An IRT-MIMIC model for the analysis of university student careers

Bruno, Bertaccini, Leonardo Grilli & Carla Rampichini

Napoli, 25-26 November 2013

Outline
- Analysis of the performance of university students
- Goals
- Data
- Models
- IRT-MIMIC models results
- Probability of passing exams
- Exam grades
- Concluding remarks and work in progress
- References

Main goals
- We analyze the performance of university students careers during the first year
  - This is important for both student tutoring and course organization.
- The main goal of the paper is to compare the exams in terms of difficulty, discrimination and use of the grades
- Moreover, the paper aims at assessing how student careers depend on student and course characteristics

Student ratings
- One of the aims is to evaluate if the course quality is associated with the performance at exams
  - The questionnaire on student satisfaction is filled before taking the exam through a web system which first requires authentication and then ensures anonymity
  - Students express ratings on a ten-point scale on several aspects of the course, including the item “Teacher’s clarity” considered here
  - All the courses have parallel classes offered to groups of students defined by the first letter of the surname; each version of the course has its own teacher with corresponding student ratings
  - Student ratings are summarized by computing the average rating for each teacher

Data
- 808 freshmen enrolled in a.y. 2011/2012, School of Economics, Univ. of Florence
  - In 1st year, each student has 7 compulsory courses

Modelling strategy
- The analysis exploits an Item Response Theory (IRT) approach
  - where exams are treated as items and the student ability is a latent variable.
  - In particular, we consider
    - the probability of passing the exams
    - the grades obtained in passed exams
  - Course characteristics directly affect the items,
  - whereas student characteristics indirectly affect the items via the latent ability, even if some direct effects are allowed by fitting a MIMIC model with DIF.

See the path diagram in slide 8
Model for passing the exams

- 2-Parameter Logistic model for the responses
  \[ \text{logit}(P(Y_{ni} = 1 | X_{ni}, \theta_i)) = \alpha_i + \beta X_{ni} + \lambda_i \theta_i \]

- \(Y_{ni}\) → binary variable =1 if exam passed by student \(n\)
- \(X_{ni}\) → course covariates (and interactions for DIF)
- \(\theta_i\) → latent variable representing student ability
- \(\alpha_i\) and \(\lambda_i\) → easiness and discrimination of exam \(i\)

Structural model for student ability (MIMIC)

- \(\theta_n\) → student covariates
- \(\epsilon_n\) → Normal error term representing residual student ability

Results: model for passing the exams

For a student with average ability, the easiest exam is History, and the hardest one is Economics EC.

The probability to pass an exam depends on the ability of the student through the discrimination parameter: e.g., Statistics has \(\lambda_i = 1.43\) → for a given increase in the student ability, the logit of the probability of success increases 43% more for Statistics than for Management (\(\lambda_i = 1\))

Effects of student characteristics on the ability to pass the exams

- HS grade and HS type have significant effects
- students with a better high school mark and coming from a Lyceum have higher ability
- Lyceum has both direct and indirect effects on the probability of passing the exams

Direct and indirect effects

- In the MIMIC model the covariate Lyceum has a significant direct effect (DIF) for some courses: the better performance of students coming from Lyceum is attenuated for Accounting, magnified for quantitative courses, i.e., Economics EC, Math EC, Math MG and Statistics.
- Example: let us consider a course with average student rating equal to 7, and a student with HS grade=80 and average residual ability
  - Probability of passing Statistics:
    - 9% if not Lyceum, 11% if Lyceum
    - Effect of Lyceum: total= +12%, indirect= +2%, direct= +10%
  - Probability of passing Accounting:
    - 9% if not Lyceum, 21% if Lyceum
    - Effect of Lyceum: total= +12%, indirect= +32%, direct= -20%

Exam grades

The observed distribution of exam scores shows peaks at the extremes and an irregular use of the scale; moreover, the maximum score is qualitative (’30 with honors’)

Thus we treat exam scores as ordinal responses, instead of continuous.

To avoid sparseness, we define the exam grade by aggregating adjacent scores →
The IRT-MIMIC model for exam grades is the same as the one for passing the exams, with the exception of the specification for the response ($Y_{ic}$ grade from 1 to 7) which is a cumulative logit, yielding an IRT Graded Response Model (Samejima 1969)

$$\logit[P(Y_{ic} \leq c | X_{ic}, \theta_i)] = \gamma_{ic} - (\beta X_{ic} + \lambda_j \theta_i), c = 1, \ldots, 6$$

(note: for any i, the intercept $\alpha_i$ is replaced by 6 thresholds $\gamma_{ic}$ )

However, the estimation datasets are different:

- The model for passing the exams is fitted using all the compulsory exams (5656 exams, 808 students)
- The model for exam grades is fitted using only passed exams (2189 exams, 615 students $\rightarrow$ 24% students excluded as they did not pass any exam)

Results: model for exam grades

- The model for passing the exams and the model for the grades yield similar results
- The estimated discrimination parameters are in good agreement (correlation = 0.75) $\rightarrow$ the implicit grade corresponding to a failed examination is assigned in a way that is somewhat consistent with the use of the grading scale for a successful examination

In the MIMIC part, the effects of Lyceum and HS grade and their contribution to the reduction of the residual variance of the ability (37%) are similar

However, the course quality assessed through the average student ratings on teacher’s clarity has a significant effect on the probability of passing the exams, but not on the grades

Effects of student characteristics on the ability

- HS grade and HS type have significant effects
- students with a better grade and coming from a Lyceum have higher ability

Main findings

- Passing an exam and the obtained grade are two distinct aspects that should be analysed separately
- The model for the grades highlights peculiarities in the use of the grading scale, e.g.
  - Anomalous proportion of lowest grades in Math MG
  - Teachers of Accounting tend to rule out extreme grades
- The High School type and HS grade are good predictors for both passing the exams and obtained grades
- The average student rating on teacher’s clarity has a significant effect on the probability of passing the exams, but not on the grades $\rightarrow$ better teachers may help reduce the drop-out rate (though we need experiments to reliably evaluate the impact of interventions)

Final remarks

- IRT-MIMIC modelling gives more insight w.r.t. traditional approaches relying on a summary measure of student performance (e.g. number of passed exams)
- It highlights peculiarities of the exams, such as different discrimination or anomalous use of the grading scale
- It estimates the effects of course and student characteristics, allowing for exam-specific effects

The results are valuable to plan interventions in the degree programme organization and to tailor the student tutoring service.