KU LEUVEN

FACULTY OF SCIENCE

Generalized Linear Latent and Mixed Models:

method, estimation procedures, advantages, and applications to educational policy.

Jose Manuel Rivera Espejo

Supervisor: Prof. Geert Molenbegrhs
Affiliation *(optional)*

Co-supervisor: Prof. Wim Van den Noortgate *(optional)*
Affiliation *(optional)*

Thesis presented in fulfillment of
the requirements for the degree of
Master of Science in Statistics and Data Science
for Social, Behavioral and Educational Sciences

Academic year 2020-2021

Dedication

To Manuel, for being my friend and father.
To Margarita, Susan, and Marysu, for their relentless encouragement.
To Ana, for showing me the value of family, here in this moorland.
To both of you, as you are always in my mind.
And to all that knowingly or not, help me to get here.
I am lucky due to all of you.
I hope I make you all proud.

A Manuel, por ser mi amigo y mi padre.
A Margarita, Susan y Marysu, por su incansable aliento.
A Ana, por mostrarme el valor de la familia, aquí en este páramo.
A ustedes dos, que siempre las tengo en mente.
Y a todos los que sabiendolo o no, me ayudaron a llegar aquí.
Soy un suertudo gracias todos ustedes.
Espero llenarlos de orgullo.

Acknowledgment

(in the works)

Abstract

(in the works)
Keywords:

Contents

List of Figures

List of Tables

Abbreviations

GLLAMM	Generalized Linear Latent and Mixed Model.
SEM	Structural Equation Model.
GLM	Generalized Linear Model.
EFA	Exploratory Factor Analysis.
CFA	Confirmatory Factor Analysis.
IRT	Item Response Theory.

Symbols

J	total number of subjects.
j	index of specific subject.
I	total number of items.
i	index for the specific item.
$M_{(l)}$	total number of latent variables at level l.
m	index for the specific latent variable.
L	total number of levels (clusters).
l	index for the specific level (cluster).
$\mathbf{V}$	vector of the linear predictors.
$\boldsymbol{\beta}$	vector of fixed effects, for the I items.
$\mathbf{X}$	design matrix for the $\boldsymbol{\beta}$ parameters.
$\eta_{mj}^{(l)}$	latent variable, at the respective indices.
$\boldsymbol{\lambda}_m^{(l)}$	vector of loadings, at the respective indices.
$\mathbf{Z}_m^{(l)}$	design matrix for the $\boldsymbol{\lambda}_m^{(l)}$ parameters.

Chapter 1

Introduction

1.1 Preliminar considerations

The short and long term benefits of effective teaching practices can be observed throughout the literature: improvements in student achievements (Rockoff; 2004; Rivkin et al.; 2005; Duflo et al.; 2009; Hanushek and Rivkin; 2012; Muralidharan and Sundararaman; 2013; Chetty et al.; 2014a; Araujo et al.; 2016); development of executive functions (Araujo et al.; 2016), increased college attendance, higher salaries, and a lower possibility of premature parenthood (Chetty et al.; 2014b), among others. Similarly, the literature has shown most of the negative impacts resulting from the presence of teacher shortages[1] (Duflo et al.; 2009; Muralidharan and Sundararaman; 2013; Chetty et al.; 2015; Ayala; 2017; Marotta; 2019) or ineffective teaching practices (Hanushek and Rivkin; 2012).

However, while the evidence have a solid methodological support, Hanushek and Rivkin (2006) have indicated that some of the proxy variables, used in the methods, are not consistently related to either teacher effectiveness or quality of instruction, examples of such are: out of field teaching[2] (Ingersoll; 1998; Dee and Cohodes; 2008; Bertoni et al.; 2020); teaching hours (Bruns et al.; 2015); years of experience or educational degree (Rockoff; 2004; Rivkin et al.; 2005; Clotfelter et al.; 2006, 2007; Hanushek and Rivkin; 2012); among others.

Consequently, given that most of the measured teaching factors are proxies, and that the effects estimated from such variables lack consistency, Hanushek and Rivkin (2012) have pointed out that the analysis of teacher effectiveness has largely turned away, from attempts to identify the teacher's specific characteristics, to focus its attention into measuring the direct relationship between them and the student outcomes[3]. For that reason, considerable uncertainty is still present in the literature, regarding exactly which aspects

[1]Bertoni et al. (2020) defined it as the context in which the teacher's supply, i.e. the number of available teachers in the system, is less than its demand. The authors further elaborate that one of the causes of these shortages is related to the applicants' lower quality or due to their faulty initial training, implying that the shortage can also be conceived as the lack of good quality teachers. The evidence of such shortage has been more prevalent, but not decisive, with temporary teachers, as they are usually associated with inferior attributes, compared to their contracted counterparts

[2]Medeiros et al. (2018) defines it as teachers that are currently teaching a subject in which they are not specialized or do not have the appropriate certificate.

[3]The method is known as value-added analysis, and it is based on the perspective that a good teacher is one who consistently gets higher achievement from students after other determinants of such are controlled for. For a more detailed explanation of the method refer to Scherrer (2011).

of teachers are key for the student's learning and whether those qualities can be measured (Rockoff; 2004; Clotfelter et al.; 2006).

However, because the evidence still largely supports the perception that teachers are the main driver behind the student's learning processes, any educational authority need to have, among their main agenda points, the design of an assessment system that can attract, select, develop, and retain the most effective ones (Elacqua et al.; 2018), and in order to do so, the definition of an Educational Performance Standard (EPS) is a necessity. With an EPS, rooted in the country's context, the authorities can now set clear expectations about what a "good" teacher should know and know to do (Cruz-Aguayo et al.; 2020).

While the specific requirements for such definition are not easy to identify, the aforementioned authors have hinted that most of them can be largely grouped into two: (i) to have the disciplinary knowledge and pedagogical practices adequate to the classroom characteristics, context and teaching level, and (ii) to display such knowledge and practices in the classroom, using the appropriate material and technological resources available.

As one can infer from the previous general conditions, and the slew evidence, the disciplinary knowledge is a relevant observable factor, consistently associated with teacher effectiveness and growth in the student's achievement (Santibañez; 2006; Clotfelter et al.; 2006, 2007; Hanushek and Rivkin; 2006; Marshall; 2009; Rockoff et al.; 2011; Kane et al.; 2010; Kane and Staiger; 2012; Ome; 2012; Metzler and Woessmann; 2012; Kane et al.; 2013; Araujo et al.; 2016; Bietenbeck et al.; 2018; Estrada; 2019); and in that sense, its measurement should be of interest for any educational authority.

The measurement of knowledge has a myriad of available tools, nevertheless, given that any educational department are bounded by budgetary constraints, valid[4] and reliable[5] standardized tests[6] stand out, not only for its cost-effectiveness, and a much simpler implementation (Cruz-Aguayo et al.; 2020); but also because, compared to other instruments, they are one of tools with less subjective scoring processes and interpretations.

However, as no instrument is perfect, the subject's knowledge scores resulting from their use will likely have two main problems. First, they could manifest measurement error (Metzler and Woessmann; 2012), which would imply that the estimates obtained from them could be an biased reflection of the true effects (Angrist and Krueger; 1999). And second, as the score is a composite value, does not allow to test which specific factors leads to better or worse teacher performance.

These two issues has direct and important policy implications, and devoting effort to appropriately assess and control them, could help the educational authorities to understand, for example: (i) the characteristics of the applicants to the public teaching carrer, (ii) to identify which teachers should be hired, and finally, once they are inside, what the authorities should do to train them (Hanushek and Rivkin; 2012), (iii) if the scores thresholds used for the selection processes are appropriately set[7], to mention a few.

[4]the extend to which a measurement tool is well-founded and accurately corresponds to the real measure (Kelley; 1927)

[5]the overall consistency of a measure under consistent conditions.

[6]Assessment instrument in which the implementation, questions, scoring processes, and interpretations are consistent with a predetermined or typified way. The instrument is usually composed of questions or items that fulfill three conditions: (i) they are polytomous, i.e. they have multiple choices, (ii) the choice categories are nominal, i.e. do not present any specific order, and (iii) there is only one "correct" category or answer (Rivera; 2019)

[7]Approximately 60% of the Caribbean and Latin American countries use standardized test scores as

In summary, teachers are one of the main drivers behind the student achievements. However, some of the evidence supporting this claim has been based on proxy variables that are not consistently related to the quality of instruction, or methods that are not concerned with the outline of the teaching factors, responsible for the student's learning. Nevertheless, while the literature still reflects considerable uncertainty on what are the "ingredients for a good teacher", a good amount of evidence has supported the disciplinary and pedagogical knowledge as relevant components of the teacher effectiveness. Finally, the literature has shown that valid and reliable standardized tests are among the best tools to assess such factors, but also have emphasized that such scores could reflect the teacher's abilities with considerable noise.

1.2 Objectives

This research will have two main goals. First, to describe the method, estimation procedures, and advantages of the Generalized Linear Latent and Mixed Modeling framework (GLLAMM), developed by Rabe-Hesketh et al. (2004a,c); Skrondal and Rabe-Hesketh (2004a); Rabe-Hesketh et al. (2012). And second, tests the policy implications of the methods, and its results, in a data composed of large repeated Teacher's standardized educational assessments from Peru.

Specifically, for the first objective of the research, the author expects to appraise:

1. If the method can provide a general framework that could serve multiple psychometric purposes, e.g. to analyze the quality of the items, to obtain a dynamical noise-free "score" for the disciplinary abilities of the teachers, among others; and

2. What are the advantages or disadvantages of such models, specially compared to factor, item-response theory and multilevel models.

For the second objective, the author expects to shed some lights about some key policy decisions related to those large evaluation processes, to mention a few:

- Are the educational authorities screening the teachers with higher disciplinary knowledge?, and in that sense, what differentiate a contract teacher from a temporary one?,

- What are the general characteristics of the teaching-career applicants?, What is the level of their disciplinary knowledge, and how it evolves?,

- Do the initial training or socioeconomic status help to explain the disciplinary knowledge profile of the applicants?

- What factors of the disciplinary knowledge are consistently related to a good performance in the classroom?

- Do the instruments guarantee a fair assessment of minority groups with different abilities?

part of or as a main teacher selection tool (Cruz-Aguayo et al.; 2020).

Given the aforementioned goals, the researcher believes the master's thesis contributes to the literature in two aspects:

1. In a the theoretical and methodological sense, as the research is focused on offering an exhaustive description and analysis of the GLLAMM framework; and

2. In a more practical sense, as it helps to provide evidence on some of key policy decisions that most of Latin America countries are currently facing.

Finally, it is important to mention, that the computational implementation of the method will be developed in `R` (R Core Team; 2015) and `WinBUGS` (Lunn et al.; 2000).

1.3 Organization

Chapter 2, The Generalized Linear Latent and Mixed Model, will describe the model, its components, characteristics, assumptions and properties, to finally assess its benefits against factor (EFA and CFA), IRT and Multilevel models.

Chapter 3, Estimation, will describe two of the methods that can be used to fit such models: Likelihood and Bayesian methods. The chapter will also present the computational implementation of the model.

Chapter 4, Application, will describe the instruments and the "dimensions" under analysis. Additionally, it will describe briefly the data collection process, the sample design, and the results of the analysis under the GLLAMM framework.

Finally, **Chapter 5, Conclusions**, will discuss the conclusion for the research, in term of the method and the policy implications derived from its implementation in a large teacher's assessment process. Finally, it will outline the path of future research that can be derived from the present effort.

Chapter 2

The Generalized Linear Latent and Mixed Model

The Generalized Linear Latent and Mixed Model (GLLAMM) is a framework that unifies a wide range of latent variable models. Developed by Rabe-Hesketh et al. (2004a,c,b); Skrondal and Rabe-Hesketh (2004a); Rabe-Hesketh et al. (2012), the method was motivated by the need of a multilevel Structural Equation Model (SEM) that accommodates for unbalanced data, noncontinuous responses and the use of cross-level effects among latent variables.

This chapter presents the definition, characteristics, assumptions and properties of such framework.

2.1 Definition

Following Rabe-Hesketh et al. (2004a, 2012), we depart from the traditional multivariate framework for formulating factor and structural models, i.e. a "wide" data format, and adopt a univariate approach, i.e. "long" or vectorized format. In that sense, for each unit, the response variables are "stacked" in a single response vector, with different variables distinguished from each other, by a design matrix. With this structure, we proceed to outline the three parts of the framework:

1. The response model,
2. The structural latent variable model, and
3. The distribution of the latent variables.

For a detailed description of some of the special cases of multilevel SEM, that can be derived with this framework, refer to Appendix A.

2.1.1 Response model

As outlined by the authors, conditional on the latent variables, the response model is a Generalized Linear Model (GLM) defined by a systematic and a distributional part. For the systematic part, a linear predictor and a link function are selected, in accordance to the characteristics of the manifest variables. On the other hand, for the distributional part, a distribution from the exponential family is selected.

In the following sections, we proceed to describe the linear predictor, the link function and the distributions accommodated by the framework.

Linear predictor

For a model with L levels and M_l latent variables at $l > 1$ levels, the linear predictor takes the following form:

$$v = \mathbf{X}\boldsymbol{\beta} + \sum_{l=2}^{L}\sum_{m=1}^{M_{(l)}} \eta_m^{(l)} \mathbf{Z}_m^{(l)} \boldsymbol{\lambda}_m^{(l)} \tag{2.1}$$

where $\mathbf{X}$ is a design matrix that maps the parameter vector $\boldsymbol{\beta}$ to the linear predictor, $\eta_m^{(l)}$ the mth latent variable at level l ($m = 1, \ldots, M_{(l)}$ and $l = 1, \ldots, L$), and $\mathbf{Z}_m^{(l)}$ a design matrix that maps the vector of loadings $\boldsymbol{\lambda}_m^{(l)}$ to the mth latent variable at level l.

Note that wo do not use subscripts for the units of observation at different levels. This decision was made with the purpose of avoiding the use of mathematical definitions with large number of subscripts. However, a careful reader should consider that equation (2.1) rest on the assumption that each unit is identified at their appropriate level. For special cases of multilevel SEM, and their use of subscripts, refer to Appendix A.

Links and Distributions

As in the GLM framework, the model "links" the expectation of the conditional response, to the linear predictor, through a inverse-link function $h(\cdot)$, in the following form:

$$\mu = E[y|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] = h(v) \tag{2.2}$$

where equation (2.2) can be re-written in terms of the link function $g(\cdot) = h^{-1}(\cdot)$:

$$g(\mu) = g(E[y|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}]) = v \tag{2.3}$$

with $\boldsymbol{\eta} = \left[\eta^{(2)T}, \ldots, \eta^{(L)T}\right]^T$ and $\mathbf{Z} = \left[\mathbf{Z}^{(2)T}, \ldots, \mathbf{Z}^{(L)T}\right]^T$, as the "stacked" vector of latent variables, and the "stacked" design matrices of explanatory variables, for all L levels, respectively. Additionally, $\boldsymbol{\eta}^{(l)} = \left[\eta_1^{(l)}, \ldots, \eta_{M_{(l)}}^{(l)}\right]^T$ and $\mathbf{Z}^{(l)} = \left[\mathbf{Z}_1^{(l)T}, \ldots, \mathbf{Z}_{M_{(l)}}^{(l)T}\right]^T$, denotes the vector of latent variables, and the "stacked" design matrix of explanatory variables, at level l, respectively.

Finally, the response model specification is complete when we select an appropriate distribution from the family of exponential distributions. The types of responses that can be accommodated by the framework are the following:

1. **Continuous:**
   It results form selecting an identity link function for the scaled mean response,

$$\begin{aligned} \mu^* &= E[y^*|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= v \end{aligned} \tag{2.4}$$

   where $\mu^* = \mu\sigma^{-1}$, $y^* = y\sigma^{-1}$, and σ denotes the standard deviation of the errors.

On the other hand, the distributional part is defined by a Standard Normal distribution $\phi(x) = (2\pi)^{-1/2} exp(-x^2/2)$,

$$\begin{aligned} f(y^*|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}) &= \phi(\mu^*)\sigma^{-1} \\ &= \phi(v)\sigma^{-1} \end{aligned} \tag{2.5}$$

Notice that the same parametrization can be achieved considering $y^* = v + \epsilon^*$, and $\epsilon^* \sim N(0,1)$. Additionally, the decision to standardize the response variables has been made with the purpose of making the estimation process easier, as such distribution is free of unknown parameters.

2. **Dichotomous:**
It results from selecting an appropriate inverse-link function for the expected value of the manifest variable, which describe the probability of endorsing one of the two available categories,

$$\begin{aligned} \mu &= E[y = 1|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= P[y = 1|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= \pi \\ &= h(\kappa - v) \end{aligned} \tag{2.6}$$

where κ is the decision threshold, and $h(\cdot)$ can be defined in three ways:

$$h(x) = \begin{cases} exp(x)[1 + exp(x)]^{-1} \\ \Phi(x) \quad \text{No closed form.} \\ exp(-exp(x)) \end{cases} \tag{2.7}$$

which corresponds to the logistic, standard normal $\Phi(x)$, and Gumbel (extreme value type I) *cumulative distributions*, respectively. In terms of link functions, the distributions corresponds to the well known logit, probit and complementary log-log link functions, respectively.

Alternatively, the same parametrization can be achieved using the concept of an underlying latent variable in the form $y^* = v + \epsilon^*$, where $y = 1$ if $y^* \geq \kappa$, and ϵ^* can have a distribution as the ones defined in equation (2.7). It is important to mention that under this parametrization, the threshold parameters κ and the $\boldsymbol{\beta}$ are confounded as they serve similar purposes, so only one would be estimated.

Finally, the distributional part is defined by a Binomial distribution,

$$\begin{aligned} f[y = 1|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] &= \binom{n}{k} \mu^k (1 - \mu)^{n-k} \\ &= \binom{n}{k} \pi^k (1 - \pi)^{n-k} \end{aligned} \tag{2.8}$$

where k denotes the number of successes in n independent Bernoulli trials.

3. **Polytomous:**
It results from selecting a generalized logistic inverse-link function (Bock; 1972) for

the expected value of the response, which in this case, describe the probability of endorsing one of the S unordered available categories,

$$\begin{aligned}\mu_s &= E[y = y_s|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= P[y = y_s|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= \pi_s \\ &= h(v_s)\end{aligned} \tag{2.9}$$

where v_s is the linear predictor for category s ($s = 1, \ldots, S$), and $h(\cdot)$ is defined as:

$$h(x) = exp(x)\left[\sum_{s=1}^{S} exp(x)\right]^{-1} \tag{2.10}$$

It is important to note that, as in the dichotomous case, the same parametrization can be achieved using the concept of underlying continuous responses in the form $y_s^* = v_s + \epsilon_s$, where $y = s$ if $y_s^* > y_k^*\ \forall s, s \neq k$, ϵ_s have a Gumbel (extreme value type I) distribution, as the one defined in equation (2.7), and y_s denotes the random utility for the s category.

Finally, the distributional part is defined by a Multinomial distribution,

$$\begin{aligned}f[y = \{y_1, \cdots, y_S\}|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] &= \frac{n!}{y_1! \cdots y_S!} \prod_{s=1}^{S} \mu_s^{y_s} \\ &= \frac{n!}{y_1! \cdots y_S!} \prod_{s=1}^{S} \pi_s^{y_s}\end{aligned} \tag{2.11}$$

where y_s denotes the number of "successes" in category s.

4. **Ordinal and discrete time duration:**
For the ordinal case, the linear predictor is "linked" to the probability of endorsing category s, against all previous categories, in the following form:

$$\begin{aligned}\mu_s &= E[y = y_s|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= P[y \leq y_s|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] - P[y \leq y_{s-1}|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= h(\kappa_s - v_s) - h(\kappa_{s-1} - v_{s-1})\end{aligned} \tag{2.12}$$

where κ_s denotes the thresholds for category s. For discrete time duration, the linear predictor is "linked" to the probability of survival, in the sth time interval, as follows:

$$\begin{aligned}\mu_s &= E[t_{s-1} \leq T \leq t_s|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= P[T \leq t_s|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] - P[T \leq t_{s-1}|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= h(v_s + t_s) - h(v_{s-1} + t_{s-1})\end{aligned} \tag{2.13}$$

where T is the unobserved continuous time, and t_s its observed discrete realization. Additionally, for both type of responses, $h(\cdot)$ can be defined as the logistic, standard normal, and Gumbel (extreme value type I) *cumulative distributions*, as in equation (2.7).

Similar to the dichotomous and polytomous case, the same parametrization can be achieved using the concept of underlying latent variables with $y_s^* = \upsilon_s + \epsilon_s$, where $y = s$ if $\kappa_{s-1} < y_s^* \leq \kappa_s$, $\kappa_0 = -\infty$, $\kappa_1 = 0$, $\kappa_S = +\infty$, ϵ_s has one of the distributions in equation (2.7), and y_s denotes the random utility for the s category.

It is important to note, for discrete time duration responses, the logit link corresponds to a *Proportional-Odds model*, while the complementary log-log link to a *Discrete Time Hazards model* (Rabe-Hesketh et al.; 2001). Other models for ordinal responses, such as the *Baseline Category Logit* or the *Adjacent Category Logit* models can be specified as special cases of the generalized logistic response function, defined in equation (2.10).

Finally, the distributional part is defined by a Multinomial distribution, as the one defined in equation (2.11).

5. **Counts and continuous time duration:**
It results from selecting an exponential inverse-link function (log link) for the expected value of the response,

$$\begin{aligned} \mu &= E[y|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= \lambda \\ &= exp(\upsilon) \end{aligned} \tag{2.14}$$

and a Poisson conditional distribution for the counts,

$$\begin{aligned} f[y|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] &= exp(-\mu)\mu^y(y!)^{-1} \\ &= exp(-\lambda)\lambda^y(y!)^{-1} \end{aligned} \tag{2.15}$$

It is important to mention that unlike the models for dichotomous, polytomous and ordinal responses, model for counts cannot be written under the random utility framework.

6. **Rankings and pairwise comparisons:**
Following Skrondal and Rabe-Hesketh (2003a), the parametrization for polytomous responses can serve as the building block for the conditional distribution of rankings. Selecting a "exploded logit" inverse-link function (Chapaaan and Staelin; 1982) for the expected value of the response, which describes the probability of the full rankings of category s,

WORK IN PROGRESS

$$\begin{aligned} \mu_s &= P[\mathbf{R}_s = \{r_s^1, \dots r_s^1\}|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= \pi_s \\ &= h(\upsilon_s) \end{aligned} \tag{2.16}$$

where υ_s is the linear predictor for category s $(s = 1, \dots, S)$, and $h(\cdot)$ is defined as:

$$h(x) = \prod_{s=1}^{S} exp(x^s) \left[\sum_{s=1}^{S} exp(x^s) \right]^{-1} \tag{2.17}$$

Again, as in specific previous cases, the same parametrization can be achieved using the concept of underlying latent variables.

Finally, the distributional part is defined by a Multinomial distribution,

$$\begin{aligned} f[y = \{y_1, \cdots, y_S\} | \mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] &= \frac{n!}{y_1! \cdots y_S!} \prod_{s=1}^{S} \mu_s^{y_s} \\ &= \frac{n!}{y_1! \cdots y_S!} \prod_{s=1}^{S} \pi_s^{y_s} \end{aligned} \tag{2.18}$$

where y_s denotes the number of "success cases" in category s.

7. **Mixtures:**
Given the previous definitions, the framework easily lends itself to model five additional settings:

   (a) **Different links and distributions for different latent variables**. This can be easily achieved by setting different links and distributions for each of the M_2 latent variables located at level 2.

   (b) **Left- or right-censored continuous responses**. Common in selection models (e.g. Heckman; 1979), they can be achieved by specifying an identity link and Normal distribution for the uncensored scaled responses, as in equations (2.4) and (2.5); and a scaled probit link and Binomial distribution otherwise, as in equations (2.7) and (2.8).

   (c) **zero-inflated count responses**. where a log link and a Poisson distribution is set for the counts, as in equations (2.14) and (2.15); and a logit link and Binomial distribution is specified to model the zero center of mass, as in equations (2.6) and (2.8).

   (d) **Measurement error in covariates**. this setting occurs when standard models use variables, with measurement error, as covariates, e.g. a logistic regression with a continuous covariate that presents measurement error. For more details on this type of setting see Rabe-Hesketh, Skrondal and Pickles (2003); Rabe-Hesketh, Pickles and Skrondal (2003), and Skrondal and Rabe-Hesketh (2003b).

   (e) **Composite links**. Useful for specifying proportional odds models for right-censored responses, for handling missing categorical covariates and many other model types. For more details on this type of settings see Skrondal and Rabe-Hesketh (2004b).

Heteroscedasticity and over-dispersion in the response

GLLAMM allows to model heteroscedasticity, and over- or under-dispersion by adding random effects to the linear predictor, at level 1. The types of responses, in which such characteristics can be modeled, are the following:

1. **Continuous:**
   We model **heteroscedasticity** in the following form:
   $$\sigma = exp(\boldsymbol{\alpha}^T \mathbf{Z^{(1)}}) \tag{2.19}$$
   Notice that the previous formula implies that equation (2.5) can be re-written in the following form:
   $$f(y^*|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}) = \phi(v + \boldsymbol{\alpha}^T \mathbf{Z^{(1)}}) \tag{2.20}$$
   where $\mathbf{Z^{(1)}}$ is the design matrix that maps the random effects $\boldsymbol{\alpha}$. Notice that equation (2.20) effectively corresponds to a model that includes random intercepts at level 1.

2. **Dichotomous:**
   In a more straightforward way, we model over- or under-dispersion by modifying equation (2.6), to include random intercepts at level 1, in the following form:
   $$\begin{aligned} \mu &= P[y = 1|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= \pi \\ &= h(\kappa - v + \boldsymbol{\alpha}^T \mathbf{Z^{(1)}}) \end{aligned} \tag{2.21}$$

3. **Ordinal, and discrete time duration:**
   Similar to the dichotomous case, by including random intercepts at level 1 in equation (2.12), we can model over- or under-dispersion:
   $$\begin{aligned} \mu_s &= P[y \leq y_s|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] - P[y \leq y_{s-1}|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= h(\kappa_s - v_s + \boldsymbol{\alpha}^T \mathbf{Z^{(1)}}) - h(\kappa_{s-1} - v_{s-1} + \boldsymbol{\alpha}^T \mathbf{Z^{(1)}}) \end{aligned} \tag{2.22}$$
   A similar parametrization can be used for discrete time duration.

4. **Counts, and continuous time duration:**
   Finally, modifying equation (2.14) allow us to model over- or under-dispersion under a counts model:
   $$\begin{aligned} \mu &= E[y|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta}] \\ &= \lambda \\ &= exp(v + \boldsymbol{\alpha}^T \mathbf{Z^{(1)}}) \end{aligned} \tag{2.23}$$

2.1.2 Structural model for the latent variables

The structural model for the latent variables has the form:

$$\boldsymbol{\eta} = \underset{(M\times M)}{\mathbf{B}} \underset{(M\times 1)}{\boldsymbol{\eta}} + \underset{(M\times Q)}{\boldsymbol{\Gamma}} \underset{(Q\times 1)}{\mathbf{W}} + \underset{(M\times 1)}{\boldsymbol{\zeta}} \tag{2.24}$$

where $\mathbf{B}$ and $\boldsymbol{\Gamma}$ are parameter matrices that maps the relationship between the latent variables $\boldsymbol{\eta}$, and the vector of "stacked" covariates $\mathbf{W}$, respectively; $\boldsymbol{\zeta}$ is a vector of errors or disturbances, and $M = \sum_l M_l$. Notice that while equation (2.24) resembles to single-level structural equation models, the main difference lies in the fact that the latent variables may vary at different levels. Additionally, considering that $\boldsymbol{\eta}$ has no feedback effects, and it is permuted and sorted according to the levels, $\mathbf{B}$ is defined as a strictly upper triangular matrix. In this regard, it is important to mention that,

1. The absence of feedback loops implies that the method deals with non-recursive models, i.e. none of the latent variables are specified as both causes and effects of each other (Kline; 2012); this in turn allows the easy estimation of the model parameters.

2. The strictly upper triangular structure reveals that the framework does not allow latent variables to be regressed on lower level latent or observed variables, as such specification is more related to the use of formative, rather than reflective, latent variables. For a detail explanation on the topic refer to Edwards and Bagozzi (2000).

Notice, however, the previous restrictions does not hinder the ability of the method to model contextual effects, after controlling the lower level compositional effects. For examples of such refer to Appendix A.

2.1.3 Distribution of the latent variables

Finally, to fully specify the framework, and provide a scale for the latent variables, we have to make assumptions for either the distribution of the disturbances $\boldsymbol{\zeta}$ or the latent variables $\boldsymbol{\eta}$. If our research interest lies in the structural equation model, it is more convenient to make assumptions for the distribution of the disturbances; otherwise, we make assumptions for the distributions for the latent variables.

Furthermore, as in the hierarchical framework, it is assumed the latent variables at different levels are independent, whereas latent variables at the same level may present dependency. In that sense, we presume all latent variables at level l to have a multivariate normal distribution with zero mean and covariance matrix Σ_l, i.e. $\boldsymbol{\eta}^{(l)} \sim MVN(\mathbf{0}, \boldsymbol{\Sigma}_l)$. It is important to emphasize that, while the multivariate normal distribution is widely used in these settings, it is not the only distribution that can be assumed. Rabe-Hesketh, Skrondal and Pickles (2003) have provided evidence that it can be even left unspecified, by using non-parametric maximum likelihood estimation.

2.2 Model identification

WORK IN PROGRESS

The structure of the latent variables is specified by the number of levels L and the number of latent variables Ml at each level. A particular level may coincide with a level of clustering in the hierarchical dataset. However, there will often not be a direct correspondence between the levels of the model and the levels of the data hierarchy.

2.3 Relationship with other modeling schemes

WORK IN PROGRESS

Under the GLM framework, the continuous response model assumes that the errors are homoscedastic, identical and independently distributed, i.e. $\epsilon^* \overset{iid}{\sim} N(0, 1)$. However, this assumption that is often violated when, for example, the units of analysis are correlated or belong to a cluster. In similar fashion, for non-continuous responses the framework fixes the relationship of the modeled dispersion to the mean value, e.g. for the dichotomous case $\mu = \pi$, and $v(\mu) = \pi(1 - \pi)$. But again, in practice, this assumption is also often violated as the data can present over- or under-dispersion.

Given the restrictions of GLM, the Generalized Linear Mixed Model (GLMM) framework was developed. Under the new framework, heteroscedasticity, over-dispersion and other characteristics of the data can be easily modeled by adding random effects to the linear predictor.

In that sense, GLLAMM is similar to the GLMM framework, as it borrow the idea of including random effects to model heteroscedasticity, and over-dispersion in the responses. The types of responses, in which such characteristics can be modeled, are the following:

Throughout The statistical literature was instigated to develop methods that can handle such data characteristics. Considering that researchers are often interested in variables that cannot be measured directly or reflect measurement error, e.g. intelligence, depression; Is in this setting that Structural Equation Models (SEM) were developed, as a framework motivated by the need of tools that can impute relationships between unobserved constructs (also known as factors, latent variables, amng others) and observable or manifest variables. Under this framework, it is assumed that such "common factors" are responsible for the variation and dependence in the manifest variables.

Use of SEM is commonly justified in the social sciences because of its ability to

Multilevel regression models were developed to handle hierarchical or clustered structure in the data, where elementary units are nested in clusters, such as students in schools, which in turn may be nested in higher-level clusters (e.g., school districts or states). The latent variables, often called "random effects" in this context, can be interpreted as the effects of unobserved covariates at different levels that induce dependence among lower-level units.

As stated by Rabe-Hesketh et al. (2012), the GLLAMM framework was motivated by the need of a multilevel Structural Equation Model (SEM) to accommodates for unbalanced data, noncontinuous responses and the use of cross-level effects among latent variables.

multilevel structural equation models represent a synthesis between multilevel regression models and structural equation models. Considering that

is a framework that unifies a wide range of latent variable models. Developed by Rabe-Hesketh et al. (2004a,c,b); Skrondal and Rabe-Hesketh (2004a); Rabe-Hesketh et al. (2012), the method was motivated by the need of a multilevel Structural Equation Model (SEM) that accommodates for unbalanced data, noncontinuous responses and the use of cross-level effects among latent variables.

2.3.1 Factor Models

2.3.2 Item Response Theory and Generalized Latent Models

2.3.3 Multilevel Models

2.4 Advantages and Disadvantages

Chapter 3

Estimation

3.1 Likelihood methods

3.1.1 Likelihood function

3.1.2 Adaptive Quadrature

3.2 Bayesian methods

3.2.1 Prior distributions

3.2.2 Initial start

3.2.3 Posterior distributions

Chapter 4

Application

4.1 Instruments

4.2 Data

4.2.1 Collection

4.2.2 Sample scheme

4.3 Results

4.3.1 Hypothesis 1:

4.3.2 Hypothesis 2:

4.3.3 Hypothesis 3:

Chapter 5

Conclusion and Discussion

5.1 Discussion

5.2 Conclusions

5.3 Future development

Appendix A

Additional Theory

A.1 Special cases for the GLAMM

Appendix B

Code

Bibliography

Angrist, J. and Krueger, A. (1999). Empirical strategies in labor economics, *in* O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, Vol. 3, Elsevier, chapter 23, pp. 1277 – 1366.
DOI: https://www.doi.org/10.1016/S1573-4463(99)03004-7.
URL: http://www.sciencedirect.com/science/article/pii/S1573446399030047.

Araujo, M., Carneiro, P., Cruz-Aguayo, Y. and Schady, N. (2016). Teacher quality and learning outcomes in kindergarten, *The Quarterly Journal of Economics* **131**(3): 1415–1453.
DOI: https://www.doi.org/10.1093/qje/qjw016.
URL: https://publications.iadb.org/publications/english/document/Teacher-Quality-and-Learning-Outcomes-in-Kindergarten.pdf.

Ayala, M. (2017). *Efecto de los docentes provisionales sobre desempeño escolar - evidencia para la educación secundaria oficial en colombia*, Master's thesis, Universidad de los Andes.
URL: http://biblioteca.uniandes.edu.co/acepto201699.php?id=11802.pdf.

Bertoni, E., Elacqua, G., Marotta, L., Martinez, M., Méndez, C., Montalva, V., Olsen, A., Santos, H. and Soares, S. (2020). Escasez de docentes en latinoamérica: ¿cómo se puede medir y que políticas están implementando los países para resolverlo?, *Technical report*, Banco Interamericano de Desarrollo.

Bietenbeck, J., Piopiunik, M. and Wiederhold, S. (2018). Africa's skill tragedy: Does teachers' lack of knowledge lead to low student performance?, *Comparative Education Review* **53**(3): 553–578.
DOI: https://www.doi.org/10.3368/jhr.53.3.0616-8002R1.
URL: http://jhr.uwpress.org/content/53/3/553.abstract.

Bock, R. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories, *Psychometrika* **37**(1).

Bruns, B., Luque, J., De Gregorio, S., Evans, D., Fernández, M., Moreno, M., Rodriguez, J. Toral, G. and Yarrow, N. (2015). Great teachers: How to raise student learning in latin america and the caribbean, *Technical report*, World Bank Group.

Chapaaan, R. and Staelin, R. (1982). Exploiting rank ordered choice set data within the stochastic utility model, *Journal of Marketing Research* **19**(3): 288–301.
DOI: https://www.doi.org/10.1177/002224378201900302.

Chetty, R., Friedman, J. and Rockoff, J. (2014a). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates, *American Economic Review* **104**(9): 2593–2632.
DOI: https://www.doi.org/10.1257/aer.104.9.2593.
URL: https://www.aeaweb.org/articles?id=10.1257/aer.104.9.2593.

Chetty, R., Friedman, J. and Rockoff, J. (2014b). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood, *American Economic Review* **104**(9): 2633–2679.
DOI: https://www.doi.org/10.1257/aer.104.9.2593.
URL: https://www.aeaweb.org/articles?id=10.1257/aer.104.9.2593.

Chetty, R., Friedman, J. and Rockoff, J. (2015). School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from kenyan primary schools, *Journal of Public Economics* **123**: 92–110.
DOI: https://www.doi.org/10.1016/j.jpubeco.2014.11.008.
URL: http://www.sciencedirect.com/science/article/pii/S0047272714002412.

Clotfelter, C., Ladd, H. and Vigdor, J. (2006). Teacher-student matching and the assessment of teacher effectiveness, *Working Paper 11936*, National Bureau of Economic Research.
DOI: https://www.doi.org/10.3386/w11936.
URL: http://www.nber.org/papers/w11936.

Clotfelter, C., Ladd, H. and Vigdor, J. (2007). How and why do teacher credentials matter for student achievement?, *Working Paper 12828*, National Bureau of Economic Research.
DOI: https://www.doi.org/10.3386/w12828.
URL: http://www.nber.org/papers/w12828.

Cruz-Aguayo, Y., Hincapié, D. and Rodríguez, C. (2020). Profesores a prueba: claves para una evaluación docente exitosa, *Technical report*, Banco Interamericano de Desarrollo.

Dee, T. and Cohodes, S. (2008). Out-of-field teachers and student achievement: Evidence from matched-pairs comparisons, *Public Finance Review* **36**(1): 7–32.
DOI: https://www.doi.org/10.1177/1091142106289330.

Duflo, E., Dupas, P. and Kremer, M. (2009). Additional resources versus organizational changes in education: Experimental evidence from kenya.

Edwards, J. and Bagozzi, R. (2000). On the nature and direction of relationships between constructs and measures, *Psychological Methods* **5**(2): 155–174.
DOI: https://www.doi.org/10.1037/1082-989X.5.2.155.

Elacqua, G., Hincapié, D., Vegas, E. and Alfonso, M. (2018). Profesión: profesor en américa latina ¿por qué se perdió el prestigio docente y cómo recuperarlo?, *Technical report*, Banco Interamericano de Desarrollo.

Estrada, R. (2019). Rules versus discretion in public service: Teacher hiring in mexico, *Journal of Labor Economics* **37**(2): 545–579.
DOI: https://www.doi.org/10.1086/700192.

Hanushek, E. and Rivkin, S. (2006). Teacher quality, *in* E. Hanushek and F. Welch (eds), *Handbook of the Economics of Education*, Vol. 2, Elsevier, chapter 18, pp. 1051 – 1078.
DOI: https://www.doi.org/10.1016/S1574-0692(06)02018-6.
URL: http://www.sciencedirect.com/science/article/pii/S1574069206020186.

Hanushek, E. and Rivkin, S. (2012). The distribution of teacher quality and implications for policy, *Annual Review of Economics* **4**(1): 131–157.
DOI: https://www.doi.org/10.1146/annurev-economics-080511-111001.

Heckman, J. (1979). Sample selection bias as a specification error, **47**(1): 153–161.
DOI: https://www.doi.org/10.2307/1912352.
URL: https://www.jstor.org/stable/1912352.

Ingersoll, R. (1998). The problem of out-of-field teaching.
URL: https://repository.upenn.edu/gse$_{p}ubs$/137.

Kane, T., McCaffrey, D., Miller, T. and Staiger, D. (2013). Have we identified effective teachers? validating measures of effective teaching using random assignment, *Research paper*, Bill Melinda Gates Foundation.
URL: https://files.eric.ed.gov/fulltext/ED540959.pdf.

Kane, T. and Staiger, D. (2012). Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains, *Research paper*, Bill Melinda Gates Foundation.
URL: https://k12education.gatesfoundation.org/download/?Num=2678filename=MET$_{G}athering_{F}$e

Kane, T., Taylor, E., Tyler, J. and Wooten, A. (2010). Identifying effective classroom practices using student achievement data, *Working Paper 15803*, National Bureau of Economic Research.
DOI: https://www.doi.org/10.3386/w15803.
URL: http://www.nber.org/papers/w15803.

Kelley, T. (1927). *Interpretation of educational measurements*, Measurement and adjustment series, World Book Co.

Kline, R. (2012). Assumptions in structural equation modeling, *in* R. Hoyle (ed.), *Handbook of Structural Equation Modeling*, The Guilford Press, chapter 7, pp. 111–125.

Lunn, D., Thomas, A., Best, N. and Spiegelhalter, D. (2000). Winbugs - a bayesian modelling framework: Concepts, structure, and extensibility, *Statistics and Computing* (10): 325–337.
DOI: https://www.doi.org/10.1023/A:1008929526011.

Marotta, L. (2019). Teachers' contractual ties and student achievement: The effect of temporary and multiple-school teachers in brazil, *Comparative Education Review* **63**(3): 356–376.
DOI: https://www.doi.org/10.1086/703981.

Marshall, J. (2009). School quality and learning gains in rural guatemala, *Economics of Education Review* **28**(2): 207–216.

DOI: https://www.doi.org/10.1016/j.econedurev.2007.10.009.
URL: http://www.sciencedirect.com/science/article/pii/S0272775708000745.

Medeiros, M., Gómez, C., Sánchez, M. and Orrego, V. (2018). Idoneidad disciplinar de los profesores y mercado de horas docentes en chile, *Calidad en la Educación* (48): 50–95.
DOI: https://www.doi.org/10.31619/caledu.n48.479.

Metzler, J. and Woessmann, L. (2012). The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation, *Journal of Development Economics* **99**(2): 486–496.
DOI: https://www.doi.org/10.1016/j.jdeveco.2012.06.
URL: https://ideas.repec.org/a/eee/deveco/v99y2012i2p486-496.html.

Muralidharan, K. and Sundararaman, V. (2013). Contract teachers: Experimental evidence from india, *Working Paper 19440*, National Bureau of Economic Research.
DOI: https://www.doi.org/10.3386/w19440.
URL: http://www.nber.org/papers/w19440.

Ome, A. (2012). The effects of meritocracy for teachers in colombia, *Research report*, Fedesarrollo.
URL: https://ideas.repec.org/p/col/000124/010260.html.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: http://www.R-project.org/.

Rabe-Hesketh, S., Pickles, A. and Skrondal, A. (2003). Correcting for covariate measurement error in logistic regression using nonparametric maximum likelihood estimation, *Statistical Modelling* **3**(3): 215–232.
DOI: https://www.doi.org/10.1191/1471082X03st056oa.

Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2003). Maximum likelihood estimation of generalized linear models with covariate measurement error, *The Stata Journal* **3**(4): 386–411.
DOI: https://www.doi.org/10.1177/1536867X0400300408.
URL: https://journals.sagepub.com/doi/pdf/10.1177/1536867X0400300408.

Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2004a). Generalized multilevel structural equation modeling, *Psychometrika* **69**(2): 167–190.
DOI: https://www.doi.org/10.1007/BF02295939.

Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2004b). *GLLAMM Manual*, UC Berkeley Division of Biostatistics.
URL: http://www.biostat.jhsph.edu/ fdominic/teaching/bio656/software-gllamm.manual.pdf.

Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2004c). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects, *Journal of Econometrics* **128**(2): 301–323.
DOI: https://www.doi.org/10.1016/j.jeconom.2004.08.017.
URL: http://www.sciencedirect.com/science/article/pii/S0304407604001599.

Rabe-Hesketh, S., Skrondal, A. and Zheng, X. (2012). Multilevel structural equation modeling, *in* R. Hoyle (ed.), *Handbook of Structural Equation Modeling*, The Guilford Press, chapter 30, pp. 512–531.

Rabe-Hesketh, S., Yang, S. and Pickles, A. (2001). Multilevel models for censored and latent responses, *Statistical Methods in Medical Research* **10**(6): 409–427.
DOI: https://www.doi.org/10.1177/096228020101000604.

Rivera, J. (2019). *El modelo de respuesta nominal: Aplicación a datos educacionales*, Master's thesis, Pontificia Universidad Católica del Peru.
URL: http://hdl.handle.net/20.500.12404/14600.

Rivkin, S., Hanushek, E. and Kain, J. (2005). Teachers, schools, and academic achievement, *Econometrica* **73**(2): 417–458.
DOI: https://www.doi.org/10.1111/j.1468-0262.2005.00584.x.
URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0262.2005.00584.x.

Rockoff, J. (2004). The impact of individual teachers on student achievement: Evidence from panel data, *The American Economic Review* **94**(2): 247–252.
URL: http://www.jstor.org/stable/3592891.

Rockoff, J., Jacob, B., Kane, T. and Staiger, D. (2011). Can you recognize an effective teacher when you recruit one?, *Education Finance and Policy* **6**(1): 43–74.
DOI: https://www.doi.org/10.1162/EDFP_a_00022.

Santibañez, L. (2006). Why we should care if teachers get a's: Teacher test scores and student achievement in mexico, *Economics of Education Review* **25**(5): 510–520.
DOI: https://www.doi.org/10.1016/j.econedurev.2005.08.001.
URL: http://www.sciencedirect.com/science/article/pii/S0272775705000804.

Scherrer, J. (2011). Measuring teaching using value-added modeling: The imperfect panacea, *NASSP Bulletin* **95**(2): 122–140.
DOI: https://www.doi.org/10.1177/0192636511410052.

Skrondal, A. and Rabe-Hesketh, S. (2003a). Multilevel logistic regression for polytomous data and rankings, *Psychometrika* **68**: 267–287.
DOI: https://www.doi.org/10.1007/BF02294801.

Skrondal, A. and Rabe-Hesketh, S. (2003b). Some applications of generalized linear latent and mixed models in epidemiology: Repeated measures, measurement error and multilevel modeling, *Norsk Epidemiologi* **13**(2): 265–278.

Skrondal, A. and Rabe-Hesketh, S. (2004a). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*, Interdisciplinary Statistics, Chapman Hall/CRC Press.

Skrondal, A. and Rabe-Hesketh, S. (2004b). Generalized linear latent and mixed models with composite links and exploded likelihoods, *in* BiggeriA., E. Dreassi, C. Lagazio and M. Marchi (eds), *Proceedings of the 19th International Workshop on Statistical Modeling*, Firenze University Press, Florence, Italy, pp. 27–39.
URL: http://www.gllamm.org/composite$_{c}onf.pdf$.

AFDELING
Straat nr bus 0000
3000 LEUVEN, BELGIË
tel. + 32 16 00 00 00
fax + 32 16 00 00 00
www.kuleuven.be