Stat. Meth. & Appl. (2007) 16:381–393
DOI 10.1007/s10260-006-0039-z

ORIGINAL ARTICLE

A multilevel multinomial logit model for the analysis of graduates' skills

Leonardo Grilli · Carla Rampichini

Accepted: 22 November 2006 / Published online: 22 December 2006

Abstract The main goal of the paper is to specify a suitable multivariate multilevel model for polytomous responses with a non-ignorable missing data mechanism in order to determine the factors which influence the way of acquisition of the skills of the graduates and to evaluate the degree programmes on the basis of the adequacy of the skills they give to their graduates. The application is based on data gathered by a telephone survey conducted, about two years after the degree, on the graduates of year 2000 of the University of Florence. A multilevel multinomial logit model for the response of interest is fitted simultaneously with a multilevel logit model for the selection mechanism by means of maximum likelihood with adaptive Gaussian quadrature. In the application the multilevel structure has a crucial role, while selection bias results negligible. The analysis of the empirical Bayes residuals allows to detect some extreme degree programmes to be further inspected.

Keywords Job skills · Multilevel models · Polytomous response · Selection bias

1 Introduction

The analysis of graduates' skills provides an important piece of information for the decisional process on the contents of the university degree programmes.

L. Grilli (✉) · C. Rampichini
Department of Statistics "G. Parenti", University of Florence, Viale Morgagni 59, 50134 Florence, Italy
e-mail: grilli@ds.unifi.it

C. Rampichini
e-mail: carla@ds.unifi.it

Nowadays this aspect is particularly relevant since in the last few years the Italian universities are becoming increasingly oriented towards the needs of the labour market. The University of Florence gathered data which allow to study the graduates' skills in some detail. The data come from a telephone survey conducted, about two years after the degree, on the graduates of year 2000 of the University of Florence. The questionnaire allows to collect information on eight skills. For each skill the questionnaire asks if it is needed and, in case of an affirmative response, where it was acquired: during the degree programme, at workplace or otherwise. The main goal of the paper is to determine the factors which influence the way of acquisition of the skills and to evaluate the degree programmes on the basis of the adequacy of the skills they give to their graduates.

The analysis of such data raises several methodological questions: (a) the response of interest might be affected by selection bias due to the design of the questionnaire: for each skill a first question asks if the graduate currently uses it, while, in case of an affirmative response, a second question asks where the skill was acquired; therefore for the graduates that do not use the skill the second question is missing, causing a potential selection bias; (b) for each skill, the second question has a polytomous response, aggregated to three categories: the skill was acquired during the degree programme, at workplace or otherwise; (c) the data have a hierarchical structure with graduates nested in degree programmes, so the responses are correlated; (d) the data are multivariate: there are eight skills, and for each skill the two questions mentioned in *a* are asked.

The present analysis focuses on only one of the eight skills, that is *professional and technical abilities*. In order to taking into account the mentioned features of the data, an adequate multilevel multinomial logit model (Skrondal and Rabe-Hesketh 2003) with a non-ignorable missing data mechanism (Heckman 1979; Little and Rubin 2002) is developed.

Section 2 describes the multilevel multinomial logit model used to analyze the polytomous response of interest, and the random utility interpretation of the model is sketched out. In Sect. 3 the outlined multilevel model is extended in order to take into account the selection mechanism. In Sect. 4 the model is applied to the data on the graduates' skills. Section 5 concludes.

2 The multilevel multinomial logit model

2.1 The GLM formulation

Statistical models for polytomous responses are standard tools in many disciplines, but most of the theory has been developed in Econometrics, under the label *discrete choice models* (McFadden 1973; Train 2003). Skrondal and Rabe-Hesketh (2003) and Hedeker (2003) give an account of the multilevel version of such models, though applications are still quite rare. The multilevel multinomial logit model is a mixed Generalized Linear Model (McCullagh and Nelder 1989) with linear predictors

$$\eta_{ij}^{(m)} = \alpha^{(m)} + \boldsymbol{\beta}^{(m)\prime}\boldsymbol{x}_{ij} + \xi_j^{(m)} + \delta_{ij}^{(m)} \tag{1}$$

and multinomial logit link

$$P(Y_{ij} = m \mid \boldsymbol{x}_{ij}, \boldsymbol{\xi}_j, \boldsymbol{\delta}_{ij}) = \frac{\exp\{\eta_{ij}^{(m)}\}}{1 + \sum_{l=2}^{M} \exp\{\eta_{ij}^{(l)}\}} \tag{2}$$

where $m = 1, 2, \ldots, M$ denotes the response category (way of acquisition of the skills), $j = 1, 2, \ldots, J$ denotes the cluster (degree programme) and $i = 1, 2, \ldots, n_j$ denotes the subject (graduate) of the j-th cluster. In order to simplify the exposition, only two levels are considered, but the model can be easily extended to many levels. The response variable Y_{ij} has (conditional on the random effects) a multinomial distribution, taking values in the set of categories $\{1, 2, \ldots, M\}$, where $m = 1$ is the reference category for which all the parameters and the random errors are set to 0 and thus the conditional probability of $Y_{ij} = 1$ is $1/(1 + \sum_{l=2}^{M} \exp[\eta_{ij}^{(l)}])$.

Note that in the linear predictors (1) there are no category specific covariates, though this is a possible extension. Each equation has specific parameters $\alpha^{(m)}$ and $\boldsymbol{\beta}^{(m)}$ $(m = 2, 3, \ldots, M)$. Finally, $\boldsymbol{\xi}_j$ and $\boldsymbol{\delta}_{ij}$ are vectors of random errors representing unobserved heterogeneity at cluster and subject level, respectively, with the following distributional assumptions:

- errors at different levels are independent;
- $\boldsymbol{\xi}_j' = (\xi_j^{(2)}, \ldots, \xi_j^{(M)})' \overset{iid}{\sim} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\xi)$;
- $\boldsymbol{\delta}_{ij}' = (\delta_{ij}^{(2)}, \ldots, \delta_{ij}^{(M)})' \overset{iid}{\sim} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\delta)$.

The parameters of the cluster-level covariance matrix $\boldsymbol{\Sigma}_\xi$ are all identified, while the parameters of the subject-level covariance matrix $\boldsymbol{\Sigma}_\delta$ are in principle identified, but prone to empirical underidentification, unless some category specific covariate is included in the model (Skrondal and Rabe-Hesketh 2003). Indeed, in the application (Sect. 4) the subject-level covariance parameters turn out to be empirically not identified, so they are omitted. Nevertheless, the subject-specific errors $\boldsymbol{\delta}_{ij}$ are considered in the theoretical treatment in this and the following section in order to give a comprehensive account of the potentialities of the model.

A property of model (1)–(2) is that the odds for two categories m and l for subject i of cluster j are

$$\frac{P(Y_{ij} = m \mid \boldsymbol{x}_{ij}, \boldsymbol{\xi}_j, \boldsymbol{\delta}_{ij})}{P(Y_{ij} = l \mid \boldsymbol{x}_{ij}, \boldsymbol{\xi}_j, \boldsymbol{\delta}_{ij})} = \exp(\eta_{ij}^{(m)} - \eta_{ij}^{(l)}), \tag{3}$$

which depends only on the linear predictors of the two involved categories and does not depend on the other categories. This property, known as Independence from Irrelevant Alternatives (IIA), is often viewed as a restrictive feature of multinomial logit models. However, it should be realized that IIA holds conditionally on all the covariates and random errors. Therefore, since IIA does not hold marginally with respect to the random errors, the introduction of random terms in the linear predictors allows to partially relax the IIA property.

The likelihood of model (1)–(2) can be written by exploiting the conditional independence following from the assumptions:

$$L(\boldsymbol{\theta}) = \prod_{j=1}^{J} \int \prod_{i=1}^{n_j} \left\{ \int P(Y_{ij} \mid \boldsymbol{x}_{ij}, \boldsymbol{\xi}_j, \boldsymbol{\delta}_{ij}) f(\boldsymbol{\delta}_{ij}) \mathrm{d}\boldsymbol{\delta}_{ij} \right\} f(\boldsymbol{\xi}_j) \mathrm{d}\boldsymbol{\xi}_j, \tag{4}$$

where $\boldsymbol{\theta}' = (\alpha^{(2)}, \ldots, \alpha^{(M)}, \boldsymbol{\beta}^{(2)}, \ldots, \boldsymbol{\beta}^{(M)}, \boldsymbol{\Sigma}_\xi, \boldsymbol{\Sigma}_\delta)$. Since the integrals involved in the likelihood do not have closed form solutions, the maximization of the likelihood requires some sort of integral approximation. In the application (Sect. 4) the maximization is performed by means of the `gllamm` command of Stata, which allows to approximate the integrals with adaptive Gaussian quadrature (Rabe-Hesketh et al. 2004).

2.2 The random utility formulation

An alternative specification of the multinomial logit model is based on the random utility model (McFadden 1973). This specification helps the interpretation of the model and leads to a straightforward definition of the Intraclass Correlation Coefficient (ICC).

Let the continuous random variables $U_{ij}^{(m)}$, $m = 1, 2, \ldots, M$, represent the individual utilities associated to the M categories; hence the utility maximization rule implies that the observed indicator Y_{ij} equals m if and only if $U_{ij}^{(m)} > U_{ij}^{(l)}$, for every $l \neq m$ (ties are ignored as they have zero probability). A common formulation is the linear random utility model

$$U_{ij}^{(m)} = \eta_{ij}^{(m)} + \varepsilon_{ij}^{(m)}, \tag{5}$$

where $\eta_{ij}^{(m)}$ is the linear predictor (1) and $\varepsilon_{ij}^{(m)}$ are independent and identically distributed errors following the Gumbel distribution. Under this specification for the random utilities, the choice probabilities deriving from the utility maximization rule are given by the multinomial logit model (McFadden 1973).

Note that the data carry information only on the utility differences. Therefore to achieve identification all the fixed and random parameters of $U_{ij}^{(1)}$ are set to 0 so that the parameters of the m-th equation in fact refer to the utility difference $U_{ij}^{(m)} - U_{ij}^{(1)}$:

$$U_{ij}^{(m)} - U_{ij}^{(1)} = \eta_{ij}^{(m)} + [\varepsilon_{ij}^{(m)} - \varepsilon_{ij}^{(1)}]$$
$$= \alpha^{(m)} + \boldsymbol{\beta}^{(m)'}\boldsymbol{x}_{ij} + \xi_j^{(m)} + \delta_{ij}^{(m)} + [\varepsilon_{ij}^{(m)} - \varepsilon_{ij}^{(1)}]. \quad (6)$$

This model formally has a three level structure: utilities nested in subjects who are nested in clusters.

The mentioned empirical identification problem with the covariance parameters at subject level arises from the difficulty to disentangle the contribution of the components $\delta_{ij}^{(m)}$ and $[\varepsilon_{ij}^{(m)} - \varepsilon_{ij}^{(1)}]$, while the variance of $\xi_j^{(m)}$ is empirically well identifiable.

According to model (6), the total covariance of the utility differences can be decomposed as follows:

$$\mathrm{Cov}[(U_{ij}^{(m)} - U_{ij}^{(1)}), (U_{i'j'}^{(l)} - U_{i'j'}^{(1)})]$$
$$= \mathrm{E}[(\xi_j^{(m)} + \delta_{ij}^{(m)} + \varepsilon_{ij}^{(m)} - \varepsilon_{ij}^{(1)})(\xi_{j'}^{(l)} + \delta_{i'j'}^{(l)} + \varepsilon_{i'j'}^{(l)} - \varepsilon_{i'j'}^{(1)})]$$
$$= \mathrm{Cov}(\xi_j^{(m)}, \xi_{j'}^{(l)})I_{[j=j']} + \mathrm{Cov}(\delta_{ij}^{(m)}, \delta_{i'j'}^{(l)})I_{[j=j',i=i']} + \frac{\pi^2}{3}I_{[j=j',i=i',m=l]}, \quad (7)$$

where $\pi^2/3$ follows from the fact that $[\varepsilon_{ij}^{(m)} - \varepsilon_{ij}^{(1)}]$ has a logistic distribution (since it is the difference of two independent Gumbel random variables). When $j = j'$, $i = i'$ and $m = l$, expression (7) is in fact a variance decomposition:

$$\mathrm{Var}(U_{ij}^{(m)}) = \mathrm{Var}(\xi_j^{(m)}) + \mathrm{Var}(\delta_{ij}^{(m)}) + \frac{\pi^2}{3}. \quad (8)$$

This leads to a definition of the Intraclass Correlation Coefficient of the m-th equation analogous to the usual definition for dichotomous logit models:

$$\mathrm{ICC}^{(m)} = \frac{\mathrm{Var}(\xi_j^{(m)})}{\mathrm{Var}(\xi_j^{(m)}) + \mathrm{Var}(\delta_{ij}^{(m)}) + \pi^2/3}. \quad (9)$$

This index is the proportion of cluster residual variance and measures the degree of homogeneity among the subjects of the same cluster, conditionally on the covariates.

If $\delta_{ij}^{(m)}$ is not included in the model its variance is absorbed in that of $[\varepsilon_{ij}^{(m)} - \varepsilon_{ij}^{(1)}]$, but since the variance of $[\varepsilon_{ij}^{(m)} - \varepsilon_{ij}^{(1)}]$ is fixed to $\pi^2/3$ for identifiability reasons, then the other model parameters, including the standard deviation of $\xi_j^{(m)}$, are implicitly rescaled (Grilli and Rampichini 2003). However the ICC (9) is not affected by such rescaling, so it can be safely calculated even if $\delta_{ij}^{(m)}$ is not included in the model.

3 A model for the selection mechanism

In general a statistical model yields valid inferences only if the units are sampled at random (Copas and Li 1997). Selection bias may arise when the selection mechanism depends on unobserved variables correlated with the error terms of the statistical model of interest. A classical way to avoid the selection bias is to add an equation which explicitly models the selection mechanism (Heckman 1979).

Let us label with P (*Principal*) the multilevel multinomial logit model, whose response variable Y_{ij}^P takes values in the set of categories $\{1, 2, \ldots, M\}$. Moreover, let us label with S (*Selection*) the multilevel dichotomous logit model for the selection mechanism, whose response variable Y_{ij}^S takes the value 1 if Y_{ij}^P is observed. The joint model, labelled as $S\&P$, is thus defined by the following equations:

$$P(Y_{ij}^S = 1 \mid \boldsymbol{x}_{ij}^S, \xi_j^S, \delta_{ij}^S) = \frac{\exp\{\,\alpha^S + \boldsymbol{\beta}^{S\,\prime} \boldsymbol{x}_{ij}^S + \xi_j^S + \delta_{ij}^S\}}{1 + \exp\{\,\alpha^S + \boldsymbol{\beta}^{S\,\prime} \boldsymbol{x}_{ij}^S + \xi_j^S + \delta_{ij}^S\}}$$

$$P(Y_{ij}^P = m \mid \boldsymbol{x}_{ij}^P, \boldsymbol{\xi}_j^P, \boldsymbol{\delta}_{ij}^P) = \begin{cases} \frac{\exp\{\eta_{ij}^{P(m)}\}}{1+\sum_{l=2}^{M} \exp\{\eta_{ij}^{P(l)}\}} & \text{if } Y_{ij}^S = 1, \\ \text{not observed} & \text{if } Y_{ij}^S = 0 \end{cases} \tag{10}$$

where $\eta_{ij}^{P(m)} = 0$ for the reference category $m = 1$, and $\eta_{ij}^{P(m)} = \alpha^{P(m)} + \boldsymbol{\beta}^{P(m)\,\prime} \boldsymbol{x}_{ij}^P + \xi_j^{P(m)} + \delta_{ij}^{P(m)}$ for the other categories $m = 2, \ldots, M$. As before $j = 1, 2, \ldots, J$ denotes the clusters and $i = 1, 2, \ldots, n_j$ denotes the subjects of the j-th cluster. Note that also the selection equation has a multilevel structure, with random terms ξ_j^S and δ_{ij}^S representing unobserved heterogeneity at cluster and subject level, respectively. The vectors of all random errors at cluster level, $\boldsymbol{\xi}_j^*$, and subject level, $\boldsymbol{\delta}_{ij}^*$, have the following distribution:

- errors at different levels are independent;
- $\boldsymbol{\xi}_j^{*\prime} = (\xi_j^S, \xi_j^{P(2)}, \ldots, \xi_j^{P(M)})' \overset{iid}{\sim} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\xi*})$;
- $\boldsymbol{\delta}_{ij}^{*\prime} = (\delta_{ij}^S, \delta_{ij}^{P(2)}, \ldots, \delta_{ij}^{P(M)})' \overset{iid}{\sim} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\delta*})$.

It is worth to note that in the multilevel case the selection mechanism can operate at different levels:

- subject level: correlations between the pairs $(\delta_{ij}^S, \delta_{ij}^{P(m)})$, $m = 2, \ldots, M$;
- cluster level: correlations between the pairs $(\xi_j^S, \xi_j^{P(m)})$, $m = 2, \ldots, M$.

The signs of the correlations may be different at the two levels, giving rise to complex selection mechanisms.

If at least one of the correlations between $(\xi_j^S, \xi_j^{P(m)})$ or $(\delta_{ij}^S, \delta_{ij}^{P(m)})$ is not null, the selection mechanism is not ignorable, so unbiased estimation requires to fit both set of equations simultaneously. If a not ignorable selection mechanism is neglected, the coefficients of the covariates included in both $\boldsymbol{x}_{ij}^S$ and $\boldsymbol{x}_{ij}^P$ are prone to bias. Note that due to the polytomous nature of the *Principal* model, the selection mechanism can be quite complex since it can act differently on the various categories.

For the model under discussion it is not possible to derive analytical expressions for the bias. However the results of Grilli and Rampichini (2005) for the random intercept linear model can help foresee the dangers of multilevel selection mechanisms.

4 Application

The model described in the previous section is used to analyze some of the data gathered by a telephone survey conducted, about two years after the degree, on all the graduates of year 2000 of the University of Florence. The response rate is about 60%, where the non response is almost entirely due to missing contact. Particularly, the interest is in the analysis of some skills which may be requested for the current job. For each skill the questionnaire asks if it is needed and, in case of an affirmative response, where it was acquired: during the degree programme, at workplace or otherwise. The main goal is to determine the factors which influence the way of acquisition of the skills and to evaluate the degree programmes on the basis of the adequacy of the skills they give to their graduates.

The data set includes 3,148 interviewed graduates, but the analysis focuses on the 2,540 employed graduates (80% of the total). The graduates are nested in 56 degree programmes, where the number of graduates per degree programme ranges from 4 to 386, with a median of 21 graduates. The covariates used in the analysis are displayed in Table 1.

The covariates are all dichotomous, except for *age at degree* (centered at 28 years), *average mark* (centered with respect to the mean of the degree programme) and *duration index* (time to graduate divided by legal duration). The covariate *short degree* is the only cluster-level covariate.

The present work focuses on the analysis of only one of the eight skills of the questionnaire, that is *professional and technical abilities*. In the sample, 91.0% of the employed graduates currently use professional and technical abilities. Among such graduates, 47.7% acquired the skills at the university, 39.7% at workplace and 12.5% otherwise. It is of interest to study how the way of acquisition of the skills is related to the covariates in Table 1. Such covariates represent factors that are not under the control of the degree programmes, so the results of the analysis can be used to build net measures for comparison purposes.

The way of acquisition is observable only for the graduates actually using the skills and there is a likely dependence between acquisition and use. Therefore,

Table 1 Sample statistics of the covariates

Covariate	Min	Max	Mean
Demographic			
Male	0.00	1.00	0.44
Age at degree centered at 28	−6.00	24.00	0.01
University career			
Centered average mark	−4.78	2.59	0.00
Honors	0.00	1.00	0.25
Duration index	0.96	7.36	1.80
Job characteristics			
Self-employed work	0.00	1.00	0.29
Managerial post	0.00	1.00	0.29
Public sector	0.00	1.00	0.19
Temporary position	0.00	1.00	0.40
Degree not required for the job	0.00	1.00	0.32
Degree programme characteristics			
Short degree	0.00	1.00	0.12

in order to prevent a possible selection bias, the two aspects should be jointly analyzed by means of the $S\,\&P$ model (10). The *Selection* response Y_{ij}^S is equal to 1 if the graduate currently uses professional and technical abilities and 0 otherwise. If $Y_{ij}^S = 1$ then the *Principal* response Y_{ij}^P is observed. In particular, Y_{ij}^P=1 if the abilities were acquired during the degree programme, i.e. at the university (reference category), Y_{ij}^P=2 if at workplace or Y_{ij}^P=3 if otherwise.

The interpretation of the polytomous response model in terms of choice and utilities is obviously not appropriate in the present application. However, the idea that the observed response is generated by a set of latent variables is still applicable if one realize that in general a skill is acquired in various ways and the response to the question simply indicates the prevalent way of acquisition. Therefore, in the present context the phrase "utility of the m-th choice" may be translated as "amount of acquisition of the skill through the m-th way".

Maximum likelihood estimation is carried out by means of the `gllamm` procedure of Stata (Rabe-Hesketh et al. 2004). The high flexibility of `gllamm` allows to fit the joint $S\&P$ model (10) without any programming. The estimation algorithm implemented in `gllamm`, namely Newton–Raphson with adaptive Gaussian quadrature, is well established. In the application eight quadrature points turn out to be sufficient for an accurate estimate. The drawback of this algorithm is the long computational time, which increases rapidly with model complexity. Many alternative estimation methods are possible, e.g. Bayesian MCMC and Maximum Simulated Likelihood (Train 2003).

As already mentioned, the subject-level covariance parameters in $\mathbf{\Sigma}_{\delta*}$ are empirically not identified, as pointed out by the high condition number and the convergence difficulties of the estimation algorithm. Therefore the random errors $\boldsymbol{\delta}_{ij}^*$ are omitted from the models. As a consequence, the residual correlation among the responses Y^S and Y^P is only due to cluster-level factors.

Table 2 Model comparison

	Model 1 Joint S&P	Model 2 Unrelated S, P
log L	−2840.28	−2843.04
No. of parameters	26	24
Random parameters		
$\mathrm{Var}\left(\xi_j^S\right)$	0.1510	0.1448
$\mathrm{Var}\left(\xi_j^{P(2)}\right)$	0.1637	0.1531
$\mathrm{Var}\left(\xi_j^{P(3)}\right)$	0.4274	0.4129
$\mathrm{Corr}\left(\xi_j^S, \xi_j^{P(2)}\right)$	−0.2401	–
$\mathrm{Corr}\left(\xi_j^S, \xi_j^{P(3)}\right)$	−0.6772	–
$\mathrm{Corr}\left(\xi_j^{P(2)}, \xi_j^{P(3)}\right)$	0.8768	0.8479

The criterion for choosing the relevant covariates is the likelihood ratio test, with a p-value threshold of 5%. Ideally the selection of the covariates should be based on the joint *S&P* model (10), but since the computational times are in terms of many hours, the following strategy is adopted:

- selection of the covariates separately for the *Selection* model and for the *Principal* model;
- refinement using the joint *S&P* model, trying to reinsert in the *P* equations the variables which were previously discarded from the *P* model, but retained in the *S* model.

In the present case the refinement step does not cause any change, suggesting that selection bias is not relevant, as confirmed by the comparison between the joint *S&P* model and the simpler model with unrelated components (Table 2). The LR test statistic is 5.52 with 2 df, yielding a p-value of 0.0633, so the two cluster-level estimated correlations among the S and P sets of equations are jointly not significant. Even if the LR test for selection bias is known to have low power (Copas and Li 1997), in the present case the tiny differences in the parameter estimates suggest that selection bias can be safely ignored.

The analysis goes on by retaining the simpler model with unrelated components whose parameter estimates are reported in Table 3. It is worth to note that the sets of covariates which enter the S and P models are quite different: there are only two common covariates, namely *degree not required* and *short degree*.

To aid the interpretation of the results, the predicted probabilities for some typical graduates and degree programmes are calculated and reported in Table 4. Each line of the table corresponds to a given set of values of the covariates and random effects and reports the probability $\mathrm{P}(Y_{ij}^S = 1 \mid \mathbf{x}_{ij}^S, \xi_j^S)$ for the *Selection* equation and the probabilities $\mathrm{P}(Y_{ij}^P = m \mid \mathbf{x}_{ij}^P, \xi_j^{P(2)}, \xi_j^{P(3)})$,

Table 3 Estimated parameters from Model 2 (unrelated S and P equations)

Parameter	Selection equation (S)	Acquisition equations (P)	
		Workplace vs. University	Otherwise vs. University
Fixed			
Intercept	2.8705	*−0.1073*	−1.5776
Male	0.5245	–	–
Average mark	0.2296	–	–
Honors	–	−0.3992	−0.4236
Self-employed work	–	−0.3040	*0.0223*
Managerial post	–	−0.2728	*−0.0228*
Public sector	–	−0.3998	*−0.0125*
Temporary position	−0.4730	–	–
Degree not required	−1.3097	0.7787	0.9983
Short degree	0.5741	*−0.3059*	−0.8889
Random			
$\mathrm{Var}(\xi_j)$	0.1448	0.1531	0.4129
$\mathrm{Corr}(\xi_j^{P(2)}, \xi_j^{P(3)})$			0.8479
ICC (%)	4.22	4.45	11.16

Values not significant at 5% in italics

Table 4 Predicted probabilities from Model 2 (unrelated S and P equations)

Degree programme and graduate[a]	$\mathrm{P}(Y_{ij}^S = 1 \mid \mathbf{x}_{ij}^S, \xi_j^S)$	$\mathrm{P}(Y_{ij}^P = m \mid \mathbf{x}_{ij}^P, \xi_j^{P(2)}, \xi_j^{P(3)})$		
		Univ	Work	Other
Average d. p., base grad.	0.946	0.475	0.427	0.098
Male	0.968			
Average mark (+1)	0.957			
Honors		0.575	0.347	0.078
Self-employed work		0.534	0.354	0.113
Managerial post		0.530	0.363	0.107
Public sector		0.554	0.333	0.113
Temporary position	0.917			
Degree not required	0.826	0.284	0.556	0.159
Short degree	0.969	0.573	0.379	0.049
Low d. p., base grad.	0.892	0.681	0.280	0.039
High d. p., base grad.	0.974	0.269	0.529	0.201

[a] *Base graduate*: $\mathbf{x}_{ij}^S = 0$ and $\mathbf{x}_{ij}^P = 0$; *average d. p.*: $\xi_j^S = 0, \xi_j^{P(2)} = 0, \xi_j^{P(3)} = 0$
low degree programme: $\xi_j^S = -2s(\xi^S), \xi_j^{P(2)} = -2s(\xi^{P(2)}), \xi_j^{P(3)} = -2s(\xi^{P(3)})$
high degree programme: $\xi_j^S = +2s(\xi^S), \xi_j^{P(2)} = +2s(\xi^{P(2)}), \xi_j^{P(3)} = +2s(\xi^{P(3)})$
$s(\cdot)$ denotes estimated standard deviation

$m = 1, 2, 3$, for the *Principal* equations. Given that the random effects are normally distributed, the fictitious 'low' ('high') degree programme is defined by setting each random effect to minus (plus) twice the corresponding estimated standard deviation.

As for the *P* model, the degree programme variances are significant and of substantive importance: in fact, the last two rows of Table 4 show that, given the observed covariates, the predicted probabilities of the categories vary considerably among the degree programmes. Moreover, since $\mathrm{Corr}(\xi_j^{P(2)}, \xi_j^{P(3)})$ is high and positive, the two categories 'at work' and 'otherwise' can be considered as jointly opposed to the first one ('at university').

The probability of acquisition during the degree programme is obviously lower in case of *degree not required*, while it is higher for graduates with *honors*, maybe because during the preparation of the thesis they refined some crucial skills relevant for the labour market or because they are willing to take a job only if it is consistent with their skills. Moreover, the probability of acquisition during the degree programme is greater for graduates with a *short degree*, which usually provides a vocational training. The job characteristics *self-employed work*, *managerial post* and *public sector* have little effect on the third category, while they substantially reduce the probability of acquisition at the workplace: this implies an increase in the probability of acquisition at the university that might be due to the high expertise required to start a *self-employed work* or to be hired for a *managerial post*, while for the *public sector* the effect is likely a consequence of the formal recruitment procedures.

The empirical Bayes predictions of the degree programme residuals (Rabe-Hesketh et al. 2004) for the *Principal* model are represented in Fig. 1, where the labels are attached only to the extreme cases. The correlation between the two residuals is high, so the degree programmes can be approximately ordered along a one-dimensional scale: in the left bottom cell there are the degree programmes with the highest conditional probability of acquisition at the university (*Business Economics*, *Electronic Engineering* and *Mechanical Engineering*), while in the top right cell there are the degree programmes with the lowest conditional probability (*Foreign Languages* and *Political Science*). In interpreting the rankings implied by the residuals, it should be considered that an adjustment is made for the observed covariates and that the two extremes of the ranking are characterized by markedly different probabilities of acquisition at the university, as suggested by the last two rows of Table 4. Specifically, the predicted probability of acquisition at the university is 66.5% for a base graduate in *Business Economics* and 29.8% for a base graduate in *Political Science*.

Figure 1 gives some useful information to the university management, which should further investigate the degree programmes in the top right cell in order to understand if such a result is due to a lack of education or to the labour market conditions.

5 Final remarks

The present work has shown how to build a complex polytomous response model for the analysis of graduates' skills. The complexity arises from the hierarchical structure of the phenomenon and from the need to adjust for a possible

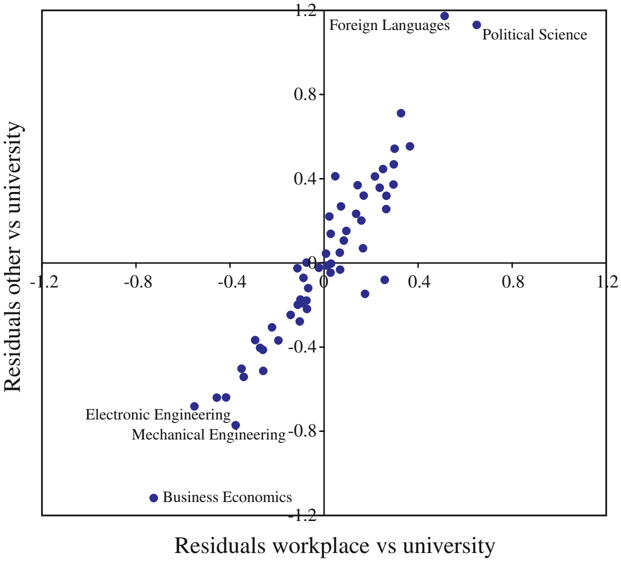

Fig. 1 Empirical Bayes predictions of degree programme residuals from the *Principal* equations of Model 2 (unrelated *S* and *P* equations)

selection bias. In the application the hierarchical structure has a crucial role, while selection bias results negligible. However the outlined methodology can be effectively used in situations where selection bias is an issue.

An alternative approach to adjust for selection bias is based on the copula method (Copas and Li 1997; Bellio and Gori 2003), which allows to perform a sensitivity analysis without relying on a single estimate for the parameters governing the selection mechanism.

Further work is needed to fully understand the implications of multilevel selection mechanisms in polytomous response models, extending the results of Grilli and Rampichini (2005) for the linear case.

The analysis described in the paper is implicity conditional on the employment status of the graduates at the interview, so the results have to be referred only to the employed graduates. In order to evaluate the degree programmes with respects to the skills they give to all the graduates, it is necessary to take into account the possible selection bias induced by the employment status.

References

Bellio R, Gori E (2003) Impact evaluation of job training programmes: Selection bias in multilevel models. J Appl Stat 30:893–907

Copas BJ, Li HG (1997) Inference for non-random samples (with discussion). J Roy Stat Soc B 59:55–95

Grilli L, Rampichini C (2003) Alternative specifications of multivariate multilevel probit ordinal response models. J Educat Behav Stat 28:31–44

Grilli L, Rampichini C (2005) Selection bias in random intercept models. Multilevel Modell Newslett 17:9–15

Heckman JJ (1979) Sample selection bias as a specificaton error. Econometrica 47:153–161

Hedeker D (2003) A mixed-effects multinomial logistic regression model. Stat Medi 22:1433–1446

Little RJA, Rubin DB (2002) Statistical analysis with missing data, 2nd edn. Wiley, New York

McCullagh P, Nelder JA (1989) Generalized linear models, 2nd edn. Chapman & Hall/CRC, London

McFadden D (1973) Conditional logit analysis of qualitative choice behavior. In: Frontiers in econometrics. Academic Press, New York

Rabe-Hesketh S, Skrondal A, Pickles A (2004) `gllamm` manual. U.C. Berkeley Division of Biostatistics Working Paper Series, p 160

Skrondal A, Rabe-Hesketh S (2003) Multilevel logistic regression for polytomous data and rankings. Psychometrika, 68:267–287

Train K (2003) Discrete choice methods with simulation. Cambridge University Press, New York