A Multilevel Chain Graph Model for the Analysis of Graduates' Employment

Anna Gottard, Leonardo Grilli, Carla Rampichini

Statistics Department "Giuseppe Parenti", University of Florence, Italy

Summary. The main goal of the present paper is the analysis of the working position of graduates using multilevel and chain graph models, extended to the case of correlated data. After a brief introduction to multilevel modelling and a description of the conditional independence implied by the model, we describe chain graphs for multilevel models. The model put forward can analyse the factors influencing the graduates' job position, using the data collected on students of the University of Florence who graduated in the year 2000.
Keywords: Chain graph models; Logistic regression; Multilevel models; University system evaluation.

1. Multilevel and chain graph modelling

The university system evaluation requires *ad hoc* methods and statistical models able to capture the complexity of the phenomenon. Such a complexity originates from many facets, such as:

(*a*) the hierarchical structure of the data, entailing a correlation among the observations and requiring the consideration of effects at different levels of the hierarchy. This hierarchical structure (students within classes, classes within study programmes, and so on) is substantial for the analysis and the underestimation of cluster effects and the fact that some of the assumption of the usual regression models are not satisfied, may lead to incorrect standard errors of the estimated coefficients;

(*b*) the presence of variables referring to different moments along the students' careers (e.g. parents education, high school grades, exam grades, graduation grades). This aspect implies a logical and temporal order among the involved variables that must be taken into consideration to shed light on the way students achieve their final result (e.g. getting a

job), and to distinguish between direct and indirect effects.

Multilevel models (Snijders & Bosker, 1999; Goldstein, 2003) allow us to cope with the intra-class correlation and to analyse in a proper way the cluster effects. For such reasons, multilevel models are widely applied in the education evaluation framework. Chain graph models (Cox & Wermuth, 1996) are a useful tool for the representation of the process described at point (b).

In the following, we propose a method for the integration of multilevel and chain graph models that allows:

(i) to properly model the relationships among the probability of finding a job after the degree, the students' careers and their individual characteristics;

(ii) to stress the contribution of the study programme to the student's success in the labour market;

(iii) to distinguish among direct and indirect effects of the background and career variables.

In Section 2, we describe the two-level linear model and its extension to a binary response. In Section 3 the multilevel graph model derived from the integration among the multilevel model and the chain graph model is illustrated. In the fourth Section, we present the data at hand and the main results of the empirical analysis, and in the fifth we conclude by giving some lines for future research.

2. The linear random intercept model

Let us consider a two-level hierarchical structure, where Y_{ij} is the response variable for the i-th subject (first level unit) of the j-th cluster (second level unit), $i=1,2,\ldots,n_j$, $j=1,2,\ldots,J$. For each subject, a vector $\mathbf{X}_{ij}$ of individual (e.g. gender, high school rank) and cluster (e.g. number of enrolled students for each program course) variables is available.

Let us assume Y_{ij} is a continuous variable. If the relationship between the response Y_{ij} and the covariates $\mathbf{X}_{ij}$ is linear, it is possible to specify the following linear random intercept model:

$$\begin{aligned} Y_{ij} &= \alpha_j + \boldsymbol{\beta}' \mathbf{X}_{ij} + \varepsilon_{ij} \\ \alpha_j &= \alpha + U_j \end{aligned} \tag{1}$$

where ε_{ij} are the first level residuals, while U_j are the second level ones. Residuals are assumed to be independent and normally distributed, with zero mean and variances $\mathrm{Var}(\varepsilon_{ij})=\sigma^2$ at subject level and $\mathrm{Var}(U_j)=\tau^2$ at cluster level. Moreover, as it is common in regression models, the correlations among residuals at both levels and the covariates are assumed null. The independence hypotheses among the observations following from this model are:

$$Y_{ij} \not\perp\!\!\!\perp Y_{i'j} \mid \mathbf{X}, \qquad \forall i \neq i', \forall j$$
$$Y_{ij} \perp\!\!\!\perp Y_{i'j'} \mid \mathbf{X}, \qquad \forall i, i', \forall j \neq j', \tag{2}$$

where $\mathbf{X}=\{\mathbf{X}_{ij} : i = 1,2,\ldots,n_j,\ j=1,2,\ldots, J\}$.

It can be seen from relationships (2), conditionally on the covariates **X**, that observations from different clusters are independent, while observations belonging to the same cluster are dependent. The intraclass correlation coefficient

$$\rho = Corr(Y_{ij}, Y_{i'j'}) = \begin{cases} 0 & \text{se } j \neq j' \\ \dfrac{\tau^2}{\tau^2 + \sigma^2} & \text{se } j = j' \end{cases}$$

measures the within-cluster dependence. Moreover, conditionally on the covariates **X** and the second-level errors U_j also the observations belonging to the same cluster are independent:

$$Y_{ij} \perp\!\!\!\perp Y_{i'j} \mid \mathbf{X}, U_j, \qquad \forall i \neq i', \forall j \tag{3}$$

For each cluster j the joint probability distribution can be factorised as follows[1]:

$$\begin{aligned} & f(\mathbf{y}_j, u_j, \mathbf{x}) = f(\mathbf{y}_j \mid u_j, \mathbf{x}) f(u_j \mid \mathbf{x}) f(\mathbf{x}) \\ & = f(\mathbf{y}_j \mid u_j, \mathbf{x}) f(u_j) f(\mathbf{x}) \\ & = \left[\prod_{i=1}^{n_j} f(y_{ij} \mid u_j, \mathbf{x}) \right] f(u_j) f(\mathbf{x}) \end{aligned} \tag{4}$$

where $\mathbf{y}'_j = \{y_{1j}, y_{2j},\ldots, y_{n_j j}\}$. Actually, from the independence among u_j and **X** it follows that $f(u_j \mid \mathbf{x}) = f(u_j)$, while for the conditional independence (3) the conditional density $f(\mathbf{y}_j \mid u_j, \mathbf{x})$ corresponds to the product of the n_j individual densities.

In general, the effect of a first-level covariate can be decomposed into two parts: within and between clusters, according to the covariate variance decomposition (Snijders & Bosker, 1999). For instance, in the linear model with only one covariate X, the OLS total coefficient $\hat{\beta}_T$ is a linear combination of the coefficient of the regression among cluster means $\hat{\beta}_B$ and of those within clusters $\hat{\beta}_W$:

$$\hat{\beta}_T = \hat{\eta}_X^2 \cdot \hat{\beta}_W + \left(1 - \hat{\eta}_X^2\right) \cdot \hat{\beta}_B \tag{5}$$

---

[1] In factorisation (4) it is the same if one considers the matrix of the covariates **X** of all the individuals or only the sub-matrix $\mathbf{X}_j$ of the covariates of the subjects of the j-th cluster, so, for the sake of simplicity, only the sub-matrix $\mathbf{X}_j$ is considered.

where $\hat{\eta}_X^2$ is the correlation ratio of X. As a result $\hat{\beta}_T$ assumes an intermediate value among $\hat{\beta}_B$ and $\hat{\beta}_W$. The *between* and *within* coefficients have a different interpretation and they can take opposite values. Thus, the total coefficient can be non-significant, whilst the *between* and *within* coefficients are significant but opposite in sign.

Therefore, it is better to specify the model in order to estimate both the *between* and the *within* coefficients of each subject-level covariate. A way to perform such an estimation is to insert in the model both the covariate X_{ij} and the cluster mean $\overline{X}_{.j}$:

$$Y_{ij} = \ldots + \beta_W X_{ij} + (\beta_B - \beta_W)\overline{X}_{.j} + \ldots \; . \tag{6}$$

In model (6), the coefficient of the cluster mean represents the difference among the *within* and *between* coefficients, so the usual test for the $\overline{X}_{.j}$ coefficient is to be interpreted as a test for the difference among the *within* and *between* coefficients. If the $\overline{X}_{.j}$ coefficient is not significant, the distinction among the *between* and *within* effects can be ignored, leaving among the predictors only the raw covariate X_{ij}.

Note that the insertion of the cluster mean of a covariate allows us to eliminate the possible correlation among the covariate and the random effects U_j (Snijders & Bosker, 1999).

When the response variable is binary, it is possible to assume the linear random intercept model (1) for a continuous latent variable that generates the observed binary variable Y_{ij}^{obs} as follows:

$$Y_{ij}^{obs} = \begin{cases} 0 & \text{if } Y_{ij} \le 0 \\ 1 & \text{if } Y_{ij} > 0 \end{cases}.$$

Assuming a standard logistic distribution for the first-level errors ε_{ij} (equation 1), the logistic random intercept model for the probability of response is:

$$P(Y_{ij}^{obs} = 1 \mid u_j, \mathbf{x}_{ij}) = \frac{1}{1 + \exp(-(\alpha + \boldsymbol{\beta}'\mathbf{x}_{ij} + u_j))} \cdot$$

Many multilevel analysis textbooks (e.g. Snijders & Bosker, 1999) describe the properties of such a model.

2.1 Graphical models for hierarchical data

Graphical models are a class of probabilistic models on a set of random variables whose conditional independence structure can be represented by a graph. A graph is a mathematical object consisting of two sets: a set of nodes and a set of undirected or directed edges (arrows) between nodes. In the graph asso-

ciated with a particular model, each node corresponds to a random variable in the model. Usually, a discrete random variable is depicted as a circle, **O**, a continuous random variable is represented by a dot, ● , an edge between two nodes stands for an association between the variables or, more precisely, the absence of a connection between two nodes indicates conditional independence between the corresponding variables.

A chain graph admits both undirected and directed edges, and (partially) directed cycles are forbidden. This implies that, starting from a node, it is not possible to go back to it through the edges and arrows of the graph.

In a chain graph, nodes can be partitioned into an ordered sequence of blocks. Nodes in a same block can be connected by undirected edges, while only arrows can connect nodes in different blocks. The arrows stay for an asymmetric relationship between the variables, while undirected edges indicate a symmetric relationship.

A chain graph model for a set of random variables is specified by assuming that their joint distribution satisfies the chain graph Markov properties. Therefore, chain graph models are a class of probabilistic models allowing for both symmetric and asymmetric relationships between variables, assuming a sort of logical and temporal order among the variables. Each variable in a block has to be considered explanatory of the variables in the following blocks.

Because of the partial ordering among the variables, it is possible to distinguish the set of pure explanatory variables (usually in the last block on the right), from the set of pure response variables (last block on the left) and from the set of intermediate variables, that are both explanatory and responses, positioned in the intermediate blocks. In this work, we refer to the chain graph Markov properties proposed by Lauritzen & Wermuth (1989) and Fridenberg (1990). These properties are usually termed LWF Markov properties.

An important Markov property for chain graphs is the global Markov property, which is based on the definition of the 'moral graph'. Starting from a given chain graph, a moral graph can be obtained by connecting parents of common children (or children belonging to the same chain component), and then converting all the arrows into undirected edges. See, for example, Figure 1, where the graph in (b) is the moral graph of the chain graph in (a). More details can be found in Lauritzen (1996).

The global Markov property combines the concept of conditional independence to that of separation between nodes in the moral graph. If, in the moral graph, a set of nodes S separates the nodes in A from the nodes in B, that is, each path from A to B passes by some node in S, then $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B | \mathbf{X}_S$, where $\mathbf{X}_k$ is the vector of random variables represented by nodes in k, $k = A, B, C$. Markov properties induce a factorization of the joint distribution of the variables in a model, which is useful for the inferential procedures (Lauritzen, 1996).

Graphical models are apt to represent the conditional independence relationships among a set of variables, if statistical units are independent. This as-

sumption is no longer valid whenever the data have a hierarchical structure. Gottard & Rampichini (2004) propose to overcome this issue by representing in a graph all the variables of a generic group j, given that the J groups are assumed independent and identically distributed. For instance, in a two-level model all the variables of the vector ($\mathbf{Y}_j$, U_j, X_{1j}, …,X_{Kj}) of the j-th group are represented in the graph. Whatever the cluster sizes n_j are, it is enough to depict the minimal sub-graph suitable to read the conditional independence structure from the graph. For instance, in the case of a two-level structure, only two elementary observations have to be included in the minimal sub-graph.

This solution implies additional definitions. An *individual node* is a node that represents a random variable for a specific statistical unit. A *grouping latent node*, is a node representing a latent random variable U_j being a separator between the individual nodes, such that, $Y_{ij} \perp\!\!\!\perp Y_{i'j} | U_j$. Such node is represented by the symbol ⊘. A *deterministic node* is a node representing a random variable whose conditional distribution is degenerate. This node is represented with a double line block.

The conditional independence structure of a two-level random intercept model can be represented by a chain graph where: the last block on the left, made up of pure response variables, contains two *individual nodes* and the second-last block contains a *grouping latent node*. LWF Markov properties can be used to encode the conditional independence structure of such a graph.

Figure 1 shows an example of a chain graph for a two-level random intercept model with only one explanatory variable. Therefore, the main advantage of this proposal is that the usual LWF Markov properties and the factorization criterion are valid in such chain graphs. For instance, in Figure 1, the pairwise chain graph Markov property suggests that the latent variable U_j is marginally independent from the explanatory variable Z_1. Moreover, due to the global Markov property, looking at the moral graph in (*b*), one can see that U_j is not independent of Z_1 conditionally on the response variables Y_j.

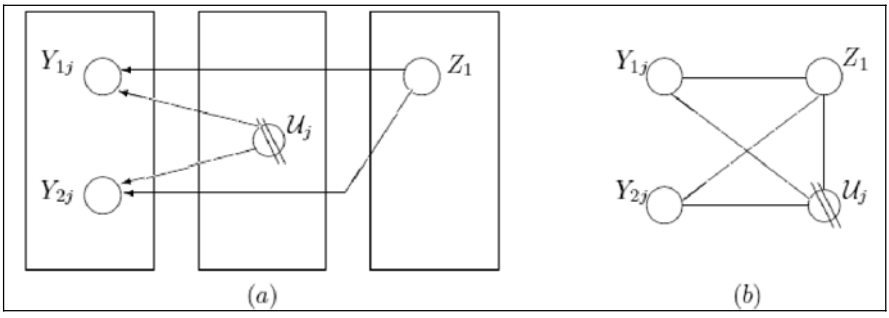

Figure 1. Example of (a) two-level random intercept graphical model, and (b) corresponding moral graph.

3. A multilevel graphical chain applied to graduates' employment

We adopted the multilevel graphical chain model described in Section 2.1 to analyse the data collected on the students who graduated in year 2000 at the University of Florence. They had been interviewed at approximately two years from the attainment of the degree. We analysed the occupational condition at the interview for the graduates who, at that time, were working or were seeking for an occupation. Our aim is to determine the factors that influence job finding, with reference to both individual characteristics and type of degree.

The data include 2,917 graduates employed or seeking a job: 46% had an occupation with tenure, the other 54% was unemployed or with a temporary occupation. The graduates under consideration had 56 types of degrees, with a number of graduates per course ranging from 4 (Chemistry) to 504 (Architecture), and a median of 22. Graduates (first level units) and degree programmes (second level units) thus characterize the hierarchical structure.

The response variable is binary with a value of 1, if the graduate has a stable occupation to the date of the interview, and 0 otherwise.

The use of a graphical model allows the study of the joint distribution of all the involved variables, bringing to light direct and indirect relationships. It is just this point that differs our contribution from others where the response variable is studied through a multilevel model, but the relationships among the explanatory variables are not modelled (e.g. Chiandotto & Bacci, 2004).

The specification of the graphical chain model requires, first, the covariates to be ordered according to the prior knowledge of the phenomenon, following a logical and/or time order. The variables used in our analysis[2] and their block ordering are shown in Table 1.

The variables in block 5 are the cluster means of the corresponding subject level variables. As shown in equation (6), the insertion of the cluster mean allows to decompose the total effect of a variable into a *between* and a *within* component. Since the conditional distribution of a cluster mean degenerates, that is $f(\overline{x}_j \mid x_{1j},\ldots,x_{n_j j}) = 1$, a cluster mean is represented in the graph as a deterministic node, located in a block following the block containing the corresponding individual-level variable and preceding the block containing the cluster latent variable.

Block 6 contains only the random effect U_j, represented as a grouping latent node, whose role is to model the variance of the response.

---

[2] The variables were selected on a logical basis and thanks to past analyses. Our aim was to design a relatively simple model able to catch the key features of the process under scrutiny.

Table 1. Block ordering of the variables. Graduates of the year 2000, University of Florence.

Block	*Variable*	*Description*
1 exogenous	MALE	Gender: 1=male, 0=female
	MOTHER EDUC.	Mother's education: secondary school or low, high school (ref. cat.), degree
2 intermediate	LICEO	High school: 1=lyceum, 0=other
	SCH. MARK	High school final mark: 36-60 (mean=48.0)
3 intermediate	DIPLOMA	Type of degree programme: 1=diploma (3 years), 0=laurea (usually 4 years)
4 intermediate	AGE	Age at graduation: 21-50 (mean=27.6)
	UNIV. MARK	Average exam mark: 18-30 (mean=26.8)
5 cluster means	c.m. SCH. MARK	Degree programme mean of SCH. MARK
	c.m. AGE	Degree programme mean of AGE
	c.m. UNIV. MARK	Degree programme mean of UNIV. MARK
6 cluster latent node	$U_j \sim N(0,\tau^2)$	Degree programme latent variable
7 response	EMPLOYED	Employment: 1=stable occupation, 0=other

The block ordering shown in Table 1 and the independence assumptions of the multilevel model imply the following factorization of the joint distribution:

$$\begin{aligned} f(\mathbf{y}_j, u_j, \mathbf{x}) &= f(\mathbf{y}_j \mid u_j, \mathbf{x})\, f(u_j)\, f(\mathbf{x}) \\ f(\mathbf{x}) &= f(\mathbf{x}_{[4]} \mid \mathbf{x}_{[3]}, \mathbf{x}_{[2]}, \mathbf{x}_{[1]})\, f(\mathbf{x}_{[3]} \mid \mathbf{x}_{[2]}, \mathbf{x}_{[1]})\, f(\mathbf{x}_{[2]} \mid \mathbf{x}_{[1]}) \end{aligned} \tag{7}$$

where $\mathbf{X}_{[k]}$ denotes the variables of the k-th block, k=1,2,3,4, for example $\mathbf{X}_{[2]}$={LICEO, SCH. MARK}.

The fitting of the multilevel graphical chain model that corresponds to factorization (7) requires the fitting of four regression models, some of which have a multivariate response.

Given the alternation between categorical and continuous variables in consecutive blocks, we adopted the estimation procedure by Cox & Wermuth (1996). The procedure consists in fitting, for every endogenous variable, a univariate regression model whose explanatory variables are those in the same block and those in the previous ones. When the endogenous variable is continuous, a linear regression is fitted, while in the binary case a logistic model is used. The multilevel (random intercept) model is the one for the response variable EMPLOYED.

All the regression models are fitted by maximum likelihood. For the random intercept logistic model, the likelihood is approximated through adaptive Gaussian quadrature using the `gllamm` command of *Stata* (Rabe-Hesketh *et al.*, 2004).

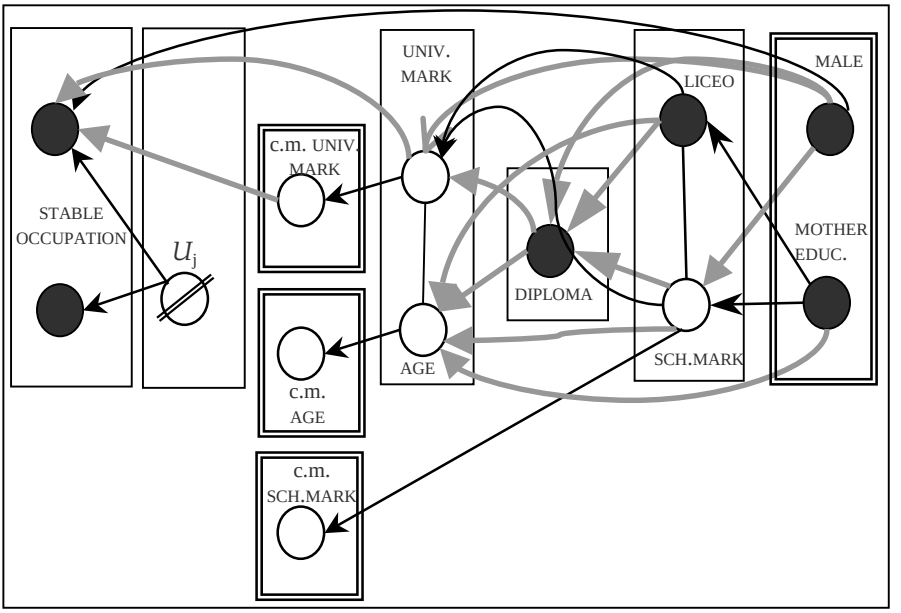

Figure 2. Multilevel graphical chain model: black (grey) arrows are positive (negative) effects at 10% significance level. Graduates of the year 2000, University of Florence.

The resulting graphical model is shown in Figure 2. The arrows are drawn when the *p*-value of the corresponding regression coefficient is less than 0.10.

In order to ease the reading of the graph the positive effects are represented by black arrows and the negative ones by grey arrows. The arrows pointing to the cluster means represent a deterministic relationship (as suggested by the double line block). The sign of the relationship between the response variable and the latent node U_j is not identifiable. Moreover, the two individual nodes in the final block are identically distributed, that is the dependence structure of the two nodes is the same even if, in order to simplify the reading of the graph, the arrows have been only traced for one of the two individual nodes.

The estimates concerning the models for the intermediate variables are not shown, as the essential information for the aims of the analysis is encapsulated in the graph of Figure 2. The estimates concerning the random intercept logistic model for the probability of a stable occupation are reported in Table 2.

Figure 2 shows that the response variable EMPLOYED directly depends only on MALE and UNIV. MARK. Therefore, the variables MALE and UNIV. MARK constitute a separator set between the response variable and the other covariates, in the sense that EMPLOYED is independent from AGE, DIPLOMA, LICEO, SCH. MARK and MOTHER EDUC. conditional on MALE and UNIV. MARK.

If only the model for the response variable was fitted, one would conclude that the factors relevant for employment were MALE and UNIV. MARK. Actu-

Table 2. Random intercept logistic model for the probability of stable occupation. Graduates of the year 2000, University of Florence.

Parameter	*Estimate*	*Standard error*	*p-value*
Intercept	4.925	6.024	0.414
MALE	0.372	0.087	0.000
MOTHER EDUC. (secondary sch.)	-0.020	0.090	0.820
MOTHER EDUC. (degree)	-0.115	0.135	0.397
LICEO	-0.089	0.087	0.308
SCH. MARK	-0.003	0.007	0.595
c.m. SCH. MARK	0.070	0.050	0.164
DIPLOMA	0.612	0.396	0.122
AGE	-0.005	0.016	0.737
c.m. AGE	-0.034	0.106	0.747
UNIV. MARK	-0.055	0.030	0.069
c.m. UNIV. MARK	-0.221	0.119	0.063
Cluster residual variance τ^2	0.510	-0.148	

ally, also other covariates influence the result, even if indirectly. This highlights the potentiality of the graphical chain model: given a block ordering of the variables, such a model allows to study the phenomenon taking into account the whole dependence structure.

The presence of an arrow between the cluster mean of UNIV. MARK and the response variable points out that UNIV. MARK has distinct *within* and *between* effects. From expression (6) it follows that the *within* effect is the coefficient of the variable at the subject level (-0.055), while the *between* effect is the sum of the coefficient of the variable at the subject level and the coefficient of the cluster mean (-0.221), and it turns out to be -0.276. Both the effects are negative, but the *between* effect is stronger, so the negative effect of the mark is largely due to the degree programme: a higher mark is associated with a lower probability of a stable occupation, because high marks are more frequent in degree programmes that usually yield modest occupational opportunities, e.g. the Humanities.

The effect at the subject level, even if significant and negative, is low. The effect at the subject level may be negative because the graduates with higher marks have greater ambitions and therefore are more demanding in job search.

If *between* and *within* effects are not disentangled, namely if the model contains UNIV. MARK without its cluster mean, the coefficient is -0.068 (s.e. 0.029). Such a coefficient is the total effect of the variable UNIV. MARK and therefore is difficult to interpret. However, a reader not accustomed to multilevel analysis is likely to misinterpret this effect as an effect at the subject level.

The unobserved factors at degree programme level are relevant: the likelihood ratio test comparing the models with and without random effects is sig-

nificant (test statistic 108.8 with 1 degree of freedom) and the intraclass correlation coefficient ρ=0.134, that is 13.4% of the unexplained variance is at the degree programme level. Such value is rather high given the kind of model and the field of application.

The probability of stable occupation $P\left(Y_{ij}=1|u_j,\mathbf{x}_{ij}\right)$ is a function of the subject level covariates $\mathbf{X}_{ij}$ and the degree programme random effect U_j. To assess the role of the random effect, let us consider a particular graduate who is a male, has a mother with high education, attended a high school other than LICEO, had a high school final mark 48, obtained a 'laurea' degree, graduated at 27, and had an average examination mark of 26. For such a graduate the probability of a stable occupation is 0.56 if graduated in a mean course (u_j = 0), 0.72 if graduated in a course yielding high occupational chances ($u_j=+2\hat{\tau}$) and 0.38 if graduated in a course yielding low occupational chances ($u_j=-2\hat{\tau}$).

After parameter estimation, the residuals $\hat{u}_j$ can be calculated with the Empirical Bayes method (Snijders & Bosker, 1999). The degree programmes with a positive (negative) residual yield graduates with a probability of a stable occupation higher (lower) than predicted. The ranking based on the residuals has Nursing in the first position and Physics in the last position.

Figure 3 can be used to compare the residuals in pairs: two residuals are significantly different (at 95% average confidence level) if and only if the corresponding intervals do not overlap. The interval length is a decreasing function of the sample size of the degree programme: two extreme examples are

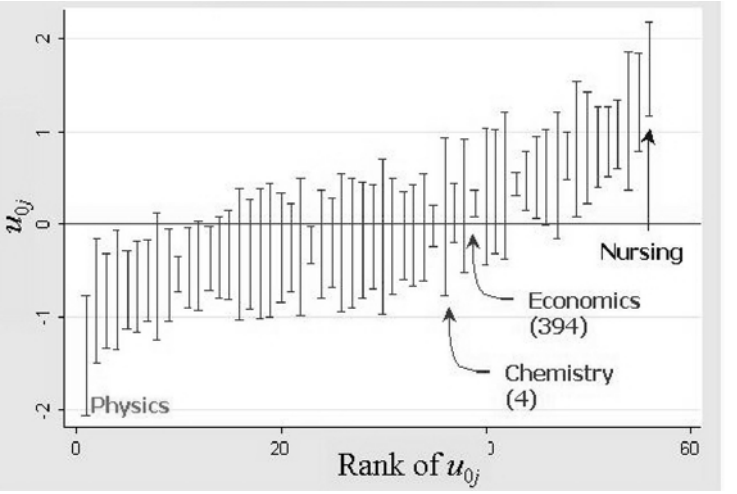

Figure 3. Intervals for pairwise comparisons between residuals at the degree programme level (95% average confidence level). Graduates of the year 2000, University of Florence.

Chemistry (4 graduates) and Economics (394 graduates). For a study programme with few graduates the considerable length of the interval hinders the comparison with the other programmes, as the differences are nearly not significant.

The residuals $\hat{u}_j$ incorporate all the unobserved factors at the study programme level, so they can be interpreted as a measure of external effectiveness of the study programme, though not adjusted for the conditions of the labour market. For example, the job opportunities yielded by Nursing depended not only on the quality of the programme, but also on the labour market needs (see, about this, Chiandotto & Grilli, 2003.

4. Final remarks

In this paper, we presented and applied a method of analysis based on the integration between chain graph and multilevel models. This method has the advantage to make explicit the assumptions on the ordering of the variables and on the conditional independences underlying the multilevel model. Moreover, the use of the graph helps to visualize the direct and indirect effects on the response variable and to read in a simple and direct way the conditional independences among the variables. It is not necessary that the variables follow a joint Gaussian distribution, nor a Conditional Gaussian one, so the estimates depend on the assumed block ordering of the variables. With a different block ordering the estimates could change and so the conditional independences. However, in our application, the adopted block ordering was plausible because it follows a logical and/or time ordering.

The potentiality of this class of models is still to be explored. It would be useful to extend the methodology in the following directions: several multilevel regressions in the same graph, modelling of the process of formation of the clusters, regression of cluster-level variables on individual-level variables.

The application based on data from the graduates employed or seeking a job, so the considered joint distribution is conditioned on this subset of graduates. For example, the relationship between the type of degree programme chosen by the student (DIPLOMA) and their characteristics (MALE, MOTHER EDUC., LICEO, SCH. MARK) is not referred to all the students enrolling in university, but only to students who eventually graduated and searched for jobs.

This choice follows the need to have a random sample from the joint distribution, as is customary in graphical models. The consideration of a wider sample, such as a cohort of freshmen, requires an adequate representation of the selection process in the graphical model: for example, for the freshmen who do not graduate, all the variables associated with graduation and successive job search are not observable, and they cannot be treated as missing at random. The development of multilevel graphical models able to represent

also the process of selection of the statistical units is an interesting topic for future search.

References

CHIANDOTTO B., BACCI S. (2004) Un modello multilivello per l'analisi della condizione occupazionale dei laureati dell'Ateneo fiorentino. In: CROCETTA C. (ed) *Modelli di statistici per l'analisi della transizione Università–lavoro*, Cleup, Padova: 211-234.

CHIANDOTTO B., GRILLI L. (2003) *La domanda di lavoro nella provincia di Firenze: Analisi integrative sui dati dell'indagine Excelsior 2003*, University of Florence (http://www.unifi.it/aut_dida/indexval.html).

COX D., WERMUTH N. (1996) *Multivariate Dependencies. Models, Analysis and Interpretation*, Chapman & Hall, London.

FRYDENBERG M. (1990) The chain graph Markov property, *Scandinavian Journal of Statistics*, **17**: 333-353.

GOLDSTEIN H. (2003) *Multilevel Statistical Models (3rd edition)*, Arnold, London.

GOTTARD A., RAMPICHINI C. (2004) Chain Graphs for Multilevel Models, *Working Paper*, Department of Statistics 'G. Parenti', n. 8/2004, Florence.

LAURITZEN S. (1996) *Graphical Models*, Clarendon Press, Oxford.

LAURITZEN S., WERMUTH N. (1989) Graphical models for the association between variables, some of which are qualitative and some quantitative, *Annals of Statistics*, **17**: 31–57.

RABE-HESKETH S., SKRONDAL A., PICKLES A. (2004) *GLLAMM Manual*, U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 160.

SNIJDERS T., BOSKER R. (1999) *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modelling*, Sage, London.