E. (1995), Psychometric latent response models. *Psychometrika*, *60*, 523–547.
McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). Academic Press.
McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. (PDF), *Econometrica*, *57*(5), 995–1026.
McKelvey, R., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, *4*, 103–120.
Moss, J., & Grønneberg, S. (2023). Partial identification of latent Correlations with ordinal data. *Psychometrika*, *88*, 241–252.
Mullahy, J. (2016). Estimation of multivariate probit models via bivariate probit. *The Stata Journal*, *16*(1), 37–51.
Muris, C. (2017). Estimation in the fixed-effects ordered logit model. *The Review of Economics and Statistics*, MIT Press, *99*(3), 465–477.
Narayanan, A. (2012). A review of eight software packages for structural equation modeling. *American Statistician - AMER STATIST.*, *66*, 129–138.
Olsson, U. (1979). Maximum likelihood estimation of the polyserial correlation coefficient.. *Psychometrika*, *44*(4), 443–460.
Roodman, D. (2007, 2020). CMP: Stata module to implement conditional (recursive) mixed process estimator. *Statistical Software Components*, S456882. Boston College Department of Economics, revised 06 Dec 2020. https://econpapers.repec.org/software/bocbocode/s456882.htm
Roodman, D. (2011). Fitting fully observed recursive mixed-process models with cmp. *The Stata Journal*, *11*(2), 159–206.
Stigler, S. M. (1981). Gauss and the invention of least squares. *The Annals of Statistics*, *9*(3), 465–474.
Van Beek, K. W. H., Koopmans, C. C., & van Praag, B. M. S. (1997). Shopping at the Labour Market: A real tale of fiction. *European Economic Review*, *41*(2), 295–317.
Van de Ven, W. P. M. M., & van Praag, B. M. S. (1981). The demand for deductibles in private health insurance: A probit model with sample selection. *Journal of econometrics*, *17*(2), 229–252

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

Verbeek, M. (2017). *A guide to modern econometrics*. (5th ed.) Wiley & Son.
Wallander, L. (2009). 25 Years of factorial surveys in sociology: A review. *Social Science Research*, *38*, 505–520.
Wiberg, M., Culpepper, S., Janssen, R., González, J., & Molenaar, D. (2019). A taxonomy of item response models in psychometrika. *Quantitative Psychology*, (265), 13–23.
World Happiness Report, Oxford University, 2024

Appendix

A. Estimation of the latent error covariance matrix Σ

For the estimation of the latent covariance matrix starting from the estimated $\widehat{b}$-estimates and the in-between covariance matrix $\widehat{\overline{\Sigma}}$ we propose the following method. We make use of the fact that in order to estimate the true error covariances $\rho_{k,k'}$ by $\widehat{\rho}_{k,k'}$ for two OP-equations we notice that for each pair k,k' we only have to look for the bivariate marginal distribution of $\varepsilon_k,\varepsilon_{k'}$. Consider for instance a cluster of three observations (1, 2, 3). The sample likelihood is $P\left(\varepsilon_1 \in S_1, \varepsilon_2 \in S_2, \varepsilon_3 \in S_3; 0, \Sigma_{1,2,3}\right)$, where $S_1 = \left(\nu_{j_i-1} - x_{i,1}\beta < \varepsilon_{i,1} \leq \nu_{j_i} - x_{i,1}\beta\right]$ and S_2 and S_3 similarly defined. However, the bivariate marginal likelihood provides the same information on $\sigma_{1,2}$ as the trivariate full likelihood. In the K-dimensional covariance matrix there are $K(K-1)/2$ different non-diagonal elements. For instance, if $K = 8$ as in the model in Section 6, this boils down to 28 simple two-dimensional estimations.

Applying the Gram-Schmidt decomposition, there holds for the pair$\varepsilon_k,\varepsilon_{k'}$.

$$\varepsilon_{i,k'} = \rho_{k,k'}.\varepsilon_{i,k} + \eta_{i,k}\sqrt{1-\rho^2_{k,k'}} \tag{A.1}$$

where $\varepsilon_{i,k},\eta_{i,k}$ are independent $N(0,1)$ drawings.[7] The coefficient $\sqrt{1-\rho^2_{k,k'}}$ guarantees that $\sigma^2(\varepsilon_{i,k'}) = 1$. We have $E(\varepsilon_{i,k}.\varepsilon_{i,k'}) = \rho_{k,k'}$. When we consider the estimated in-between error covariance $\overline{\rho}_{k.k'} = E(\overline{\varepsilon}_{i,k},\overline{\varepsilon}_{i,k'})$, then its value depends only on the two one-dimensional grids $\mathbb{C}_{ik}$ and $\mathbb{C}_{i,k'}$ and $\rho_{k,k'}$. We may now simulate $\varepsilon_{i,k},\varepsilon_{i,k'}$ for various values of $\rho_{k,k'}$ by means of (A.1), and calculate the corresponding in-between covariances $\overline{\rho}_{k,k'} = \frac{1}{N}\sum_{i=1}^{N}\overline{e}_{i,k}.\overline{e}_{i,k'}$, corresponding to the grids $\mathbb{C}_{ik}$ and $\mathbb{C}_{i,k'}$. We notice that the dependent variable vector $Y_{i,k}$ is observed according to a uniform i-independent grid$\mathbb{C}_k = \left\{\left(\nu^{(k)}_{j-1}, \nu^{(k)}_j\right]\right\}_{j=1}^{J_k} = \left\{\mathbb{S}^{(k)}_j\right\}_{j=1}^{J_k}$, while the errors $\varepsilon_{i,k}$ are observed according to i-dependent individual grids

$$\mathrm{X}_{ik} = \left\{\left(\nu^{(k)}_{j-1} - x_{i,k}\beta, \nu^{(k)}_j - x_{i,k}\beta\right]\right\}_{j=1}^{J_k} = \left\{\mathbb{S}^{(k)}_j - x_{i,k}\beta\right\}_{j=1}^{J_k}.$$

We consider the set of all two-dimensional grids for all observation units. $\left\{\mathbb{C}_{ik}, \mathbb{C}_{i,k'}\right\}_{i=1}^{N} = \mathbb{C}_{k,k'}$
We write for the sample in-between covariance

$$\overline{\rho}_{k,k'} = \frac{1}{N}\sum_{i=1}^{N}\overline{e}_{i,k}.(\rho_{k,k'}).\overline{e}_{i,k'}.(\rho_{k,k'}) \overset{def}{=} f(\rho_{k,k'}).$$

We estimate the value of the latent $\rho_{k,k'}$ by comparing the observed sample in-between covariance $\widehat{\overline{\rho}}_{k,k'}$ with its simulated counterpart $\overline{\rho}_{k,k'}$ for various values of the latent $\rho_{k,k'}$. It appears that there is one value $\widehat{\rho}_{k,k'}$ solving $f(\rho_{k,k'}) = \widehat{\overline{\rho}}_{k,k'}$. In order to gain insight into the relationship between the latent $\rho_{k,k'}$ and the corresponding in-between covariance $\overline{\rho}_{k,k'}$ we did some simulation experiments for three different two-dimensional grids. In Table A1 we present three different two-dimensional grids with different $\overline{\rho}_{k,k'}$ values for different values of $\rho_{k,k'}$. We describe one example in detail. Consider the two-dimensional grid with $\mathbb{C}_1 = \{(-\infty,0], (0,\infty)\}$, $\mathbb{C}_2 = \{(-\infty,0], (0,\infty)\}$ in the middle part of Table A1. We simulate a sample from a two-dimensional normal distribution with a correlation with $\rho = 0.1$. We find an in-between covariance $\overline{\rho} = 0.043$. For $\rho = 0.2$ we find a corresponding value of $\overline{\rho} = 0.087$, and so on. In Table A1 we present the relationship between ρ and $\overline{\rho}$ for three different two-dimensional grids. We found that the function $f(\rho_{k,k'}) = \overline{\rho}_{k,k'}$ is monotonically increasing in $\rho_{k,k'}$ for all grids we tested. See also Aitkin(1964). We conclude that for a given grid $(\mathbb{C}_1,\mathbb{C}_2)$ the function $f(\rho) = \overline{\rho}$ is monotonically increasing in ρ.

The solution $\widehat{\rho}_{k,k'}$ of $f(\rho_{k,k'}) = \widehat{\overline{\rho}}_{k,k'}$ is, using Slutsky's Law, a consistent estimator $\widehat{\rho}_{k,k'}$ of the population parameter $\rho_{k,k'}$. Doing this for each non-diagonal element of Σ, we estimate the underlying non-diagonal elements of the full correlation matrix Σ. The diagonal elements are equal to one by assumption. This yields a consistent estimator $\widehat{\Sigma}_\varepsilon$ of the error covariance matrix Σ_ε. Confidence intervals can be found by the delta method. Fig. 1 shows the graph of the function $f(\rho_{k,k'})$ for the left grid in Table A1.

In Table 4 we presented the full correlation matrices calculated by SUOP and by `cmp`. In Table A2. below we present the in-between covariance matrix and the full covariance matrix for the five-year panel considered in Section 5.1. We notice the fact

[7] We use standard normal draws for $\varepsilon_{i,k},\eta_{i,k'}$, because in (A.1) the simulated errors have to be $N(0,1)$.

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

Table A1. Relation between covariance ρ and in-between covariance $\overline{\rho}$.

grid	ρ	$\overline{\rho}$	grid	ρ	$\overline{\rho}$	grid	ρ	$\overline{\rho}$
$\nu_{11} = -0.5$	0.1	0.075	$\nu_1 = 0.0$	0.1	0.043	$\nu_{11} = -1.0$	0.1	0.032
$\nu_{12} = 0.0$	0.2	0.141		0.2	0.087	$\nu_{12} = 2.0$	0.2	0.051
$\nu_{13} = 0.75$	0.3	0.198		0.3	0.128		0.3	0.092
	0.4	0.290	$\nu_2 = 0.0$	0.4	0.167	$\nu_{21} = -2.0$	0.4	0.119
$\nu_{21} = -0.75$	0.5	0.342		0.5	0.224	$\nu_{22} = 1.0$	0.5	0.140
$\nu_{22} = -0.5$	0.6	0.451		0.6	0.264		0.6	0.167
$\nu_{23} = 0.5$	0.7	0.505		0.7	0.321		0.7	0.189
N = 10,000	0.8	0.601	N = 10,000	0.8	0.367	N = 10,000	0.8	0.217
	0.9	0.681		0.9	0.457		0.9	0.219

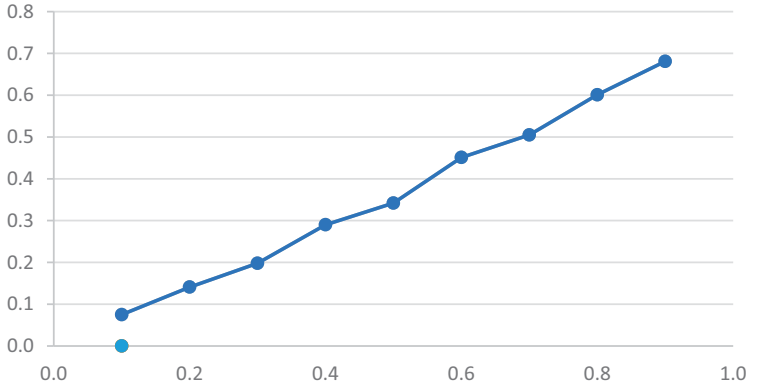

Figure. A1. $\overline{\rho}$ as a function of ρ.

that the grid-wise observation causes a considerable loss of information. About 60% of the information is lost. The estimates are not based on structural assumptions like "random effects" or errors where the correlations are specific functions of the time difference. Actually, this result could be used to estimate and test specific functional specifications of the covariance matrix. From Table A2. it is already obvious that "random effects" are to be rejected for this panel structure because in that case, the non-diagonal elements would have to be roughly equal to each other.

Table A2. In-between and full covariance matrices for the five-year panel.

	In-between covariance matrix					Full covariance matrix				
	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
2009	0.6434					1.0000				
2010	0.4852	0.6292				0.8899	1.0000			
2011	0.4123	0.4841	0.6289			0.7984	0.8945	1.0000		
2012	0.3771	0.4210	0.4793	0.6277		0.7490	0.8133	0.8910	1.0000	
2013	0.3413	0.3725	0.4156	0.4944	0.6321	0.7064	0.7482	0.7970	0.9270	1.0000

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

B. Table B1 is complete

Table B1. Comparison of the parameter estimates and their s.e.'s for Ordered Probit, Method of Moments, and Maximum Likelihood.

McK-Z R^2	0.1489		0.1481		0.1154	
Health Satisfaction	β_{OP}	σ_{OP}	β_{SUOP}	σ_{SUOP}	β_{ML}	σ_{ML}
Health: AGE	−0.0695	0.0046	−0.0693	0.0048	−0.0651	0.0045
Health: AGE^2	0.0006	0.00005	0.0006	0.00005	0.0006	0.00005
Health: D_FEMALE	−0.0182	0.0175	−0.0181	0.0174	−0.0134	0.0166
Health: D_SINGLE	−0.0514	0.0291	−0.0510	0.0293	−0.0570	0.0278
Health: D_SEPARATED	−0.1241	0.0240	−0.1238	0.0243	−0.1108	0.0230
Health: Ln_LABOURINC	0.0289	0.0025	0.0288	0.0026	0.0227	0.0022
Health: CHILDREN	0.0357	0.0123	0.0356	0.0126	0.0308	0.0114
Health: D_EAST	−0.1678	0.0201	−0.1670	0.0196	−0.1402	0.0193
Health: DISABLE	−0.7991	0.0272	−0.7971	0.0268	−0.5825	0.0232
Health: Cut point 1	−1.8277	0.0184	−1.8240	0.0191	−1.8163	0.0181
Health: Cut point 2	−1.1161	0.0131	−1.1094	0.0134	−1.1122	0.0130
Health: Cut point 3	−0.3432	0.0109	−0.3415	0.0109	−0.3399	0.0107
Health: Cut point 4	0.9468	0.0123	0.9442	0.0124	0.9409	0.0122
McK-Z R^2	0.0278		0.0278		0.0196	
Sleep Satisfaction	β_{OP}	σ_{OP}	β_{SUOP}	σ_{SUOP}	β_{ML}	σ_{ML}
Sleep: AGE	−0.0380	0.0041	−0.0381	0.0042	−0.0348	0.0041
Sleep: AGE^2	0.0004	0.00004	0.0004	0.00004	0.0003	0.00004
Sleep: D. FEMALE	−0.1213	0.0172	−0.1213	0.0172	−0.1204	0.0164
Sleep: D. SINGLE	0.0508	0.0302	0.0505	0.0307	0.0594	0.0284
Sleep: D. SEPARATED	−0.0762	0.0253	−0.0762	0.0260	−0.0658	0.0238
Sleep: Ln(HH.LABOURINC)	0.0067	0.0021	0.0066	0.0020	0.0062	0.0018
Sleep: # of CHILDREN	0.0219	0.0121	0.0220	0.0122	0.0164	0.0113
Sleep: YEARS of EDUCATION	0.0316	0.0032	0.0315	0.0031	0.0117	0.0028
Sleep: D. EAST GERMANY	−0.1063	0.0199	−0.1065	0.0196	−0.0763	0.0192
Sleep: Cut point 1	−1.6918	0.0175	−1.6912	0.0174	−1.7036	0.0174
Sleep: Cut point 2	−0.9791	0.0123	−0.9800	0.0123	−0.9831	0.0123
Sleep: Cut point 3	−0.9251	0.0106	−0.2970	0.0106	−0.2940	0.0104
Sleep: Cut point 4	0.7286	0.0114	0.7269	0.0115	0.7150	0.0113

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

Table B1. Continued

McK-Z R^2	0.1345		0.1342		0.1133	
Household Income Satisfaction	β_{OP}	σ_{OP}	β_{SUOP}	σ_{SUOP}	β_{ML}	σ_{ML}
HH inc.: AGE	−0.0648	0.0045	−0.0649	0.0048	−0.0593	0.0045
HH inc.: AGE^2	0.0008	0.00005	0.0008	0.00005	0.0007	0.00005
HH inc.: D. FEMALE	0.0837	0.0175	0.0836	0.0175	0.0764	0.0167
HH inc.: D. SINGLE	−0.1398	0.0304	−0.1404	0.0308	−0.1326	0.0279
HH inc.: D. SEPARATED	−0.2582	0.0255	−0.2583	0.0264	−0.2404	0.0237
HH inc.: Ln(LABOURINC)	0.0396	0.0026	0.0395	0.0027	0.0349	0.0025
HH inc.: Ln(HH.LABOURINC)	0.0261	0.0021	0.0261	0.0021	0.0262	0.0014
HH inc.: # of CHILDREN	−0.0088	0.0122	−0.0085	0.0122	−0.0212	0.0095
HH inc.: YEARS of EDUCATION	0.0808	0.0033	0.0806	0.0034	0.0693	0.0032
HH inc.: D. EAST GERMANY	−0.3862	0.0201	−0.3867	0.0200	−0.3472	0.0185
HH-inc.: Cut point 1	−1.7878	0.0179	−1.7106	0.0174	−1.7012	0.0175
HH-inc.:Cut point 2	−1.0544	0.0129	−1.0564	0.0128	−1.0323	0.0126
HH-inc.:Cut point 3	−0.2814	0.0108	−0.2858	0.0107	−0.2743	0.0103
HH-inc.:Cut point 4	0.9480	0.0123	0.9442	0.0126	0.9020	0.0121
McK-Z R^2	0.1503		0.1495		0.1306	
Individual Income Satisfaction	β_{OP}	σ_{OP}	β_{SUOP}	σ_{SUOP}	β_{ML}	σ_{ML}
Ind. inc.: AGE	−0.0562	0.0045	−0.0562	0.0048	−0.0565	0.0045
Ind. inc.: AGE^2	0.0008	0.00005	0.0008	0.00005	0.0007	0.00005
Ind. inc.: D. FEMALE	−0.1239	0.0174	−0.1239	0.0173	−0.1180	0.0169
Ind. inc.: D. SINGLE	−0.0595	0.0290	−0.0602	0.0292	−0.0613	0.0277
Ind. inc.: D. SEPARATED	−0.1016	0.0240	−0.1016	0.0245	−0.0885	0.0233
Ind. inc.: Ln(LABOURINC)	0.0727	0.0026	0.0725	0.0028	0.0728	0.0025
Ind. inc.: # of CHILDREN	0.0311	0.0121	0.0314	0.0119	0.0277	0.0102
Ind. inc.: YEARS of EDUCATION	0.0752	0.0033	0.0749	0.0034	0.0670	0.0032
Ind. inc.: D. EAST GERMANY	−0.2912	0.0200	−0.2916	0.0198	−0.2561	0.0189
Ind. inc.: DISABLILITY RATE	−0.1510	0.0270	−0.1513	0.0272	0.0430	0.0177
Ind. Inc.: Cut point 1	−1.3425	0.0147	−1.3428	0.0144	−1.2938	0.0141
Ind. Inc.: Cut point 2	−0.7372	0.0117	−0.7401	0.0116	−0.7361	0.0114
Ind. inc.: Cut point 3	−0.0300	0.0107	−0.0358	0.0106	−0.0442	0.0103
Ind. inc.: Cut point 4	1.1096	0.0129	1.1043	0.0133	1.0914	0.0130

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

Table B1. Continued

McK-Z R^2	0.0566		0.0564		0.0475	
Dwelling Satisfaction	β_{OP}	σ_{OP}	β_{SUOP}	σ_{SUOP}	β_{ML}	σ_{ML}
Dwelling: AGE	−0.0432	0.0047	−0.0424	0.0049	−0.0397	0.0046
Dwelling: AGE^2	0.0006	0.00005	0.0006	0.00005	0.0005	0.00005
Dwelling: D. FEMALE	0.0333	0.0183	0.0332	0.0182	0.0409	0.0168
Dwelling: D. SINGLE	−0.2409	0.0313	−0.2414	0.0322	−0.2312	0.0282
Dwelling: D. SEPARATED	−0.2384	0.0265	−0.2384	0.0266	−0.2126	0.0242
Dwelling: Ln(LABOURINC)	0.0179	0.0027	0.0178	0.0027	0.0151	0.0026
Dwelling: Ln(HH.LABOURINC)	0.0134	0.0022	0.0133	0.0022	0.0142	0.0020
Dwelling: # of CHILDREN	−0.0079	0.0127	−0.0077	0.0127	−0.0159	0.0109
Dwelling: YEARS of EDUC.	0.0313	0.0034	0.0312	0.0034	0.0213	0.0033
Dwelling: DISABLILITY RATE	−0.1317	0.0283	−0.1320	0.0287	0.0258	0.0245
Dwelling: Cut point 1	−2.1862	0.0261	−2.1806	0.0256	−2.1637	0.0252
Dwelling: Cut point 2	−1.6332	0.0172	−1.6321	0.0170	−1.6211	0.0169
Dwelling: Cut point 3	−0.9408	0.0123	−0.9453	0.0122	−0.9364	0.0121
Dwelling: Cut point 4	0.2504	0.0106	0.2477	0.0107	0.2360	0.0105
McK-Z R^2	0.0902		0.0901		0.0940	
Leisure Time Satisfaction	β_{OP}	σ_{OP}	β_{SUOP}	σ_{SUOP}	β_{ML}	σ_{ML}
Leisure: AGE	−0.0433	0.0043	−0.0432	0.0044	−0.0406	0.0042
Leisure: AGE^2	0.0006	0.00005	0.0006	0.00005	0.0005	0.00005
Leisure: D. FEMALE	−0.0353	0.0175	−0.0352	0.0175	−0.0257	0.0160
Leisure: Ln(LABORINC)	−0.0273	0.0026	−0.0272	0.0028	−0.0299	0.0025
Leisure: Ln(HH.LABOURINC)	0.0047	0.0019	0.0047	0.0020	0.0051	0.0018
Leisure: # of CHILDREN	−0.0847	0.0114	−0.0848	0.0116	−0.0896	0.0102
Leisure: YEARS of EDUCATION	0.0055	0.0033	−0.0056	0.0032	−0.0043	0.0031
Leisure: D. EAST GERMANY	−0.1293	0.0202	−0.1288	0.0200	−0.0921	0.0185
Leisure: DISABLE RATE	−0.1039	0.0277	−0.1036	0.0287	0.0498	0.0249
Leisure: Cut point 1	−1.8463	0.0194	−1.8454	0.0200	−1.8090	0.0188
Leisure: Cut point 2	−1.2129	0.0136	−1.2092	0.0138	−1.2051	0.0133
Leisure: Cut point 3	−0.4716	0.0109	−0.4687	0.0110	−0.4846	0.0108
Leisure: Cut point 4	0.6486	0.0113	.6499	0.0113	0.6359	0.0114

https://doi.org/10.1017/psy.2024.19 Published online by Cambridge University Press

Table B1. Continued

McK-Z R^2	0.0834		0.0834		0.0763	
Family Life Satisfaction	β_{OP}	σ_{OP}	β_{SUOP}	σ_{SUOP}	β_{ML}	σ_{ML}
Family life: AGE	−0.0572	0.0048	−0.0573	0.0048	−0.0534	0.0046
Family life: AGE^2	0.0006	0.00005	0.0006	0.00005	0.0005	0.00005
Family life: D. SINGLE	−0.4871	0.0297	−0.4880	0.0295	−0.4665	0.0270
Family life: D. SEPARATED	−0.5665	0.0260	−0.5666	0.0267	−0.5420	0.0236
Family life: Ln(LABOURINC)	0.0013	0.0027	0.0012	0.0028	−0.0012	0.0025
Fam. life: Ln(HH.LABOURINC)	0.0167	0.0022	0.0166	0.0022	0.0180	0.0020
Family life: YEARS of EDUC.	0.0040	0.0034	0.0038	0.0034	−0.0084	0.0032
Family life: D. EAST GERMANY	−0.1370	0.0209	−0.1376	0.0208	−0.1049	0.0192
Family life: DISABLE RATE	−0.1290	0.0281	−0.1293	0.0289	0.0321	0.0239
Family life: Cut point 1	−2.1949	0.0250	−2.1905	0.0244	−2.1609	0.0240
Family life: Cut point 2	−1.6631	0.0172	−1.6638	0.0169	−1.6476	0.0168
Family life: Cut point 3	−0.9763	0.0125	−0.9813	0.0124	−0.9741	0.0123
Family life: Cut point 4	0.1778	0.0106	0.1749	0.0107	0.1609	0.0105
McK-Z R^2	0.1193		0.1187		0.0965	
Standard of Living Satisfaction	β_{OP}	σ_{OP}	β_{SUOP}	σ_{SUOP}	β_{ML}	σ_{ML}
Stand.living: AGE	−0.0760	0.0047	−0.0758	0.0048	−0.0699	0.0045
Stand.living: AGE^2	0.0008	0.00005	0.0008	0.00005	0.0008	0.00005
Stand.living: D. FEMALE	0.1111	0.0180	0.1108	0.0180	0.1113	0.0155
Stand.living: D. SINGLE	−0.2982	0.0293	−0.2984	0.0292	−0.2762	0.0266
Stand.living: D. SEPARATED	−0.3720	0.0259	−0.3712	0.0265	−0.3415	0.0234
Stand.living: Ln(LABOURINC)	0.0312	0.0026	0.0310	0.0027	0.0269	0.0025
Std.living: Ln(HH.LABOURINC)	0.0225	0.0021	0.0225	0.0022	0.0228	0.0017
Stand.living: YEARS of EDUC.	0.0714	0.0034	0.0711	0.0034	0.0576	0.0032
Stand.living: D. EAST GERM.	−0.3197	0.0205	−0.3199	0.0201	−0.2755	0.0180
Stand.living: Cut point 1	−2.2120	0.0254	−2.1991	0.0251	−2.2096	0.0242
Stand.living: Cut point 2	−1.6048	0.0167	−1.6018	0.0164	−1.5980	0.0163
Stand.living: Cut point 3	−0.8760	0.0119	−0.8219	0.0118	−0.8087	0.0117
Stand.living: Cut point 4	0.5312	0.0111	0.5264	0.0113	0.5095	0.0109

Cite this article: van Praag, B. M. S., Hop, J. P. and Greene, W. H. (2025). Estimation of Linear Models from Coarsened Observations: A Method of Moments Approach. *Psychometrika*, 870–903. https://doi.org/10.1017/psy.2024.19