PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

FACULTAD DE MATEMÁTICAS
DEPARTAMENTO DE ESTADÍSTICA

TOWARD A ROBUST MEASURE OF SCHOOL EFFECTIVENESS THROUGH VALUE-ADDED MODEL: CONSIDERING PROBLEMS OF ENDOGENEITY AND DEPENDENCE ON TIME.

by
MARÍA INÉS GODOY ÁVILA

Tesis presentada al Departamento de Estadístca, Facultad de Matemáticas de la Pontificia Universidad Católica de Chile, para optar al título de Doctor en Estadística

Academic advisors:
Dr. Ernesto SAN MARTÍN

November
Santiago, Chile

A mis hijos
Josefa y Enrique.

Acknowledgements

Es inevitable al finalizar esta tesis doctoral pensar en todas esas personas que de una u otra manera aportaron sus granitos de arena en el desarrollo y término de este proyecto.

En primer lugar quisiera expresar mi agradecimiento a mi profesor guía Ernesto San Martín, sin lugar a duda sin su apoyo y confianza esto no hubiese sido posible, le agradezco infintitamente su generosidad y amabilidad en la entrega de conocimientos y por las opotunidades profesionales brindadas durante estos años.

Mi más sincero agradecimiento al Dr. Sebastien Van Bellegem por su importante aporte y participación en este proyecto. A su vez le doy las gracias por su invitación a desarrollar parte de esta tesis en Louvain-la-Neuve, Bélgica, le agradezco su amabilidad y apoyo entregado.

No puedo dejar de mencionar a Catherine Germain, una maravillosa persona que hizo que fuese más fácil nuestra estadía en Louvain-la-Neuve. A Martha Artunduaga y sus hermosas hijas quienes con su amistad y cariño nos hicieron sentir en casa. A Joniada Milla y Cristían Vasquez por su amistad y aporte al desarrollo de este trabajo.

Por supuesto que no puedo dejar de mencionar a mis padres que han sido desde siempre un pilar fundamental en mi vida, gracias a ellos por su incondicional apoyo. Si no fuese por sus esfuerzos de vida mis estudios y esta tesis no hubiesen sido posible.

Por último, a mi compañero de vida, Enrique, de quien no sólo he recibido su apoyo y amor incondicional, también se embarcó en mis sueños haciéndolos suyos. Gracias, por tu paciencia, comprensión, cariño y sentido del humor. Espero poder acompañarte en tus futuros proyectos tal y como tu lo has hecho conmigo.

A todos muchas GRACIAS.

Contents

Abstract

For many countries is nowadays a common issue is searching for an improvement of their educational systems. A solid and well established educational system will undoubtedly lead to highly qualified students and to improve the quality of education. Although assessment is often seen as a tool to measure the progress of individual students, it also allows individuals, communities, and countries to track the quality of schools and educational systems Cohen, Bloom, and Malin (1996). Having accurate and reliable measures of schools performance is neccesary for to decide among educational options. A growing emphasis is being placed upon measures of school performance , they are central to school improvement efforts, system of accountability and school choice, and broader educational policies OECD (2008). In this context, many countries have implemented accountability systems which aim to monitor school to be accountable for their functioning. The accountability system involve a process of evaluating school performance on the basis of students' performance measures. The obtained information is used by policy makers, teachers and principals, and parents for developing educational policies, improve professional practice, and school choice, respectively.

According to Mortimore (1991): "An effective school is one in which students progress further than might be expected from consideration of its intake". In educational research studies, the school effectiveness is assessed through value-added techniques. Broadly speaking the value added of a school is "the contribution of a school to students progress towards stated or prescribed education objectives, the contribution is net of other factors that contribute to students educational progress" OECD (2008); see also Braun, Chudowsky, and Koening (2010) and Baker et al. (2010). Thus school effectiveness seeks to identify the 'Value Added' by schools to student outcomes. Value-Added models are used to estimate the contribution of teachers, educational programs or schools to student

achievement. From a methodological point of view, this can be achieved by modeling student's scores taking into account the differences in prior achievements and possibly other measured characteristics in the form of covariates at both the school level and the student level; see Braun et al. (2010); OECD (2008); Raudenbush (2004) and Timmermans, Doolaard, and Wolf (2011). The role of the covariates is to characterize a school of reference with respect to which the Value-Added is substantively interpreted. As a matter of fact, the Value-Added of a school is a comparison between the conditional expected scores in a given school and the conditional expected score in the school of reference: if the covariates are modified, the school of reference is also modified and the meaning of the Value-Added changes. Because the interest is to know the net contribution of a school, and the covariates have an influence on the student performance, an important requirement for the covariates is that they have to be unrelated to the internal pedagogical processes performed by a school. Using the econometric jargon, it is said that the covariates are *exogenous* with respect to the school.

Due to the hierarchical structure of the data where students are nested into schools, a standard approach to model the school Value-Added is the use of hierarchical linear models (HLM), or multilevel models Goldstein (2002); Snijders and Bosker (1999). Under this approach, students scores are explained by their previous achievement, some covariates and a random effect representing the school effect. A measure of Value-Added has been typically obtained as the prediction of the random school effect Aitkin and Longford (1986); Longford (2012); Raudenbush and Willms (1995); Tekwe et al. (2004). As every statistical model, some assumptions need to be met for the inferences to be valid. For instance, the above mentioned requirement of exogeneity of covariates is key to isolate the net effect of the school represented by the random effect in the model, yet it is not considered by most of the models currently in used for the estimation of value-added measures. This can have serious consequences as the reported measures of value added could be misleading. Despite this problems, hierarchical linear models are widely accepted for estimating Value-Added. Although constantly the students of schools are evaluated across time, such that, they have information coming from more of a cohort of students, understood by cohort, the group of students that were measured on two occasions (previous and current score) and the set of them are different. When there are several cohorts is not quite right considered independence among cohorts, because it is the same school that treats students, the value added model are not considering an association among the

schools effects across the time. These are only one example of various problems and challenges seen in some educational data sets.

Many problems involving the value added models are due to the structured of educational data, educational system, calculating measures of value added forgetting some of the problems, could, however, lead to erroneous conclusions about the effectiveness of a school. In this dissertation we addressed two problems; the endogeneity and dependence across the time, these are challenges seen in some educational data sets, particularly in the Chile case. The Chilean educational systems do not ensure the exogeneity (i.e., the fact that covariates and the school effect in the model are uncorrelated) of the prior scores with respect to the school effect, because students are typically educated by the same school during the two test occasions, meaning that the prior score already contains the effect of the school. Absence of exogeneity makes impossible to isolate the net contribution of a school to student achievement. The Chilean educational systems do not ensure the exogeneity (i.e., the fact that covariates and the school effect in the model are uncorrelated) of the prior scores with respect to the school effect, because students are typically educated by the same school during the two test occasions, meaning that the prior score already contains the effect of the school. Absence of exogeneity makes impossible to isolate the net contribution of a school to student achievement. Moreover, Chile has the information of more a cohort of students on different times , the consequence of not consider a relation across the time among the schools effect can lead to erroneous conclusions, a school may worsen compared to a reference school but improve in comparison with itself.

This dissertation named; Toward a robust measure of school effectiveness through value-added model: it is considered both problems of endogeneity as dependence on time, separately and not jointly, however, it is possible and not so complicated, link these methodologies. Our proposals addressed developing of two value added models, two robust measures under the hierarchical linear mixed models and an application to real data, specifically in Chilean educational data.

Introduction

Introduction

Nowadays countries seek to improve their educational systems since a solid and well established educational system will undoubtedly lead to highly qualified students and will improve the quality of education. Furthermore a growing emphasis is being placed upon measures of school performance. Accurate and reliable measures of schools performance are central to school improvement efforts, for systems of school accountability and to broader educational policies. In this context, many countries have implemented accountability systems which aim to monitor schools to be accountable for their functioning. The accountability system involves a process of the school performance on the basis of student performance measures. The obtained information is used by policy makers, teachers, principals and parents for developing educational policies, improve professional practice and school choice. It also allows to track the quality of schools and educational systems Cohen et al. (1996).

According to Mortimore (1991): "An effective school is one in which students progress further than might be expected from consideration of its intake". Thus, school effectiveness seeks to identify the 'Value Added' by schools to student outcomes. Value-Added models are used to estimate the contribution of teachers, educational programs or schools to student achievement. From a methodological point of view, this can be achieved by modeling student's scores taking into account the differences in prior achievements and possibly other measured characteristics in the form of covariates at both the school level and the student level; see Braun et al. (2010); OECD (2008); Raudenbush (2004) and Timmermans et al. (2011). The role of the covariates is to characterize a school of reference with respect to which the Value-Added is substantively interpreted. As a matter of fact, the Value-Added of a school is a comparison between the conditional expected scores in a given school and the conditional expected score in the school of reference: if the covariates are modified, the school of reference is also modified and the meaning of the Value-Added changes. Because the interest is to know the net contribution of a school, and the covariates have an influence on the student performance, an important requirement for the covariates is that they have to be unrelated to the internal pedagogical processes performed by a school. Using the econometric jargon, it is said that the covariates are *exogenous* with respect to the school. A standard approach to model the school Value-Added is the use of hierarchical linear models (HLM), or multilevel models Goldstein (2002); Snijders and Bosker (1999), due to the hierarchical structure of the data where students are nested into schools (see chapter 1). Under this approach, students scores are explained by their previous achievement, some covariates and a random effect representing the school effect. A measure of Value-Added has been typically obtained as the prediction of the random school effect Aitkin and Longford (1986); Longford (2012); Raudenbush and Willms (1995); Tekwe et al. (2004).

As every statistical model, some assumptions need to be made for the inferences to be valid. For instance, the above mentioned requirement of exogeneity of covariates is key to isolate the net effect of the school represented by the random effect in the model, yet it is not considered by most of the models currently in use for the estimation of value-added measures. This can have serious consequences as the reported measures of value added could be misleading. Despite these problems, hierarchical linear models are widely accepted for estimating Value-Added.

Chilean context

The Chilean educational system is not exempted of this reality. For instance, every year when results of educational tests are released, a remarkable difference by social economic status (SES), and types of schools (i.e., private or public) is seen not only at the scores level but also in terms of variability. The model used to calculate measures of value added should accordingly take into account this type of heterogeneity. Also, chilean educational systems do not ensure the exogeneity (i.e., the fact that covariates and the school effect in the model are uncorrelated) of the prior scores with respect to the school effect, because students are typically educated by the same school during the two test occasions, meaning that the prior score already contains the effect of the school.

Absence of exogeneity makes impossible to isolate the net contribution of a school to student achievement.

Over the last decades, Chile has been consolidating a national large-scale standardized test called SIMCE (Sistema de Medición de la Calidad de la Educación, System Measurement of Quality of Education). Since its creation at the end of 80s, it has had a strong, increasing and persistent development as a key component of Chilean education policies Manzi and Preiss (2013); Meckes and Carrasco (2010). SIMCE is a census national tests that measures three subjects (language, mathematics and science), and it is administered at 2nd, 4th, 6th, 8th, 10th grade levels. For 2nd and 4th grade, the subjects of language and mathematics are tested annually.

Between 2007 and 2009, the SIMCE office from the Chilean Ministry of Education commissioned three Value-Added studies. One national Value-Added study was conducted using the 2004 and 2006 SIMCE applications, using score information of the 8th and 10th level in schools (see Del Pino, San Martín, Manzi, González, and Taut (2008)) and two Value-Added analyses at the Metropolitan Region level (see Del Pino, González, Manzi, and San Martín (2009); Del Pino, San Martín, de la Cruz, et al. (2008)). These studies are relevant not only for being the first national Value-Added analyses performed in Chile with governmental support, but also because they showed that the ranking of schools induced by Value-Added indicators is dramatically different from the ranking induced by the average SIMCE - scores. Currently, the schedule for SIMCE administrations considers 6 out of the 12 levels of Chilean education.

In 2009, Chile passed a law introducing a National Accountability System based fundamentally on SIMCE results, but not exclusive. Accordingly, it was created the National Agency for Quality (where the SIMCE office is now administratively located) whose fundamental task is to classify schools according to an official methodology of classification. This classification is not neutral for schools: schools poorly performing will be closed after three years ranked in the category at the bottom of the four categories introduced by the new classification school, whereas the practices of schools highly performing will be transferred to the all of the system. The methodology of school classification is essentially based on a linear regression, although the Law under which such an accountability system is founded contemplates the possibility of using Value-Added models. Note that because every year a measure of value-added is calculated for each school, it would be of interest for schools to assess whether the value-added is stable or it changes across the years. Knowing how the value added evolves over time is crucial information for schools decision-making.

The Chilean educational system shows particular features which are reflected in the scores of SIMCE tests. For instance, every year when results are released, a remarkable difference by social economic status (SES), and types of schools (i.e., private or public) is seen not only at the scores level but also in terms of variability. Moreover, the Chilean available longitudinal data at the student level, does not ensure the exogeneity of the prior scores with respect to the school effect, because students were educated by the same school during the two test occasions, therefore, the prior score measure contains the effect of the school Manzi, San Martín, and Van Bellegem (2014).

Although these data have been used for some studies, apart from government decision-making, the research agenda on School Effectiveness Research follows its own path. Studies using Value-Added in Chile are practically non-existent with notable recent exceptions as, Carrasco and San Martín (2012), San Martín and Carrasco (2012), González, San Martín, Manzi, and Del Pino (2010, Agosto); Page, Orellana, San Martín, and González (2015); Santelices, Galleguillos, González, and Taut (2015),Thieme, Prior, Tortosa-Ausina, and Gempp (2013), this dissertation.

Motivation

To develop a robust measure of school effectiveness through value-added model for a fairer classification of schools, has been the principal motivation for this dissertation and it is certainly a great challenge. Commonly, as previously mentioned, the estimation of value-added of a school is determined through standard hierarchical lineal models, but in some educational data, these models have some problems to consider. For example in the chilean case, the first measurement of SIMCE is taken when the student was already educated by the same school, i.e. the exogeneity in model not met, as the students are educated by the same school during the two SIMCE test occasions. Moreover, for the modeling of value added across the time, it is at least arguable that currently is not considered a dependency in time and the added value of each cohort is considered independent (understood by cohort, the group of students that were measured on two occasions, for example the students give a test SIMCE in 4th grade and four years after in 8th grade). Under this discussion is possible that a school does not improve when it is compared with a school of reference, but it does when it is compared with itself.

This way the problems associated in modeling educational data, leads us to a twofold motivation, i) Develop an extension for model of value added developed by Manzi et al. (2014), that consider the endogenous problem, and ii) Develop a model of value added that include a dependence across the time.

Endogenous problem in data educational

The endogenous problem can be represented as in Figure 1, where students of a school j are evaluated by SIMCE on two different occasions, 4th and 8th grades. In both measurements the students belong to school j, therefore the results obtained are achieved from human capital of students and the school effect that it is the same for both SIMCE. Thus the school effect has a partnership between the prior and post scores of SIMCE test.

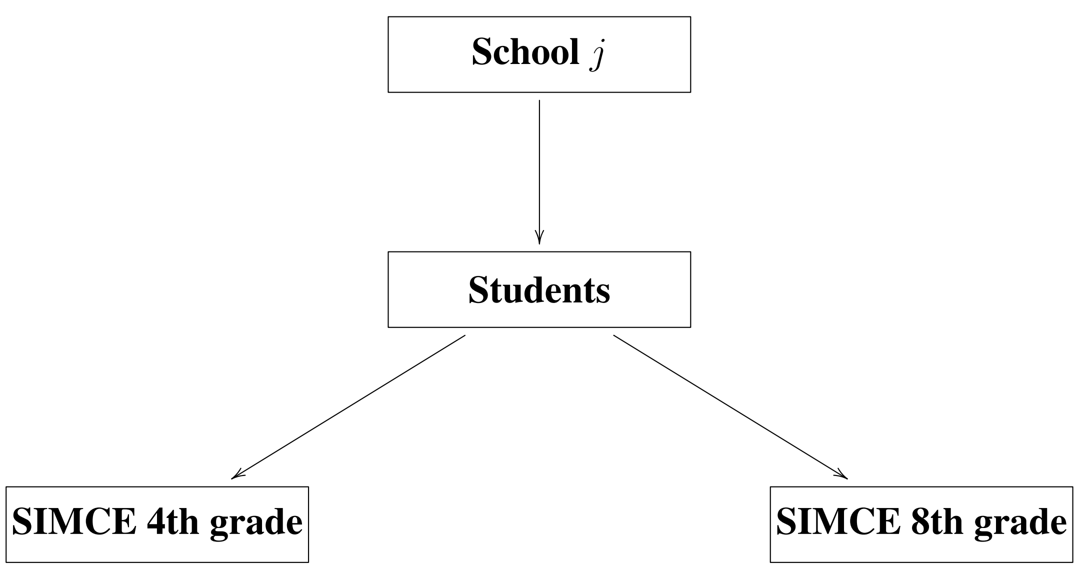


Figure 1: Example of endogenous problem in chilean context

Longitudinal Value Added Models

The problem of dependence across time can be represented as shown in Figure 2, where the students of a school j are evaluated by SIMCE on two different occasions, for example 4th and 8th grades, but constantly the school j is being evaluated. Though we talk about different students, they belong to the same school, then it is intuitive to think that the school effect of the first cohort is associated with the school effect of second cohort, assume independence is at least questionable.

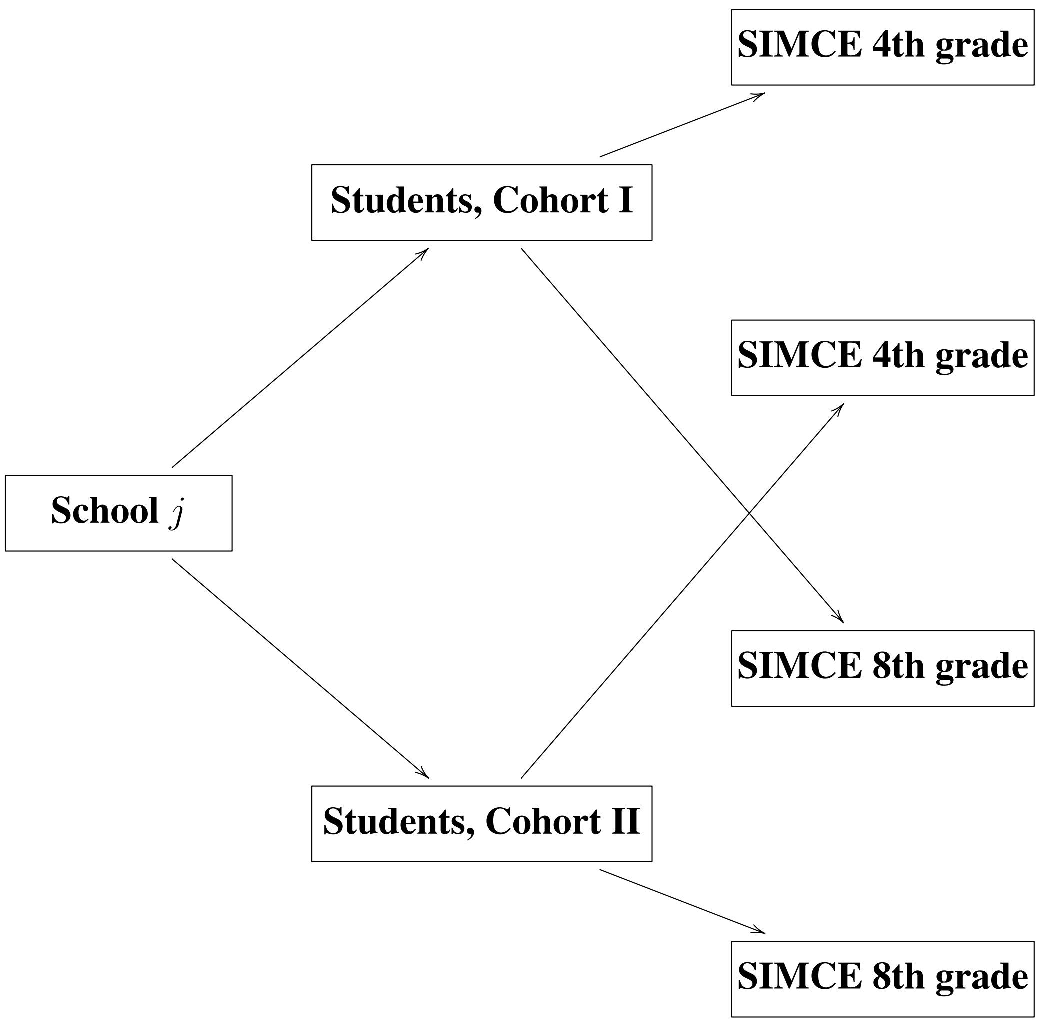


Figure 2: Example of dependence across the time in chilean context

Organization of the thesis

For convenience of the reader, some concepts may be repeated on the later chapter. This dissertation is organized in four chapters.

The first chapter focuses on the concept of school effectiveness. How the effectiveness of a school is estimated. Recall the notation of value added under free-model and it is described the standard methodology for estimate of school effectiveness. Finally its difficulties and associated problems are mentioned it.

In chapter 2, we propose an extension of the endogenous value-added model as introduced by Manzi et al. (2014). The extension essentially consists in specifying a different correlation between an endogenous covariate and the school effect according to the subgroup to which a school belongs. We derive its statistical properties and the identification of interest parameter. Also, we show a study of simulation, and an application in the chilean context.

In chapter 3 we propose a model of value added with dependence across the time. We show the specification of model, the identification of interest parameters and its process of estimation. Also we present, a study of simulation and an application to educational data, in the chilean context.

Chapter 4 discusses the advantages and limitations of two value-added models proposed, and of course possible extensions, improvements and future works.

Chapter 1:

School Effectiveness

Chapter 1:
School Effectiveness

Contents

List of Figures

Chapter 1: School Effectiveness

This chapter focuses on what we know about the concept of school effectiveness, how the effectiveness of a school is estimated, and what methodology is used commonly, with their difficulties and problems associated.

1.1 School Effectiveness

The "school effectiveness" is an enduring problem in research (onward SER[1]). In general terms, it tries to identify the "value added" from school on the outcomes of their students. From a methodological point of view, this can be achieved by modelling student's scores, taking into account the differences in prior achievements and possibly other measured characteristics in the form of covariates at both the school level and the student level; see Raudenbush (2004), OECD (2008), Braun et al. (2010) and Timmermans et al. (2011).

Moreover, it is very important to determine the school effectiveness consider the structure of the educational data. These present evidently a hierarchical structure, because each student belongs to one school and only one, in other words, students are grouped within the schools, or students are nested in schools. Thus, the educational data have at least 2 levels corresponding to students and schools, such that the lowest level are students and the schools are on the top, as shown in Figure 1.1, borrowed from Goldstein (2002).

[1] SER : School Effectiveness Research.

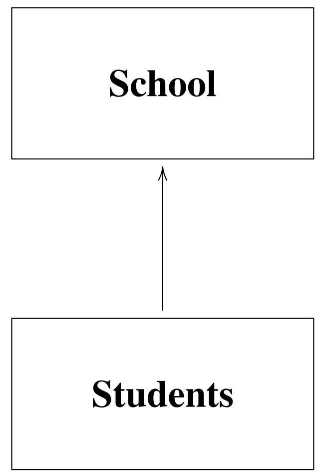


Figure 1.1: Hierarchical relationship among units

Thus, following the structure of educational data, we denote by;

- Y_{ij} score (outcome in a standardized test) of student i belonging to school j.
- $\boldsymbol{X}_{ij}$ be a vector of covariates at both the individual level and the school level; we denote by $\boldsymbol{X}_j$ the set of all the covariates $\boldsymbol{X}_{ij}$s.
- The school effect θ_j is an unobserved random variable defined by two conditions:

Condition 1 *For each school j, the scores $\{Y_{1j}, \ldots, Y_{n_j j}\}$ are mutually independent conditionally on both the school effect θ_j and the set of all the covariates $\boldsymbol{X}_j$.*

Condition 2 *For each student i belonging to school j, the conditional distribution of the score Y_{ij} depends on both the school effect θ_j and the covariates $\boldsymbol{X}_{ij}$.*

1.1.1 Definitions of school effect

Let us begin by defining the concept of school effect in a *model-free way*. By a model-free way we mean that the basic concept of school effect should be defined without resorting on a particular statistical specification (such as a hierarchical linear mixed model).

The condition 1 previously mentioned, identifies the factor by which the scores

$\{Y_{1j}, \ldots, Y_{n_j j}\}$ are related between them. This factor is no longer captured by the covariates, but it corresponds to the unobserved school effect. Heuristically speaking, this means that a feature common to all the students of a given school (which it is supposed to be captured by the scores Y_{ij}s) is due to the school effectθ_j, and not to the set of covariates $\boldsymbol{X}_j$. In a more technical fashion, condition 1 implies the following statements:

1. Conditionally on the covariates $\boldsymbol{X}_j$, all the within relationships between the scores $\{Y_{1j}, \ldots, Y_{n_j j}\}$, should be accounted for by the way in which each score, Y_{ij}, alone is related to the latent variable, θ_j.

2. The school effect θ_j captures, therefore, the heterogeneity that is present in the scores $\{Y_{1j}, \ldots, Y_{n_j j}\}$ and that is not fully explained by the covariates $\boldsymbol{X}_j$.

The condition 2 means that the score Y_{ij} of each student is stochastically generated by two factors: a common one, the school effect; and a student-specific factor, the covariate $\boldsymbol{X}_{ij}$. It is allowed to include in $\boldsymbol{X}_{ij}$ covariates defined at the school level, as for instance the socio-economic group of the school (they are the same for each pupil). However, these observed covariates defined at the school level are different from the school effect in the sense that the firsts do not explain the heterogeneity that is present in the scores $\{Y_{1j}, \ldots, Y_{n_j j}\}$.

It should be remarked that this definition of school effect underlies the multilevel models typically used in SER, such as two-level or three-level models, including or not random slopes, and so on. However, the proposed model-free definition is relevant for practitioners because it allows to answering the following question: What kind of internal school dynamics are we representing when we adopt the concept of school effect introduced above? The answer to this question is quadruple:

(a) We analyze a school in a specific context explicitly characterized by the covariates we put in the model: if we modify the context (by including or excluding covariates), we modify what we are studying regarding the school and its students.

(b) Given a context, we are representing a school which produces within-heterogeneity in the scores of its students.

(c) Each of those scores is equally affected by the school effect.

(d) The covariates (at both individual and school level) do not explain the school effect, neither the heterogeneity we (suppose to) observe in the school, although such covariates have an impact on the scores.

If some of these implications is not pertinent to a specific school context, then the concept of school effect as defined above is inadequate; other approach needs to be adopted.

Identify school effectiveness under a accountability system is extremely important. It is relevant to know to what extend a specific school should be accounted by the score obtained by their students in a national assessment. The answer to this question should satisfy an ethical constraint: it is not fair that a school be accountable by factors which affecting the score of a student and are far from the control of the school.

Heuristically speaking, the strategy to define a school effect consists in decomposing the score of a student i,belonging to school j, Y_{ij} , into components, and thereafter to identify which of these components are under the control of the school. From a modelling point of view, each score Y_{ij} is affected by both the covariates and the school effect. Therefore, it can be decomposed into three components, each of them being represented in terms of conditional expectations, as follows:

$$Y_{ij} = \underbrace{\mathrm{E}\left(Y_{ij}|_{ij}\right)}_{\text{Component 1}} + \underbrace{\{\mathrm{E}\left(Y_{ij}|\boldsymbol{X}_{ij}, \theta_j\right) - \mathrm{E}\left(Y_{ij}|\boldsymbol{X}_{ij}\right)\}}_{\text{Component 2}} + \underbrace{\{Y_{ij} - \mathrm{E}\left(Y_{ij}|\boldsymbol{X}_{ij}, \theta_j\right)\}}_{\text{Component 3}} \qquad (1.1)$$

for the semantic meaning of the conditional expectation, see Appendix A.1.2. The three components of decomposition (1.1) are by construction uncorrelated between them. Therefore, each of them represents a specific contribution to the students score Y_{ij}. Each of these contributions has a specific meaning:

1. The first component measures the contribution of the covariates vector $\boldsymbol{X}_{ij}$ on the score Y_{ij}.

2. The second component corresponds to the contribution of the school effect θ_j on Y_{ij} after taking into account the contribution of the covariates vector $\boldsymbol{X}_{ij}$ on Y_{ij}.

3. The third component corresponds to the idiosyncratic error, that is, the "part" of Y_{ij} which is not statistically explained either by the school effect θ_j, nor by the covariates vector $\boldsymbol{X}_{ij}$.

1.1.2 A model-free definition of school value-added

From decomposition (1.1), it is palatable that the second component is the one that depends on the school and, consequently, it is under its control. This leads to define the value-added of school j as the average of such a component,see (Manzi et al., 2014), namely

$$\mathrm{VA}_j(\boldsymbol{X}_j) = \frac{1}{n_j}\sum_{i=1}^{n_j}\left\{\underbrace{\mathrm{E}\left(Y_{ij}|\boldsymbol{X}_{ij},\theta_j\right)}_{\text{Term 1}} - \underbrace{\mathrm{E}\left(Y_{ij}|\boldsymbol{X}_{ij}\right)}_{\text{Term 2}}\right\} \tag{1.2}$$

The first term represents an average of the expected score in a specific school, after controlling by the covariates. The second term corresponds to an average of the expected score in the reference school, after controlling by the covariates; this last interpretation rests on the following general property: $E(Y_{ij}|\boldsymbol{X}_{ij}) = E[E(Y_{ij}|\boldsymbol{X}_{ij},\theta_j)|\boldsymbol{X}_{ij}]$, that is, the school effect is integrating out with respect to its distribution.

In the SER literature, school value added (onward VA[2]) is defined in terms of an average or reference school; see, among many others, (Gray, Goldstein, & Thomas, 2003; OECD, 2008; Raudenbush, 2004; Raudenbush & Willms, 1995; Timmermans et al., 2011). Definition (1.2) not only provides an accurate meaning of the "average" or reference school, but also shows in which sense the meaning of school value added is covariates-dependent.

Covariates

Given the definition of VA introduced in the previous section, equation (1.2), in very simple words, one wants compare the expected outcomes of a standardized test of a school specific and the expected results of a reference school, also called as "average school".

[2]VA: Value Added

Let us illustrate this aspect with an example (taken from (Milla, San Martín, & Van Bellegem, 2014)). Suppose we are considering three different secondary schools. For each of them, the prior attainment score was measured by the end of primary school, whereas the contemporaneous score was measured at the end of second year of secondary school. In Figure 1.2, each school is represented for different colours, so the names of schools are "Green-School", "Blue-School", and "Red-School". All the students belonging school are represented by points green, blue and red, respectively.

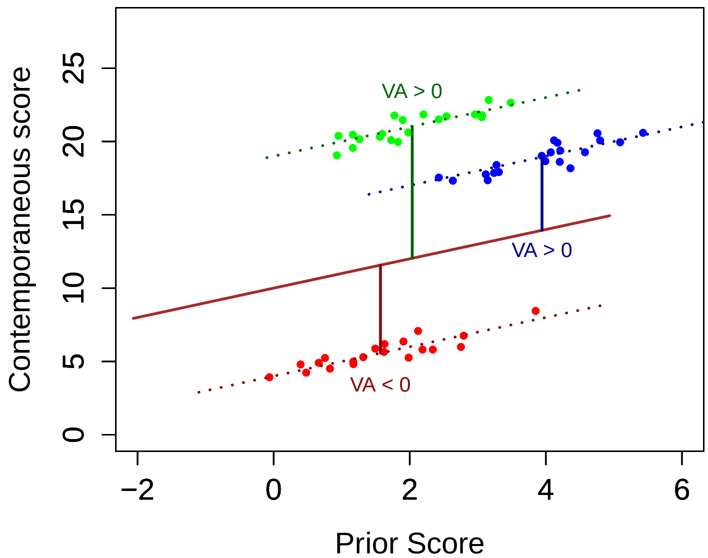


Figure 1.2: **Value Added using as only covariate "Prior Score"**. The dotted lines green blue and red represent the expected contemporaneous scores in each school respectively, after controlling for the prior scores. The continuous dark red line represents the expected contemporaneous scores in the reference school, after controlling for the prior scores.

Therefore, the value added of each school corresponds, to the difference between those expected scores. Thus, in Figure 1, the school represented by the "green School" and "blue School" have a positive value added, whereas the "red School" has a negative value added.

This example shows that the value added is a concept relative to a reference (the dark red line in Figure 1), which in turn is characterized by the covariates that are included in the value added analysis. Therefore, it should be asked what is it the impact of the covariates on the value added indicators. In this example, the reference school was computed using the information of all students. However, it should be asked whether the definition of such a reference based on the prior attainment score is fair. This question should be answered taking into account a specific school context. For instance, it would be asked whether is it fair to define the “average” school without considering the different geographical provinces in a country, or if is it fair to define such an “average” school without distinguishing “highly selective schools” from “moderate selective schools”. These questions lead to conclude that the reference or average school is not a unique entity, but it is characterized in different ways by covariates related to those and similar questions.

In order to illustrate this aspect, let us consider now a value added model in which we control by the prior score and one characteristic of school. These characteristic of school can be for example monthly cost, social economic group or the school composition effect, etc. The compositional effect can be simply the average at school level of the prior scores, or also with a correction, such that, the compositional effect of a student i is defined as the average at school level of the prior scores without the student i, ie if Y_{ij1} then the compositional effect is defined as,

$$\bar{Y}_{j,(-i)1} = \frac{1}{n_j - 1} \sum_{i \neq j}^{n_j - 1} Y_{ij1} = \frac{1}{n_j - 1} \left\{ n_j \bar{Y}_{.j1} - Y_{ij1} \right\}$$

In order to illustrate this aspect, let us consider now a value added model in which we control by both the prior score and the compositional effect defined as the average at school level of the prior scores.

In Figure 1.3, it considers a value added model in which we control by both the prior score and the compositional effect. Such that, four schools are represented and grouped into two groups: the two green schools (points and triangles) have the same compositional effect (which is positive), whereas the two schools blue (squares and diamonds) also have the same compositional effect, but a negative one. If one characteristic of school is included as a control variable, changes the reference school, as shown in example, where the value added of school “triangle green” is negative under this school (continuous line dark green). Nevertheless, its value added is positive if compared with other reference school, where only the prior scores is using as control (dotted line dark red).

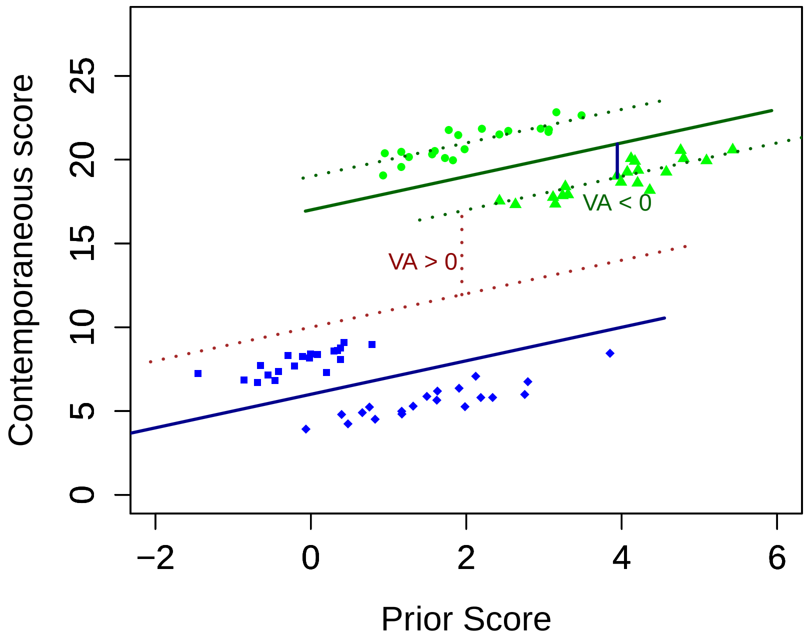


Figure 1.3: **Value Added using as covariates "Prior Score" and characteristic of school**. The dotted lines green represent the expected contemporaneous scores in each school respectively, after controlling for the prior scores. The continuous dark green line represents the expected contemporaneous scores in the reference school, after controlling for the prior scores and a characteristic of school. While, the dotted lines dark red is the expected contemporaneous scores in the reference school, after controlling only by prior scores

From this illustration, we conclude that the school value added not only changes when the covariates are modified, but also shows how its meaning is modified. As a matter of fact, continuing with the illustration, the following interpretation of the different value added indicators can be proposed:

1. If both the individual prior score and the compositional effect are included in the value added model, then we are not only controlling by the initial level of each student, but also for the quality of the group of students. Therefore, this type of value added model measures the effectiveness of a program to educate students after taking into account not only the individual initial achievement level, but also the group initial level. Consequently, the policy usability of this type of models is related to the possible difference between pedagogical methods.

2. If the individual prior score is only included, then the value added measure the effectiveness of a program to educate students after controlling for their initial

prior scores. The effectiveness of the program is only related to the students and, therefore, this type of indicators would be valuable for parents when they want to know if their children are educated in a school producing the largest effect on the contemporaneous performance.

Summarizing, the meaning of a value added indicator depends on the control variable we put in the model. This is due to the fact that the "average" or reference school is characterized by those covariates. Before performing a value added analysis, it is desirable to make explicit the policy purposes of doing value added. The covariates should be chosen according to those purposes.

1.2 Standard hierarchical linear mixed model, two levels

A standard approach to model the school value-added are the hierarchical linear mixed models (onward HLM[3]), or multilevel models Goldstein (2002); Raudenbush and Bryk (2002); Snijders and Bosker (1999). This class of models fits a specific feature of the educational data typically used to perform value-added analysis, namely the hierarchical structure of the data where students are nested into schools. In this context, the value-added of a school is, on the one hand, modelled as a random effect and, on the other hand, calculated as the estimated prediction of it (Aitkin & Longford, 1986; Longford, 2012; Raudenbush & Willms, 1995; Tekwe et al., 2004).

1.2.1 Model specification

From the perspective of the HLM model and following with above notation, Y_{ij} is the score of student i belongs school j in a particular area, for example mathematics, language, etc., such that $i \in \{1, \ldots, n_j\}$, with n_j is the number of students in the school j and $j \in \{1, \ldots, J\}$. Further, it is possible to identify a latent source of variation in the results of students at the level of school, ie these are affected by a school effect, denoted by θ_j. At the same time, this observed test score is explained by some covariables, denoted as $\boldsymbol{X}_{ij}$, to individual level (level 1), as the prior score, and to school level (level 2), as a

[3]HLM: Hierarchical Lineal Models

group effect captured by the average of the prior score, see Figure 1.4.

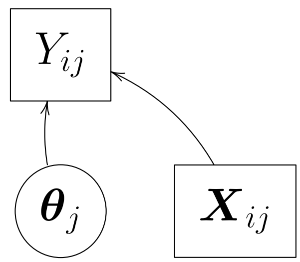


Figure 1.4: Relationship among Post score (Y_{ij}), Covariates ($\boldsymbol{X}_{ij}$) and school effect (θ_j) in a standard HLM model

Moreover, the random school effect is not correlated with the covariables. Consequently, we assume that the test scores between students given the school effect are independent, this assumption is called axiom of local independence (Lazarsfeld, 1950). This way, if denote by $\boldsymbol{X}_j^\top = (\boldsymbol{X}_{1j}, \ldots, \boldsymbol{X}_{n_j j})^\top$ as the design matrix covariates of dimension $n_j \times K$, where K is the number of covariates, then the axiom of local independence is represented by,

$$\underset{1 \leq i \leq n_j}{\perp\!\!\!\perp} Y_{ij} \mid \boldsymbol{X}_j \ , \ \theta_j$$

for each school j. The model structure is linear, then it is assumed that the expected final individual score depend linearly of covariables and the school effect, i.e, for each student i belonging to school j, it is assumed that exists a vector of k parameter, denoted by β, such that,

$$\mathrm{E}(Y_{ij}|\theta_j) \ = \ \boldsymbol{X}_{ij}\boldsymbol{\beta} + \theta_j$$

which is equivalent to writing

$$Y_{ij} \ = \ \boldsymbol{X}_{ij}\boldsymbol{\beta} + \theta_j + \epsilon_{ij}$$

where ϵ_{ij} is the idiosyncratic error, that are independent across students and school.

Structural model

Denote by $\boldsymbol{Y}_j^\top = (Y_{1j}, \ldots, Y_{n_j j})$ the vector of scores in outcome of school j, then the model multidimensional of school j is write as,

$$\boldsymbol{Y}_j = \boldsymbol{X}_j\boldsymbol{\beta} + \boldsymbol{\imath}_{n_j}\theta_j + \boldsymbol{\epsilon}_j$$

where $\boldsymbol{\imath}_{n_j} = (1, \ldots, 1)^\top$ is a vector of ones of dimension n_j. Now, the model multidimensional of all school can be written as,

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{L}\theta + \boldsymbol{\epsilon}$$

such that, $\boldsymbol{L}$ is a matrix of dimension $N \times J$ with entries 0 and 1 (see Appendix), where $N = \sum_{i=1}^{J} n_j$ total number of students, $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_J)^\top$ is the vector of scores in outcome of all student. Similarly, $\boldsymbol{X} = (\boldsymbol{X}_1^\top, \ldots, \boldsymbol{X}_J^\top)^\top$ is the matrix of covariates of dimension $N \times K$, and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_J)^\top$ is a vector of dimension J. Thereby, the structure of model assumes the following conditions;

Assumption 1 ***: Exogeneity,***

The covariates matrix $\boldsymbol{X}$ is independent of vector random effects $\boldsymbol{\theta}$.

$$\mathit{Cov}(\boldsymbol{X}, \boldsymbol{\theta}) = \boldsymbol{0}$$

Assumption 2 ***: Independence of random effect,***

The θ_j's are mutually independent and their distribution for each j is,

$$\theta_j \sim N(0\,, \tau^2)$$

Assumption 3 ***: Distribution and Homoscedasticity of the idiosyncratic error,***

The ϵ_j's for each j are mutually independent with common distribution,

$$\epsilon_{ij} \sim N(0\,, \boldsymbol{I}_N\sigma^2)$$

with $\sigma^2 > 0$*.*

Assumption 4 ***: Local Independence***

The local independence corresponds to

$$\underset{1 \le i \le n_j}{\perp\!\!\!\perp} Y_{ij} \mid \boldsymbol{X}_j \;,\; \theta_j,$$

as it was seen previously.

Thus considering these four assumptions, HLM model is specified as follows,

$$Y_{ij} \sim \mathcal{N}\left(\boldsymbol{X}_{ij}\boldsymbol{\beta} + \theta_j; \sigma^2\right), \quad i = 1, \ldots, n_j, \tag{1.3a}$$

where $\boldsymbol{X}_{ij}$ is a covariates matrix of dimension $n_j \times K$, $\boldsymbol{\beta} \in \mathbb{R}^K$, n_j is the number of students in the school j and $\sigma^2 > 0$.

$$\text{(i)}\, \theta_j \perp\!\!\!\perp \boldsymbol{X}_{ij} \quad \text{(ii)}\, \theta_j \sim\sim \mathcal{N}\left(0; \tau^2\right) \tag{1.3b}$$

with $\tau^2 > 0$ and $(\boldsymbol{X}_{ij})$ is left unspecified.

1.2.2 Parameter identification

The structural of model given by (1.3a) and (1.3b) explains the data generating process that is characterized by the conditional distribution of $(\boldsymbol{Y}_j)$ given $\boldsymbol{X}_j$ for $j \in J$, which correspond to a n_j-multivariate normal distribution, such that,

$$\boldsymbol{Y}_j|\boldsymbol{X}_j \;\sim\; \mathcal{N}\left(\boldsymbol{X}_j\boldsymbol{\beta} \;;\; \sigma^2\boldsymbol{I}_{n_j} + \tau^2\boldsymbol{\imath}_{n_j}\boldsymbol{\imath}_{n_j}^\top\right)$$

Recalling that the mean and the variance-covariance matrix of a multivariate normal distribution are identified, we have that,

1. As $\mathrm{E}\left(\boldsymbol{Y}_j|\boldsymbol{X}_j\right)$ is identified then $\boldsymbol{X}_j\boldsymbol{\beta}$ is identified, this way if the matriz $\boldsymbol{X}_j$ is the complete rank, ie $\mathrm{r}\left(\boldsymbol{X}_j\right) = K$, then the parameter $\boldsymbol{\beta}$ is identified.
2. From $\mathrm{Var}\left(\boldsymbol{Y}_{ij}|\boldsymbol{X}_j\right)$ and $\mathrm{Cov}\left(\boldsymbol{Y}_{ij}, \boldsymbol{Y}_{kj}|\boldsymbol{X}_j\right)$ for $i \neq k$ τ^2 and σ^2 become identified.

1.2.3 Definition of Value Added

Using the definition of value-added in a model, equation (1.2),

$$VA_j = \frac{1}{n_j}\sum_{i=1}^{n_j} \mathrm{E}\left(Y_{ij}|\boldsymbol{X}_j, \theta_j\right) - \frac{1}{n_j}\sum_{i=1}^{n_j} \mathrm{E}\left(Y_{ij}|\boldsymbol{X}_j\right)$$

under the HLM model, $\mathrm{E}\left(Y_{ij}|\boldsymbol{X}_j, \theta_j\right) = \boldsymbol{X}_{ij}^{\top}\boldsymbol{\beta} + \theta_j$, and $\mathrm{E}\left(Y_{ij}|\boldsymbol{X}_j,\right) = \boldsymbol{X}_{ij}^{\top}\boldsymbol{\beta}$.

This way, the value added of a school j under HLM models is equal to school effect,

$$\mathrm{VA}_j = \theta_j$$

Estimation of school effect

For the prediction of random effect is used the empirical Bayes prediction in which the unknown parameters are replaced by their estimators (see next section 1.2.4, for the estimation of parameters). It is used the following identity, equation (1.4), which is valid under the linearity assumption of the conditional expectations; $\mathrm{E}\left(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\theta}\right)$ and $\mathrm{E}\left(\boldsymbol{\theta}\right)$, (Florens, Marimoutou, & Péguin-Feissolle, 2007, see)

$$\mathrm{E}\left(\boldsymbol{\theta}_j|\boldsymbol{Y}_j, \boldsymbol{X}_j\right) = \mathrm{E}\left(\boldsymbol{\theta}_j|\boldsymbol{X}_j\right) + \mathrm{Cov}\left(\boldsymbol{\theta}_j, \boldsymbol{Y}_j|\boldsymbol{X}_j\right)\left[\mathrm{V}\left(\boldsymbol{Y}_j|\boldsymbol{X}_j\right)\right]^{-1}\left(\boldsymbol{Y}_j - \mathrm{E}\left(\boldsymbol{Y}_j|\boldsymbol{X}_j\right)\right) \tag{1.4}$$

Therefore, using the prior expression (1.4), the prediction the school effect correspond to,

$$\mathrm{E}\left(\boldsymbol{\theta}_j|\boldsymbol{Y}_j, \boldsymbol{X}_j\right) = \left[\frac{\tau^2}{\sigma^2 + n_j\tau^2}\right]\boldsymbol{\imath}_{n_j}^{\top}\left(\boldsymbol{Y}_j - \boldsymbol{X}_j\boldsymbol{\beta}\right) \tag{1.5}$$

1.2.4 Parameter Estimation

To facilitate the calculations of the parameters estimates, we rewrite the conditional distribution of score given the covariates as follows,

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}, \quad \text{where} \quad \mathbf{e} = \mathbf{L}_{\underline{n}}\theta + \epsilon \tag{1.6}$$

Let $\mathrm{V}(\mathbf{e}|\mathbf{X}) = \mathbf{V}$, and $\boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_{n_1} & & \\ & \ddots & \\ & & \boldsymbol{V}_{n_J} \end{pmatrix}$, such that $\boldsymbol{V}_{n_j} = \boldsymbol{\imath}_{n_j}\boldsymbol{\imath}_{n_j}^\top \tau^2 + \sigma^2 \boldsymbol{I}_{n_j}$

for each $j = 1, \dots, J$, such that

$$\boldsymbol{V} = \boldsymbol{Q}\sigma^2 + \boldsymbol{P}\boldsymbol{D}^2\boldsymbol{P}^\top\tau^2 + \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^\top\sigma^2$$

so that $\boldsymbol{Q}$ is called Within-operator, and $\boldsymbol{P}$ is called Between-operator, see Appendix A.1.

1) Fixed effects estimators, β

(a) Ordinal least squared (OLS)

Using the equation (1.6) and OLS,

$$\widehat{\boldsymbol{\beta}}_{OLS} = \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\left(\mathbf{X}^\top\mathbf{Y}\right). \tag{1.7}$$

But, the two following properties are necessary for use OLS;

Assumption OLS.1 $E(\mathbf{X}^\top\mathbf{e}) = 0$

Assumption OLS.2 *rank*$(\mathbf{X}^\top\mathbf{X}) = K$ *is of completed rank*

Assumption OLS.3 $E(\mathbf{e}\mathbf{e}^\top) = \sigma^2\boldsymbol{I}_N$

Properties of $\widehat{\boldsymbol{\beta}}_{OLS}$

(i) $\widehat{\boldsymbol{\beta}}_{OLS} = \boldsymbol{\beta} + \left(\frac{1}{N}\mathbf{X}^\top\mathbf{X}\right)^{-1}\left(\frac{1}{N}\mathbf{X}^\top\mathbf{e}\right)$

using Assumption OLS.1, $\mathbf{X}^\top\mathbf{X}$ is not singular then $\left(\frac{1}{N}\mathbf{X}^\top\mathbf{X}\right)^{-1} \xrightarrow{P} A^{-1}$, where $A = \mathrm{E}\left(\mathbf{X}^\top\mathbf{X}\right)$. Further, under Assumption OLS.2, $plim\left(\frac{1}{N}\mathbf{X}^\top\mathbf{e}\right) = \mathrm{E}\left(\mathbf{X}^\top\mathbf{e}\right) = 0$, $plim$ means probability limit. Therefore, by Slutskys theorem

$$\widehat{\boldsymbol{\beta}}_{OLS} \xrightarrow{P} \boldsymbol{\beta} + A^{-1}0$$

Then, $\widehat{\boldsymbol{\beta}}_{OLS}$ is a consistent estimator of $\boldsymbol{\beta}$.

(ii) The asymptotic distribution is derived by

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}\right) = \left(\frac{1}{N}\mathbf{X}^\top\mathbf{X}\right)^{-1}\left(\frac{1}{\sqrt{N}}\mathbf{X}^\top\mathbf{e}\right)$$

where, $\left(\frac{1}{N}\mathbf{X}^\top\mathbf{X}\right)^{-1} - A^{-1} = op(1)$, and by theorem central limit

$\left(\frac{1}{\sqrt{N}}\mathbf{X}^\top\mathbf{e}\right) \xrightarrow{D} \mathrm{N}\left(\mathbf{0}\,;\, B\right)$. This way,

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}\right) = A^{-1}\left(\frac{1}{\sqrt{N}}\mathbf{X}^\top\mathbf{e}\right) + op(1)$$

because $\left(\frac{1}{\sqrt{N}}\mathbf{X}^\top\mathbf{e}\right) op(1) = op(1)$. Now, if Assumption OLS.3 is satisfied, i.e. $\mathbf{V} = \mathbf{I}\sigma^2$ then,

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}\right) \overset{a}{\sim} \mathrm{N}\left(\mathbf{0}\,;\, A^{-1}BA^{-1}\right)$$

where $B \equiv \mathrm{V}\left(\frac{1}{\sqrt{N}}\mathbf{X}^\top\mathbf{e}\right) = \frac{1}{N}\mathbf{X}^\top\mathbf{X}$,

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}\right) \overset{a}{\sim} \mathrm{N}\left(\mathbf{0}\,;\, A^{-1}\right)$$

this way

$$\widehat{\boldsymbol{\beta}}_{OLS} \overset{a}{\sim} \mathrm{N}\left(\boldsymbol{\beta}\,;\, \boldsymbol{X}^\top\boldsymbol{X}\right)$$

However, the Assumption OLS.3 is not satisfied in this model, the data set is heteroscedastic thus to estimator of effects fixed by OLS is not suitable.

(b) Generalized Least Squares (GLS)

This methods of estimation have the following Assumptions,

Assumption GLS.1 $E(\mathbf{X}^\top \mathbf{e}) = 0$, *that each element of* **e** *is uncorrelated with each element of* **X**.

Assumption GLS.2 *Let,* $\boldsymbol{V} \equiv E(\mathbf{e}\mathbf{e}^\top)$, *where* $\boldsymbol{V}$ *is positive definite and* $E(\mathbf{X}\boldsymbol{V}^{-1}\mathbf{X}^\top)$ *is not singular.*

This way,

$$\widehat{\boldsymbol{\beta}}_{GLS} = \left(\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{Y}\right) \tag{1.8}$$

Properties of $\widehat{\boldsymbol{\beta}}_{GLS}$

(i) $\widehat{\boldsymbol{\beta}}_{GLS} = \boldsymbol{\beta} + \left(\frac{1}{N}\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{X}\right)^{-1}\left(\frac{1}{N}\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{e}\right)$

using Assumption GLS.2, $\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{X}$ is not singular, then $\left(\frac{1}{N}\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{X}\right)^{-1} \xrightarrow{P} A^{-1}$, where $A = \mathrm{E}\left(\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{X}\right)$. Further, under Assumption GLS.1, $plim\left(\frac{1}{N}\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{e}\right) = \mathrm{E}\left(\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{e}\right) = 0$, $plim$ means probability limit. Therefore, by Slutskys theorem

$$\widehat{\boldsymbol{\beta}}_{GLS} \xrightarrow{P} \boldsymbol{\beta} + A^{-1}0$$

Then, $\widehat{\boldsymbol{\beta}}_{GLS}$ is a consistent estimator of $\boldsymbol{\beta}$.

(ii) The asymptotic distribution is derived by

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{GLS} - \boldsymbol{\beta}\right) = \left(\frac{1}{N}\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{X}\right)^{-1}\left(\frac{1}{\sqrt{N}}\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{e}\right)$$

where, $\left(\frac{1}{N}\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{X}\right)^{-1} - A^{-1} = op(1)$, and by theorem central limit

$\left(\frac{1}{\sqrt{N}}\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{e}\right) \xrightarrow{D} \mathrm{N}\left(\mathbf{0}\,;\,B\right)$, with $B = \frac{1}{N}\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X}$. This way,

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{GLS} - \boldsymbol{\beta}\right) = A^{-1}\left(\frac{1}{\sqrt{N}}\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{e}\right) + op(1)$$

because $\left(\frac{1}{\sqrt{N}}\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{e}\right) op(1) = op(1)$. Therefore,

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{GLS} - \boldsymbol{\beta}\right) \overset{a}{\sim} \mathrm{N}\left(\mathbf{0}\,;\,A^{-1}\right)$$

Obtaining the GLS estimator $\boldsymbol{\beta}$ requires knowing $\boldsymbol{V}$, but often this variance covariance is unknown.

(c) Feasible Generalized Least Squares (FGLS)

In practice, $\mathbf{V}$ is typically unknown so that the GLS estimator is not available. Substituting an consistent estimator $\widehat{\mathbf{V}}$ for $\mathbf{V}$, which is readily computed from data, then

$$\widehat{\boldsymbol{\beta}}_{FGLS} = \left(\mathbf{X}^\top\widehat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}^\top\widehat{\mathbf{V}}^{-1}\mathbf{Y}\right) \tag{1.9}$$

where $\widehat{\mathbf{V}} \xrightarrow{P} \mathbf{V}$. Consider the following assumptions,

Assumption FGLS.1 $E(\mathbf{X}^\top\mathbf{e}) = 0$, *that each element of* $\mathbf{e}$ *is uncorrelated with each element of* $\mathbf{X}$.

Assumption FGLS.2 *Let,* $\boldsymbol{V} \equiv E(\mathbf{e}\mathbf{e}^\top)$, *where* $\boldsymbol{V}$ *is positive definite and*

$E(\mathbf{X}\boldsymbol{V}^{-1}\mathbf{X}^\top)$ *is not singular.*

By the Weak Law of Large Numbers (WLLN), $plim\left(\frac{1}{N}\boldsymbol{e}\boldsymbol{e}^\top\right) = \mathbf{V}$, the its natural parameter is $\widehat{\mathbf{V}} = \left(\frac{1}{N}\widehat{\boldsymbol{e}}\widehat{\boldsymbol{e}}^\top\right)$.

Note that under this standard HLM model the $\widehat{\mathbf{V}}$ estimator depends of σ^2 and τ^2 estimators.

Asymptotic Properties of $\widehat{\beta}_{FGLS}$

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{FGLS}-\boldsymbol{\beta}\right)=\left(\frac{1}{N}\mathbf{X}^{\top}\widehat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-1}\left(\frac{1}{\sqrt{N}}\mathbf{X}^{\top}\widehat{\mathbf{V}}^{-1}\mathbf{e}\right)$$

but, $\left(\frac{1}{\sqrt{N}}\mathbf{X}^{\top}\widehat{\mathbf{V}}^{-1}\mathbf{e}\right)=\left(\frac{1}{\sqrt{N}}\mathbf{X}^{\top}\mathbf{V}^{-1}\mathbf{e}\right)+op(1)$, then with $A=\left(\frac{1}{N}\mathbf{X}^{\top}\mathbf{V}^{-1}\mathbf{X}\right)$,

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{FGLS}-\boldsymbol{\beta}\right)=A^{-1}\left(\frac{1}{\sqrt{N}}\mathbf{X}^{\top}\mathbf{V}^{-1}\mathbf{e}\right)+op(1)$$

This way,using Assumption FGLS.2 and Slutsky's theorem $\widehat{\boldsymbol{\beta}}_{FGLS}$ is a consistent estimator of $\boldsymbol{\beta}$ and moreover as we saw in GLS,

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{FGLS}-\boldsymbol{\beta}\right)\overset{a}{\sim}\mathrm{N}\left(\mathbf{0}\,;\,A^{-1}\right)$$

As we obtain an estimator of asymptotic variance, Avar($\widehat{\boldsymbol{\beta}}_{GLS}$), by using the consistent estimator of A; $\widehat{\text{Avar}(\widehat{\boldsymbol{\beta}}_{FGLS})}=\widehat{A}^{-1}/N=\left(\mathbf{X}^{\top}\widehat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-1}$.However with heterocedasticity this is not robust, Thus, a robust estimator of the asymptotic variance is

$$\widehat{\text{Avar}(\widehat{\boldsymbol{\beta}}_{FGLS})}=\left(\mathbf{X}^{\top}\widehat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}^{\top}\widehat{\mathbf{V}}^{-1}\widehat{\mathbf{e}}\widehat{\mathbf{e}}^{\top}\widehat{\mathbf{V}}^{-1}\mathbf{X}\right)\left(\mathbf{X}^{\top}\widehat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-1}$$

2) Estimation variance components

The estimation of variance components is through the method of moments. Then,

Estimation of σ^2

Apply the W-operator to in the equation (1.6), we obtain the following (see Appendix A.1 for matricial notation),

$$\boldsymbol{QY}=\boldsymbol{QX\beta}+\boldsymbol{QL}_{\underline{n}}\boldsymbol{\theta}+\boldsymbol{Q\epsilon}$$

but $\boldsymbol{Q}$ and $\boldsymbol{L}_{\underline{n}}$ are orthogonal. So the within regression correponds to,

$$\boldsymbol{QY}=\boldsymbol{QX\beta}+\boldsymbol{Q\epsilon} \tag{1.10}$$

where,

(a) $\text{Var}\,(\boldsymbol{Q\epsilon}|\boldsymbol{X}) = \boldsymbol{QVQ} = \sigma^2\boldsymbol{Q}$

(b) $\hat{\boldsymbol{\beta}}^{\boldsymbol{w}} = (\boldsymbol{X}'\boldsymbol{QX})^{-1}\boldsymbol{X}'\boldsymbol{QY}$

Now, if be defined $\widehat{\boldsymbol{\epsilon}}^{w} = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\boldsymbol{w}})$, then

$$\begin{aligned}\boldsymbol{Q}\widehat{\boldsymbol{\epsilon}}^{w} &= \boldsymbol{Q}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\boldsymbol{w}}) \\ &= (\boldsymbol{Q} - \boldsymbol{QX}(\boldsymbol{X}'\boldsymbol{QX})^{-1}\boldsymbol{X}'\boldsymbol{Q})\boldsymbol{Y}) \\ &= \boldsymbol{MY}\end{aligned}$$

where $\boldsymbol{M} = (\boldsymbol{Q} - \boldsymbol{QX}(\boldsymbol{X}'\boldsymbol{QX})^{-1}\boldsymbol{X}'\boldsymbol{Q})$, such that $\boldsymbol{MM} = \boldsymbol{M}$ and $\boldsymbol{M}^\top\boldsymbol{M} = \boldsymbol{M}$. Then, $\widehat{\boldsymbol{\epsilon}}^{w\top}\boldsymbol{Q}\widehat{\boldsymbol{\epsilon}}^{w} = \boldsymbol{\epsilon}^\top\boldsymbol{M\epsilon}$. This way,

$$\begin{aligned}\text{E}\left(\widehat{\boldsymbol{\epsilon}}^\top\boldsymbol{M}\widehat{\boldsymbol{\epsilon}}\right) &= \text{tr}\,(\boldsymbol{M}\text{V}(\boldsymbol{\epsilon})) \\ &= \text{tr}\left(\boldsymbol{QV} - \boldsymbol{QX}(\boldsymbol{X}'\boldsymbol{QX})^{-1}\boldsymbol{X}'\boldsymbol{QV}\right) \\ &= \text{tr}\left(\boldsymbol{Q}\sigma^2 - \boldsymbol{QX}(\boldsymbol{X}'\boldsymbol{QX})^{-1}\boldsymbol{X}'\boldsymbol{Q}\sigma^2\right) \\ &= \sigma^2\,(N - J) - \sigma^2\,(K^*)\end{aligned}$$

where $K^* \leq K$ is the number of covariates non-zero in the within regression. Therefore, the estimation of σ^2 is given by,

$$\hat{\sigma^2} = \frac{\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\boldsymbol{w}}\right)^\top\boldsymbol{Q}\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\boldsymbol{w}}\right)}{N - J - K^*} \tag{1.11}$$

Estimation of τ^2

The between-regression is formulated of the following way,

$$\boldsymbol{PY} = \boldsymbol{PX\beta} + \boldsymbol{PL}_{\underline{n}}\boldsymbol{\theta} + \boldsymbol{P\epsilon}$$

where,

(a) $\text{Var}\left(\boldsymbol{PL}_{\underline{n}}\boldsymbol{\theta} + \boldsymbol{P\epsilon}|\boldsymbol{X}\right) = \boldsymbol{P}\ \boldsymbol{V}\ \boldsymbol{P}^\top = \tau^2\boldsymbol{I}_J + \sigma^2\boldsymbol{D}^{-1}$

(b) $\hat{\boldsymbol{\beta}}^{\boldsymbol{b}} = (\boldsymbol{X}^\top \boldsymbol{P}^\top \boldsymbol{P}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{P}^\top \boldsymbol{P}\boldsymbol{Y}$

Now, if be defined $\widehat{\boldsymbol{e}}^{\boldsymbol{b}} = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\boldsymbol{b}})$, then

$\boldsymbol{P}\widehat{\boldsymbol{e}}^{\boldsymbol{b}} = \boldsymbol{T}\boldsymbol{P}(\boldsymbol{L}_{\underline{n}}\boldsymbol{\theta} + \boldsymbol{\epsilon})$, where $\boldsymbol{T} = (\boldsymbol{I}_J - \boldsymbol{P}\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{P}^\top \boldsymbol{P}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{P}^\top)$, $\boldsymbol{T}\boldsymbol{T} = \boldsymbol{T}$, and $\boldsymbol{T}^\top \boldsymbol{T} = \boldsymbol{T}$. Then, $\widehat{\boldsymbol{e}}^{\boldsymbol{b}\top} \boldsymbol{P}^\top \boldsymbol{P}\widehat{\boldsymbol{e}}^{\boldsymbol{b}} = (\boldsymbol{L}_{\underline{n}}\boldsymbol{\theta} + \boldsymbol{\epsilon})^\top \boldsymbol{P}^\top \boldsymbol{T}\boldsymbol{P}(\boldsymbol{L}_{\underline{n}}\boldsymbol{\theta} + \boldsymbol{\epsilon})$ and

$$\begin{aligned}
\mathrm{E}\left(\widehat{\boldsymbol{e}}^{\boldsymbol{b}\top} \boldsymbol{P}^\top \boldsymbol{P}\widehat{\boldsymbol{e}}^{\boldsymbol{b}}\right) &= \mathrm{tr}\left(\boldsymbol{T}\mathrm{Var}(\boldsymbol{P}(\boldsymbol{L}_{\underline{n}}\boldsymbol{\theta} + \boldsymbol{\epsilon}))\right) = \tau^2 \mathrm{tr}\left(\boldsymbol{T}\right) + \sigma^2 \mathrm{tr}\left(\boldsymbol{T}\boldsymbol{D}^{-1}\right) \\
&= \tau^2 (J - K^{**}) + \sigma^2 \left(\sum_{j=1}^{J} (1/n_j) - K^{***}\right)
\end{aligned}$$

Thus be have,

$$\hat{\tau^2} = \frac{\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\boldsymbol{b}}\right)^\top \boldsymbol{P}^\top \boldsymbol{P}\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\boldsymbol{b}}\right) - \sigma^2 \left(\sum_{j=1}^{J}(1/n_j) - K^{***}\right)}{J - K^{**}} \tag{1.12}$$

where,
$K^{***} = \mathrm{tr}\left\{(\boldsymbol{X}^\top \boldsymbol{P}^\top \boldsymbol{P}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{P}^\top \boldsymbol{D}^{-1}\boldsymbol{P}\boldsymbol{X}\right\}$, and
$K^{**} = \mathrm{tr}\left\{(\boldsymbol{X}^\top \boldsymbol{P}^\top \boldsymbol{P}\boldsymbol{X})^{-1}(\boldsymbol{X}^\top \boldsymbol{P}^\top \boldsymbol{P}\boldsymbol{X})\right\}$

1.2.5 Summary of the Estimation Process

1. When the process starts $\boldsymbol{\beta}$ is estimated by equation (1.7), OLS.

$$\widehat{\beta}_{OLS} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}\left(\mathbf{X}^\top \mathbf{Y}\right).$$

2. Completed steps 1, be must estimate the Within Residual of the model (1.6). Then, it is estimated σ^2 by the equation (1.11), ie

$$\hat{\sigma^2} = \frac{\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\boldsymbol{w}}\right)^\top \boldsymbol{Q}\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\boldsymbol{w}}\right)}{N - J - K^*}$$

3. Completed steps 1–2, be must estimate the Between Residual of the model (1.6). Then, $\widehat{\tau^2}$ is estimated by the equation (1.12), ie

$$\hat{\tau^2} = \frac{\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\boldsymbol{b}}\right)^\top \boldsymbol{P}^\top \boldsymbol{P}\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\boldsymbol{b}}\right) - \sigma^2\left(\sum_{j=1}^{J}(1/n_j) - K^{***}\right)}{J - K^{**}}$$

4. Completed the steps 1–3, it is possoble calculate the estimation of $\widehat{\mathbf{V}}$.

$$\widehat{\boldsymbol{V}} = \boldsymbol{Q}\widehat{\sigma^2} + \boldsymbol{P}\boldsymbol{D}^2\boldsymbol{P}^\top\widehat{\tau^2} + \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^\top\widehat{\sigma^2}$$

5. Using the results of step 4, the estimation of $\boldsymbol{\beta}$ is recalculated by equation (1.9), i.e. by feasible generalized least squared,

$$\widehat{\boldsymbol{\beta}}_{FGLS} = \left(\mathbf{X}^\top\widehat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}^\top\widehat{\mathbf{V}}^{-1}\mathbf{Y}\right).$$

1.3 Some comments

In the context of school improvement is possible and important develop new sophisticated techniques on modeling of value added, considering the associated problems in the fit of educational data and thereby get results more accurate and and reliable not forgetting of ensure the performance of its assumptions for a correct interpretation. Nevertheless the modeling is not the only relevant since the covariates play an important role in any current methods associated to value added school.

The choice of covariates as well as their meaning is highly dependent on the policy context under which a value added analysis is required, as well as on the social context in which an educational system is organized. The consequences of this paradigm are at least the following:

1. The criteria to choosing covariates should be policy-driven. There does not exist universal covariates that we need to include in a value added model.

2. Covariates are not included in a value added model to ensuring a more reliable estimation of the effectiveness of a school. Covariates define the context under

which a school effectiveness needs to be understood.

3. The role of the covariates in a value added model needs to be delimited with respect to the school effect. Their endogeneity or exogeneity character depends on the school context (in particular, the type of available data).

4. Covariates also allow us to define subgroups of schools. These subgroups need to be considering when the school value added is estimated. The way in which this can be done is through heterocedastic models.

5. Practices of effective schools are not automatically transferred. Although we control by similar characteristics, the level of such characteristics need to be taken into account before arguing transferability.

Chapter 2:

Endogenous Value-Added Models for Subgroups of Schools

Chapter 2: Endogenous Value-Added Models for Subgroups of Schools

Contents

Figures

Tables

Chapter 2: Endogenous Value-Added Models for Subgroups of Schools

This chapter focuses on the problem of endogeneity in some educational data, where it is developed the general structure of a endogenous value-added model and the justification of its specification, beside to the identification the interest parameters, and also it is explained estimation process. Furthermore, this chapter shows a simulation study which compares this model proposed with a standard HLM model. Finally an application in chilean educational data, SIMCE 2009-2013 with 4th and 8th grade respectively.

2.1 Introduction

The Value-added models are typically specified through hierarchical linear mixed (HLM) models (this is discussed in the Chapter 1, Introduction), where the dependent variable corresponds to examinees' test scores (called *contemporaneous scores*), and the vector of covariates or explanatory factors is assumed to be exogenous with respect to the school effect. This exogeneity condition is characterized through a zero correlation between the school effect and the explanatory factors; see Snijders and Bosker (1999). A lagged score is typically included among the exogenous explanatory factors. The school effect is in turn characterized by means of the Axiom of Local Independence, namely, for each school, the examinees' contemporaneous scores are mutually independent conditionally on both the covariates and the school effect. This last condition means that the school effect explains the heterogeneity of the contemporaneous test scores that it is not explained by the covariates.

Although in practice this type of models is widely used, not necessarily the exogeneity of the covariates is guaranteed in each real situation. Thus, for instance, as argued by Spencer and Fielding (2002) and Manzi et al. (2014), the lagged score can be endogenous by design in the sense that this score is measured *after* the school has treated students. In this case, the lagged score measure already contains the school effect and, therefore, the non-correlation between the lagged score and the school effect is questionable. For real examples, beside the case study reported in this chapter (section 2.4.2) , see Gansle, Noell, and Burns (2012). When this is the case, HLM models need to be extended to what we call *endogenous value-added models*.

Before making explicit the underlying structure of endogenous value-added models, let us make precise the endogeneity problem sketched above. The contemporaneous score of pupil i belonging to school j is denoted as Y_{ij}, whereas the corresponding lagged score is denoted as Z_{ij}. Assuming that a school j groups n_j pupils, let $\boldsymbol{Y}_j = (Y_{1j}, \ldots, Y_{n_jj})^\top$ and $\boldsymbol{Z}_j = (Z_{1j}, \ldots, Z_{n_jj})^\top$. Similarly, let $\boldsymbol{X}_j = (\boldsymbol{X}_{1j}, \ldots, \boldsymbol{X}_{n_jj})^\top$, where $\boldsymbol{X}_{ij}$ corresponds to a vector of explanatory exogenous variables related to pupil i belonging to school j. Finally, let θ_j be the school effect corresponding to school $j \in \{1, \ldots, J\}$. The structure of a standard value-added model is summarized by the following conditions:

1. For each school j, $\{Y_{ij} : i = 1, \ldots, n_j\}$ are mutually independent conditionally on $(\boldsymbol{Z}_j, \boldsymbol{X}_j, \theta_j)$.
2. For each school j and each pupil i in school j, the distribution of Y_{ij} depends on $(\boldsymbol{Z}_j, \boldsymbol{X}_j, \theta_j)$ through $(Z_{ij}, \boldsymbol{X}_{ij}, \theta_j)$ only. This condition, along with the previous one, are known as the Axiom of Local Independence.
3. For each school j, the explanatory factors $(\boldsymbol{X}_j, \boldsymbol{Z}_j)$ are uncorrelated with the school effect θ_j.

The endogeneity problem arises when by design the lagged score Z_{ij} is correlated with the school effect. In this case, Z_{ij} is termed an *endogenous variable*.

It is known that an endogeneity problem leads to biased estimators. A standard strategy to solve the problem is through an instrumental variable approach Kim and Frees

(2007); Wooldridge (2008). An instrumental variable $\boldsymbol{W}_{ij}$ is typically characterized by the following three conditions:

i $\boldsymbol{W}_{ij}$ is uncorrelated with the school effect θ_j;

ii $\boldsymbol{W}_{ij}$ is correlated with the endogenous variable Z_{ij};

iii $\boldsymbol{W}_{ij}$ is uninformative to explain Y_{ij} once we condition on $(Z_{ij}, \boldsymbol{X}_{ij}, \theta_j)$

To ensure model identification, the number of instrumental variables should be at least equal to the number of endogenous variables. For a general discussion on instrumental variables, see Angrist and Krueger (2001).

2.1.1 General Structure of Endogenous Value Added Models

A standard approach to specify structural models (as value-added models) is through a recursive decomposition of the underlying global mechanism into an ordered sequence of simpler sub-mechanisms, each one involving an endogenous variable and exogenous variables. In a first stage of statistical modeling, this recursive decomposition should be based on conditional distributions rather than on (linear) equations relating variables. By doing so, not only it is shown why “endogeneity” and “exogeneity” are not “permanent” features of variables, but it also displayed the level at which maintained hypotheses should be introduced. These hypotheses, as the order in which the recursive decomposition should be performed, are justified by contextual considerations; for details, see Mouchart, Russo, and Wunsch (2010) and Wunsh, Mouchart, and Russo (2014).

Let us make precise those considerations in the context of value-added models as sketched in the previous section. The aim is to show that the instrumental variable approach used to solve the endogeneity problem, corresponds to a new model specification that takes into account that both lagged and contemporaneous scores are generated by both the school effect and (possibly) other exogenous covariates. We assume that $\{(\boldsymbol{Y}_j, \boldsymbol{Z}_j, \boldsymbol{X}_j, \boldsymbol{W}_j, \theta_j)) : j = 1, \ldots, J\}$ are mutually independent, which means to consider schools acting independently between them. Each vector $(\boldsymbol{Y}_j, \boldsymbol{Z}_j, \boldsymbol{X}_j, \boldsymbol{W}_j, \theta_j)$ is specified by the family of multivariate distributions

$$\{p^{\alpha}(\boldsymbol{Y}_j, \boldsymbol{Z}_j, \boldsymbol{X}_j, \boldsymbol{W}_j, \theta_j) \,:\, \alpha \in \boldsymbol{A}\}, \quad j = 1, \ldots, J \tag{2.13}$$

representing the underlying global mechanism of interest; here, α denotes the parameter indexing such distributions, while $\boldsymbol{A}$ denotes the corresponding parameter space.

A recursive decomposition of (2.13) requires substantive considerations to select the sub-mechanisms of the decomposition. In the case of school effectiveness, both the prior scores $\boldsymbol{Z}_j$ and the contemporaneous scores $\boldsymbol{Y}_j$ are due to two sources: the school and the covariates that are far from the control of the school. Consequently, (2.13) should be decomposed as follows: for each school j,

$$p^{\alpha}(\boldsymbol{Y}_j, \boldsymbol{Z}_j, \boldsymbol{X}_j, \boldsymbol{W}_j, \theta_j) = p^{\alpha_1}(\boldsymbol{Y}_j, \boldsymbol{Z}_j \mid \boldsymbol{X}_j, \boldsymbol{W}_j, \theta_j)\, p^{\alpha_2}(\theta_j \mid \boldsymbol{X}_j, \boldsymbol{W}_j)\, p^{\alpha_3}(\boldsymbol{X}_j, \boldsymbol{W}_j), \tag{2.14}$$

each component of the right hand side having mutually independent parameters, that is,

$$\alpha = (\alpha_1, \alpha_2, \alpha_3) \in \boldsymbol{A}_1 \times \boldsymbol{A}_2 \times \boldsymbol{A}_3. \tag{2.15}$$

Under condition (2.15), the decomposition (2.14) corresponds to a recursive decomposition and, therefore, the conditioning variables of each term are the exogenous variables of their corresponding component. More precisely, condition (2.15) is a condition ensuring that the parameters α_1, α_2 and α_3 are *variation-free*, that is, no restriction links those parameters. It ensures, therefore, that the inference on the parameter α_1 can be based on the conditional distribution $p^{\alpha_1}(\boldsymbol{Y}_j, \boldsymbol{Z}_j \mid \boldsymbol{X}_j, \boldsymbol{W}_j, \theta_j)$ only, being the distribution generating $(\boldsymbol{X}_j, \boldsymbol{W}_j, \theta_j)$ noninformative for such an inference; for details, see (Barndorff-Nielsen, "1978, Chapter 4). In other words, the generating mechanism of $(\boldsymbol{X}_j, \boldsymbol{W}_j, \theta_j)$ does not contain any relevant information about the parameter α_1, which means that $(\boldsymbol{X}_j, \boldsymbol{W}_j, \theta_j)$ are "exogenous" with respect to the parameter α_1. In a similar sense, the variation-free condition between α_2 and α_3 ensures that $(\boldsymbol{X}_j, \boldsymbol{W}_j)$ are exogenous with respect to the parameter α_2 indexing the conditional distribution $p^{\alpha_2}(\theta_j \mid \boldsymbol{X}_j, \boldsymbol{W}_j)$. It should be remarked that model specification should be performed in such a way that the validity of condition (2.15) be guaranteed by construction.

Remark 1 The concept of *variation-free parameterization* has been used to define exogeneity in a rigorous way; for details and discussion, see, among others,Engle, Hendry, and Richard (1983), Hendry and Richard (1983), Florens and Mouchart (1985), Florens et al. (2007), Spanos (1994).

Model specification is completed by assuming the Axiom of Local Independence, namely that $\{(Y_{ij}, Z_{ij}) : i = 1, \ldots, n_j\}$ are mutually independent conditionally on $(\boldsymbol{X}_j, \boldsymbol{W}_j, \theta_j)$, and that, for each pupil i, the distribution of (Y_{ij}, Z_{ij}) depends on $(\boldsymbol{X}_j, \boldsymbol{W}_j, \theta_j)$ through $(\boldsymbol{X}_{ij}, \boldsymbol{W}_{ij}, \theta_j)$ only.

One component of the previous specification, namely $p^{\alpha_1}(\boldsymbol{Y}_j, \boldsymbol{Z}_j \mid \boldsymbol{X}_j, \boldsymbol{W}_j, \theta_j)$, is still decomposed in two subcomponents because in value-added analysis the lagged score is typically considered as an exogenous explanatory factor of the contemporaneous score. More precisely, for each school j,

$$p^{\alpha_1}(\boldsymbol{Y}_j, \boldsymbol{Z}_j \mid \boldsymbol{X}_j, \boldsymbol{W}_j, \theta_j) = p^{\alpha_1^1}(\boldsymbol{Y}_j \mid \boldsymbol{Z}_j, \boldsymbol{X}_j, \boldsymbol{W}_j, \theta_j)\, p^{\alpha_1^2}(\boldsymbol{Z}_j \mid \boldsymbol{X}_j, \boldsymbol{W}_j, \theta_j), \quad (2.16)$$

where

$$\alpha_1 = (\alpha_1^1, \alpha_1^2) \in \boldsymbol{A}_1^1 \times \boldsymbol{A}_1^2. \quad (2.17)$$

In order to make explicit the specific role of the explanatory factors $\boldsymbol{X}_j$ and $\boldsymbol{W}_j$, the following maintained hypothesis is typically introduced:

$$\boldsymbol{Y}_j \perp\!\!\!\perp \boldsymbol{W}_j \mid \boldsymbol{Z}_j, \boldsymbol{X}_j, \theta_j, \quad j = 1, \ldots, J; \quad (2.18)$$

that is, once we condition on $(\boldsymbol{Z}_j, \boldsymbol{X}_j, \theta_j)$, the explanatory factor $\boldsymbol{W}_j$ is redundant to explain the contemporaneous scores $\boldsymbol{Y}_j$. Under this condition, (2.16) is equivalently rewritten as

$$p^{\alpha_1}(\boldsymbol{Y}_j, \boldsymbol{Z}_j \mid \boldsymbol{X}_j, \boldsymbol{W}_j, \theta_j) = p^{\alpha_1^1}(\boldsymbol{Y}_j \mid \boldsymbol{Z}_j, \boldsymbol{X}_j, \theta_j)\, p^{\alpha_1^2}(\boldsymbol{Z}_j \mid \boldsymbol{X}_j, \boldsymbol{W}_j, \theta_j) \quad (2.19)$$

for each $j = 1, \ldots, J$, and the Axiom of Local Independence above introduced is equivalent to the following conditions:

ALI1. For each school j, $\{Y_{ij} : i = 1, \ldots, n_j\}$ are mutually independent conditionally on $(\boldsymbol{Z}_j, \boldsymbol{X}_j, \theta_j)$; and for each pupil i in school j, the distribution of Y_{ij} depends on $(\boldsymbol{Z}_j, \boldsymbol{X}_j, \theta_j)$ through $(Z_{ij}, \boldsymbol{X}_{ij}, \theta_j)$. This corresponds to assume the Axiom of Local Independence for the conditional model $(\boldsymbol{Y}_j \mid \boldsymbol{Z}_j, \boldsymbol{X}_j, \theta_j)$.

ALI2. For each school j, $\{Z_{ij} : i = 1, \ldots, n_j\}$ are mutually independent conditionally on $(\boldsymbol{W}_j, \boldsymbol{X}_j, \theta_j)$; and for each pupil i in school j, the distribution of Z_{ij} depends on $(\boldsymbol{W}_j, \boldsymbol{X}_j, \theta_j)$ through $(\boldsymbol{W}_{ij}, \boldsymbol{X}_{ij}, \theta_j)$. This corresponds to assume the Axiom of

Local Independence for the conditional model $(\boldsymbol{Z}_j \mid \boldsymbol{X}_j, \boldsymbol{W}_j, \theta_j)$.

Summarizing, the general structure of endogenous value-added models is made explicit in the following recursive decomposition: for each school j,

$$
\begin{aligned}
p^{\alpha}(\boldsymbol{Y}_j, \boldsymbol{Z}_j, \boldsymbol{X}_j, \boldsymbol{W}_j, \theta_j) = \\
= p^{\alpha_1^1}(\boldsymbol{Y}_j \mid \boldsymbol{Z}_j, \boldsymbol{X}_j, \theta_j)\, p^{\alpha_1^2}(\boldsymbol{Z}_j \mid \boldsymbol{X}_j, \boldsymbol{W}_j, \theta_j)\, p^{\alpha_2}(\theta_j \mid \boldsymbol{X}_j, \boldsymbol{W}_j)\, p^{\alpha_3}(\boldsymbol{X}_j, \boldsymbol{W}_j),
\end{aligned}
\tag{2.20}
$$

where

$$
\alpha = (\alpha_1^1, \alpha_1^2, \alpha_3, \alpha_4) \in \boldsymbol{A}_1^1 \times \boldsymbol{A}_1^2 \times \boldsymbol{A}_3 \times \boldsymbol{A}_4. \tag{2.21}
$$

Model specification is completed by conditions **ALI1** and **ALI2**. The recursive decomposition (2.20)-(2.21) can graphically be depicted through the following directed graph:

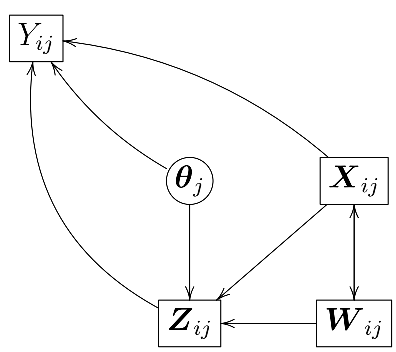


Figure 2.1: Structure of endogeneity problem using an instrumental variable approach

Following the structural equation modeling approach, the observed or manifest variables are represented by rectangular or square boxes, whereas the unobservable variables are represented by circles or ellipses; see Bollen (1989).

The so-called instrumental variable approach can be embedded in this structure. In particular, the three conditions defining an instrumental variable are captured in a general form. In fact, the exogenous character of the instrumental variable with respect to the school effect is captured by the variation-free property (2.21); the dependency between the instrumental variable and the endogenous variable Z_{ij} is captured by the second component at the right hand side of decomposition (2.20); and, finally, the non-informativeness of the instrumental variable with respect to Y_{ij} once we "control" by $(Z_{ij}, \boldsymbol{X}_{ij}, \theta_j)$ is captured by condition (2.18). Consequently, an instrumental variable approach consists in specifying a value-added model taking into account a substantive aspect, namely that both the lagged score and the contemporaneous score are generated by both the school and factors that are far from the control of the school.

Model Specification

For the practical computation of the school value-added, it is necessary to specify the different components of the recursive decomposition (2.20)-(2.21). Special attention should be focused on the compatibility between the corresponding sub-models specifying each component of decomposition (2.20) and the variation-free parametrization property (2.21). Manzi et al. (2014) accordingly propose the following specification for heteroscedastic hierarchical linear instrumental mixed (HHLIM) models, where the heteroscedasticity intends to capture specific social and/or contextual characteristics grouping schools, as for instance the socio-economic status of a school: it is firstly assumed that, for each school j, **ALI1** and **ALI2** are valid. Under these structural hypotheses, for each school j,

$$(Y_{ij} \mid \boldsymbol{Z}_{ij}, \boldsymbol{X}_{ij}, \boldsymbol{W}_{ij}, \theta_j) \sim \mathcal{N}\left(\boldsymbol{X}_{ij}^{\top}\boldsymbol{\beta} + \boldsymbol{Z}_{ij}^{\top}\boldsymbol{\gamma} + \theta_j, \sigma^2_{\rho(j)}\right), \quad i = 1, \ldots, n_j, \tag{2.22a}$$

for some $\boldsymbol{\beta} \in \mathbb{R}^{K_1}$, $\boldsymbol{\gamma} \in \mathbb{R}^{K_2}$ and $\sigma^2_{\rho(j)} > 0$. The function $\rho(\cdot)$ groups the school according to a social or contextual characteristic. In the case study discussed in Section 2.4.2, this characteristic corresponds to the socio-economic status (SES) of the school that it is divided in 5 categories, from A to E being A the lowest SES. Thus, the function $\rho(\cdot)$ is defined as follows: $\rho(j) = A$ if the SES of school j is A; $\rho(j) = B$ if the SES of school j is B; and so on. In the sequel, the function $\rho(\cdot)$ will be called *the grouping school function*. Note that in this specification we are assuming that the explanatory factor $\boldsymbol{Z}_{ij}$ is K_2-dimensional, whereas in the conceptual discussion developed previously it was assumed to be unidimensional;

$$(\boldsymbol{Z}_{ij} \mid \boldsymbol{X}_{ij}, \boldsymbol{W}_{ij}, \theta_j) \sim \mathcal{N}_{K_2}\left(\boldsymbol{A}^\top \boldsymbol{X}_{ij} + \boldsymbol{H}^\top \boldsymbol{W}_{ij} + \delta\theta_j\, \imath_{K_2}, \boldsymbol{\Phi}_{\rho(j)}\right), \quad i = 1, \ldots, n_j \tag{2.22b}$$

with $\boldsymbol{A} = (\alpha_{kl})$ a $K_1 \times K_2$ matrix of real coefficients, $\boldsymbol{H} = (\eta_{lk})$ a $L \times K_2$ matrix of real coefficients, $\delta \geq 0$, $\boldsymbol{\Phi}_{\rho(j)} \doteq \text{diag}(\phi^2_{1\rho(j)}, \cdots, \phi^2_{K_2\rho(j)})$ is a $K_2 \times K_2$ diagonal matrix ($\doteq$ is used to denote a definition), and $\imath_{K_2}$ a K_2-dimensional column vector of 1's;

$$\text{(i)} \quad \theta_j \perp\!\!\!\perp (\boldsymbol{X}_j, \boldsymbol{W}_j); \quad \text{(ii)} \quad \theta_j \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \tau^2_{\rho(j)}\right), \tag{2.22c}$$

with $\tau^2_{\rho(j)} > 0$; and, finally, the distribution of $(\boldsymbol{X}_j, \boldsymbol{W}_j)$ is left unspecified.

Let us explain how the variation-free property is ensured in this specification:

1. Condition (2.22c.i) automatically implies the variation-free property between the parameters indexing the distribution of $(\boldsymbol{X}_j, \boldsymbol{W}_j)$ and those indexing the distribution of θ_j. This not only ensures that $(\boldsymbol{X}_j, \boldsymbol{W}_j)$ are exogenous with respect to the school effect θ_j, but also explains why their distribution is typically left unspecified.

2. Conditions (2.22b) and (2.22c) correspond to a marginal-conditional decomposition of the joint distribution of $(\boldsymbol{Z}_{ij}, \theta_j)$ conditionally on $(\boldsymbol{X}_j, \boldsymbol{W}_j)$. In this case, it can be verified that the parameters $(\boldsymbol{A}, \boldsymbol{H}, \delta, \boldsymbol{\Phi}_{\rho(j)})$ and $\tau_{\rho(j)}$ are in variation-free; for a proof, see (Engle et al., 1983, Example 3.1).

3. By a similar argument, it can be stated that the parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2_{\rho(j)})$ and $(\boldsymbol{A}, \boldsymbol{H}, \delta,$ $\boldsymbol{\Phi}_{\rho(j)})$ are in variation-free and, therefore, $(\boldsymbol{Z}_{ij}, \boldsymbol{X}_{ij}, \boldsymbol{W}_{ij}, \theta_j)$ are exogenous explanatory factors of Y_{ij}.

Remark 2 Specification (2.22a), (2.22b) and (2.22c) is not the only one which ensures a recursive decomposition of the joint distribution generating $(\boldsymbol{Y}_j, \boldsymbol{Z}_{ij}, \boldsymbol{X}_{ij}, \boldsymbol{W}_{ij}, \theta_j)$. Two other possible specifications are the following:

1. Condition (2.22c) can be replaced by assuming that the conditional distribution of θ_j given $(\boldsymbol{X}_j, \boldsymbol{W}_j)$ is normally distributed and that $(\boldsymbol{X}_j, \boldsymbol{W}_j)$ is also normally distributed.

2. Specification (2.22a), (2.22b) and (2.22c) can be relaxed to a specification assuming that conditional expectations are linear functions of the conditioning variables. In this

case, the marginal-conditional decomposition of $(\boldsymbol{Y}_j, \boldsymbol{Z}_{ij}, \boldsymbol{X}_{ij}, \boldsymbol{W}_{ij}, \theta_j)$ reduces to compute both conditional expectations and conditional variances of the sub-components at the right hand side of decomposition (2.20). The properties of linear conditional expectations ensure a free-variation parametrization; for an exposition of those properties, see (Florens et al., 2007, Chapter 7).

Value Added Under HHLIM Specification

Manzi et al. (2014) define the value-added of a school as the difference between two conditional expectations: the first one corresponds to the conditional expectation of the contemporaneous score Y_{ij} given the exogenous explanatory factors and the school effect. The second one corresponds to the conditional expectation of the contemporaneous score Y_{ij} given the exogenous explanatory factors only; it is obtained after integrating out the first conditional expectation with respect to the school effect. This explains why the second conditional expectation corresponds to the expected score in the "average" school. Using the specification (2.22a), (2.22b) and (2.22c), the value added of school j is accordingly given by

$$\begin{aligned}
\text{VA}_j &\doteq \frac{1}{n_j}\sum_{i=1}^{n_j} E(Y_{ij} \mid \boldsymbol{Z}_{ij}, \boldsymbol{X}_{ij}, \boldsymbol{W}_{ij}, \theta_j) - \frac{1}{n_j}\sum_{i=1}^{n_j} E(Y_{ij} \mid \boldsymbol{Z}_{ij}, \boldsymbol{X}_{ij}, \boldsymbol{W}_{ij}) \\
&= \theta_j - \frac{1}{n_j}\sum_{i=1}^{n_j} E(\theta_j \mid \boldsymbol{Z}_{ij}, \boldsymbol{X}_{ij}, \boldsymbol{W}_{ij})
\end{aligned}$$

because

$$\begin{aligned}
E(Y_{ij} \mid \boldsymbol{Z}_{ij}, \boldsymbol{X}_{ij}, \boldsymbol{W}_{ij}) &= E\left\{E(Y_{ij} \mid \boldsymbol{Z}_{ij}, \boldsymbol{X}_{ij}, \boldsymbol{W}_{ij}, \theta_j) \mid \boldsymbol{Z}_{ij}, \boldsymbol{X}_{ij}, \boldsymbol{W}_{ij}\right\} \\
&= \boldsymbol{X}_{ij}^\top \boldsymbol{\beta} + \boldsymbol{Z}_{ij}^\top \boldsymbol{\gamma} + E(\theta_j \mid \boldsymbol{Z}_{ij}, \boldsymbol{X}_{ij}, \boldsymbol{W}_{ij}).
\end{aligned}$$

It can be seen that the value-added of a school is not longer equivalent to the school effect, but corresponds to the school effect corrected by an additive term. This term is the "part" of the school effect that is explained by the covariates. Consequently, the school value-added is precisely the "part" of the school effect that is not explained by the covariates: not only by those covariates that are under the control of the school (as the factor $\boldsymbol{Z}_{ij}$ in (2.22b)), but also by those covariates that are far from the control of the school (as the factors $(\boldsymbol{X}_{ij}, \boldsymbol{W}_{ij})$ in (2.22c)). This additive term is given by

$$E(\theta_j \mid \boldsymbol{Z}_{ij}, \boldsymbol{X}_{ij}, \boldsymbol{W}_{ij}) = \delta\tau^2_{\rho(j)}\imath^\top_{K_2}\left[\delta^2\tau^2_{\rho(j)}\imath_{K_2}\imath^\top_{K_2} + \boldsymbol{\Phi}_{\rho(j)}\right]^{-1}\left(\boldsymbol{Z}_{ij} - \boldsymbol{A}^\top\boldsymbol{X}_{ij} - \boldsymbol{H}^\top\boldsymbol{W}_{ij}\right).$$

Using Randolph (1988)'s approach, it is possible to compute explicitly the inverse of $\delta^2\tau^2_{\rho(j)}\imath_{K_2}\imath^\top_{K_2} + \boldsymbol{\Phi}_{\rho(j)}$. By doing so, the value-added of school j reduces to the following expression:

$$\mathrm{VA}_j = \theta_j - \frac{\delta\tau^2_{\rho(j)}}{1 + \delta^2\tau^2_{\rho(j)}\sum_{k=1}^{K_2}\phi^{-2}_{k\rho(j)}}\frac{1}{n_j}\sum_{i=1}^{n_j}\imath^\top_{K_2}\boldsymbol{\Phi}^{-1}_{\rho(j)}\left(\boldsymbol{Z}_{ij} - \boldsymbol{A}^\top\boldsymbol{X}_{ij} - \boldsymbol{H}^\top\boldsymbol{W}_{ij}\right). \tag{2.23}$$

The parameter δ corresponds to the marginal effect of the school effect on $\boldsymbol{Z}_{ij}$. This parameter plays a role in the way in which the school effect should be corrected in order to obtain the school value-added. In particular, under (2.22b), $\boldsymbol{Z}_{ij} \perp\!\!\!\perp \theta_j \mid \boldsymbol{X}_{ij}, \boldsymbol{W}_{ij}$ if and only if $\delta = 0$. This last condition, along with condition (2.22c.i), are equivalent to $\theta_j \perp\!\!\!\perp (\boldsymbol{Z}_{ij}, \boldsymbol{X}_{ij}, \boldsymbol{W}_{ij})$, which in turn implies the following statements:

1. $E(\theta_j \mid \boldsymbol{Z}_{ij}, \boldsymbol{X}_{ij}, \boldsymbol{W}_{ij}) = E(\theta_j) = 0$ and, therefore, $\mathrm{VA}_j = \theta_j$. The same conclusion is reached by putting $\delta = 0$ in (2.23).

2. $Y_{ij} \perp\!\!\!\perp \boldsymbol{W}_{ij} \mid \boldsymbol{Z}_{ij}, \boldsymbol{X}_{ij}$; that is, at the observed level, the exogenous explanatory factor $\boldsymbol{W}_{ij}$ becomes uninformative to explain the contemporaneous score Y_{ij} once we condition on $(\boldsymbol{Z}_{ij}, \boldsymbol{X}_{ij})$ only. Therefore, the standard HHLM model nested into the HHLIM is specified by the following conditions: for each school j,

$$(Y_{ij} \mid \boldsymbol{Z}_{ij}, \boldsymbol{X}_{ij}, \theta_j) \sim \mathcal{N}\left(\boldsymbol{X}^\top_{ij}\boldsymbol{\beta} + \boldsymbol{Z}^\top_{ij}\boldsymbol{\gamma} + \theta_j, \sigma^2_{\rho(j)}\right), \quad i = 1, \ldots, n_j; \tag{2.24a}$$

$$\text{(i)} \quad \theta_j \perp\!\!\!\perp (\boldsymbol{X}_j, \boldsymbol{Z}_j); \quad \text{(ii)} \quad \theta_j \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \tau^2_{\rho(j)}\right). \tag{2.24b}$$

2.2 Extended HHLIM Models

The grouping school function $\rho(\cdot)$ intends to incorporate characteristics common to a group of schools in the computation of the corresponding value-added. The specific way in which those characteristics are incorporated is though variances $\sigma^2_{\rho(j)}$, $\mathbf{\Phi}_{\rho(j)}$ and $\tau^2_{\rho(j)}$, the rationale being that characteristics common to a group of schools are specified through both within-school and between-school heterogeneity. In the same sense, it should be asked whether the marginal effect of the school effect on $\boldsymbol{Z}_{ij}$ depends or not on those characteristics. Thus, for instance, it can be asked if the relationship between the school effect and the lagged score (which is captured by the parameter δ) depends on the grouping school function $\rho(\cdot)$. From a substantive point of view, this seems to be relevant because, in a first approach, the way in which a school affects the lagged scores can be viewed as a characteristic of a group of schools.

In applications, the variable $\boldsymbol{Z}_{ij}$ is not restricted to a unidimensional factor. Thus, for instance, in the case study reported in Section 2.4.2, $\boldsymbol{Z}_{ij}$ is a two-dimensional vector, the first coordinate being the lagged score of pupil i belonging to school j, the second coordinate being the group effect captured by the average at the school level of lagged scores. Similarly to the previous questions, it should be asked whether the dependency between the school effect and each coordinate of $\boldsymbol{Z}_{ij}$ are equal or not.

These considerations lead to extend HHLIM models by considering a multidimensional δ-parameter that depends on the grouping school function $\rho(\cdot)$. In what follows, we discuss the specification of the model, its identifiability and estimability.

2.2.1 Model Specification

The grouping school function classifies schools into S mutually disjoints groups. For instance, schools can be grouped according to their socio-economic status, or to a geographical zone. If the set of groups' labels is denoted by $\mathcal{S}$, the grouping school function is defined from $\{1,\ldots,J\}$ to $\{1,\ldots,S\}$ as $j \in \{1,\ldots,J\} \longmapsto \rho(j) = s \in \{1,\ldots,S\}$. Therefore, the function $\rho(\cdot)$ induces a partition $\{\mathcal{J}_1,\ldots,\mathcal{J}_S\}$ on $\{1,\ldots,J\}$, where $S = \text{card}(\mathcal{S})$ and $J_s = \text{card}(\mathcal{J}_s)$, namely

$$\bigcup_{s=1}^{S} \mathcal{J}_s = \mathcal{J}, \quad \mathcal{J}_s \cap \mathcal{J}_{s'} = \emptyset \quad \text{when } s \neq s'.$$

Accordingly, let $Y_{ij}^{(s)}$ be the contemporaneous test score of student i belonging to school $j \in \mathcal{J}_s$. Similarly, we denote as $(\boldsymbol{X}_{ij}^{(s)}, \boldsymbol{W}_{ij}^{(s)}, \boldsymbol{Z}_{ij}^{(s)}) \in \mathbb{R}^{K_1} \times \mathbb{R}^{L} \times \mathbb{R}^{K_2}$ the explanatory factors associated to pupil i belonging to school $j \in \mathcal{J}_s$. The school effect of school $j \in \mathcal{J}_s$ is denoted as $\theta_j^{(s)}$. For each school $j \in \mathcal{J}_s$ with $s = 1, \ldots, S$.

Extended HHLIM models are specified as follows: It is firstly assumed that, for each school $j \in \mathcal{J}_s$, with $s = 1, \ldots, S$, **ALI1** and **ALI2** are valid. Under these structural hypotheses, for each school $j \in \mathcal{J}_s$, with $s = 1, \ldots, S$,

$$(Y_{ij}^{(s)} \mid \boldsymbol{Z}_{ij}^{(s)}, \boldsymbol{X}_{ij}^{(s)}, \boldsymbol{W}_{ij}^{(s)}, \theta_j^{(s)}) \sim \mathcal{N}\left(\boldsymbol{X}_{ij}^{(s)\top}\boldsymbol{\beta} + \boldsymbol{Z}_{ij}^{(s)\top}\boldsymbol{\gamma} + \theta_j^{(s)}, \sigma_s^2\right), \quad i = 1, \ldots, n_j^{(s)}, \tag{2.25a}$$

for $\boldsymbol{\beta} \in \mathbb{R}^{K_1}$, $\boldsymbol{\gamma} \in \mathbb{R}^{K_2}$ and $\sigma_s^2 > 0$;

$$(\boldsymbol{Z}_{ij}^{(s)} \mid \boldsymbol{X}_{ij}^{(s)}, \boldsymbol{W}_{ij}^{(s)}, \theta_j^{(s)}) \sim \mathcal{N}_{K_2}\left(\boldsymbol{A}^\top \boldsymbol{X}_{ij}^{(s)} + \boldsymbol{H}^\top \boldsymbol{W}_{ij}^{(s)} + \theta_j^{(s)}\boldsymbol{\delta}_s, \boldsymbol{\Phi}_s\right), \quad i = 1, \ldots, n_j^{(s)}, \tag{2.25b}$$

with $\boldsymbol{A} = (\alpha_{kl})$ a $K_1 \times K_2$ matrix of real coefficients, $\boldsymbol{H} = (\eta_{lk})$ a $L \times K_2$ matrix of real coefficients, $\boldsymbol{\delta}_s$ is a K_2-dimensional vector with non-negative components, $\boldsymbol{\Phi}_s \doteq \mathrm{diag}(\phi_{1s}^2, \cdots, \phi_{K_2 s}^2)$ is a $K_2 \times K_2$ diagonal matrix;

$$\text{(i)} \quad \theta_j^{(s)} \perp\!\!\!\perp (\boldsymbol{X}_j^{(s)}, \boldsymbol{W}_j^{(s)}); \quad \text{(ii)} \quad \theta_j^{(s)} \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \tau_s^2\right), \tag{2.25c}$$

with $\tau_s^2 > 0$. Finally, the distribution of $(\boldsymbol{X}_j^{(s)}, \boldsymbol{W}_j^{(s)})$ is left unspecified.

In this specification, the k-th coordinate of the vector $\boldsymbol{\delta}_s$ ($k = 1, \ldots, K_2$) captures the marginal effect of school $j \in \mathcal{J}_s$ on the factor $\boldsymbol{Z}_{ijk}^{(s)}$. This marginal effect is specific to each group of schools, that is, $\boldsymbol{\delta}_s$ is in principle different from $\boldsymbol{\delta}_{s'}$ for $s \neq s'$. It should be remarked that if $\boldsymbol{\delta}_s = \delta\, \imath_{K_2}$ for all s, specification (2.25a), (2.25b) and (2.25c) reduces to the HHLIM given by (2.22a), (2.22b) and (2.22c). Similarly, if $\boldsymbol{\delta}_s = 0\, \imath_{K_2}$, specification (2.25a), (2.25b) and (2.25c) reduces to the HHLM given by (2.24a) and (2.24b).

Under specification (2.25a), (2.25b) and (2.25c), the value-added of school $j \in \mathcal{J}_s$

is given by

$$\mathrm{VA}_j^{(s)} = \theta_j^{(s)} - \tau_s^2 \boldsymbol{\delta}_s \left(\tau_s^2 \boldsymbol{\delta}_s \boldsymbol{\delta}_s^\top + \boldsymbol{\Phi}_s\right)^{-1} \frac{1}{n_j^{(s)}} \sum_{i=1}^{n_j^{(s)}} \left(\boldsymbol{Z}_{ij}^{(s)} - \boldsymbol{A}^\top \boldsymbol{X}_{ij}^{(s)} - \boldsymbol{H}^\top \boldsymbol{W}_{ij}^{(s)}\right). \quad (2.26)$$

2.2.2 Parameter Identification

Following the conceptual discussion developed in Section 2.1.1,the structural model given by (2.25a), (2.25b) and (2.25c) intends to explain the generation of both the contemporaneous and lagged scores conditionally on the exogenous explanatory factors. More precisely, the data generating process is characterized by the conditional distribution of $\left(\boldsymbol{Y}_j^{(s)\top}, (\mathrm{vec}(\boldsymbol{Z}_j^{(s)\top}))^\top\right)^\top$ given $(\boldsymbol{X}_j^{(s)}, \boldsymbol{W}_j^{(s)})$ for $j \in \mathcal{J}_s$, which corresponds to a $n_j^{(s)}(K_2+1)$-multivariate normal distribution of conditional mean equal to

$$\begin{pmatrix} \boldsymbol{X}_j^{(s)}\boldsymbol{\beta} + (\boldsymbol{X}_j^{(s)}\boldsymbol{A} + \boldsymbol{W}_j^{(s)}\boldsymbol{H})\gamma \\ \\ \mathrm{vec}\left(\boldsymbol{A}^\top \boldsymbol{X}_j^{(s)\top}\right) + \mathrm{vec}\left(\boldsymbol{H}^\top \boldsymbol{W}_j^{(s)\top}\right) \end{pmatrix}, \quad (2.27)$$

and conditional variance-covariance matrix

$$\begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{Y}_j^{(s)}\boldsymbol{Y}_j^{(s)}} & \boldsymbol{\Sigma}_{\boldsymbol{Y}_j^{(s)}\mathrm{vec}(\boldsymbol{Z}_j^{(s)\top})} \\ \\ \boldsymbol{\Sigma}_{\mathrm{vec}(\boldsymbol{Z}_j^{(s)\top})\boldsymbol{Y}_j^{(s)}} & \boldsymbol{\Sigma}_{\mathrm{vec}(\boldsymbol{Z}_j^{(s)\top})\,\mathrm{vec}(\boldsymbol{Z}_j^{(s)\top})} \end{pmatrix} \quad (2.28)$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{Y}_j^{(s)}\boldsymbol{Y}_j^{(s)}} = (\boldsymbol{\gamma}^\top \boldsymbol{\Phi}_s \boldsymbol{\gamma} + \sigma_s^2)\boldsymbol{I}_{n_j^{(s)}} + \tau_s^2(\boldsymbol{\delta}_s^\top \boldsymbol{\gamma} + 1)^2 \boldsymbol{J}_{n_j^{(s)}}; \quad (2.29a)$$

$$\boldsymbol{\Sigma}_{\mathrm{vec}(\boldsymbol{Z}_j^{(s)\top})\,\mathrm{vec}(\boldsymbol{Z}_j^{(s)\top})} = \boldsymbol{I}_{n_j^{(s)}} \otimes \boldsymbol{\Phi}_s + \boldsymbol{J}_{n_j^{(s)}} \otimes \tau_s^2 \boldsymbol{\delta}_s \boldsymbol{\delta}_s^\top; \quad (2.29b)$$

$$\boldsymbol{\Sigma}_{\mathrm{vec}(\boldsymbol{Z}_j^{(s)\top})\boldsymbol{Y}_j^{(s)}} = \boldsymbol{I}_{n_j^{(s)}} \otimes \boldsymbol{\Phi}_s \boldsymbol{\gamma} + \boldsymbol{J}_{n_j^{(s)}} \otimes \tau_s^2(\boldsymbol{\delta}_s^\top \boldsymbol{\gamma} + 1)\,\boldsymbol{\delta}_s; \quad (2.29c)$$

and $\boldsymbol{J}_{n_j^{(s)}} = \imath_{n_j^{(s)}} \imath_{n_j^{(s)}}^\top$. For the computation of $E[\mathrm{vec}(\boldsymbol{Z}_j^{(s)\top}) \mid \boldsymbol{X}_j^{(s)}, \boldsymbol{W}_j^{(s)}]$, we use the identity $\text{vec}(BCD) = (D^\top \otimes B)\text{vec}(C)$ with $B = \boldsymbol{A}^\top$, $C = \boldsymbol{X}^{(s)\top}$ and $D = \boldsymbol{I}_{n_j^{(s)}}$.

The parameters to be estimated are the regression coefficients $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{A}, \boldsymbol{H}, \boldsymbol{\delta}_s)$ for $s = 1, \ldots, S$, and the variances $(\sigma_s^2, \tau_2^2, \boldsymbol{\Phi}_s)$ for $s = 1, \ldots, S$. In order to show that these parameters are identified, recall that the mean and the variance-covariance matrix of a multivariate normal distribution are identified. Therefore, it is enough to write $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{A}, \boldsymbol{H}, \sigma_s^2, \tau_2^2, \boldsymbol{\Phi}_s, \boldsymbol{\delta}_s)$ for $s = 1, \ldots, S$ as functions of them. As a matter of fact, for each school $j \in \mathcal{J}_s$, with $s = 1, \ldots, S$.

1. By standard arguments, from $E(\boldsymbol{Y}_j \mid \boldsymbol{X}_j^{(s)}, \boldsymbol{W}_j^{(s)})$, we identify $\boldsymbol{\beta} + \boldsymbol{A}\boldsymbol{\gamma}$ and $\boldsymbol{H}\boldsymbol{\gamma}$ provided that $r[(\boldsymbol{X}_j^{(s)} \quad \boldsymbol{W}_j^{(s)})] = K_1 + L$. Similarly, from $E[\text{vec}(\boldsymbol{Z}_j^{(s)\top}) \mid \boldsymbol{X}_j^{(s)}, \boldsymbol{W}_j^{(s)}]$, we identify $\boldsymbol{A}$ and $\boldsymbol{H}$. Therefore, $\boldsymbol{\gamma}$ is identified provided that $r(\boldsymbol{H}) = K_2$. Taking into account that $\boldsymbol{\gamma}$ and $\boldsymbol{A}$ are identified, $\boldsymbol{\beta}$ is finally identified. Note that the condition $r(\boldsymbol{H}) = K_2$ implies that $K_2 \leq L$, that is, the number of instrumental variable is at least equal to the number of endogenous variables, as it is typically required in econometrics Wooldridge (2008).

2. From $Var(\text{vec}(\boldsymbol{Z}_{ij}^{(s)\top}) \mid \boldsymbol{X}_j^{(s)}, \boldsymbol{W}_j^{(s)})$ and $cov(\text{vec}(\boldsymbol{Z}_{ij}^{(s)\top}), \text{vec}(\boldsymbol{Z}_{kj}^{(s)\top}) \mid \boldsymbol{X}_j^{(s)}, \boldsymbol{W}_j^{(s)})$ for $i \neq k$, $\tau_s^2 \boldsymbol{\delta}_s \boldsymbol{\delta}_s^\top$ and $\boldsymbol{\Phi}_s$ become identified.

3. From $Var(Y_{ij}^{(s)} \mid \boldsymbol{X}_j^{(s)}, \boldsymbol{W}_j^{(s)})$ and $cov(Y_{ij}^{(s)}, Y_{kj}^{(s)} \mid \boldsymbol{X}_j^{(s)}, \boldsymbol{W}_j^{(s)})$ for $i \neq k$, and using the identifiability of $\boldsymbol{\gamma}$ and $\boldsymbol{\Phi}_s$, it follows that σ_s^2 is identified.

4. From $cov(\text{vec}(\boldsymbol{Z}_{ij}^{(s)\top}), \boldsymbol{Y}_{ij}^{(s)} \mid \boldsymbol{X}_j^{(s)}, \boldsymbol{W}_j^{(s)})$, and using the identifiability of $\tau_s^2 \boldsymbol{\delta}_s \boldsymbol{\delta}_s^\top$, $\boldsymbol{\Phi}_s$ and $\boldsymbol{\gamma}$, it follows that $\tau_s^2 \boldsymbol{\delta}_s$ is identified.

5. Using the identifiability of $\tau_s^2 \boldsymbol{\delta}_s$ and $\tau_s^2 \boldsymbol{\delta}_s \boldsymbol{\delta}_s^\top$, and assuming that $\boldsymbol{\delta}_s \neq 0\imath_{K_2}$, it follows that
$$\delta_{sk} = \frac{\tau_s^2 \delta_{sk}^2}{\tau_s^2 \delta_{sk}}, \quad k = 1, \ldots K_2$$
are identified. The identification of τ_s^2 is a direct consequence.

6. Finally, the parameter $\boldsymbol{\delta}_s = 0\,\imath_{K_2}$ is identified in the sense that the multivariate normal distribution of $(\boldsymbol{Y}_j^{(s)\top}, (\text{vec}(\boldsymbol{Z}_j^{(s)\top}))^\top)^\top$ given $(\boldsymbol{X}_j^{(s)}, \boldsymbol{W}_j^{(s)})$ under the constraints $\boldsymbol{\delta}_s = 0\,\imath_{K_2}$ is different from the multivariate normal distribution under the constraints $\boldsymbol{\delta}_s \neq 0\imath_{K_2}$.

This identification analysis shows that the statistical model corresponding to the conditional multivariate normal distribution of mean given by (2.27) and variance-covariance matrix given by (2.28) can be used to estimating $\boldsymbol{\delta}_s$'s such that $\boldsymbol{\delta}_s \neq 0\,\imath_{K_2}$. For the case $\boldsymbol{\delta}_s = 0\,\imath_{K_2}$, the corresponding nested model is given by the specification (2.24a) and (2.24b).

2.2.3 Parameter Estimation in HHLIM Models

(i) **Fixed Effects Estimators, second level of hierarchy ($\boldsymbol{A}$ and $\boldsymbol{H}$)**

From the equation (2.25b), one can write the following equality

$$\boldsymbol{Z}_{ij}^{(s)} = \boldsymbol{A}^\top \boldsymbol{X}_{ij}^{(s)} + \boldsymbol{H}^\top \boldsymbol{W}_{ij}^{(s)} + \theta_j^{(s)} \boldsymbol{\delta}_s, + \boldsymbol{\epsilon}_{ij}^{(s)}$$

where, $\boldsymbol{\epsilon}_{ij}^{(s)} \sim \mathcal{N}_{K_2}\left(\mathbf{0}, \boldsymbol{\Phi}_s\right), \quad i = 1, \ldots, n_j^{(s)}$ and $\theta_j^{(s)} \sim \mathcal{N}\left(0, \tau_s^2\right)$. Now, by school

$$\boldsymbol{Z}_j^{(s)} = \boldsymbol{X}_j^{(s)} \boldsymbol{A} + \boldsymbol{W}_j^{(s)} \boldsymbol{H} + \imath_{n_j^{(s)}} \otimes \theta_j^{(s)} \boldsymbol{\delta}_s^\top + \boldsymbol{\epsilon}_j^{(s)}$$

with $\boldsymbol{\epsilon}_j^{(s)} = (\boldsymbol{\epsilon}_{1j}^{(s)}, \ldots, \boldsymbol{\epsilon}_{n_j^{(s)} j}^{(s)})^\top$. Finally, we can write,

$$\boldsymbol{Z} = \boldsymbol{\mathcal{W}}\, \boldsymbol{F} + \boldsymbol{\eta}, \quad \text{where} \quad \boldsymbol{\eta} = \begin{pmatrix} \boldsymbol{\omega}(\boldsymbol{\theta}^{(1)}) \\ \vdots \\ \boldsymbol{\omega}(\boldsymbol{\theta}^{(S)}) \end{pmatrix} + \boldsymbol{\epsilon}, \tag{2.30}$$

such that $\boldsymbol{\omega}(\boldsymbol{\theta}^{(s)}) = \begin{pmatrix} \theta_1^{(s)} \boldsymbol{I}_{n_1^{(s)}} & & \\ & \ddots & \\ & & \theta_{J_s}^{(s)} \boldsymbol{I}_{n_{J_s}^{(s)}} \end{pmatrix} \left(\imath_{N_s} \otimes \boldsymbol{\delta}_s^\top\right)$, $\boldsymbol{F} = (\boldsymbol{A}^\top, \boldsymbol{H}^\top)^\top$, $\boldsymbol{\epsilon} \perp\!\!\!\perp \boldsymbol{\theta}$, and $\boldsymbol{\mathcal{W}}$ was defined in the previous section (ii). Furthermore, $\boldsymbol{\eta}_s = \boldsymbol{\omega}(\boldsymbol{\theta}^{(s)}) + \boldsymbol{\epsilon}_s$, whose expected value is zero and variance $\boldsymbol{R}_s$, where

$$\begin{aligned} \boldsymbol{R}_s &= \boldsymbol{I}_{N_s} \otimes \boldsymbol{\Phi}_s + \boldsymbol{P}_s^\top \boldsymbol{D}_s^2 \boldsymbol{P}_s \otimes \left(\tau_s^2 \boldsymbol{\delta}_s \boldsymbol{\delta}_s^\top\right) \\ &= \boldsymbol{Q}_s \otimes \boldsymbol{\Phi}_s + \boldsymbol{P}_s^\top \boldsymbol{D}_s \boldsymbol{P}_s \otimes \boldsymbol{\Phi}_s + \boldsymbol{P}_s^\top \boldsymbol{D}_s^2 \boldsymbol{P}_s \otimes \left(\tau_s^2 \boldsymbol{\delta}_s \boldsymbol{\delta}_s^\top\right) \end{aligned} \tag{2.31}$$

Then

$$\text{vec}\left(\boldsymbol{Z}^\top\right) = \text{vec}\left(\boldsymbol{F}^\top \boldsymbol{\mathcal{W}}^\top\right) + \text{vec}\left(\boldsymbol{\eta}^\top\right), \tag{2.32}$$

where vec $\left(\boldsymbol{\eta}^\top\right)$ is normally distributed with mean equal to $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{R} = \text{diag}\left(\boldsymbol{R}_1, \cdots, \boldsymbol{R}_s\right)$.

Using the identity $\text{vec}(AB) = (B^\top \otimes I)\,\text{vec}(A)$, (2.32) can equivalently be rewritten as

$$\text{vec}\left(\boldsymbol{Z}^\top\right) = \left(\boldsymbol{\mathcal{W}} \otimes \boldsymbol{I}_{K_2}\right) \text{vec}\left(\boldsymbol{F}^\top\right) + \text{vec}\left(\boldsymbol{\eta}^\top\right). \tag{2.33}$$

Under the restriction that $r(\boldsymbol{\mathcal{W}}) = K_1 + L$, $\text{vec}\left(\boldsymbol{F}^\top\right)$ can, therefore, be estimated as

$$\text{vec}\left(\widehat{\boldsymbol{F}}^\top\right) = \left[\left(\boldsymbol{\mathcal{W}}^\top \otimes \boldsymbol{I}_{K_2}\right) \boldsymbol{R}^{-1} \left(\boldsymbol{\mathcal{W}} \otimes \boldsymbol{I}_{K_2}\right)\right]^{-1} \left(\boldsymbol{\mathcal{W}}^\top \otimes \boldsymbol{I}_{K_2}\right) \boldsymbol{R}^{-1} \text{vec}\left(\boldsymbol{Z}^\top\right).$$

Asymptotic Properties of $\widehat{\boldsymbol{F}}^\top$

In the estimation process of F through generalized least squares (GLS), it was assumed that,

Assumption 3 *Since equation (2.32), we can say the orthogonality condition for consistent estimate of vec $\left(\boldsymbol{F}^\top\right)$ by GLS.*

$$E\left(\left(\boldsymbol{\mathcal{W}}^\top \otimes \boldsymbol{I}_{K_2}\right) \otimes vec\left(\boldsymbol{\eta}^\top\right)\right) = \mathbf{0}.$$

this means that each element of vec $(\boldsymbol{\eta})$ is uncorrelated with each element of $\boldsymbol{\mathcal{W}}$.

Assumption 4 *$\boldsymbol{R}$ is positive definite, $E\left(\left(\boldsymbol{\mathcal{W}}^\top \otimes \boldsymbol{I}_{K_2}\right)^\top \boldsymbol{R}^{-1} \left(\boldsymbol{\mathcal{W}}^\top \otimes \boldsymbol{I}_{K_2}\right)\right)$ is nonsingular and its rank is $K_1 + L$*

Under Assumption3 and 4 one can write,

$$\begin{aligned}\text{vec}\left(\widehat{\boldsymbol{F}}^\top\right) &= E\left(\left(\boldsymbol{\mathcal{W}}^\top \otimes \boldsymbol{I}_{K_2}\right)^\top \boldsymbol{R}^{-1} \left(\boldsymbol{\mathcal{W}}^\top \otimes \boldsymbol{I}_{K_2}\right)\right)^{-1} \\ &\quad E\left(\left(\boldsymbol{\mathcal{W}}^\top \otimes \boldsymbol{I}_{K_2}\right)^\top \boldsymbol{R}^{-1} \text{vec}\left(\boldsymbol{Z}^\top\right)\right).\end{aligned}$$

Then,

$$
\begin{aligned}
\operatorname{vec}\left(\widehat{\boldsymbol{F}}^{\top}\right) &= \left(\sum_{j=1}^{J}\left(\boldsymbol{\mathcal{W}}_j^{(s)\top}\otimes \boldsymbol{I}_{K_2}\right)^{\top}\boldsymbol{R}_j^{-1}\left(\boldsymbol{\mathcal{W}}_j^{(s)\top}\otimes \boldsymbol{I}_{K_2}\right)\right)^{-1} \\
&\quad \left(\sum_{j=1}^{J}\left(\boldsymbol{\mathcal{W}}_j^{(s)\top}\otimes \boldsymbol{I}_{K_2}\right)^{\top}\boldsymbol{R}_j^{-1}\operatorname{vec}\left(\boldsymbol{Z}_j^{(s)\top}\right)\right) \\
&= \operatorname{vec}\left(\boldsymbol{F}^{\top}\right) + \left(N^{-1}\sum_{j=1}^{J}\left(\boldsymbol{\mathcal{W}}_j^{(s)\top}\otimes \boldsymbol{I}_{K_2}\right)^{\top}\boldsymbol{R}_j^{-1}\left(\boldsymbol{\mathcal{W}}_j^{(s)\top}\otimes \boldsymbol{I}_{K_2}\right)\right)^{-1} \\
&\quad \left(N^{-1}\sum_{j=1}^{J}\left(\boldsymbol{\mathcal{W}}_j^{(s)\top}\otimes \boldsymbol{I}_{K_2}\right)^{\top}\boldsymbol{R}_j^{-1}\operatorname{vec}\left(\boldsymbol{\eta}_j^{(s)\top}\right)\right). \qquad (2.34)
\end{aligned}
$$

Now, if we define $A \equiv E\left(\left(\boldsymbol{\mathcal{W}}^{\top}\otimes \boldsymbol{I}_{K_2}\right)^{\top}\boldsymbol{R}^{-1}\left(\boldsymbol{\mathcal{W}}^{\top}\otimes \boldsymbol{I}_{K_2}\right)\right)$, then by Weak Law of Large Numbers (WLLN)

$$
\left(N^{-1}\sum_{j=1}^{J}\left(\boldsymbol{\mathcal{W}}_j^{(s)\top}\otimes \boldsymbol{I}_{K_2}\right)^{\top}\boldsymbol{R}_j^{-1}\left(\boldsymbol{\mathcal{W}}_j^{(s)\top}\otimes \boldsymbol{I}_{K_2}\right)\right) \xrightarrow{P} A, \quad \text{when} \quad N\to\infty
$$

and by Assumption 4 and Slutsky's theorem,

$$
\left(N^{-1}\sum_{j=1}^{J}\left(\boldsymbol{\mathcal{W}}_j^{(s)\top}\otimes \boldsymbol{I}_{K_2}\right)^{\top}\boldsymbol{R}_j^{-1}\left(\boldsymbol{\mathcal{W}}_j^{(s)\top}\otimes \boldsymbol{I}_{K_2}\right)\right)^{-1} \xrightarrow{P} A^{-1}, \quad \text{when} \quad N\to\infty
$$

similarly, under Assumption 3 it can be shown that when $N\to\infty$

$$
\left(N^{-1}\sum_{j=1}^{J}\left(\boldsymbol{\mathcal{W}}_j^{(s)\top}\otimes \boldsymbol{I}_{K_2}\right)^{\top}\boldsymbol{R}_j^{-1}\operatorname{vec}\left(\boldsymbol{\eta}_j^{(s)\top}\right)\right) \xrightarrow{P} E\left(\left(\boldsymbol{\mathcal{W}}^{\top}\otimes \boldsymbol{I}_{K_2}\right)^{\top}\boldsymbol{R}^{-1}\operatorname{vec}\left(\boldsymbol{\eta}^{\top}\right)\right) = 0
$$

because,

$$\begin{aligned}
\operatorname{vec}\left(E\left(\left(\boldsymbol{\mathcal{W}}^{\top}\otimes\boldsymbol{I}_{K_2}\right)^{\top}\boldsymbol{R}^{-1}\operatorname{vec}\left(\boldsymbol{\eta}^{\top}\right)\right)\right) &= \\
&= E\left(\operatorname{vec}\left(\boldsymbol{\eta}\top\right)^{\top}\otimes\left(\boldsymbol{\mathcal{W}}^{\top}\otimes\boldsymbol{I}_{K_2}\right)^{\top}\right)\operatorname{vec}\left(\boldsymbol{R}^{-1}\right) \\
&= E\left(\left(\operatorname{vec}\left(\boldsymbol{\eta}\top\right)\otimes\left(\boldsymbol{\mathcal{W}}^{\top}\otimes\boldsymbol{I}_{K_2}\right)\right)^{\top}\right)\operatorname{vec}\left(\boldsymbol{R}^{-1}\right) \\
&= 0 \quad \text{under Assumption 3.}
\end{aligned}$$

Therefore

$$\operatorname{vec}\left(\widehat{\boldsymbol{F}}^{\top}\right)\xrightarrow{P}\operatorname{vec}\left(\boldsymbol{F}^{\top}\right),\quad \text{when}\quad N\to\infty$$

Asymptotic Normality

Using equation (C.1) one obtains the following

$$\begin{aligned}
\sqrt{N}\left(\operatorname{vec}\left(\widehat{\boldsymbol{F}}^{\top}\right)-\operatorname{vec}\left(\boldsymbol{F}^{\top}\right)\right) &= \\
&= \left(N^{-1}\sum_{j=1}^{J}\left(\boldsymbol{\mathcal{W}}_j^{(s)\top}\otimes\boldsymbol{I}_{K_2}\right)^{\top}\boldsymbol{R}_j^{-1}\left(\boldsymbol{\mathcal{W}}_j^{(s)\top}\otimes\boldsymbol{I}_{K_2}\right)\right)^{-1} \\
&\quad\left(N^{-\frac{1}{2}}\sum_{j=1}^{J}\left(\boldsymbol{\mathcal{W}}_j^{(s)\top}\otimes\boldsymbol{I}_{K_2}\right)^{\top}\boldsymbol{R}_j^{-1}\operatorname{vec}\left(\boldsymbol{\eta}_j^{(s)^{\top}}\right)\right)
\end{aligned}$$

but, by theorem central limit (TCL) $\left(N^{-\frac{1}{2}}\sum_{j=1}^{J}\left(\boldsymbol{\mathcal{W}}_j^{(s)\top}\otimes\boldsymbol{I}_{K_2}\right)^{\top}\boldsymbol{R}_j^{-1}\operatorname{vec}\left(\boldsymbol{\eta}_j^{(s)^{\top}}\right)\right)\xrightarrow{D}$ $\mathcal{N}\left(\mathbf{0},\boldsymbol{B}\right)$, where

$\boldsymbol{B} = E\left(\left(\boldsymbol{\mathcal{W}}^{\top}\otimes\boldsymbol{I}_{K_2}\right)^{\top}\boldsymbol{R}^{-1}\operatorname{vec}\left(\boldsymbol{\eta}^{\top}\right)\operatorname{vec}\left(\boldsymbol{\eta}^{\top}\right)^{\top}\boldsymbol{R}^{-1}\left(\boldsymbol{\mathcal{W}}^{\top}\otimes\boldsymbol{I}_{K_2}\right)\right)$ and as we saw previously $\left(N^{-1}\sum_{j=1}^{J}\left(\boldsymbol{\mathcal{W}}_j^{(s)\top}\otimes\boldsymbol{I}_{K_2}\right)^{\top}\boldsymbol{R}_j^{-1}\left(\boldsymbol{\mathcal{W}}_j^{(s)\top}\otimes\boldsymbol{I}_{K_2}\right)\right)^{-1}\xrightarrow{P}A^{-1}$,

then $\left(N^{-1}\sum_{j=1}^{J}\left(\boldsymbol{\mathcal{W}}_j^{(s)\top}\otimes\boldsymbol{I}_{K_2}\right)^{\top}\boldsymbol{R}_j^{-1}\left(\boldsymbol{\mathcal{W}}_j^{(s)\top}\otimes\boldsymbol{I}_{K_2}\right)\right)^{-1}-A^{-1}=O_p(1)$

$$\sqrt{N}\left(\mathrm{vec}\left(\widehat{\boldsymbol{F}}^{\top}\right)-\mathrm{vec}\left(\boldsymbol{F}^{\top}\right)\right) =$$
$$= A^{-1}\left(N^{-\frac{1}{2}}\sum_{j=1}^{J}\left(\boldsymbol{\mathcal{W}}_j^{(s)\top}\otimes \boldsymbol{I}_{K_2}\right)^{\top}\boldsymbol{R}_j^{-1}\mathrm{vec}\left(\boldsymbol{\eta}_j^{(s)\top}\right)\right)+O_p(1)$$

Thus,

$$\sqrt{N}\left(\mathrm{vec}\left(\widehat{\boldsymbol{F}}^{\top}\right)-\mathrm{vec}\left(\boldsymbol{F}^{\top}\right)\right) \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, A^{-1}\boldsymbol{B}A^{-1}\right)$$

and the asymptotic variance of vec $\left(\widehat{\boldsymbol{F}}^{\top}\right)$ corresponds to $Avar\left(\mathrm{vec}\left(\widehat{\boldsymbol{F}}^{\top}\right)\right) = A^{-1}\boldsymbol{B}A^{-1}/N$

Feasible GLS estimation of $\boldsymbol{A}$ and $\boldsymbol{H}$

But often one do not know the variance-covariance matrix $\boldsymbol{R}$, then, one replace the unknown matrix with a consistent estimator of $\boldsymbol{R}$.

Assumption 5 $\widehat{\boldsymbol{R}} \xrightarrow{P} \boldsymbol{R}$*, when* $N\to\infty$*, is a consistent estimator.*

To obtain a feasible GLS estimator first thing you should do is an estimate by OLS of vec $\left(\boldsymbol{F}^{\top}\right)$, which we will denote by vec $\left(\widehat{\boldsymbol{F}}^{\top}\right)_{\text{OLS}}$ that is a consistent estimator of vec $\left(\boldsymbol{F}^{\top}\right)$ under Assumptions 3 and $E\left(\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{\mathcal{W}}\right)$ is nonsingular and its rank is K_1+L. Now, by WLLN, when $N\to\infty$

$$\left(N^{-1}\sum_{j=1}^{J}\mathrm{vec}\left(\boldsymbol{\eta}_j^{(s)\top}\right)\mathrm{vec}\left(\boldsymbol{\eta}_j^{(s)\top}\right)^{\top}\right)\xrightarrow{P}\boldsymbol{R}$$

then an estimator for $\boldsymbol{R}$ is $\widehat{\boldsymbol{R}} = \left(N^{-1}\sum_{j=1}^{J}\mathrm{vec}\left(\widehat{\boldsymbol{\eta}_j^{(s)\,\top}}\right)_{\text{OLS}}\mathrm{vec}\left(\widehat{\boldsymbol{\eta}_j^{(s)\,\top}}\right)_{\text{OLS}}^{\top}\right)$, where vec $\left(\widehat{\boldsymbol{\eta}_j^{(s)}}\right)_{\text{OLS}} = \mathrm{vec}\left(\boldsymbol{Z}^{\top}\right)-\left(\boldsymbol{\mathcal{W}}\otimes\boldsymbol{I}_{K_2}\right)\mathrm{vec}\left(\widehat{\boldsymbol{F}}^{\top}\right)_{\text{OLS}}$ are the OLS residuals.

This way, given $\widehat{\boldsymbol{R}}$, the estimation of vec $\left(\boldsymbol{F}^\top\right)$ is

$$\text{vec}\left(\widehat{\boldsymbol{F}}^\top\right) = \left[\left(\boldsymbol{\mathcal{W}}^\top \otimes \boldsymbol{I}_{K_2}\right)\widehat{\boldsymbol{R}}^{-1}\left(\boldsymbol{\mathcal{W}} \otimes \boldsymbol{I}_{K_2}\right)\right]^{-1}\left(\boldsymbol{\mathcal{W}}^\top \otimes \boldsymbol{I}_{K_2}\right)\widehat{\boldsymbol{R}}^{-1}\text{vec}\left(\boldsymbol{Z}^\top\right).$$

we show that GLS is a consistent estimator, now as $\widehat{\boldsymbol{R}}$ converges to $\boldsymbol{R}$, this last estimator is also consistent estimator, see Wooldridge (2002).

(ii) **Fixed Effects Estimators, first level of hierarchy (β and γ)**

Equation (2.25a) can equivalently be rewritten as

$$Y_{ij}^{(s)} = \boldsymbol{X}_{ij}^{(s)\top}\boldsymbol{\beta} + \boldsymbol{Z}_{ij}^{(s)\top}\boldsymbol{\gamma} + \theta_j^{(s)} + \boldsymbol{u}_{ij}^{(s)}$$

where $\boldsymbol{u}_{ij}^{(s)} \sim \mathcal{N}\left(0, \sigma_s^2\right)$, and $\theta_j^{(s)} \sim \mathcal{N}\left(0, \tau_s^2\right)$. Now, by school the data vector corresponding to $\boldsymbol{Y}_j^{(s)} = \boldsymbol{X}_j^{(s)}\boldsymbol{\beta} + \boldsymbol{Z}_j^{(s)}\boldsymbol{\gamma} + \imath_{n_j^{(s)}}\theta_j^{(s)} + \boldsymbol{u}_j^{(s)}$, with $\boldsymbol{e}_j^{(s)} = (\boldsymbol{u}_{1j}^{(s)}, \ldots, \boldsymbol{u}_{n_j^{(s)}j}^{(s)})^\top$, whit $\boldsymbol{u}_j^{(s)} \perp\!\!\!\perp \theta_j^{(s)}$. Thus one can write,

$$\boldsymbol{Y} = \boldsymbol{\mathcal{X}}\boldsymbol{\pi} + \boldsymbol{e}, \tag{2.35}$$

such that, $\boldsymbol{\pi} = (\boldsymbol{\beta}^\top\, \boldsymbol{\gamma}^\top)^\top$ and $\boldsymbol{\mathcal{X}} = (\boldsymbol{X}\, \boldsymbol{Z})$. In this way we can define a new model,

$$\boldsymbol{\mathcal{W}}^\top\mathbf{Y} = \boldsymbol{\mathcal{W}}^\top\boldsymbol{\mathcal{X}}\boldsymbol{\pi} + \boldsymbol{\mathcal{W}}^\top\mathbf{e}$$

Then, to estimate $\boldsymbol{\pi}$ it is necessary to specify some restrictions:

Assumption 6 *The model specification implies the following system of $K_1 + L$ moment restrictions:*

$$E\left[\boldsymbol{\mathcal{W}}^\top\left(\boldsymbol{Y} - \boldsymbol{\mathcal{X}}\boldsymbol{\zeta}\right)\right] = 0.$$

That is orthogonality condition.

Assumption 7 *As Assumption 6 in not enough to identify $\boldsymbol{\pi}$. An condition sufficient for identification is the rank condition. Then,*

$r(\boldsymbol{\mathcal{W}}^\top\boldsymbol{\mathcal{X}}) = K_1 + K_2$, $r(\boldsymbol{\mathcal{W}}^\top\boldsymbol{\mathcal{W}}) = K_1 + L$, therefore, necessarily the rank

condition correspond to $K_2 \leq L$.

These conditions are ensured by the identification analysis; see Section 2.2.2.

Considering the Assumption 6 and 7, one can use the White's (1980) approach to estimate $\boldsymbol{\pi}$,

$$\widehat{\boldsymbol{\pi}} = \left(\boldsymbol{\mathcal{X}}^{\top}\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{\mathcal{X}}\right)^{-1} \boldsymbol{\mathcal{X}}^{\top}\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{Y} \tag{2.36}$$

where $\boldsymbol{\mathcal{S}}^{-1} = \mathrm{V}(\boldsymbol{\mathcal{W}}^{\top}\mathbf{e}) = \boldsymbol{\mathcal{W}}^{\top}\mathrm{V}(\boldsymbol{e}|\mathbf{X},\mathbf{W})\boldsymbol{\mathcal{W}} = \boldsymbol{\mathcal{W}}^{\top}\boldsymbol{V}\boldsymbol{\mathcal{W}}$ and $\boldsymbol{\mathcal{X}}^{\top}\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{\mathcal{X}}$ is nonsingular, and identifiability condition is that the range of $\boldsymbol{\mathcal{X}}^{\top}\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{\mathcal{X}}$ is complete. It should be remarked that $\boldsymbol{\mathcal{S}}$ depends on $\tau_s^2, \sigma_s^2, \boldsymbol{\delta}_s, \boldsymbol{\Phi}_s$ for $s = 1, \ldots, S$, and $\boldsymbol{\gamma}$, since

$$\mathrm{Var}\left(\boldsymbol{e}^{(s)}\right) = \boldsymbol{Q}_s(\gamma^{\top}\boldsymbol{\Phi}_s\gamma+\sigma_s^2)+\boldsymbol{P}_s\boldsymbol{D}_s\boldsymbol{P}_s^{\top}(\gamma^{\top}\boldsymbol{\Phi}_s\gamma+\sigma_s^2)+\boldsymbol{P}_s\boldsymbol{D}_s^2\boldsymbol{P}_s^{\top}\tau_s^2(\boldsymbol{\delta}_s^{\top}\gamma+1)^2 \tag{2.37}$$

Asympotic Properties of $\widehat{\boldsymbol{\pi}}$

As we saw in the item i, The estimation of $\boldsymbol{\pi}$, is obtained by a estimation of $\boldsymbol{\mathcal{S}}$, such that

$$\widehat{\boldsymbol{\pi}} = \left(\boldsymbol{\mathcal{X}}^{\top}\boldsymbol{\mathcal{W}}\widehat{\boldsymbol{\mathcal{S}}}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{\mathcal{X}}\right)^{-1} \boldsymbol{\mathcal{X}}^{\top}\boldsymbol{\mathcal{W}}\widehat{\boldsymbol{\mathcal{S}}}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{Y}$$

Assumption 8 $\hat{\boldsymbol{\mathcal{S}}} \xrightarrow{P} \boldsymbol{\mathcal{S}}$*, when* $N\to\infty$*, and* $\boldsymbol{\mathcal{S}}$ *is a nonrandom, symmetric,* $(L + K_1) \times (L + K_1)$ *positive definite matrix.*

Then, by Assumption 6, 7 and 8, $\widehat{\boldsymbol{\pi}} \xrightarrow{P} \boldsymbol{\pi}$. Since Assumption 6 $E\left[\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{e}\right] = 0$, then $N^{-1}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{e} \to 0$, when $N \to \infty$, by Assumption 7, $\boldsymbol{C} \equiv E\left(\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{\mathcal{X}}\right)$ has rank $K_1 + K_2$, and adding Assumption 8 $\boldsymbol{C}^{\top}\boldsymbol{\mathcal{S}}\boldsymbol{C}$ also has rank $K_1 + K_2$. This way as

$$\begin{aligned}
\widehat{\boldsymbol{\pi}} &= \left(\boldsymbol{\mathcal{X}}^{\top}\boldsymbol{\mathcal{W}}\widehat{\boldsymbol{\mathcal{S}}}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{\mathcal{X}}\right)^{-1} \boldsymbol{\mathcal{X}}^{\top}\boldsymbol{\mathcal{W}}\widehat{\boldsymbol{\mathcal{S}}}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{Y} \\
&= \boldsymbol{\pi} + \left(\boldsymbol{\mathcal{X}}^{\top}\boldsymbol{\mathcal{W}}\widehat{\boldsymbol{\mathcal{S}}}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{\mathcal{X}}\right)^{-1} \boldsymbol{\mathcal{X}}^{\top}\boldsymbol{\mathcal{W}}\widehat{\boldsymbol{\mathcal{S}}}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{e} \\
\widehat{\boldsymbol{\pi}} &\xrightarrow{P} \boldsymbol{\pi} + \left(\boldsymbol{C}^{\top}\boldsymbol{\mathcal{S}}\boldsymbol{C}\right)^{-1} \boldsymbol{C}^{\top}\boldsymbol{\mathcal{S}}\mathrm{lim}_{N\to\infty}N^{-1}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{e} \\
&= \boldsymbol{\pi}
\end{aligned}$$

Therefore,

$$\widehat{\boldsymbol{\pi}} \xrightarrow{P} \boldsymbol{\pi}, \text{ when } N \to \infty$$

Asymptotic Normality

Analogous to that developed of the previous section, one find that

$$\sqrt{N}\left(\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}\right) = \left(\left(N^{-1}\boldsymbol{\mathcal{X}}^{\top}\boldsymbol{\mathcal{W}}\right)\widehat{\boldsymbol{\mathcal{S}}}\left(N^{-1}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{\mathcal{X}}\right)\right)^{-1}\left(N^{-1}\boldsymbol{\mathcal{X}}^{\top}\boldsymbol{\mathcal{W}}\right)\widehat{\boldsymbol{\mathcal{S}}}\left(N^{-1/2}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{e}\right)$$

where $\left(N^{-1/2}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{e}\right) \xrightarrow{d} N(0, \boldsymbol{\mathcal{W}}^{\top}\boldsymbol{V}\boldsymbol{\mathcal{W}}) \equiv N(0, \boldsymbol{\mathcal{S}}^{-1})$. Then,

$$\text{Avar}\left(\boldsymbol{\pi}\right) = \left(\boldsymbol{CSC}\right)^{-1}$$

Thus,

$$\sqrt{N}\left(\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}\right) \xrightarrow{d} N(0, (\boldsymbol{CSC})^{-1})$$

(iii) **Estimators of variance components, Second level of Hierarchy**

Estimation of Φ_s

We apply the within-operator $\boldsymbol{Q}$ to equation (2.32) and we obtain

$$\begin{aligned}
&(\boldsymbol{Q} \otimes \boldsymbol{I}_{K_2}) \text{ vec}\left(\boldsymbol{Z}^{\top}\right) = \\
&\qquad = (\boldsymbol{Q} \otimes \boldsymbol{I}_{K_2})\ (\boldsymbol{\mathcal{W}} \otimes \boldsymbol{I}_{K_2}) \text{ vec}\left(\boldsymbol{F}^{\top}\right) + (\boldsymbol{Q} \otimes \boldsymbol{I}_{K_2}) \text{ vec}\left(\boldsymbol{\eta}^{\top}\right) \\
&\qquad = (\boldsymbol{Q}\boldsymbol{\mathcal{W}} \otimes \boldsymbol{I}_{K_2}) \text{ vec}\left(\boldsymbol{F}^{\top}\right) + (\boldsymbol{Q} \otimes \boldsymbol{I}_{K_2}) \text{ vec}\left(\boldsymbol{\eta}^{\top}\right).
\end{aligned}$$

note that $(\boldsymbol{Q} \otimes \boldsymbol{I}_{K_2}) \text{ vec}\left(\boldsymbol{Z}^{\top}\right) = \text{vec}\left(\boldsymbol{Z}^{\top}\boldsymbol{Q}^{\top}\right)$. From this equation we obtain an OLS estimator of $\boldsymbol{F}$, that is,

$$\begin{aligned}
&\text{vec}\left(\widehat{\boldsymbol{F}}^{\top}\right)_{\text{OLS}} = \\
&\quad = \left[(\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{Q} \otimes \boldsymbol{I}_{K_2})(\boldsymbol{Q}\boldsymbol{\mathcal{W}} \otimes \boldsymbol{I}_{K_2})\right]^{-1}(\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{Q} \otimes \boldsymbol{I}_{K_2})(\boldsymbol{Q} \otimes \boldsymbol{I}_{K_2})\text{vec}\left(\boldsymbol{Z}^{\top}\right) \\
&\quad = \left[(\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{Q}\boldsymbol{\mathcal{W}})^{-1}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{Q} \otimes \boldsymbol{I}_{K_2}\right] \text{vec}\left(\boldsymbol{Z}^{\top}\right).
\end{aligned}$$

Note that

$$
\begin{aligned}
\operatorname{vec}\left(\widehat{\boldsymbol{F}}^{\top}\right)_{\text{OLS}} &= \left[(\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{Q}\boldsymbol{\mathcal{W}})^{-1}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{Q}\otimes \boldsymbol{I}_{K_2}\right]\left[(\boldsymbol{\mathcal{W}}\otimes \boldsymbol{I}_{K_2})\operatorname{vec}(\boldsymbol{F}^{\top}) + \operatorname{vec}(\boldsymbol{\eta}^{\top})\right] \\
&= \operatorname{vec}(\boldsymbol{F}^{\top}) + \left[(\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{Q}\boldsymbol{\mathcal{W}})^{-1}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{Q}\otimes \boldsymbol{I}_{K_2}\right]\operatorname{vec}(\boldsymbol{\eta}^{\top})
\end{aligned}
$$

We define the following residual:

$$
\begin{aligned}
\operatorname{vec}(\widehat{\boldsymbol{\eta}}^{\top})^{\text{w}} &= \operatorname{vec}\left(\boldsymbol{Z}^{\top}\right) - (\boldsymbol{\mathcal{W}}\otimes \boldsymbol{I}_{K_2})\operatorname{vec}\left(\widehat{\boldsymbol{F}}^{\top}\right)_{\text{OLS}} \\
&= (\boldsymbol{\mathcal{W}}\otimes \boldsymbol{I}_{K_2})\operatorname{vec}(\boldsymbol{F}^{\top}) - (\boldsymbol{\mathcal{W}}\otimes \boldsymbol{I}_{K_2})\operatorname{vec}\left(\widehat{\boldsymbol{F}}^{\top}\right)_{\text{OLS}} + \operatorname{vec}(\boldsymbol{\eta}^{\top}) \\
&= (\boldsymbol{\mathcal{W}}\otimes \boldsymbol{I}_{K_2})\left[\operatorname{vec}(\boldsymbol{F}^{\top}) - \operatorname{vec}\left(\widehat{\boldsymbol{F}}^{\top}\right)_{\text{OLS}}\right] + \operatorname{vec}(\boldsymbol{\eta}^{\top}) \\
&= (\boldsymbol{\mathcal{W}}\otimes \boldsymbol{I}_{K_2})\left[-(\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{Q}\boldsymbol{\mathcal{W}})^{-1}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{Q}\otimes \boldsymbol{I}_{K_2}\operatorname{vec}(\boldsymbol{\eta}^{\top})\right] + \operatorname{vec}(\boldsymbol{\eta}^{\top}) \\
&= \left[\boldsymbol{I}_{NK_2} - \boldsymbol{\mathcal{W}}(\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{Q}\boldsymbol{\mathcal{W}})^{-1}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{Q}\otimes \boldsymbol{I}_{K_2}\right]\operatorname{vec}\left(\boldsymbol{\eta}^{\top}\right).
\end{aligned}
$$

Therefore

$$
\begin{aligned}
&(\boldsymbol{Q}\otimes \boldsymbol{I}_{K_2})\operatorname{vec}(\widehat{\boldsymbol{\eta}}^{\top})^{\text{w}} = \\
&\qquad = \left[(\boldsymbol{Q}\otimes \boldsymbol{I}_{K_2}) - \boldsymbol{Q}\boldsymbol{\mathcal{W}}(\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{Q}\boldsymbol{\mathcal{W}})^{-1}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{Q}\otimes \boldsymbol{I}_{K_2}\right]\operatorname{vec}\left(\boldsymbol{\eta}^{\top}\right) \\
&\qquad = \left[\boldsymbol{Q} - \boldsymbol{Q}\boldsymbol{\mathcal{W}}(\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{Q}\boldsymbol{\mathcal{W}})^{-1}\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{Q}\right]\otimes \boldsymbol{I}_{K_2}\operatorname{vec}\left(\boldsymbol{\eta}^{\top}\right) \\
&\qquad = \left[\left(\boldsymbol{I}_N - \boldsymbol{Q}\boldsymbol{\mathcal{W}}(\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{Q}\boldsymbol{\mathcal{W}})^{-1}\boldsymbol{\mathcal{W}}^{\top}\right)\otimes \boldsymbol{I}_{K_2}\right](\boldsymbol{Q}\otimes \boldsymbol{I}_{K_2})\operatorname{vec}\left(\boldsymbol{\eta}^{\top}\right) \\
&\qquad \doteq \left[\boldsymbol{M}_N\otimes \boldsymbol{I}_{K_2}\right](\boldsymbol{Q}\otimes \boldsymbol{I}_{K_2})\operatorname{vec}(\boldsymbol{\eta}^{\top}),
\end{aligned}
$$

where $\boldsymbol{M}_N = \left(\boldsymbol{I}_N - \boldsymbol{Q}\boldsymbol{\mathcal{W}}(\boldsymbol{\mathcal{W}}^{\top}\boldsymbol{Q}\boldsymbol{\mathcal{W}})^{-1}\boldsymbol{\mathcal{W}}^{\top}\right)$ is an idempotent matrix.

Using the equation (2.31) is obtain that,

$$
\begin{aligned}
V\left((\boldsymbol{Q}\otimes \boldsymbol{I}_{K_2})\operatorname{vec}(\boldsymbol{\eta}^{\top}),\right) &= (\boldsymbol{Q}\otimes \boldsymbol{I}_{K_2})V\left(\operatorname{vec}(\boldsymbol{\eta}^{\top})\right)(\boldsymbol{Q}\otimes \boldsymbol{I}_{K_2}) \\
&= \operatorname{diag}\left\{(\boldsymbol{Q}_s\otimes \boldsymbol{I}_{K_2})\boldsymbol{R}_s(\boldsymbol{Q}_s\otimes \boldsymbol{I}_{K_2})\right\}_{s=1}^{S} \\
&= \operatorname{diag}\left\{(\boldsymbol{Q}_s\otimes \boldsymbol{I}_{K_2})(\boldsymbol{Q}_s\otimes \boldsymbol{\Phi}_s)(\boldsymbol{Q}_s\otimes \boldsymbol{I}_{K_2})\right\}_{s=1}^{S} \\
&= \operatorname{diag}\left\{(\boldsymbol{Q}_s\otimes \boldsymbol{\Phi}_s)\right\}_{s=1}^{S}
\end{aligned}
$$

In order to use the residual above defined to estimate $\boldsymbol{\Phi}_s$, it is necessary to consider the coordinates of such a residual which are related to the k-th endogenous variable.

Thereby we define $\boldsymbol{e}_k$ as the k-th vector of the canonical basis of $\mathbb{R}^{NK_2}$. Then, the k-th endogenous variable of $\text{vec}(\boldsymbol{Z}^\top)$ is given by the following matrix $\boldsymbol{\Pi}_k$,

$$\boldsymbol{\Pi}_k = \begin{bmatrix} \boldsymbol{e}_k^\top \\ \boldsymbol{e}_{k+K_2}^\top \\ \boldsymbol{e}_{k+2K_2}^\top \\ \vdots \\ \boldsymbol{e}_{k+(N-1)K_2}^\top \end{bmatrix}$$

So, as an example $\boldsymbol{\Pi}_1\text{vec}(\boldsymbol{Z}^\top)$ is the first endogenous variable. On the other hand,

$$\boldsymbol{\Pi}_k^{(s)} = \begin{bmatrix} \boldsymbol{e}_k^{(s)\top} \\ \boldsymbol{e}_{k+K_2}^{(s)\top} \\ \boldsymbol{e}_{k+2K_2}^{(s)\top} \\ \vdots \\ \boldsymbol{e}_{k+(N_s-1)K_2}^{(s)\top} \end{bmatrix}$$

$\boldsymbol{e}_k^{(s)}$ as the k-th vector of the canonical basis of $\mathbb{R}^{N_sK_2}$ and $\boldsymbol{\Pi}_k^{(s)}\text{vec}(\boldsymbol{Z}^{(s)\top})$ corresponds to the endogenous variable k-th of group s.

Let us consider the residuals corresponding to both the k-th and l-th endogenous variable of group s, then as $(\boldsymbol{Q}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}_s^\top)^{\text{w}} = [\boldsymbol{M}_{N_s} \otimes \boldsymbol{I}_{K_2}]\ (\boldsymbol{Q}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\boldsymbol{\eta}_s^\top)$, with $\boldsymbol{M}_{N_s} = \left(\boldsymbol{I}_{N_s} - \boldsymbol{Q}_s\boldsymbol{\mathcal{W}}_s(\boldsymbol{\mathcal{W}}^\top\boldsymbol{Q}\boldsymbol{\mathcal{W}})^{-1}\boldsymbol{\mathcal{W}}_s^\top\right)$. Thus,

$$\Pi_k^{(s)}(\boldsymbol{Q}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}_s^\top)^{\text{w}} = \Pi_k^{(s)}\,[\boldsymbol{M}_{N_s} \otimes \boldsymbol{I}_{K_2}]\ (\boldsymbol{Q}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\boldsymbol{\eta}_s^\top)$$

Now

$$E\left\{ \left[\Pi_k^{(s)}(\boldsymbol{Q}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}_s^\top)^{\text{w}}\right]^\top \left[\Pi_l^{(s)}(\boldsymbol{Q}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}_s^\top)^{\text{w}}\right]\right\} =$$

$$= E\left\{ \left[\Pi_k^{(s)}\left[\boldsymbol{M}_{N_s} \otimes \boldsymbol{I}_{K_2}\right] (\boldsymbol{Q}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\boldsymbol{\eta}_s^\top)\right]^\top \right.$$
$$\left.\left[\Pi_l^{(s)}\left[\boldsymbol{M}_{N_s} \otimes \boldsymbol{I}_{K_2}\right] (\boldsymbol{Q}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\boldsymbol{\eta}_s^\top)\right]\right\}$$

$$= E\left\{\left[(\boldsymbol{Q}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\boldsymbol{\eta}_s^\top)\right]^\top \left[\boldsymbol{M}_{N_s} \otimes \boldsymbol{I}_{K_2}\right]^\top \boldsymbol{\Pi}_k^{(s)\top}\boldsymbol{\Pi}_l^{(s)}\right.$$
$$\left.\left[\boldsymbol{M}_{N_s} \otimes \boldsymbol{I}_{K_2}\right]\left[(\boldsymbol{Q}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\boldsymbol{\eta}_s^\top)\right]\right\}$$

$$= \text{tr}\left[\left[\boldsymbol{M}_{N_s} \otimes \boldsymbol{I}_{K_2}\right]^\top \boldsymbol{\Pi}_k^{(s)\top}\boldsymbol{\Pi}_l^{(s)}\left[\boldsymbol{M}_{N_s} \otimes \boldsymbol{I}_{K_2}\right] (\boldsymbol{Q}_s \otimes \boldsymbol{\Phi}_s)\right]$$

$$= \text{tr}\left[\boldsymbol{\Pi}_k^{(s)\top}\boldsymbol{\Pi}_l^{(s)} \,(\boldsymbol{M}_{N_s}\boldsymbol{Q}_s\boldsymbol{M}_{N_s}^\top) \otimes \boldsymbol{\Phi}\right]$$

Therefore,

$$\text{tr}\left[\boldsymbol{\Pi}_k^{(s)\top}\boldsymbol{\Pi}_l^{(s)} \,(\boldsymbol{M}_{N_s}\boldsymbol{Q}_s\boldsymbol{M}_{N_s}^\top) \otimes \boldsymbol{\Phi}\right] \quad = \quad \phi_{kl}\,\text{tr}\left[\boldsymbol{M}_{N_s}\boldsymbol{Q}_s\boldsymbol{M}_{N_s}^\top\right]$$

Thus, an estimator of ϕ_s^{kl} is given by

$$\widehat{\phi}_s^{kl} = \frac{\left[\Pi_k^{(s)}(\boldsymbol{Q}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}_s^\top)^{\text{w}}\right]^\top \left[\Pi_L^{(s)}(\boldsymbol{Q}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}_s^\top)^{\text{w}}\right]}{\text{tr}\left[\boldsymbol{M}_{N_s}\boldsymbol{Q}_s\boldsymbol{M}_{N_s}^\top\right]}$$

This result can be extended to,

$$\widehat{\boldsymbol{\Phi}}_s^{\,kl} = \frac{\left[(\boldsymbol{Q}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}_s^\top)^{\text{w}}\right]^\top \left[(\boldsymbol{Q}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}_s^\top)^{\text{w}}\right]}{\text{tr}\left[\boldsymbol{M}_{N_s}\boldsymbol{Q}_s\boldsymbol{M}_{N_s}^\top\right]} \tag{2.38}$$

Estimation of $\tau_s^2 \boldsymbol{\delta}_s \boldsymbol{\delta}_s^\top$

We apply the between-operator $\boldsymbol{P}$ to equation (2.32) and we obtain

$$\begin{aligned}
\text{vec}\left(\boldsymbol{Z}^\top \boldsymbol{P}^\top\right) &= (\boldsymbol{P} \otimes \boldsymbol{I}_{K_2})\ \text{vec}\left(\boldsymbol{Z}^\top\right) \\
&= (\boldsymbol{P} \otimes \boldsymbol{I}_{K_2})\ (\boldsymbol{\mathcal{W}} \otimes \boldsymbol{I}_{K_2})\ \text{vec}\left(\boldsymbol{F}^\top\right) + (\boldsymbol{P} \otimes \boldsymbol{I}_{K_2})\ \text{vec}\left(\boldsymbol{\eta}^\top\right) \\
&= (\boldsymbol{P}\boldsymbol{\mathcal{W}} \otimes \boldsymbol{I}_{K_2})\ \text{vec}\left(\boldsymbol{F}^\top\right) + (\boldsymbol{P} \otimes \boldsymbol{I}_{K_2})\ \text{vec}\left(\boldsymbol{\eta}^\top\right).
\end{aligned}$$

From this equation we obtain an OLS estimator of $\boldsymbol{F}$, that is,

$$\text{vec}\left(\widehat{\boldsymbol{F}}^\top\right)_{\text{OLS}} = \left[(\boldsymbol{\mathcal{W}}^\top \boldsymbol{P}^\top \boldsymbol{P}\boldsymbol{\mathcal{W}})^{-1} \boldsymbol{\mathcal{W}}^\top \boldsymbol{P}^\top \boldsymbol{P} \otimes \boldsymbol{I}_{K_2}\right] \text{vec}(\boldsymbol{Z}^\top).$$

Note that

$$\begin{aligned}
\text{vec}\left(\widehat{\boldsymbol{F}}^\top\right)_{\text{OLS}} &= \left[(\boldsymbol{\mathcal{W}}^\top \boldsymbol{P}^\top \boldsymbol{P}\boldsymbol{\mathcal{W}})^{-1} \boldsymbol{\mathcal{W}}^\top \boldsymbol{P}^\top \boldsymbol{P} \otimes \boldsymbol{I}_{K_2}\right] \\
&\quad \left\{(\boldsymbol{\mathcal{W}} \otimes \boldsymbol{I}_{K_2})\text{vec}(\boldsymbol{F}^\top) + \text{vec}(\boldsymbol{\eta}^\top)\right\} \\
&= \text{vec}(\boldsymbol{F}^\top) + \left[(\boldsymbol{\mathcal{W}}^\top \boldsymbol{P}^\top \boldsymbol{P}\boldsymbol{\mathcal{W}})^{-1} \boldsymbol{\mathcal{W}}^\top \boldsymbol{P}^\top \boldsymbol{P} \otimes \boldsymbol{I}_{K_2}\right] \text{vec}(\boldsymbol{\eta}^\top).
\end{aligned}$$

We define the following residual:

$$\begin{aligned}
\text{vec}(\widehat{\boldsymbol{\eta}}^\top)^{\text{b}} &= \text{vec}\left(\boldsymbol{Z}^\top\right) - (\boldsymbol{\mathcal{W}} \otimes \boldsymbol{I}_{K_2})\text{vec}\left(\widehat{\boldsymbol{F}}^\top\right)_{\text{OLS}} \\
&= (\boldsymbol{\mathcal{W}} \otimes \boldsymbol{I}_{K_2})\text{vec}(\boldsymbol{F}^\top) - (\boldsymbol{\mathcal{W}} \otimes \boldsymbol{I}_{K_2})\text{vec}\left(\widehat{\boldsymbol{F}}^\top\right)_{\text{OLS}} + \text{vec}(\boldsymbol{\eta}^\top) \\
&= (\boldsymbol{\mathcal{W}} \otimes \boldsymbol{I}_{K_2})\left[\text{vec}(\boldsymbol{F}^\top) - \text{vec}\left(\widehat{\boldsymbol{F}}^\top\right)_{\text{OLS}}\right] + \text{vec}(\boldsymbol{\eta}^\top) \\
&= (\boldsymbol{\mathcal{W}} \otimes \boldsymbol{I}_{K_2})\left\{-\left[(\boldsymbol{\mathcal{W}}^\top \boldsymbol{P}^\top \boldsymbol{P}\boldsymbol{\mathcal{W}})^{-1} \boldsymbol{\mathcal{W}}^\top \boldsymbol{P}^\top \boldsymbol{P} \otimes \boldsymbol{I}_{K_2}\right] \text{vec}(\boldsymbol{\eta}^\top)\right\} + \\
&\quad + \text{vec}(\boldsymbol{\eta}^\top) \\
&= \left[\boldsymbol{I}_{NK_2} - \boldsymbol{\mathcal{W}}(\boldsymbol{\mathcal{W}}^\top \boldsymbol{P}^\top \boldsymbol{P}\boldsymbol{\mathcal{W}})^{-1} \boldsymbol{\mathcal{W}}^\top \boldsymbol{P}^\top \boldsymbol{P} \otimes \boldsymbol{I}_{K_2}\right] \text{vec}(\boldsymbol{\eta}^\top).
\end{aligned}$$

Therefore,

$$\begin{aligned}
(\boldsymbol{P}\otimes\boldsymbol{I}_{K_2})\mathrm{vec}(\widehat{\boldsymbol{\eta}}^\top)^\flat &= \Big[(\boldsymbol{P}\otimes\boldsymbol{I}_{K_2}) - \boldsymbol{P}\boldsymbol{\mathcal{W}}(\boldsymbol{\mathcal{W}}^\top\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{\mathcal{W}})^{-1}\boldsymbol{\mathcal{W}}^\top\boldsymbol{P}^\top\boldsymbol{P}\otimes\boldsymbol{I}_{K_2}\Big] \\
&\quad \mathrm{vec}(\boldsymbol{\eta}^\top) \\
&= \Big[\boldsymbol{P} - \boldsymbol{P}\boldsymbol{\mathcal{W}}(\boldsymbol{\mathcal{W}}^\top\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{\mathcal{W}})^{-1}\boldsymbol{\mathcal{W}}^\top\boldsymbol{P}^\top\boldsymbol{P}\Big]\otimes\boldsymbol{I}_{K_2}\mathrm{vec}(\boldsymbol{\eta}^\top) \\
&= \Big[(\boldsymbol{I}_J - \boldsymbol{P}\boldsymbol{\mathcal{W}}(\boldsymbol{\mathcal{W}}^\top\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{\mathcal{W}})^{-1}\boldsymbol{\mathcal{W}}^\top\boldsymbol{P}^\top)\otimes\boldsymbol{I}_{K_2}\Big] \\
&\quad (\boldsymbol{P}\otimes\boldsymbol{I}_{K_2})\,\mathrm{vec}(\boldsymbol{\eta}^\top) \\
&\doteq (\boldsymbol{T}_J\otimes\boldsymbol{I}_{K_2})(\boldsymbol{P}\otimes\boldsymbol{I}_{K_2})\,\mathrm{vec}(\boldsymbol{\eta}^\top),
\end{aligned}$$

where $\boldsymbol{T}_J = \boldsymbol{I}_J - \boldsymbol{P}\boldsymbol{\mathcal{W}}(\boldsymbol{\mathcal{W}}^\top\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{\mathcal{W}})^{-1}\boldsymbol{\mathcal{W}}^\top\boldsymbol{P}^\top$ is a symmetric idempotent matrix. Using the equation (2.31) is obtain that,

$$\begin{aligned}
V\left((\boldsymbol{P}\otimes\boldsymbol{I}_{K_2})\mathrm{vec}(\boldsymbol{\eta}^\top),\right) &= (\boldsymbol{P}\otimes\boldsymbol{I}_{K_2})V\left(\mathrm{vec}(\boldsymbol{\eta}^\top)\right)(\boldsymbol{P}\otimes\boldsymbol{I}_{K_2})^\top \\
&= \mathrm{diag}\left\{(\boldsymbol{P}_s\otimes\boldsymbol{I}_{K_2})\boldsymbol{R}_s(\boldsymbol{P}_s^\top\otimes\boldsymbol{I}_{K_2})\right\}_{s=1}^{S} \\
&= \mathrm{diag}\left\{(\boldsymbol{P}_s\otimes\boldsymbol{I}_{K_2})\left(\boldsymbol{P}_s^\top\boldsymbol{D}_s\boldsymbol{P}_s\otimes\boldsymbol{\Phi}_s + \boldsymbol{P}_s^\top\boldsymbol{D}_s^2\boldsymbol{P}_s\otimes\right.\right. \\
&\quad \left.\left.\left(\tau_s^2\boldsymbol{\delta}_s\boldsymbol{\delta}_s^\top\right)\right)\left(\boldsymbol{P}_s^\top\otimes\boldsymbol{I}_{K_2}\right)\right\}_{s=1}^{S} \\
&= \mathrm{diag}\left\{\boldsymbol{D}_s^{-1}\otimes\boldsymbol{\Phi}_s + \boldsymbol{I}_{J_s}\otimes\left(\tau_s^2\boldsymbol{\delta}_s\boldsymbol{\delta}_s^\top\right)\right\}_{s=1}^{S}
\end{aligned}$$

Therefore, analogous to the calculation of $\boldsymbol{\Phi}_s$, it is necessary to consider the coordinates of such a residual which are related to the k-th endogenous variable. We define the matrix $\boldsymbol{\Gamma}_k$ of $J\times JK_2$ as follows:

$$\boldsymbol{\Gamma}_k = \begin{pmatrix} \boldsymbol{e}_k^\top \\ \boldsymbol{e}_{k+J}^\top \\ \vdots \\ \boldsymbol{e}_{k+J(K_2-1)}^\top \end{pmatrix}$$

where $\boldsymbol{e}_k$ is the vector of the canonical base of $\mathbf{R}^{JK_2}$. But as we are concerned the variance between each group, we define

$$\boldsymbol{\Gamma}_k^{(s)} = \begin{pmatrix} \boldsymbol{e}_k^{(s)\ \top} \\ \boldsymbol{e}_{k+K_2}^{(s)\ \top} \\ \vdots \\ \boldsymbol{e}_{k+K_2(J_s-1)}^{(s)\ \top} \end{pmatrix}$$

where $\boldsymbol{e}_k^{(s)}$ is the vector of the canonical base of $\mathbf{R}^{J_s K_2}$.

Now consider the residuals corresponding to both the k-th and l-th endogenous variable of group s, and $(\boldsymbol{P}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}_s^\top)^{\text{w}} = [\boldsymbol{T}_{J_s} \otimes \boldsymbol{I}_{K_2}]\ (\boldsymbol{P}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\boldsymbol{\eta}_s^\top)$, with $\boldsymbol{T}_{J_s} = \boldsymbol{I}_{J_s} - \boldsymbol{P}_s \boldsymbol{\mathcal{W}}_s (\boldsymbol{\mathcal{W}}^\top \boldsymbol{P}^\top \boldsymbol{P} \boldsymbol{\mathcal{W}})^{-1} \boldsymbol{\mathcal{W}}_s^\top \boldsymbol{P}_s^\top$ a symmetric idempotent matrix. Thus,

$$E\left\{ \left[\boldsymbol{\Gamma}_k^{(s)}(\boldsymbol{P}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}^{(s)\ \top})^{\text{b}}\right]^\top \left[\boldsymbol{\Gamma}_l^{(s)}(\boldsymbol{P} \otimes \boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}^{(s)\ \top})^{\text{b}}\right]\right\} =$$

$$= E\left\{ \left[\boldsymbol{\Gamma}_k^{(s)}(\boldsymbol{T}_{J_s} \otimes \boldsymbol{I}_{K_2})(\boldsymbol{P}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\boldsymbol{\eta}^{(s)\ \top})\right]^\top \right.$$
$$\left. \left[\boldsymbol{\Gamma}_l^{(s)}(\boldsymbol{T}_{J_s} \otimes \boldsymbol{\Phi}_s)(\boldsymbol{P}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\boldsymbol{\eta}^{(s)\ \top})\right]\right\}$$

$$= E\left\{\left[(\boldsymbol{P}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\boldsymbol{\eta}^{(s)\ \top})\right]^\top (\boldsymbol{T}_{J_s} \otimes \boldsymbol{I}_{K_2})\boldsymbol{\Gamma}_k^{(s)\ \top}\boldsymbol{\Gamma}_l^{(s)} \right.$$
$$\left. (\boldsymbol{T}_{J_s} \otimes \boldsymbol{I}_{K_2})\left[(\boldsymbol{P}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\boldsymbol{\eta}^{(s)\ \top})\right]\right\}$$

$$= \text{tr}\left[(\boldsymbol{T}_{J_s} \otimes \boldsymbol{I}_{K_2})\boldsymbol{\Gamma}_k^{(s)\ \top}\boldsymbol{\Gamma}_l^{(s)}(\boldsymbol{T}_{J_s} \otimes \boldsymbol{I}_{K_2}) \right.$$
$$\left. \left\{\boldsymbol{D}_s^{-1} \otimes \boldsymbol{\Phi}_s + \boldsymbol{I}_{J_s} \otimes \tau_s^2 \boldsymbol{\delta}_s \boldsymbol{\delta}_s^\top\right\}\right]$$

$$= \text{tr}\left[\boldsymbol{\Gamma}_k^{(s)\ \top}\boldsymbol{\Gamma}_l^{(s)}[(\boldsymbol{T}_{J_s}\boldsymbol{D}_s^{-1}\boldsymbol{T}_{J_s}) \otimes \boldsymbol{\Phi}_s]\right] +$$
$$+\text{tr}\left[\boldsymbol{\Pi}_k^\top \boldsymbol{\Pi}_l (\boldsymbol{T}_{J_s}\boldsymbol{T}_{J_s} \otimes \tau_s^2 \boldsymbol{\delta}_s \boldsymbol{\delta}_s^\top)\right]$$

Note that $\boldsymbol{\Gamma}_k^\top \boldsymbol{\Gamma}_l$ is a $JK_2 \times JK_2$ matrix. Therefore, it is composed of $J \times J$ blocks of size $K_2 \times K_2$: in each of these blocks there are a 1 in position (k, l). using the same arguments as in $\boldsymbol{\Phi}_S$. Thus, an estimator of $\delta_s^k \delta_s^l$ is given by

$$\begin{aligned}
\widehat{\tau_s^2 \delta_s^k \delta_s^l} &= \frac{\left[\boldsymbol{\Gamma}_k^{(s)}(\boldsymbol{P}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}^{(s)\,\top})^\flat\right]^\top \left[\boldsymbol{\Gamma}_l^{(s)}(\boldsymbol{P} \otimes \boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}^{(s)\,\top})^\flat\right]}{\text{tr}(\boldsymbol{T}_{J_s}\boldsymbol{T}_{J_s})} \\
&\quad - \frac{\widehat{\phi_s}^{kl}\text{tr}(\boldsymbol{T}_{J_s}\boldsymbol{D}_s^{-1}\boldsymbol{T}_{J_s})}{\text{tr}(\boldsymbol{T}_{J_s}\boldsymbol{T}_{J_s})}
\end{aligned}$$

This result can be extended to,

$$\begin{aligned}
\widehat{\tau_s^2 \boldsymbol{\delta}_s \boldsymbol{\delta}_s^\top} &= \frac{\left(\boldsymbol{P}_s \otimes \boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}^{(s)\,\top})^\flat\right]^\top \left[(\boldsymbol{P} \otimes \boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}^{(s)\,\top})^\flat\right]}{\text{tr}(\boldsymbol{T}_{J_s}\boldsymbol{T}_{J_s})} \\
&\quad - \frac{\widehat{\boldsymbol{\Phi}}_s\text{tr}(\boldsymbol{T}_{J_s}\boldsymbol{D}_s^{-1}\boldsymbol{T}_{J_s})}{\text{tr}(\boldsymbol{T}_{J_s}\boldsymbol{T}_{J_s})}
\end{aligned} \tag{2.39}$$

(iv) **Estimators of variance components, first level of hierarchy**

Estimation of σ_s^2

We apply the within-operator $\boldsymbol{Q}$ to equation (2.35),

$$\boldsymbol{Q}\boldsymbol{Y} = \boldsymbol{Q}\boldsymbol{\mathcal{X}}\boldsymbol{\pi} + \boldsymbol{Q}\boldsymbol{e}.$$

It should be noticed that $\boldsymbol{Q}$ is an instrumental variable because in the previous equation $\boldsymbol{\theta}$ has any role (that is, there is any endogeneity problem). Then

$$\widehat{\boldsymbol{\pi}}_{\text{OLS}}^{\text{w}} = \left(\boldsymbol{\mathcal{X}}^\top\boldsymbol{Q}\boldsymbol{\mathcal{X}}\right)^{-1}\boldsymbol{\mathcal{X}}^\top\boldsymbol{Q}\boldsymbol{Y},$$

and we define the following residual:

$$\begin{aligned}
\widehat{e}^{\text{w}} &= \boldsymbol{Y} - \boldsymbol{\mathcal{X}}\widehat{\boldsymbol{\pi}}_{\text{OLS}}^{\text{w}} \\
&= \left(\boldsymbol{I}_N - \boldsymbol{\mathcal{X}}\left(\boldsymbol{\mathcal{X}}^\top\boldsymbol{Q}\boldsymbol{\mathcal{X}}\right)^{-1}\boldsymbol{\mathcal{X}}^\top\boldsymbol{Q}\right)\boldsymbol{Y}.
\end{aligned}$$

Therefore

$$\begin{aligned}
\boldsymbol{Q}\widehat{\boldsymbol{e}}^{\text{w}} &= \left(\boldsymbol{I}_N - \boldsymbol{Q}\boldsymbol{\mathcal{X}}\left(\boldsymbol{\mathcal{X}}^\top\boldsymbol{Q}\boldsymbol{\mathcal{X}}\right)^{-1}\boldsymbol{\mathcal{X}}^\top\right)\boldsymbol{Q}\boldsymbol{Y} \\
&\doteq \boldsymbol{M}\boldsymbol{Q}\boldsymbol{Y}, \quad \boldsymbol{M} \text{ is idempotent} \\
&= \boldsymbol{M}(\boldsymbol{Q}\boldsymbol{\mathcal{X}}\boldsymbol{\pi} + \boldsymbol{Q}\boldsymbol{e}) \\
&= \boldsymbol{M}\boldsymbol{Q}\boldsymbol{e}.
\end{aligned}$$

It follows that

$$\begin{aligned}
E\left(\widehat{\boldsymbol{e}}^{\text{w}\top}\boldsymbol{Q}\widehat{\boldsymbol{e}}^{\text{w}}\right) &= E\left(\boldsymbol{e}^\top\boldsymbol{Q}\boldsymbol{M}\boldsymbol{Q}\boldsymbol{e}\right) \\
&= \text{tr}(\boldsymbol{M}\text{Var}(\boldsymbol{Q}\boldsymbol{e})).
\end{aligned}$$

But, $\text{Var}(\boldsymbol{Q}\boldsymbol{e}) = \text{diag}\left(\text{Var}(\boldsymbol{Q}_1\boldsymbol{e}^{(1)}), \ldots, \text{Var}(\boldsymbol{Q}_S\boldsymbol{e}^{(S)})\right)$, where given the equation (2.37) $\text{Var}(\boldsymbol{Q}_s\boldsymbol{e}^{(s)}) = \boldsymbol{Q}_s(\boldsymbol{\gamma}^\top\boldsymbol{\Phi}_s\boldsymbol{\gamma} + \sigma_s^2)$.

Thus, $E\left(\widehat{\boldsymbol{e}}^{(s)\,\text{w}\top}\boldsymbol{Q}_s\widehat{\boldsymbol{e}}^{(s)\,\text{w}}\right) = (\boldsymbol{\gamma}^\top\boldsymbol{\Phi}_s\boldsymbol{\gamma} + \sigma_s^2)\text{tr}(\boldsymbol{M}_s\boldsymbol{Q}_s)$,

such that $\boldsymbol{M}_s\boldsymbol{Q}_s = \left(\boldsymbol{Q}_s - \boldsymbol{Q}_s\boldsymbol{\mathcal{X}}^{(s)}\left(\boldsymbol{\mathcal{X}}^\top\boldsymbol{Q}\boldsymbol{\mathcal{X}}\right)^{-1}\boldsymbol{\mathcal{X}}^{(s)\top}\boldsymbol{Q}_s\right)$. This way, we obtain

$$\begin{aligned}
E\left(\widehat{\boldsymbol{e}}^{(s)\,\text{w}\top}\boldsymbol{Q}_s\widehat{\boldsymbol{e}}^{(s)\,\text{w}}\right) &= (\boldsymbol{\gamma}^\top\boldsymbol{\Phi}_s\boldsymbol{\gamma} + \sigma_s^2)\text{tr}(\boldsymbol{M}_s\boldsymbol{Q}_s) \\
&= (\boldsymbol{\gamma}^\top\boldsymbol{\Phi}_s\boldsymbol{\gamma} + \sigma_s^2)\,(N_s - J_s - \\
&\quad \text{tr}\left(\left(\boldsymbol{\mathcal{X}}^\top\boldsymbol{Q}\boldsymbol{\mathcal{X}}\right)^{-1}\boldsymbol{\mathcal{X}}^{(s)\top}\boldsymbol{Q}_s\boldsymbol{\mathcal{X}}^{(s)}\right))
\end{aligned}$$

where $\text{tr}\left(\left(\boldsymbol{\mathcal{X}}^\top\boldsymbol{Q}\boldsymbol{\mathcal{X}}\right)^{-1}\boldsymbol{\mathcal{X}}^{(s)\top}\boldsymbol{Q}_s\boldsymbol{\mathcal{X}}^{(s)}\right) \leq K_2$. Therefore

$$\widehat{\sigma_s^2} = \frac{\widehat{\boldsymbol{e}}^{(s)\,\text{w}\top}\boldsymbol{Q}_s\widehat{\boldsymbol{e}}^{(s)\,\text{w}}}{N_s - J_s - \text{tr}\left(\left(\boldsymbol{\mathcal{X}}^\top\boldsymbol{Q}\boldsymbol{\mathcal{X}}\right)^{-1}\boldsymbol{\mathcal{X}}^{(s)\top}\boldsymbol{Q}_s\boldsymbol{\mathcal{X}}^{(s)}\right)} - \widehat{\boldsymbol{\gamma}}^\top\widehat{\boldsymbol{\Phi}_s}\widehat{\boldsymbol{\gamma}} \qquad (2.40)$$

Estimation of $\tau_s^2(\boldsymbol{\delta}_s^\top\boldsymbol{\gamma}+1)^2$

We apply the between-operator $\boldsymbol{P}$ to equation (2.35), $\boldsymbol{PY} = \boldsymbol{P}\boldsymbol{\mathcal{X}}\boldsymbol{\pi} + \boldsymbol{Pe}$. This equation still suffers from an endogeneity problem and therefore we apply the instrument $\boldsymbol{P}\boldsymbol{\mathcal{W}}$ and therefore we obtain

$$\boldsymbol{\pi}^{\mathrm{b}}_{\mathrm{OLS}} = \left(\boldsymbol{\mathcal{X}}^\top\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{W}}^\top\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{\mathcal{X}}\right)^{-1}\boldsymbol{\mathcal{X}}^\top\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{W}}^\top\boldsymbol{P}^\top\boldsymbol{PY},$$

and we define the following residual:

$$\begin{aligned}\widehat{\boldsymbol{e}^{\mathrm{b}}} &= \boldsymbol{Y} - \boldsymbol{\mathcal{X}}\boldsymbol{\pi}^{\mathrm{b}}_{\mathrm{OLS}}\\ &= \left(\boldsymbol{I}_N - \boldsymbol{\mathcal{X}}\left(\boldsymbol{\mathcal{X}}^\top\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{W}}^\top\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{\mathcal{X}}\right)^{-1}\boldsymbol{\mathcal{X}}^\top\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{W}}^\top\boldsymbol{P}^\top\boldsymbol{P}\right)\boldsymbol{Y}.\end{aligned}$$

Therefore

$$\begin{aligned}\boldsymbol{P}\widehat{\boldsymbol{e}^{\mathrm{b}}} &= \left(\boldsymbol{I}_J - \boldsymbol{P}\boldsymbol{\mathcal{X}}\left(\boldsymbol{\mathcal{X}}^\top\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{W}}^\top\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{\mathcal{X}}\right)^{-1}\boldsymbol{\mathcal{X}}^\top\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{W}}^\top\boldsymbol{P}^\top\right)\boldsymbol{PY}\\ &= \left(\boldsymbol{I}_J - \boldsymbol{P}\boldsymbol{\mathcal{X}}\left(\boldsymbol{\mathcal{X}}^\top\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{W}}^\top\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{\mathcal{X}}\right)^{-1}\boldsymbol{\mathcal{X}}^\top\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{W}}^\top\boldsymbol{P}^\top\right)\boldsymbol{Pe}\\ &\doteq \boldsymbol{TPe},\end{aligned}$$

where $\boldsymbol{T}$ is idempotent. It follows that

$$\begin{aligned}E\left(\widehat{\boldsymbol{e}^{\mathrm{b}\top}}\boldsymbol{P}^\top\boldsymbol{P}\widehat{\boldsymbol{e}^{\mathrm{b}}}\right) &= E\left(\boldsymbol{e}^\top\boldsymbol{P}^\top\boldsymbol{T}^\top\boldsymbol{TPe}\right)\\ &= \mathrm{tr}[\boldsymbol{T}^\top\boldsymbol{T}\,\mathrm{Var}(\boldsymbol{Pe})].\end{aligned}$$

But, $\mathrm{Var}(\boldsymbol{Pe}) = \mathrm{diag}\left(\mathrm{Var}(\boldsymbol{P}_1\boldsymbol{e}^{(1)}),\ldots,\mathrm{Var}(\boldsymbol{P}_S\boldsymbol{e}^{(S)})\right)$, such that using (2.37), we obtain $\mathrm{Var}(\boldsymbol{P}_s\boldsymbol{e}^{(s)}) = \boldsymbol{D}_s^{-1}(\gamma^\top\boldsymbol{\Phi}_s\gamma+\sigma_s^2) + \boldsymbol{I}_{J_s}\tau_s^2(\boldsymbol{\delta}_s^\top\gamma+1)^2$. Therefore, if we defined, $\boldsymbol{T}_s = \boldsymbol{I}_{J_s} - \boldsymbol{P}_s\boldsymbol{\mathcal{X}}^{(s)}\left(\boldsymbol{\mathcal{X}}^\top\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{W}}^\top\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{\mathcal{X}}\right)^{-1}\boldsymbol{\mathcal{X}}^{(s)\top}\boldsymbol{P}_s^\top\boldsymbol{P}_s\boldsymbol{\mathcal{W}}^{(s)}\boldsymbol{\mathcal{W}}^{(s)\top}\boldsymbol{P}_s^\top$, then $E\left(\widehat{\boldsymbol{e}^{\mathrm{b}\top}}\boldsymbol{P}^\top\boldsymbol{P}\widehat{\boldsymbol{e}^{\mathrm{b}}}\right) = (\boldsymbol{\gamma}^\top\boldsymbol{\Phi}\boldsymbol{\gamma}+\sigma^2)\mathrm{tr}[\boldsymbol{T}_s^\top\boldsymbol{T}_s\boldsymbol{D}_s^{-1}] + \tau^2(\boldsymbol{\delta}^\top\boldsymbol{\gamma}+1)^2\mathrm{tr}[\boldsymbol{T}_s^\top\boldsymbol{T}_s]$,

Thus the following estimator is obtained

$$\tau_s^2\widehat{(\boldsymbol{\delta}_s^\top\boldsymbol{\gamma}+1)^2} = \frac{\widehat{\boldsymbol{e}^{(s)\mathrm{b}}}^\top\boldsymbol{P}_s^\top\boldsymbol{P}_s\widehat{\boldsymbol{e}^{(s)\mathrm{b}}} - (\widehat{\gamma}^\top\widehat{\boldsymbol{\Phi}}_s\widehat{\gamma}+\widehat{\sigma_s^2})\mathrm{tr}[\boldsymbol{T}_s^\top\boldsymbol{T}_s\boldsymbol{D}_s]}{\mathrm{tr}[\boldsymbol{T}_s^\top\boldsymbol{T}_s]} \tag{2.41}$$

(v) **Estimators of Covariance Components Between Hierarchies**

Estimation of $\tau_s^2\boldsymbol{\delta}_s^\top$

In the previous section iv, and iii we define the following

(a) $\boldsymbol{P}\widehat{\boldsymbol{e}^\flat} = \boldsymbol{T}\boldsymbol{P}\boldsymbol{e}$,

(b) $(\boldsymbol{P}\otimes\boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}^\top)^\flat = (\boldsymbol{T}_J\otimes\boldsymbol{I}_{K_2})(\boldsymbol{P}\otimes\boldsymbol{I}_{K_2})\,\text{vec}(\boldsymbol{\eta}^\top)$

This way, if to consider the residuals corresponding to both the k-th endogenous variable

$$
\begin{aligned}
&\text{Cov}\left(\boldsymbol{\Gamma}_k(\boldsymbol{P}\otimes\boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}^\top)^\flat, \boldsymbol{P}\widehat{\boldsymbol{e}^\flat}\right) = \\
&\quad = E\left\{\left[\boldsymbol{\Gamma}_k(\boldsymbol{P}\otimes\boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}^\top)^\flat\right]^\top \boldsymbol{P}\widehat{\boldsymbol{e}^\flat}\right\} \\
&\quad = E\left[\left((\boldsymbol{P}\otimes\boldsymbol{I}_{K_2})\text{vec}(\boldsymbol{\eta}^\top)^\flat\right)^\top(\boldsymbol{T}_J\otimes\boldsymbol{I}_{K_2})\boldsymbol{\Gamma}_k\boldsymbol{T}\boldsymbol{P}\boldsymbol{e}\right] \\
&\quad = \text{tr}\left[(\boldsymbol{T}_J\otimes\boldsymbol{I}_{K_2})\boldsymbol{\Gamma}_k^\top\boldsymbol{T}\text{Cov}\left(\left((\boldsymbol{P}\otimes\boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}^\top)^\flat\right)^\top, \boldsymbol{P}\boldsymbol{e}\right)\right]
\end{aligned}
$$

But,

$\text{Cov}\left(\left((\boldsymbol{P}\otimes\boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}^\top)^\flat\right)^\top, \boldsymbol{P}\boldsymbol{e}\right) = \left\{\boldsymbol{D}_s^{-1}\otimes\boldsymbol{\gamma}^\top\boldsymbol{\Phi}_s + \boldsymbol{I}_{J_s}\otimes\tau_s^2\boldsymbol{\delta}_s^\top(\boldsymbol{\gamma}^\top\boldsymbol{\delta}_s+1)\right\}_{s=1}^S$.
Thus, we can find the estimators for each of the groups $s = 1, \ldots, S$, then

$$
\begin{aligned}
&\text{Cov}\left(\boldsymbol{\Gamma}_k^{(s)}(\boldsymbol{P}_s\otimes\boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}^{(s)\top})^\flat, \boldsymbol{P}_s\,\widehat{\boldsymbol{e}^{(s)\flat}}\right) = \\
&\quad = \text{tr}\left[(\boldsymbol{T}_{J_s}\otimes\boldsymbol{I}_{K_2})\boldsymbol{\Gamma}_k^{(s)\top}\boldsymbol{T}_s\left(\boldsymbol{D}_s^{-1}\otimes\boldsymbol{\gamma}^\top\boldsymbol{\Phi}_s + \boldsymbol{I}_{J_s}\otimes\tau_s^2\boldsymbol{\delta}_s^\top(\boldsymbol{\gamma}^\top\boldsymbol{\delta}_s+1)\right)\right] \\
&\quad = \text{tr}\left[\boldsymbol{T}_s\boldsymbol{D}_s^{-1}\boldsymbol{T}_{J_s}\boldsymbol{\gamma}^\top\boldsymbol{\Phi}_s^k + \boldsymbol{T}_s\boldsymbol{T}_{J_s}\tau_s^2\boldsymbol{\delta}_s^{k\top}(\boldsymbol{\gamma}^\top\boldsymbol{\delta}_s+1)\right]
\end{aligned}
$$

But, how $\tau_s^2\boldsymbol{\delta}_s^{k\top}(\boldsymbol{\gamma}^\top\boldsymbol{\delta}_s+1) = \tau_s^2\boldsymbol{\gamma}^\top\boldsymbol{\delta}_s\boldsymbol{\delta}_s^{k\top} + \tau_s^2\boldsymbol{\delta}_s^{k\top}$, it corresponds to the $k-th$ coordinate, which is associated with the k-th endogenous variable.

This way,

$$
\widehat{\tau_s^2\boldsymbol{\delta}_s^{k\top}} = \frac{\boldsymbol{\Gamma}_k^{(s)}(\boldsymbol{P}_s\otimes\boldsymbol{I}_{K_2})\text{vec}(\widehat{\boldsymbol{\eta}}^{(s)\top})^\flat\boldsymbol{P}_s\,\widehat{\boldsymbol{e}^{(s)\flat}} - \text{tr}\left[\boldsymbol{T}_s\boldsymbol{D}_s^{-1}\boldsymbol{T}_{J_s}\right](\widehat{\boldsymbol{\gamma}}^\top\widehat{\boldsymbol{\Phi}})_s^k}{\text{tr}\left[\boldsymbol{T}_s\boldsymbol{T}_{J_s}\right]} - \widehat{\boldsymbol{\gamma}}^\top\widehat{\tau_s^2\boldsymbol{\delta}_s\boldsymbol{\delta}_s^{k\top}}
$$

This result can be extended to,

$$\widehat{\tau_s^2\boldsymbol{\delta}_s^\top} = \frac{\boldsymbol{\Gamma}_k^{(s)}(\boldsymbol{P}_s\otimes\boldsymbol{I}_{K_2})\mathrm{vec}(\widehat{\boldsymbol{\eta}}^{(s)\top})^{\mathrm{b}}\boldsymbol{P}_s\,\widehat{\boldsymbol{e}^{(s)\,\mathrm{b}}} - \mathrm{tr}\left[\boldsymbol{T}_s\boldsymbol{D}_s^{-1}\boldsymbol{T}_{J_s}\right](\widehat{\boldsymbol{\gamma}}^\top\widehat{\boldsymbol{\Phi}})_s}{\mathrm{tr}\left[\boldsymbol{T}_s\boldsymbol{T}_{J_s}\right]} - \widehat{\boldsymbol{\gamma}}^\top\widehat{\tau_s^2\boldsymbol{\delta}_s\boldsymbol{\delta}_s^\top} \tag{2.42}$$

2.2.4 Summary of the Estimation Process

1. Estimation of $\boldsymbol{F}$, determined at item i. The estimate of this parameter consists of two steps.

   (a) When the process starts is estimated by OLS, ie

   $$\mathrm{vec}\left(\widehat{\boldsymbol{F}}^\top\right) = \left[\left(\boldsymbol{\mathcal{W}}^\top\otimes\boldsymbol{I}_{K_2}\right)(\boldsymbol{\mathcal{W}}\otimes\boldsymbol{I}_{K_2})\right]^{-1}\left(\boldsymbol{\mathcal{W}}^\top\otimes\boldsymbol{I}_{K_2}\right)\mathrm{vec}\left(\boldsymbol{Z}^\top\right).$$

   (b) When you have already made steps 1-8, F is estimated again, but now with Feasible GLS, ie,

   $$\mathrm{vec}\left(\widehat{\boldsymbol{F}}^\top\right) = \left[\left(\boldsymbol{\mathcal{W}}^\top\otimes\boldsymbol{I}_{K_2}\right)\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{\mathcal{W}}\otimes\boldsymbol{I}_{K_2})\right]^{-1}\left(\boldsymbol{\mathcal{W}}^\top\otimes\boldsymbol{I}_{K_2}\right)\widehat{\boldsymbol{R}}^{-1}\mathrm{vec}\left(\boldsymbol{Z}^\top\right).$$

2. Completed step 1, be must estimate the Within Residual of the model (2.25b). Then, it is estimated $\boldsymbol{\Phi}_s$ by the equation (2.38), ie

   $$\widehat{\boldsymbol{\Phi}}_s = \frac{\left[(\boldsymbol{Q}_s\otimes\boldsymbol{I}_{K_2})\mathrm{vec}(\widehat{\boldsymbol{\eta}}_s^\top)^{\mathrm{w}}\right]^\top\left[(\boldsymbol{Q}_s\otimes\boldsymbol{I}_{K_2})\mathrm{vec}(\widehat{\boldsymbol{\eta}}_s^\top)^{\mathrm{w}}\right]}{\mathrm{tr}\left[\boldsymbol{M}_{N_s}\boldsymbol{Q}_s\boldsymbol{M}_{N_s}^\top\right]}$$

3. Completed steps 1–2, be must estimate the Between Residual of the model (2.25b). Then, it is estimated $\widehat{\tau_s^2\delta_s^k\delta_s^l}$ by the equation (2.39), ie

   $$\widehat{\tau_s^2\boldsymbol{\delta}_s\boldsymbol{\delta}_s^\top} = \frac{\left[(\boldsymbol{P}_s\otimes\boldsymbol{I}_{K_2})\mathrm{vec}(\widehat{\boldsymbol{\eta}}^{(s)\,\top})^{\mathrm{b}}\right]^\top\left[(\boldsymbol{P}\otimes\boldsymbol{I}_{K_2})\mathrm{vec}(\widehat{\boldsymbol{\eta}}^{(s)\,\top})^{\mathrm{b}}\right] - \widehat{\boldsymbol{\Phi}}_s\mathrm{tr}(\boldsymbol{T}_{J_s}\boldsymbol{D}_s^{-1}\boldsymbol{T}_{J_s})}{\mathrm{tr}(\boldsymbol{T}_{J_s}\boldsymbol{T}_{J_s})}$$

4. The Estimation of $\boldsymbol{\pi}$, determined at section ii

5. Completed steps 1–4, be must estimate the Within Residual of the model (2.25a). Then, it is estimated σ_s^2 by the equation (2.40), ie

$$\widehat{\sigma_s^2} = \frac{\widehat{\boldsymbol{e}}^{(s)\,\mathrm{w}\,\top}\boldsymbol{Q}_s\widehat{\boldsymbol{e}}^{(s)\,\mathrm{w}}}{N_s - J_s - \mathrm{tr}\left(\left(\boldsymbol{\mathcal{X}}^\top\boldsymbol{Q}\boldsymbol{\mathcal{X}}\right)^{-1}\boldsymbol{\mathcal{X}}^{(s)\top}\boldsymbol{Q}_s\boldsymbol{\mathcal{X}}^{(s)}\right)} - \widehat{\boldsymbol{\gamma}}^\top\widehat{\boldsymbol{\Phi}_s}\widehat{\boldsymbol{\gamma}}$$

6. Completed steps 1–5, be must estimate the Between Residual of the model (2.25a). Then, $\tau_s^2\widehat{(\boldsymbol{\delta}_s^\top\boldsymbol{\gamma}+1)^2}$ is estimated by the equation (2.41), ie

$$\tau_s^2\widehat{(\boldsymbol{\delta}_s^\top\boldsymbol{\gamma}+1)^2} = \frac{\widehat{\boldsymbol{e}^{(s)\,\mathrm{b}}}^\top\boldsymbol{P}_s^\top\boldsymbol{P}_s\widehat{\boldsymbol{e}^{(s)\,\mathrm{b}}} - (\widehat{\boldsymbol{\gamma}}^\top\widehat{\boldsymbol{\Phi}}_s\widehat{\boldsymbol{\gamma}} + \widehat{\sigma_s^2})\mathrm{tr}[\boldsymbol{T}_s^\top\boldsymbol{T}_s\boldsymbol{D}_s]}{\mathrm{tr}[\boldsymbol{T}_s^\top\boldsymbol{T}_s]}$$

7. Completed steps 1–6, are used the Between Residuals estimation of model (2.25a) and model (2.25b). This way $\tau_s^2\boldsymbol{\delta}_s^{k\,\top}$ is estimated by the equation (2.42),

$$\widehat{\tau_s^2\boldsymbol{\delta}_s^\top} = \frac{\boldsymbol{\Gamma}_k^{(s)}(\boldsymbol{P}_s\otimes\boldsymbol{I}_{K_2})\mathrm{vec}(\widehat{\boldsymbol{\eta}}^{(s)\top})^{\mathrm{b}}\boldsymbol{P}_s\,\widehat{\boldsymbol{e}^{(s)\,\mathrm{b}}} - \mathrm{tr}\left[\boldsymbol{T}_s\boldsymbol{D}_s^{-1}\boldsymbol{T}_{J_s}\right](\widehat{\boldsymbol{\gamma}}^\top\widehat{\boldsymbol{\Phi}})_s}{\mathrm{tr}\left[\boldsymbol{T}_s\boldsymbol{T}_{J_s}\right]} - \widehat{\boldsymbol{\gamma}}^\top\widehat{\tau_s^2\boldsymbol{\delta}_s\boldsymbol{\delta}_s^\top}$$

8. Then, using the results of steps 3 and 7, can be calculated $\boldsymbol{\delta}_s^k$ as,

$$\boldsymbol{\delta}_s = \frac{\mathrm{diag}\left\{\widehat{\tau_s^2\boldsymbol{\delta}_s\boldsymbol{\delta}_s^\top}\right\}}{\widehat{\tau_s^2\boldsymbol{\delta}_s^\top}}$$

9. Substituting the estimates obtained in step 8 in equation of step 6, we can get an estimate of τ_s^2.

2.2.5 Prediction of the school effect and estimation of the value added

In order to estimate the value added of a school $j \in \mathcal{J}_s$, it is necessary to predict the school effect $\theta_j^{(s)}$. The specification (2.25a), (2.25b) and (2.25c) imply that the joint distribution of $(\boldsymbol{Y}_j^{(s)\top}, \boldsymbol{Z}_j^{(s)\top}, \theta_j^{(s)})^\top$ conditionally on $\boldsymbol{X}_j^{(s)}, \boldsymbol{W}_j^{(s)})$ is a multivariate normal distribution of mean

$$\begin{pmatrix} \boldsymbol{X}_j^{(s)}\boldsymbol{\beta} + (\boldsymbol{X}_j^{(s)}\boldsymbol{A}^\top + \boldsymbol{W}_j^{(s)}\boldsymbol{H}^\top)\boldsymbol{\gamma} \\ (\boldsymbol{I}_{n_j^{(s)}} \otimes \boldsymbol{A}^\top)\boldsymbol{X}_j^{(s)} + (\boldsymbol{I}_{n_j^{(s)}} \otimes \boldsymbol{H}^\top)\boldsymbol{W}_j^{(s)} \\ 0 \end{pmatrix},$$

and conditional variance-covariance matrix

$$\begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{Y}_j^{(s)}\boldsymbol{Y}_j^{(s)\top}} & \boldsymbol{\Sigma}_{\boldsymbol{Y}_j^{(s)}\boldsymbol{Z}_j^{(s)\top}} & \boldsymbol{\Sigma}_{\boldsymbol{Y}_j^{(s)}\theta_j^{(s)}} \\ \boldsymbol{\Sigma}_{\boldsymbol{Z}_j^{(s)}\boldsymbol{Y}_j^{(s)\top}} & \boldsymbol{\Sigma}_{\boldsymbol{Z}_j^{(s)}\boldsymbol{Z}_j^{(s)\top}} & \boldsymbol{\Sigma}_{\boldsymbol{Z}_j^{(s)}\theta_j^{(s)}} \\ \boldsymbol{\Sigma}_{\theta_j^{(s)}\boldsymbol{Y}_j^{(s)\top}} & \boldsymbol{\Sigma}_{\theta_j^{(s)}\boldsymbol{Z}_j^{(s)\top}} & \boldsymbol{\Sigma}_{\theta_j^{(s)}\theta_j^{(s)}} \end{pmatrix},$$

where $\boldsymbol{\Sigma}_{\boldsymbol{Y}_j^{(s)}\boldsymbol{Y}_j^{(s)\top}}$ is given by (2.29a), $\boldsymbol{\Sigma}_{\boldsymbol{Y}_j^{(s)}\boldsymbol{Z}_j^{(s)\top}}$ is given by (2.29c), $\boldsymbol{\Sigma}_{\boldsymbol{Z}_j^{(s)}\boldsymbol{Z}_j^{(s)\top}}$ is given by (2.29b),

$$\boldsymbol{\Sigma}_{\boldsymbol{Y}_j^{(s)}\theta_j^{(s)}} = \tau_s^2(\boldsymbol{\delta}_s^\top\boldsymbol{\gamma} + 1)\imath_{n_j^{(s)}}; \tag{2.43a}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{Z}_j^{(s)}\theta_j^{(s)}} = \imath_{n_j^{(s)}} \otimes \tau_s^2\boldsymbol{\delta}_s; \tag{2.43b}$$

$$\boldsymbol{\Sigma}_{\theta_j^{(s)}\theta_j^{(s)}} = \tau_s^2. \tag{2.43c}$$

The prediction of $\theta_j^{(s)}$ is, therefore, given by

$$\begin{aligned} \widehat{\theta}_j^{(s)} &= \begin{pmatrix} \widehat{\boldsymbol{\Sigma}}_{\theta_j^{(s)}\boldsymbol{Y}_j^{(s)\top}} & \widehat{\boldsymbol{\Sigma}}_{\theta_j^{(s)}\boldsymbol{Z}_j^{(s)\top}} \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}_j^{(s)}\boldsymbol{Y}_j^{(s)\top}} & \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}_j^{(s)}\boldsymbol{Z}_j^{(s)\top}} \\ \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Z}_j^{(s)}\boldsymbol{Y}_j^{(s)\top}} & \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Z}_j^{(s)}\boldsymbol{Z}_j^{(s)\top}} \end{pmatrix}^{-1} \times \\ & \left[\begin{pmatrix} \boldsymbol{Y}_j^{(s)} \\ \boldsymbol{Z}_j^{(s)} \end{pmatrix} - \begin{pmatrix} \boldsymbol{X}_j^{(s)}\widehat{\boldsymbol{\beta}} + (\boldsymbol{X}_j^{(s)}\widehat{\boldsymbol{A}}^\top + \boldsymbol{W}_j^{(s)}\widehat{\boldsymbol{H}}^\top)\widehat{\boldsymbol{\gamma}} \\ (\boldsymbol{I}_{n_j^{(s)}} \otimes \widehat{\boldsymbol{A}}^\top)\boldsymbol{X}_j^{(s)} + (\boldsymbol{I}_{n_j^{(s)}} \otimes \widehat{\boldsymbol{H}}^\top)\boldsymbol{W}_j^{(s)} \end{pmatrix} \right]. \end{aligned}$$

Once the school effect has been predicted, the school value added is estimated using (2.26).

2.3 Likelihood

Above, one saw that the conditional distribution is given by,

$$\begin{pmatrix} \boldsymbol{Y}_j^{(s)} \\ \mathrm{vec}\left(\boldsymbol{Z}_j^{(s)\top}\right) \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{X}_j^{(s)}\boldsymbol{\beta} + (\boldsymbol{X}_j^{(s)}\boldsymbol{A} + \boldsymbol{W}_j^{(s)}\boldsymbol{H})\boldsymbol{\gamma} \\ \mathrm{vec}\left(\boldsymbol{A}^\top \boldsymbol{X}_j^{(s)\top}\right) + \mathrm{vec}\left(\boldsymbol{H}^\top \boldsymbol{W}_j^{(s)\top}\right) \end{pmatrix}, \right.$$

$$\left. \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{Y}_j^{(s)}\boldsymbol{Y}_j^{(s)}} & \boldsymbol{\Sigma}_{\boldsymbol{Y}_j^{(s)}\mathrm{vec}(\boldsymbol{Z}_j^{(s)\top})} \\ \boldsymbol{\Sigma}_{\mathrm{vec}(\boldsymbol{Z}_j^{(s)\top})\boldsymbol{Y}_j^{(s)}} & \boldsymbol{\Sigma}_{\mathrm{vec}(\boldsymbol{Z}_j^{(s)\top})\,\mathrm{vec}(\boldsymbol{Z}_j^{(s)\top})} \end{pmatrix}\right) \tag{2.44}$$

such that the its expectation corresponds to the equation (2.27) and the elements of variance are given in equations (2.29a), (2.29b) and (2.29c).

The likelihood function is determined by the joint distribution of $\left(\boldsymbol{Y}, \mathrm{vec}\left(\boldsymbol{Z}^\top\right)\right)$, where the parameters associated to the distribution are $\vartheta = (\beta, \gamma, \mathbf{A}, \mathbf{H}, \sigma_s^2, \tau_s^2, \Phi_s, \delta_s)$ para cada $s = 1, \ldots, S$. Then,

$$L(\vartheta|\boldsymbol{Y}, \mathrm{vec}\left(\boldsymbol{Z}^\top\right)) = \prod_{s=1}^{S}\prod_{j=1}^{J_s} f\left(\begin{matrix} \boldsymbol{Y}_j^{(s)} \\ \mathrm{vec}\left(\boldsymbol{Z}_j^{(s)\top}\right) \end{matrix} \middle| \boldsymbol{X}_j^{(s)}, \boldsymbol{W}_j^{(s)}, \vartheta \right)$$

where f corresponds to the density function of the normal distribution defined in the equation (2.44).

So, also the log-likelihood function is obtained, as following

$$l = \sum_{s=1}^{S}\sum_{j=1}^{J_s} \log\left(L\left(\vartheta|\boldsymbol{Y}, \mathrm{vec}\left(\boldsymbol{Z}^\top\right)\right)\right)$$

2.3.1 Likelihood Ratio Tests

In the specification (2.25a), (2.25b) and (2.25c), the K_2-dimensional parameter $\boldsymbol{\delta}_s$ depends on the grouping school function $\rho(\cdot)$ and, therefore, $\boldsymbol{\delta}_s \neq \boldsymbol{\delta}_{s'}$ for $s \neq s'$. Moreover, as it was remarked at Section 2.2.1, if $\boldsymbol{\delta}_s = \delta\,\imath_{K_2}$ for all $s = 1, \ldots, S$, specification (2.25a), (2.25b) and (2.25c) reduces to the HHLIM given by (2.22a), (2.22b) and (2.22c). Similarly, if $\boldsymbol{\delta}_s = 0\,\imath_{K_2}$, specification (2.25a), (2.25b) and (2.25c) reduces to the HHLM given by (2.24a) and (2.24b). Therefore, it is of interest to explore if those equalities are statistically significant. This leads to test the following hypotheses:

(1.) $\mathrm{H}_0 : \boldsymbol{\delta}_s = \boldsymbol{\delta}_{s'}$ for $s \neq s'$; versus $\mathrm{H}_1 : \boldsymbol{\delta}_s \neq \boldsymbol{\delta}_{s'}$ for $s \neq s'$. By doing so, we can test if two groups of schools have different $\boldsymbol{\delta}_s$ parameters.

(2). $\mathrm{H}_0 : \boldsymbol{\delta}_s = \boldsymbol{\delta}$ for all $s = 1, \ldots, S$; versus $\mathrm{H}_1 : \boldsymbol{\delta}_s \neq \boldsymbol{\delta}$ for some $s = 1, \ldots, S$. By doing so, we test if the parameter $\boldsymbol{\delta}_s$ is the same for all schools groups, although its coordinates could be different.

(3). $\mathrm{H}_0 : \boldsymbol{\delta}_s = \delta\imath_{K_2}$ for all $s = 1, \ldots, S$; versus $\mathrm{H}_1 : \boldsymbol{\delta}_s \neq \delta\imath_{K_2}$ for some $s = 1, \ldots, S$. By doing so, we test if the parameter $\boldsymbol{\delta}_s$ is not only equal for all schools groups, but also for all the explanatory factors $\boldsymbol{Z}_{ijk}$ with $k = 1, \ldots, K_2$.

(4). $\mathrm{H}_0 : \boldsymbol{\delta}_s = 0$ for all $s = 1, \ldots, S$; versus $\mathrm{H}_1 : \boldsymbol{\delta}_s \neq 0$ for some $s = 1, \ldots, S$. By doing so, we test if the extended HHLIM model reduces to a standard HHLM model.

The standard strategy to contrast these hypotheses is a likelihood ratio test. More specifically, for cases (1), (2) and (3), the likelihood ratio is based on the conditional distribution of $(\boldsymbol{Y}_j^{(s)\top}, \mathrm{vec}(\boldsymbol{Z}_j^{(s)\top})^\top)^\top$ given $(\boldsymbol{X}_j^{(s)}, \boldsymbol{W}_j^{(s)})$, which corresponds to a multivariate normal distribution with a mean given by (2.27) and a variance-covariance matrix given by (2.28). If we denote by $\Lambda(\boldsymbol{Y}, \boldsymbol{Z})$ the likelihood ratio, then the rejecting zones are respectively the following:

(1). $-2\ln(\Lambda(\boldsymbol{Y}, \boldsymbol{Z})) \sim \chi^2_{K_2}$ and the null hypotheses is rejected if
$P(-2\ln(\Lambda(\boldsymbol{Y}, \boldsymbol{X}) > \chi^2_{K_2}) \leq \alpha)$.

(2). $-2\ln(\Lambda(\boldsymbol{Y}, \boldsymbol{Z})) \sim \chi^2_{K_2(S-1)}$ and the null hypotheses is rejected if
$P(-2\ln(\Lambda(\boldsymbol{Y}, \boldsymbol{X}) > \chi^2_{K_2(S-1)}) \leq \alpha)$.

(3). $-2\ln(\Lambda(\boldsymbol{Y}, \boldsymbol{Z})) \sim \chi^2_{SK_2-1}$ and the null hypotheses is rejected if

$P(-2\ln(\Lambda(\boldsymbol{Y},\boldsymbol{X}) > \chi^2_{SK_2-1}) \leq \alpha)$.

(4). $-2\ln(\Lambda(\boldsymbol{Y},\boldsymbol{Z})) \sim \chi^2_{K_2(K_2S+K_1+L)}$ and the null hypotheses is rejected if $P(-2\ln(\Lambda(\boldsymbol{Y},\boldsymbol{X}) > \chi^2_{K_2(K_2S+K_1+L)}) \leq \alpha)$.

2.3.2 Information criteria

Always it is interest for a dataset choose the "best model", the better fit. For this reason, presented below three measures of commonly used for the choice of the model. (a) Akaike information criterion, AIC, (b) AIC with a correction, AICc and (c) Bayesian information criterion, BIC. Note that the number of parameter the HHLIM model is

$$P = K_1 + K_2 + 2S + K_2(L + K_1 + SK_2).$$

(a) AIC

$$AIC \;=\; -2 * \sum_{s=1}^{S}\sum_{j=1}^{J_s} \log\left(L\left(\vartheta|\boldsymbol{Y}, \mathrm{vec}\left(\boldsymbol{Z}^\top\right)\right)\right) \;+\; 2 * P$$

(b) AICc

$$AIC \;=\; -2 * \sum_{s=1}^{S}\sum_{j=1}^{J_s} \log\left(L\left(\vartheta|\boldsymbol{Y}, \mathrm{vec}\left(\boldsymbol{Z}^\top\right)\right)\right) \;+\; 2 * P \;+\; \frac{2 * P * (P+1)}{N - P - 1}$$

(c) BIC

$$BIC \;=\; -2 * \sum_{s=1}^{S}\sum_{j=1}^{J_s} \log\left(L\left(\vartheta|\boldsymbol{Y}, \mathrm{vec}\left(\boldsymbol{Z}^\top\right)\right)\right) \;+\; P * log(N)$$

2.4 Aplication of HHLIM to educational data

2.4.1 Simulation Study

In this simulation study we will generate data from HHLIM defined in equations, (2.25a), (2.25b), (2.25c). To generate a simulation of a educational database, we consider that the population comes from five different groups $S = 5$, such that the groups are made up of card$(\mathcal{J}_s) = 150$ schools respectively, and in each of them there are 20 to 30 students. The simulation procedure is detailed below,

1. Do $i = 1$ to S,
   - Generate $n_j^{(s)}$ for each $j \in \mathcal{J}_s$, where $20 \leq n_j^{(s)} \leq 30$
   - $\theta_j^{(s)} \sim \mathcal{N}(0, \tau_s^2)$ for each $j = 1, \ldots, \text{card}(\mathcal{J}_s)$.

   Where $\tau_s^2, s = 1, \ldots, S$ are fixed by $\tau_1^2 = 28, \tau_2^2 = 15, \tau_3^2 = 32, \tau_4^2 = 22$ and $\tau_5^2 = 11$

2. Generate L instrumental variables, these may be generated from any distribution. In this particular case, these are assumed $\boldsymbol{W}_{ij}^{(s)} \sim \mathcal{N}(0, 2)$ for each scenario.

3. Generate K_1 exogenous variables, also these can be generated from any distribution. For each scenario are assumed $\boldsymbol{X}_{ij}^{(s)} \sim N(0, 1)$.

4. Generate K_2 error of endogenous variables associated to the equation (2.25b). Then, simulate $\boldsymbol{\eta}_{ij}^{(s)} \sim \mathcal{N}_{K_2}(0; \boldsymbol{\Phi}_s)$. Such that are fixed each $\boldsymbol{\Phi}_s$ con $s = 1, \ldots, S$.

   $$\boldsymbol{\Phi}_1 = \begin{pmatrix} 7 & 3 \\ 3 & 4 \end{pmatrix} \boldsymbol{\Phi}_2 = \begin{pmatrix} 3.5 & 1.3 \\ 1.3 & 6 \end{pmatrix} \boldsymbol{\Phi}_3 = \begin{pmatrix} 9 & 4.1 \\ 4.1 & 6.5 \end{pmatrix} \boldsymbol{\Phi}_4 = \begin{pmatrix} 6 & 1.5 \\ 1.5 & 4 \end{pmatrix}$$
   $$\boldsymbol{\Phi}_5 = \begin{pmatrix} 5.3 & 2.9 \\ 2.9 & 3.8 \end{pmatrix}$$

5. Once the steps 1 – 4 were performed, must be simulated the K_2 endogenous variates by, $\boldsymbol{Z}_{ij}^{(s)} = \boldsymbol{A}^\top \boldsymbol{X}_{ij}^{(s)} + \boldsymbol{H}^\top \boldsymbol{W}_{ij}^{(s)} + \theta_j^{(s)} \boldsymbol{\delta}_s + \boldsymbol{\eta}_{ij}^{(s)}$. Where $\boldsymbol{A}$ and $\boldsymbol{H}$ are the same for each scenario, fixed in $\boldsymbol{A} = \begin{pmatrix} 6.8 & 5.2 \\ -0.9 & 6.1 \end{pmatrix}$ and $\boldsymbol{H} = \begin{pmatrix} 3.6 & 2.1 \\ 0.5 & 2.3 \\ 4.8 & 2.9 \end{pmatrix}$

But, $\boldsymbol{\delta}_s, s = 1, \ldots, S$ are fixed in three different scenarios.

6. Generate a error associated to the equation (2.25a). So, $\epsilon_{ij} \sim \mathcal{N}(0; \sigma_s^2)$, Where $\sigma_s^2, s = 1, \ldots, S$ are fixed by $\sigma_1^2 = 38, \sigma_2^2 = 25, \sigma_3^2 = 42, \sigma_4^2 = 32$ and $\sigma_5^2 = 21$.

7. Once the steps 5 – 6 were performed, must be simulated the dependent variables by, $\boldsymbol{Y}_{ij}^{(s)} = \boldsymbol{X}_{ij}^{(s)\top}\boldsymbol{\beta} + \boldsymbol{Z}_{ij}^{(s)\top}\boldsymbol{\gamma} + \theta_j^{(s)} + \epsilon_{ij}$. Where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the same for each scenario, fixed in $\boldsymbol{\beta} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ and $\boldsymbol{\gamma} = \begin{pmatrix} 7 \\ 4 \end{pmatrix}$

Simulation Results

We study the performance of our estimation methods under the model defined above. We consider three scenarios: (I) parameters of association ($\boldsymbol{\delta}_s$), between endogenous variables and school effect, different for each groups and endogenous variables, i.e, we have $S * K_2$ parameters, (II) parameters of association ($\boldsymbol{\delta}$), between endogenous variables and school effect, only different for each endogenous variables, i.e, we have K_2 parameters, and (III) parameters of association (δ), between endogenous variables and school effect, is same for each groups and endogenous variables, i.e, we have 1 parameters.

In the tables B.7, B.14 and B.21 we show te true value and Monte Carlo estimation of the parameter $\boldsymbol{\delta}_s$ for each scenarios, for samples of 50, 100, 500 and 1000. In each one scenario and samples of Monte Carlo, the same trend is reflected, a proper and satisfactory estimate of parameters, where the Monte Carlo error is low (this value is presented in the tables between parentheses). Similary the estimates of the other parameters ($\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{A}$, $\boldsymbol{H}$, $\boldsymbol{\Phi}_s$, τ_s^2 and σ_s^2), have estimates close to the real value, being the variances τ_s^2 and σ_s^2 that has a higher error Monte Carlo. See Annex B.

Moreover, to measure the accuracy in estimating the school effect and value added, is calculated the average of the squared differences in each of the samples Monte Carlo, thus the Tables B.1, B.8 and B.15 are a summary of what happened in the Monte Carlo simulation. Thereby for different values of $\boldsymbol{\delta}_s$ the accuracy in the estimation of both indicators (value added and school effect) have an averaged low squared error, value around of 0.003 in the case of value added and 0.30 in the school effect.

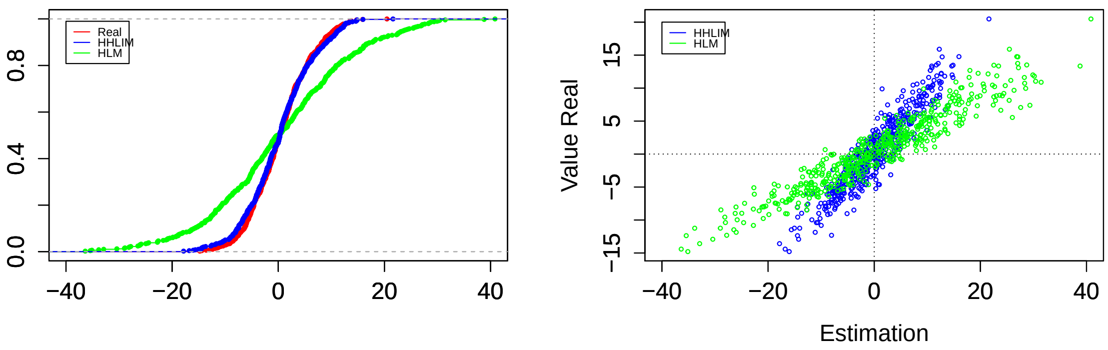
Comparation among models

This section aims to illustrate through a simulation of educational data the differences in fit between a standard hierarchical linear mixed model and the model proposed in this chapter.

The Figures 2.2a, 2.3a and 2.4a shown three curves of cumulative distribution corresponding to the real value added (red colour), the estimation of value added by means of a standard HLM model (green curve) and the estimation of value added by HHLIM (blue curves). We conclude that (to compare the red and green curves) in the three figures mentioned, the distribution of estimations by a standard HLM is different to the distribution of real values (curve red), its distribution are not close to the real and further it has tails heavy. This means that in the presence of one o more endogenous variables classify the school through a HLM model is unfair because this has serious problems in the estimation of value added. In contrary, if we compare the distributions of real value added with the estimation obtained through our proposed model (red and blue curves), where we saw in the previous section that the Monte Carlo errors are low, we conclude that this model generates estimates close to reality when there are some grade of endogeneity.

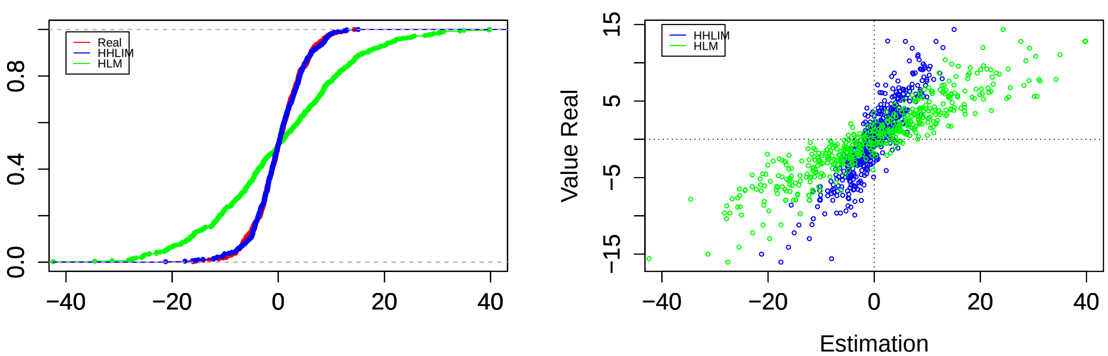
Moreover, in the Figures 2.2b, 2.3b and 2.4b of three scenarios one can observe that the estimates of standard HLM model (green points) are concentrated arround zero and increases dispersion on the tails, while in the estimation of HHLIM model (blue points) its dispersion is constant.

It is important not to forget this possible endogeneity, because in the presence of this the standard models have problems in the estimates of value added, therefore the schools can be classified unfairly.



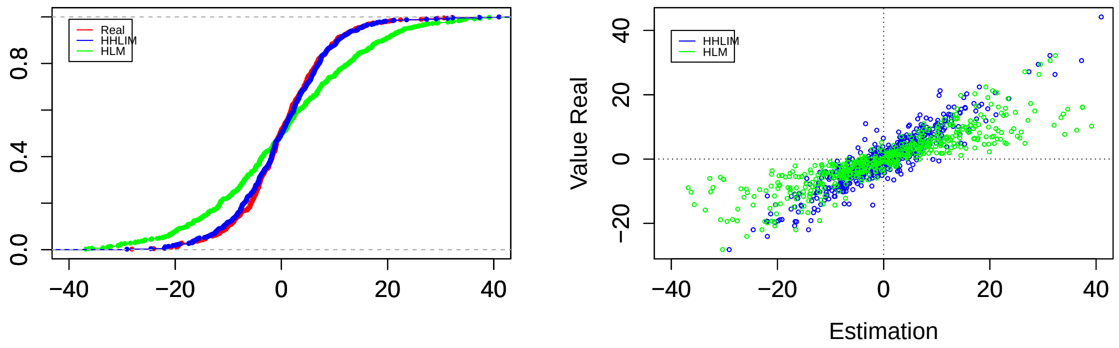
(a) Cumulative distribution (b) Scatter plot

Figure 2.2: Comparison between standard HLM and HHLIM model, Scenario I. (a) Real value of Value Added (red) and estimation by HHLIM (blue) and standard HLM (green). (b) Axis Y: Real value of Value Added, Axis X: Estimation of Value Added by HHLIM (blue) and standard HLM (green).



(a) Cumulative distribution (b) scatter plot

Figure 2.3: Comparison between standard HLM and HHLIM, Scenario II. (a) Real value of Value Added (red) and estimation by HHLIM (blue) and standard HLM (green). (b) Axis Y: Real value of Value Added, Axis X: Estimation of Value Added by HHLIM (blue) and standard HLM (green).



(a) Cumulative distribution

(b) scatter plot

Figure 2.4: Comparison between standard HLM and HHLIM, Scenario III. (a) Real value of Value Added (red) and estimation by HHLIM (blue) and standard HLM (green). (b) Axis Y: Real value of Value Added, Axis X: Estimation of Value Added by HHLIM (blue) and standard HLM (green).

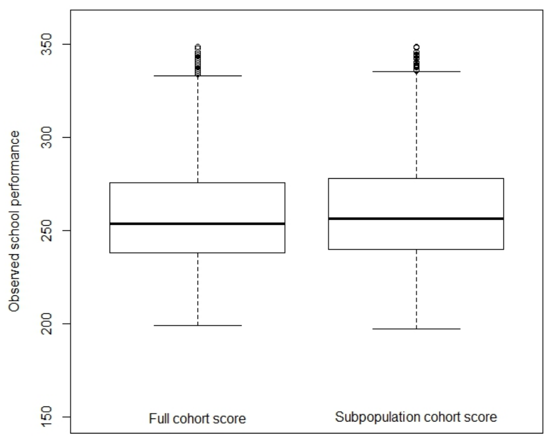
2.4.2 Data application

In order to apply the endogenous value-added model for subgroups of schools, we use the data set SIMCE applications in Mathematics 2007 and 2011. Then after merging both data sets and eliminating the schools with less than 20 students, we obtain a data set containing 157,737 students belonging to 3,305 schools. Table 2.1 summarizes the SIMCE Mathematics scores at the school level controlling by SES.

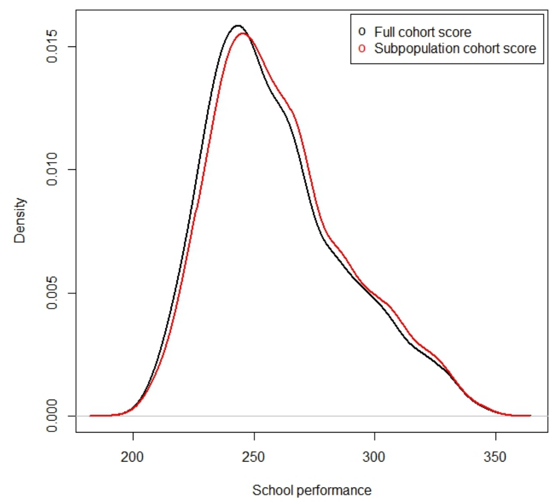
Table 2.1: SIMCE Mathematics scores at the school level controlling by SES

SES	Number of schools	SIMCE Math 2007		SIMCE Math 2011	
		Mean	Std. Dev.	Mean	Std. Dev.
A	261	236.88	15.63	220.29	23.18
B	1150	241.19	17.50	230.22	21.52
C	1074	258.07	19.15	246.83	22.72
D	548	281.67	20.78	269.61	21.88
E	272	311.08	20.37	294.14	18.36

Taking into account that our main methodological concern deals with the endogeneity of the prior score, the value added analysis will be focused on a subsample of schools defined by the following condition: the percentage of students in a school in 2011 who was in a different school in 2007 is at most 50%. Thus, the subsample includes 2,880 schools (that is, the 87% of the population of schools), which gather 139,071 students. Let us mention that the distribution of the 2011 SIMCE Mathematics scores of these schools is practically the same as the distribution of the 2011 SIMCE Mathematics scores in the full sample of schools, see Figures 2.5a and 2.5b.



(a) Distribution of SIMCE test 2007 in all the population and the subpopulation



(b) Density of SIMCE test 2007 in all the population and the subpopulation

Figure 2.5: From both figures we conclude that study population has the same distribution and density of a subpopulation, this last considers only the schools that at least 50% of its students remained in both measurements SIMCE 2007-2011.

The endogenous value added model for subgroups of schools and students is constructed using the following variables:

1. Dependent variable: the 2011 SIMCE Mathematics score, denoted as Math11_{ij}.
2. Endogenous variables:
   (a) Prior score: the 2007 SIMCE Mathematics score, denoted as Math07_{ij}.
   (b) Compositional effect: the mean, at the school level, of Math07_{ij}, denoted as AMath07_{j}.
3. Exogenous variable:
   (a) Student selection: along with the standardized SIMCE test, a survey is administered to parents. One of its questions deals with the three possible selection mechanisms applied by the school to eventually choose a student: a play session; a cognitive test; and an interview. We define a binary variable of

selection: it is equal to 1 if one of these mechanisms is declared by parents; and it is zero otherwise. A logistic regression was fitted and used to impute missing responses; the covariates used in the logistic regression were the SES of the school, the administrative dependency of the school (public, subsidized without co-payment, subsidized with copayment), a vulnerability index of the school, rurality, and if the student have or not brothers in the same school. The selection variable was defined as the probability to be selected by the school; it is denoted as Select_{ij}.

4. Instrumental variables:

   (a) Mother educational level and father educational level, denoted as Mother_{ij} and Father_{ij}. These are exogenous factors because the intra-school practices are affected either by the national curriculum or by internal pedagogical organization. In Chile, parents have not a relevant active role inside of school practices.

   (b) The number of persons living in the same house with the student, denoted as Hab_{ij}. This is an exogenous variable because it is a factor uncontrolled by the school.

Model Fit

Model (2.25a), (2.25b) and (2.25c) was fitted to the 2007-2011 SIMCE data using the estimation procedure developed in this dissertation. In order to summarize the results, let us begin by the degree of endogeneity, which is represented by a two-dimensional parameter $\delta^{(s)}$, where s takes the values A, B, C, D or E, depending on the SES level of the school. For the data set under analysis, it should be said that the following two null hypotheses are rejected ($p-value = 0.000$): that the degrees of endogeneity are unique irrespective of the SES (Socioeconomic status) of schools as well as of the endogenous variables $Math07_{ij}$ and $AMath07_j$; and that the degrees of endogeneity are unique irrespective of the SES of schools. The estimation of the degrees of endogeneity by both SES of the school and endogenous variables are reported in Table 2.2.

Table 2.2: Estimations of degree of endogeneity and of component variances

SES	Endogeneity degree related to $Mat07$	Endogeneity degree related to $AMat07$	σ_s^2	τ_s^2
A	4.916	5.914	822.322	26.356
B	4.230	5.556	816.742	35.879
C	2.664	3.576	936.926	65.176
D	1.945	2.546	999.591	96.423
E	2.674	2.298	913.834	58.994

These results suggest to testing that the degrees of endogeneity associated to $Math07_{ij}$ is equal for the SES C and E. The null hypothesis is accepted ($p-value = 0.87$), which leads to conclude that the degree of endogeneity of schools of SES levels C or E is different from schools of levels A or B or D. This feature is taken into account to estimating the school value added.

Regarding the between-school and within-school variances, Table 2.2 shows the estimations. The variability between schools of SES level D is the largest one, whereas the variability between schools of SES level A is the smaller. For the within-variances, the larger are those of students belonging to schools of SES level D, whereas the smaller is the variance of students belonging to schools of SES level A.

The estimations of the marginal effects of the explanatory factors in equation (2.25a) are the following: for the exogenous individual factor of selectivity, it is equal to 61.8; for the endogenous individual factor $Math07_{ij}$, it is equal to 0.57; for the group endogenous factor $AMath07_j$, it is -0.128.

Finally, let us report the value added indicators by comparing the effectiveness of schools according to school characteristics other than those considered for estimation purposes (degree of endogeneity by SES; within- and between-variability by SES). Taking into account the Chilean policy context, we consider the following two characteristics:

1. The administrative dependency. In Chile, the administration of the schools is fourfold: public schools administered by a county corporation (MC); public schools administered directly by the county (MD); subsidized schools administered by private providers, each of them receiving the voucher from the state (PS); and private schools administered by private providers, not economically supported by the state (PP). In the sample under analysis, there are 15.7% of MC public schools, 30.4% of

MD public schools, 46.3% of subsidized schools and 7.6% of private schools.

2. Level of vulnerability[4]. As explained above, this index measures the percentage of students at social risk. We define four groups of school according to the level of vulnerability as shown in Table 2.3.

Table 2.3: Schools classified by vulnerability index and administrative dependency

	MC	MD	PS	PP	Total
IVE 0-25%	17	11	320	219	577 (20 %)
IVE 25% -50%	105	137	622	1	865 (30 %)
IVE 50% -75%	183	285	253	0	721 (25 %)
IVE 75%-100%	148	445	129	0	722 (25%)
Total	453 (15.7%)	878 (30.4%)	1334 (7.6%)	220 (46.3%)	2885

Figure 2.6 shows the box plot of the value added indicators for each school administrated dependency. It can be concluded that the dispersion of effectiveness in subsidized schools (PS) is greater that PP, MC and MD schools. Furthermore, the effectiveness of the MD schools is not quite different from the effectiveness of MC public schools.

[4]The vulnerability index (IVE) is constructed by the Junta Nacional de Auxilio Escolar y Becas (JUNAEB, National Scholarship and School Aid Board). JUNAEB estimates IVE on the basis of the results of a parent survey conducted by schools. These surveys provide information about the student's background. The IVE has a minimum value of 0, which represents 0 percent of children at social risk, and goes up to a maximum value of 100, indicating the most disadvantaged schools (100 percent social risk).

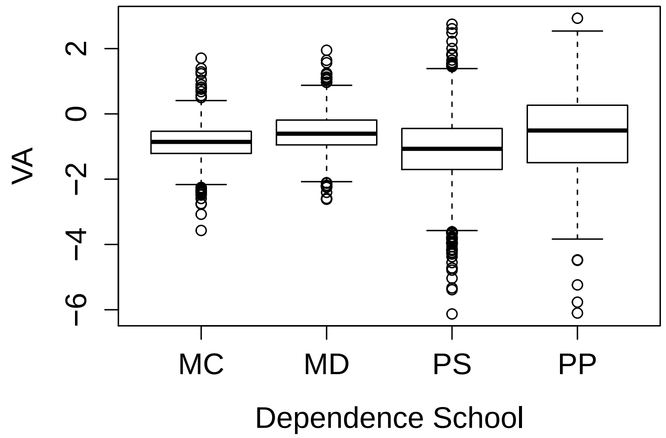


Figure 2.6: Distribution of Value added by dependence school

Figure 2.7 shows the box plot of the value added indicators for each group of schools according to vulnerability level. It can be see that there are not an association between effectiveness and vulnerability index, this results suggests to investigate why the presence of students with social risk seems to not dramatically affect the effectiveness.

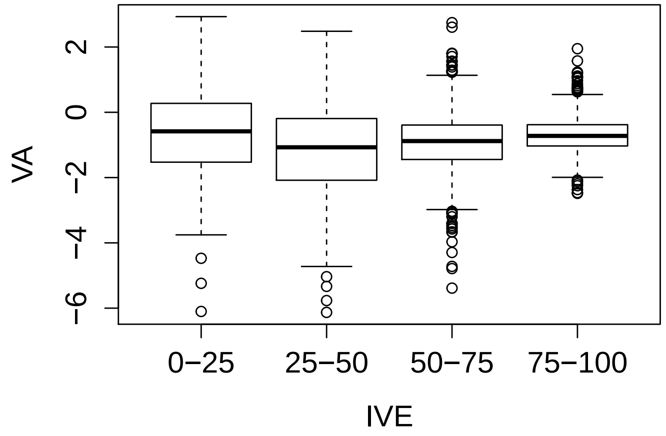


Figure 2.7: Distribution of Value added by categories of IVE

2.5 Final remarks

We proposed an extension to model presented by Manzi et al. (2014) that consist in specify the association parameter no longer as a scalar, if not as a vector. It can be different for each endogenous variables and school effect. Addition, if the educational data set is heteroscedastic, means it is coming from a mixture distribution, then it is probably that the vector of association parameter is different for each group of the population ($\boldsymbol{\delta}_s$).

From a methodological point of view, the novelty of this paper is the specification and estimation of a value added model under endogeneity. Specifically,

- The model estimates the fixed effects as well as the component variances in an unbiased way.
- The model estimates the correlation between the school effect and each endogenous variable.
- The model estimates the value added by correcting the prediction of the school effect.

From an empirical point of view, the novelties are the following:

- To perform a value added analysis of the Chilean educational system controlling endogeneity problems by design.
- To make a comparison between the previous analysis and standard value added analysis.

As a final comment, our model does not have a computational cost, its convergence is quick, even when is used on SIMCE educational data that it has a large volume of students.

Chapter 3:

On the modeling of school improvement through a time dependent value-added model

Chapter 3: On the modeling of school improvement through a time dependent value-added model

Contents

Figures

Chapter 3: On the modeling of school improvement through a time dependent value-added model

This chapter focuses in the school improvement problem through of value added trajectories. In this part of dissertation we proposed a model with dependence across the time that included in its structure an association parameter between the school effects of two cohorts of students different (under the methodology proposed can be extended our design to more cohorts). In addition to developing this model we present a simulation study and an application in the Chilean educational data, using the SIMCE tests in cohorts 4th-8th grade students years 2007-2011 and 2009-2013.

3.1 Introduction

In this chapter, we have not intention of perform a literature review on the school effectiveness, since these have been reviewed in the previous chapters and the literatures have been regularly and systematically reviewed by others authors, see, for example Teddlie and Reynolds (2000).

Recalling that typically the school effectiveness is assessed through a value added model, technique widely accepted by School Effectiveness Research (SER) . The general definition of value added is the contribution of a school to students progress towards stated or prescribed education objectives OECD (2008) . This contribution is net of other factors that contribute to students' educational progress, see also Braun et al. (2010) and Baker et al. (2010). This type of statements leads to conclude that differences in school effectiveness have important consequences for student progress.

A standard approach to model the school value-added are the hierarchical linear mixed models, or multilevel models Goldstein (2002); Snijders and Bosker (1999) due to the structure of educational data where students are nested into schools. In this context, students scores are explained by their previous achievement, some covariates and a random effect representing the school effect. A measure of Value-Added has been typically obtained as the prediction of the random school effect Aitkin and Longford (1986); Longford (2012); Raudenbush and Willms (1995); Tekwe et al. (2004).

OECD (2008) report recognizes that in the countries a growing emphasis is being placed upon measures of school performance as they are central to school improvement efforts, systems of school accountability and school choice, and broader educational policies" OECD (2008). Alongside to growth of political interest in school measurements and accountability system, also the volume of educational data increased, nowadays is possible and becoming more common, find data-sets from of more one cohort of students. This way the concept of school improvement has been defined in terms of trajectories of indicators of added value, it is necessary to have of two measurements (pre and post test) for those indicators.

With the trajectories of value added is of interest know if a school improved (or not) across time, we make an implicit assumption, that the past of a school determines (in some degree) the evolution future of the same. Thereby, our objective is to evaluate the school improvement when different cohorts of students are treated by the same school. In this work our attention is in two cohorts being possible easily generalize to more cohorts. Specifically we consider two cohorts such that for each one them the outcome of two test are available (post-test and outcomes contemporaneous, i.e pre-test). This situation can be illustrated graph the following diagram.

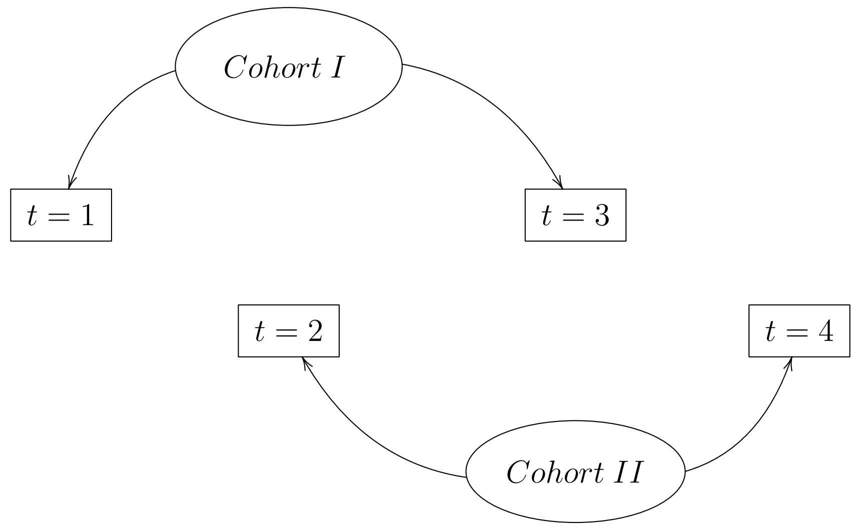


Figure 3.1: Example of Chilean case, two cohorts

Investigating the evolution of value-added measures over time is not new and it has been considered as an important topic in SER. The question on how the performance of schools across time should be measured has been studied in several references. Among others:

- Gray, Goldstein, and Jesson (1996) considered several cohorts of students, using a different intercept for each cohort in a joint analysis.
  - School effects are modelled via a linear trend or rolling averages (which combine the results of 3 consecutive cohorts).
  - Students (level 1) were nested in schools (level 2); the effect of each school for each year is allowed to vary.
  - 3-level model in which students (level 1) are grouped by year (level 2) within school (level 3). At level 3 both year and prior achievement effects are allowed to vary randomly across schools, as well as the intercept (sic).
- Thomas (2001), Gray, Goldstein, and Thomas (2001) showed that various methods of analysis can result in different estimates of stability or instability in school effects over the time. See also the debate in Pugh and Mangan (2003).

- Briggs and Weeks (2011) used longitudinal data and propose a way to model persistence of VA over time using an accumulation model:

$$\begin{aligned} Y_{j2} &= \mu_2 + \theta_2 + \epsilon_{j2} \\ Y_{j3} &= \mu_3 + \theta_2 + \theta_3 + \epsilon_{j3} \\ &\vdots \end{aligned}$$

- This idea has been followed in many studies, including recent studies e.g. on teacher VA persistence (e.g. Kinsler (2012); Rothstein (2010)).
- Recently, models have been proposed to explicitly account for within-group dynamics (Bauer, Gottfredson, Dean, and Zucker (2013); Steele, Rasbash, and Jenkins (2013))

None of these approaches, however, consider the fact that the school value added calculated at time $t-1$, do influence the one calculated at time t. Ignoring this fact would means that if a school has implemented changes in the internal educational policies to improve student performance, its effects are not being reflected in the new measure of value added at time t, see . This, in this research we propose to considered a longitudinal structure such that the school effect at time t is calculated conditioning on the school effect at time $t-1$. This strategy will be used to assess the evolution of the proposed robust measure of value added described above.

3.2 Dinamycal Models, Two Cohorts

3.2.1 Model Specification

In this section we provide the structural model that it is used to describe the educational data, where each student is submitted to two test in some moment of their school life, for example in 4th and 8th grade. But the schools have several outcomes on the time, as constantly students are evaluated, this way the schools of a *school system* have several outcomes on the time.

Now, the school improvement depend of trajectories of value added models. Therefore in order to specify a VA model for two cohorts, the school effects should be dependent, one them depends on time moments t=1 and t=3, and the other one depends by time moment t=2 and t=4, schematically, this can be illustrated as follows,

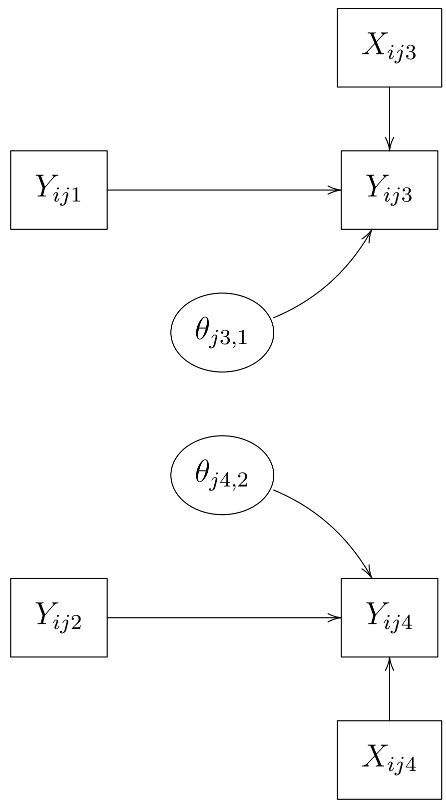


Figure 3.2: Preliminary specification of VA model for two cohorts

Where, we denote $\boldsymbol{Y}_{ijt}$ as test outcomes of student i belonging to school j on the time t, with $i = 1, \ldots, n_{jt}$, $j \in \mathcal{J}$ and $t = 1, 2, 3, 4$, $\mathcal{J}$ is the set of all school and n_{jt} is the number of students in the school j, $\boldsymbol{X}_{ij3}$ and $\boldsymbol{X}_{ij4}$ are covariates that impact on the post test score (that influence in $\boldsymbol{X}_{ij3}$ and $\boldsymbol{X}_{ij4}$) respectively. $\theta_j 3$ and $\theta_j 4$ are the schools effect. Those effects are depend of the axiom of local independence (....), namely

1. Conditionally on $(\boldsymbol{X}_{j3}, \boldsymbol{Y}_{j1}, \theta_{j3,1})$, $\{Y_{ij3} \; : \; i = 1 \ldots, n_{j3}\}$ are mutually independent.
2. $(Y_{ij3} \,|\, \boldsymbol{X}_{j3}, \boldsymbol{Y}_{j1}, \theta_{j3,1}) \equiv (Y_{ij3} \,|\, \boldsymbol{X}_{ij3}, Y_{ij1}, \theta_{j3,1})$.

a relationship between both school effects, it is shown in the following diagram,

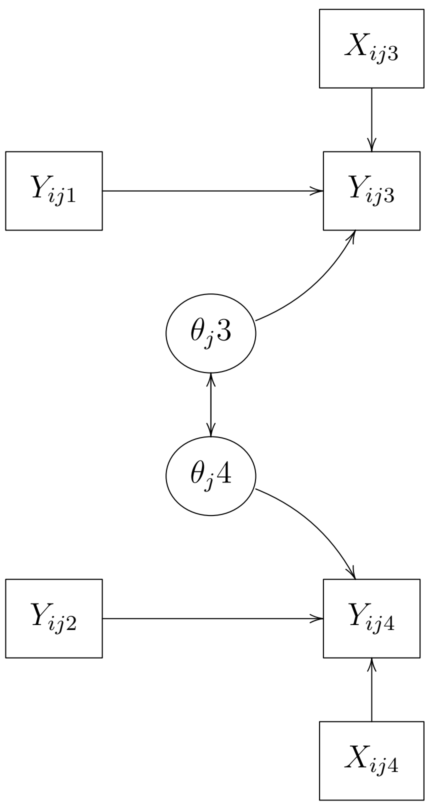


Figure 3.3: Specification of VA model for two cohorts

Strucural model

For simplicity of notation, you should note that $\boldsymbol{X}_{4j}$ and $\boldsymbol{X}_{3j}$ are covariates associated at both cohorts, then from this moment we consider that the prior score of student ($\boldsymbol{Y}_{j3}$ and $\boldsymbol{Y}_{1j}$) is a covariate in the model VA, ie $\boldsymbol{Y}_{j3} \in \boldsymbol{X}_{4j}$ and $\boldsymbol{Y}_{1j} \in \boldsymbol{X}_{3j}$. In this chapter will not be discussed if the previous score is a exogenous or endogenous variable, however this can be addressed and discuss in a future work, for the moment it is considered an exogenous variable at school effect.

Taking into account that there are J schools different, it is reasonable to assume that the set of school are independent, ie $\{(\boldsymbol{Y}_{4j},\ \boldsymbol{Y}_{3j},\ \boldsymbol{X}_{4j},\ \boldsymbol{X}_{3j},\ \theta_{j4,2},\ \theta_{j3,1}) : j = 1, \dots, J\ \}$ are mutually independent, the results of school j are not related to the results of school j'. This way, for each school j be have the following,

1. $(\boldsymbol{Y}_{j4}|\boldsymbol{Y}_{j3}, \boldsymbol{X}_{j4}, \boldsymbol{X}_{j3}, \theta_{4j}, \theta_{j3,1}\ ) \equiv (\boldsymbol{Y}_{j4}|\boldsymbol{X}_{j4}, \theta_{j4,2}\ )$. This means that the students' scores of the second cohort is independent of the set of "scores", "covariates" and "school effect" of the first cohort given the "covariates" and "school effect" of the same cohort,
$$\boldsymbol{Y}_{j4} \perp\!\!\!\perp (\boldsymbol{Y}_{j3}, \boldsymbol{X}_{j3}, \theta_{j3,1}) \mid (\boldsymbol{X}_{j4}, \theta_{j4,2})$$

2. $\underset{1\,\le\, i\, \le n_{j4}}{\perp\!\!\!\perp} \boldsymbol{Y}_{j4}|\boldsymbol{X}_{j4}, \theta_{j4,2}$. Then, by axiom of local independence

   - $Y_{ij4}|\,(\boldsymbol{X}_{ij4}, \theta_{j4,2}) \sim N(\boldsymbol{X}_{ij4}\boldsymbol{\beta}_4 + \theta_{j4}, \sigma_4^2)$
   - $\boldsymbol{Y}_{j4}|\,(\boldsymbol{X}_{j4}, \theta_{j4,2}) \sim N(\boldsymbol{X}_{j4}\boldsymbol{\beta}_4 + \theta_{j4}\iota_{n_{j4}}, \sigma_4^2\boldsymbol{I}_{n_{j4}})$

3. $\theta_{j4,2}|(\boldsymbol{Y}_{j3}, \boldsymbol{X}_{j4}, \boldsymbol{X}_{j3}, \theta_{j3,1}) \equiv \theta_{j4,2}|\theta_{j3,1}) \sim N(\mu\theta_{j3,1}, \tau_4^2)$. This means that,
$$\theta_{j4,2} \perp\!\!\!\perp (\boldsymbol{Y}_{j3}, \boldsymbol{X}_{j4}, \boldsymbol{X}_{j3})|\theta_{j3,1}$$

4. $(\boldsymbol{Y}_{j3}|\boldsymbol{X}_{j4}, \boldsymbol{X}_{j3}, \theta_{j3,1}\ ) \equiv (\boldsymbol{Y}_{j3}|\boldsymbol{X}_{j3}, \theta_{j3,1}\ )$, i.e
$$\boldsymbol{Y}_{j3} \perp\!\!\!\perp (\boldsymbol{X}_{j4}) \mid (\boldsymbol{X}_{j3}, \theta_{j3,1})$$

5. $\underset{1\leq l\leq n_{j3}}{\perp\!\!\!\perp} Y_{ij3}|\boldsymbol{X}_{ij3}, \theta_{j3,1}$, then the following distribution is assumed

   - $Y_{ij3}|\,(\boldsymbol{X}_{ij3}, \theta_{j3,1}) \sim N(\boldsymbol{X}_{ij3}\boldsymbol{\beta}_3 + \theta_{j3}, \sigma_3^2)$
   - $\boldsymbol{Y}_{j3}|\,(\boldsymbol{X}_{j3}, \theta_{j3,1}) \sim N(\boldsymbol{X}_{j3}\boldsymbol{\beta}_3 + \theta_{j3}\iota_{n_{j3}}, \sigma_3^2\boldsymbol{I}_{n_{j3}})$

6. $(\theta_{j3,1}|\boldsymbol{X}_{j4}, \boldsymbol{X}_{j3}) \sim N(0, \tau_3^2)$, ie,

$$\theta_{j3,1}\perp\!\!\!\perp\boldsymbol{X}_{j4}, \boldsymbol{X}_{j3}$$

7. $(\mathbf{X}_{j4}, \mathbf{X}_{j3})$ are left unspecified since they are assumed to be exogenous variables.

This way, under this structure the dynamic value-added model of two cohorts, is written as,

$$Y_{ij3} = \boldsymbol{X}_{ij3}\boldsymbol{\beta}_3 + \theta_{j3} + \epsilon_{ij3}, \quad \text{where,} \quad \epsilon_{ij3} \sim N(0, \sigma_3^2) \tag{3.1a}$$

with with $\sigma_3^2 > 0$. The equation (3.1a) correspond to the lineal regression of student's score i on first cohort, such that this dependent variable is explained through; the covariates by the fixed effect $\mathbf{3}\beta_{\in}\mathbb{R}^{K_3}$ with an common school effect by all students of school j and the idiosyncratic errors ϵ_{ij3}, addition recalling that the number of covariates, K_3, including the prior score, Y_{ij1}. Moreover,

$$\theta_{j3} \sim N(0, \tau_3^2) \tag{3.1b}$$

with $\tau_3^2 > 0$. Now, the lineal regression of student's score i on the second cohort correspond to the equation (3.1c), this variable is explained through the covariates by the fixed effect $\boldsymbol{\beta}_4 \in \mathbb{R}^{K_4}$ with an common school effect by all students of school j and the idiosyncratic errors ϵ_{ij4},

$$Y_{ij4} = \boldsymbol{X}_{ij4}\boldsymbol{\beta}_4 + \theta_{j4} + \epsilon_{ij4}, \quad \text{where,} \quad \epsilon_{ij4} \sim N(0, \sigma_4^2) \tag{3.1c}$$

with $\sigma_4^2 > 0$. The school effect in this second cohort is not centered in zero, because this effect depends of occurred in the past, this way it is assumed that,

$$\theta_{j4} \sim N(\mu\theta_{j3,1}, \tau_4^2) \tag{3.1d}$$

with $\tau_4^2 > 0$.

In summary, the equations of model (3.1a–3.1d) assumes the following,

Assumption 1 ***: Exogeneity,***

The covariates matrix $\boldsymbol{X}_4$ is independent of vector random effects $\boldsymbol{\theta}_4$, and $\boldsymbol{X}_3$ is independent of vector random effects $\boldsymbol{\theta}_3$,

$$Cov(\boldsymbol{X}_4, \boldsymbol{\theta}_4) = \mathbf{0}, \quad \text{and} \quad Cov(\boldsymbol{X}_3, \boldsymbol{\theta}_3) = \mathbf{0}$$

Assumption 2 ***: Independence of random effect,***

The θ_{4j}'s are mutually independent for each j, and θ_{3j}'s also are mutually independent for each j, but not between them.

Assumption 3 ***: Distribution and Homoscedasticity of the idiosyncratic error,***

The ϵ_{j4}'s and ϵ_{j3}'s for each j are mutually independent.

Assumption 4 ***: Local Independence***

The local independence corresponds to,

$$\underset{1 \leq i \leq n_{j4}}{\perp\!\!\!\perp} Y_{ij4} \mid \boldsymbol{X}_{j4}\ ,\ \theta_{j4,2}, \quad \text{and} \quad \underset{1 \leq i \leq n_{j3}}{\perp\!\!\!\perp} Y_{ij3} \mid \boldsymbol{X}_{j3}\ ,\ \theta_{j3,1}.$$

Joint Distribution

Thus we can write the joint distribution of each school, as following (see demonstration in Annex C.1.2);

$$\left(\begin{array}{c|c} \boldsymbol{Y}_{j4} & \\ \boldsymbol{Y}_{j3} & \boldsymbol{X}_{j4}, \boldsymbol{X}_{j3} \\ \theta_{j4,2} & \\ \theta_{j3,1} & \end{array}\right) \sim N\left(\left(\begin{array}{c} \mu_{j4} \\ \mu_{j3} \\ \mu_{\theta_4} \\ \mu_{\theta_3} \end{array}\right); \left(\begin{array}{cccc} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} & \Sigma_{14} \\ & \Sigma_{22} & \Sigma_{23} & \Sigma_{24} \\ & & \Sigma_{33} & \Sigma_{34} \\ & & & \Sigma_{44} \end{array}\right)\right) \tag{3.2}$$

where,

1. $\left(\begin{array}{c} \mu_{j4} \\ \mu_{j3} \\ \mu_{\theta_4} \\ \mu_{\theta_3} \end{array}\right) = \left(\begin{array}{c} \boldsymbol{X}_{j4}\beta_4 \\ \boldsymbol{X}_{j3}\beta_3 \\ 0 \\ 0 \end{array}\right)$,
   is of dimension $(n_{j4} + n_{j3} + 1 + 1) \times 1$

2. $\left(\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ & \Sigma_{22} \end{array}\right) = \left(\begin{array}{cc} \sigma_4^2 \boldsymbol{I}_{n_{j4}} + (\tau_4^2 + \mu^2\tau_3^2)\boldsymbol{J}_{n_{j4}} & \iota_{n_{j4}}\iota_{n_{j3}}^\top \mu\tau_3^2 \\ & \sigma_3^2 \boldsymbol{I}_{n_{j3}} + \tau_3^2 \boldsymbol{J}_{n_{j3}} \end{array}\right)$,
   is of dimension $(n_{j4} + n_{j3}) \times (n_{j4} + n_{j3})$

3. $\left(\begin{array}{cc} \Sigma_{13} & \Sigma_{14} \\ \Sigma_{23} & \Sigma_{24} \end{array}\right) = \left(\begin{array}{cc} \iota_{n_{j4}}(\tau_4^2 + \mu^2\tau_3^2) & \iota_{n_{j4}}\mu\tau_3^2 \\ \iota_{n_{j3}}\mu\tau_3^2 & \iota_{n_{j3}}\tau_3^2 \end{array}\right)$,
   is of dimension $(n_{j4} + n_{j3}) \times 2$

4. $\left(\begin{array}{cc} \Sigma_{33} & \Sigma_{34} \\ & \Sigma_{44} \end{array}\right) = \left(\begin{array}{cc} \tau_4^2 + \mu^2\tau_3^2 & \mu\tau_3^2 \\ & \tau_3^2 \end{array}\right)$,
   is of dimension 2×2

3.2.2 Parameter Identification

The identification of the parameters is a requirement necessary for parameter estimation. The parameters of interest are K_4 coefficients in vector β_4, K_3 coefficients in the vector β_3, μ and the variance components σ_4^2, σ_3^2, τ_4^2 and τ_3^2. To verify the identification of these parameters, it is necessary integrate the schools effect in the joint distribution expressed in equation (3.2). Then, the distribution corresponds to,

$$\left(\begin{array}{c|c} \boldsymbol{Y}_{j4} & \multirow{2}{*}{$\boldsymbol{X}_{j4}, \boldsymbol{X}_{j3}$} \\ \boldsymbol{Y}_{j3} & \end{array}\right) \sim N\left(\begin{pmatrix} \boldsymbol{X}_{j4}\beta_4 \\ \boldsymbol{X}_{j3}\beta_3 \end{pmatrix}; \begin{pmatrix} \sigma_4^2\boldsymbol{I}_{n_{j4}} + (\tau_4^2 + \mu^2\tau_3^2)\boldsymbol{J}_{n_{j4}} & \iota_{n_{j4}}\iota_{n_{j3}}^\top\mu\tau_3^2 \\ & \sigma_3^2\boldsymbol{I}_{n_{j3}} + \tau_3^2\boldsymbol{J}_{n_{j3}} \end{pmatrix}\right) \quad (3.3)$$

Given the equation (3.3) of joint distribution, it is possible to analyse the parameter identification by the following standard arguments,

- By standard arguments from $\mathbf{E}\left(\boldsymbol{Y}_3|\boldsymbol{X}_4, \boldsymbol{X}_3\right)$, is identify β_3, with the condition that the rank of the matrix, $\boldsymbol{X}_3^\top\boldsymbol{X}_3$, is complete.
- The same way from $\mathbf{E}\left(\boldsymbol{Y}_4|\boldsymbol{X}_4, \boldsymbol{X}_3\right)$, is identify β_4, with the condition that the rank of the matrix, $\boldsymbol{X}_4^\top\boldsymbol{X}_4$, is complete.
- From $\mathrm{V}\left(\boldsymbol{Y}_3|\boldsymbol{X}_4, \boldsymbol{X}_3\right)$, is identify σ_3^2 and τ_3^2
- From $\mathbf{Cov}\left(\boldsymbol{Y}_4, \boldsymbol{Y}_3|\boldsymbol{X}_4, \boldsymbol{X}_3\right)$, is identify $\mu\tau_3^2$, but as τ_3^2 was identified in the last step, so μ is identified.
- From $\mathrm{V}\left(\boldsymbol{Y}_4|\boldsymbol{X}_4, \boldsymbol{X}_3\right)$, is identify σ_4^2 and $\tau_4^2 + \mu^2\tau_3^2$, as μ and τ_3^2 were identified, also τ_4 is identified.

This way all de parameters, $\beta_4, \beta_3, \tau_3^2, \tau_4^2, \sigma_3^2, \sigma_4^2$ and μ, are identified.

3.2.3 Prediction of the school effect and estimation of the value added

Following J. Gray, Hopkins, Reynolds, Wilcox, and Farrell (1999), J. Gray, Goldstein, and Jesson (2012), we characterize school improvement in terms of trajectories of school value added over time. The value added of a school at periods t and $t+1$ (denoted as $\mathrm{VA}_{t,t+1}$) corresponds to a difference of two conditional expected scores: the first one is the expected student's score (at period $t+1$) conditionally on both a set of covariates (including the prior attainment score measured at time t) and the school effect; the second one is the expected score (at period $t+1$) of the student who would be treated by an "average" school having a similar set of covariates; see Carrasco and San Martín (2012); Goldstein (8); Timmermans et al. (2011) and Manzi et al. (2014)[section 2]. The improvement of a school is consequently characterized by the trajectory defied by VA_{12}, VA_{23}, VA_{34}, and so on. Then in this dynamic model of two cohorts its value-added is,

$$
\begin{aligned}
\mathrm{VA}_{j42} =& \frac{1}{n_{j4}}\sum_{i=1}^{n_{j4}} \mathrm{E}\left(\boldsymbol{Y}_{ij4}|\boldsymbol{X}_{ij,4},\boldsymbol{X}_{ij,3},\theta_{j4,2},\theta_{j3,1}\right) \; - \; \frac{1}{n_{j4}}\sum_{i=1}^{n_{j4}} \mathrm{E}\left(\boldsymbol{Y}_{ij4}|\boldsymbol{X}_{ij4},\boldsymbol{X}_{ij3}\right) \\
=& \frac{1}{n_{j4}}\sum_{i=1}^{n_{j4}} \boldsymbol{X}_{ij4}\beta_4 + \theta_{j4,2} - \frac{1}{n_{j4}}\sum_{i=1}^{n_{j4}} \boldsymbol{X}_{ij4}\beta_4 \\
=& \theta_{j4,2}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{VA}_{j31} =& \frac{1}{n_{j3}}\sum_{i=1}^{n_{j3}} \mathrm{E}\left(\boldsymbol{Y}_{ij3}|\boldsymbol{X}_{ij,4},\boldsymbol{X}_{ij,3},\theta_{j4,2},\theta_{j3,1}\right) \; - \; \frac{1}{n_{j3}}\sum_{i=1}^{n_{j3}} \mathrm{E}\left(\boldsymbol{Y}_{ij3}|\boldsymbol{X}_{ij4},\boldsymbol{X}_{ij3}\right) \\
=& \frac{1}{n_{j3}}\sum_{i=1}^{n_{j3}} \boldsymbol{X}_{ij3}\beta_3 + \theta_{j3,1} - \frac{1}{n_{j3}}\sum_{i=1}^{n_{j3}} \boldsymbol{X}_{ij3}\beta_3 \\
=& \theta_{j3,1}
\end{aligned}
$$

Estimation of value added

The estimate of value added corresponds to the effect school estimation, ie

$$
\begin{pmatrix} \widehat{\mathrm{VA}}_{j42} \\ \widehat{\mathrm{VA}}_{j31} \end{pmatrix} = \begin{pmatrix} \widehat{\theta}_{j42} \\ \widehat{\theta}_{j31} \end{pmatrix}
$$

Then, as

- $\left(\begin{array}{c|c} \theta_{j4,2} & \\ \theta_{j3,1} & \end{array} \boldsymbol{Y}_{j4},\boldsymbol{Y}_{j3}\boldsymbol{X}_{j4},\boldsymbol{X}_{j3} \right) \sim \mathcal{N}_2\left(\alpha_j;\Lambda_j\right)$, where

- $\alpha_j := \begin{pmatrix} \phi_{11}\iota_{n_{j4}}^{\top}(\boldsymbol{Y}_{j4} - \boldsymbol{X}_{j4}\beta_4) + \phi_{12}\iota_{n_{j3}}^{\top}(\boldsymbol{Y}_{j3} - \boldsymbol{X}_{j3}\beta_3) \\ \phi_{21}\iota_{n_{j4}}^{\top}(\boldsymbol{Y}_{j4} - \boldsymbol{X}_{j4}\beta_4) + \phi_{22}\iota_{n_{j3}}^{\top}(\boldsymbol{Y}_{j3} - \boldsymbol{X}_{j3}\beta_3) \end{pmatrix}$
- $\Lambda_j := \dfrac{1}{l_2 l_4 - n_{j3}n_{j4}\mu^2\tau_3^4} \begin{pmatrix} \sigma_4^2(\tau_4^2 + \mu\tau_3^2)(l_2 - n_{j3}\mu^2\tau_3^4) & \sigma_3^2\sigma_4^2\tau_3^2\mu \\ \sigma_3^2\sigma_4^2\tau_3^2\mu & \tau_3^2\sigma_3^2\left(\sigma_4^2 + \tau_4^2 n_{j4}\right) \end{pmatrix}$

where

$$
\begin{aligned}
\phi_{11} =& (1/(l_2 l_4 - n_{j3}n_{j4}\mu^2\tau_3^4))\left\{(\tau_4^2 + \mu^2\tau_3^2)l_2 - n_{j3}\mu^2\tau_3^4)\right\} \\
\phi_{12} =& (\sigma_4^2\mu\tau_3^2)/(l_2 l_4 - n_{j3}n_{j4}\mu^2\tau_3^4) \\
\phi_{21} =& (\mu\tau_3^2\sigma_3^2)/(l_2 l_4 - n_{j3}n_{j4}\mu^2\tau_3^4) \\
\phi_{22} =& \tau_3^2(\sigma_4^2 + \tau_4^2 n_{j4})/(l_2 l_4 - n_{j3}n_{j4}\mu^2\tau_3^4) \\
l_4 =& \sigma_4^2 + (\tau_4^2 + \mu^2\tau_3^2)n_{j4} \\
l_2 =& \sigma_3^2 + \tau_3^2 n_{j3}
\end{aligned}
$$

Therefore the expression for predict the value-added in the two cohorts is,

$$
\begin{pmatrix} \widehat{\text{VA}}_{j42} \\ \widehat{\text{VA}}_{j31} \end{pmatrix} = \begin{pmatrix} \widehat{\phi_{11}}\iota_{n_{j4}}^{\top}\left(\boldsymbol{Y}_{j4} - \boldsymbol{X}_{j4}\widehat{\beta_4}\right) + \widehat{\phi_{12}}\iota_{n_{j3}}^{\top}\left(\boldsymbol{Y}_{j3} - \boldsymbol{X}_{j3}\widehat{\beta_3}\right) \\ \widehat{\phi_{21}}\iota_{n_{j4}}^{\top}\left(\boldsymbol{Y}_{j4} - \boldsymbol{X}_{j4}\widehat{\beta_4}\right) + \widehat{\phi_{22}}\iota_{n_{j3}}^{\top}\left(\boldsymbol{Y}_{j3} - \boldsymbol{X}_{j3}\widehat{\beta_3}\right) \end{pmatrix}
$$

As specified in the model, the school effects of two cohorts (therefore also the values added) are correlated between them. It is important mentioned that the estimation of the a value added θ_{jt} not only depends on information of its cohort, but also of the information of other cohort.

Now consider the case when the school effects across the time are mutually independent, that is the covariace matrix of equation (3.2) is a diagonal matrix with zero covariance elements. In that case $\mu = 0$ and the school effects are equal to the estimation obtained in a standard HLM, ie

$$
\begin{pmatrix} \widehat{\text{VA}}_{j42} \\ \widehat{\text{VA}}_{j31} \end{pmatrix} = \begin{pmatrix} \iota_{n_{j4}}^{\top}\widehat{\tau_4}^2\left(\widehat{\sigma_4}^2\boldsymbol{I}_{n_{j4}} + \widehat{\tau_4}^2\boldsymbol{J}_{n_{j4}}\right)^{-1}\left(\boldsymbol{Y}_{j4} - \boldsymbol{X}_{j4}\widehat{\beta_4}\right) \\ \iota_{n_{j3}}^{\top}\widehat{\tau_3^2}\left(\widehat{\sigma_3}^2\boldsymbol{I}_{n_{j3}} + \widehat{\tau_3}^2\boldsymbol{J}_{n_{j3}}\right)^{-1}\left(\boldsymbol{Y}_{j3} - \boldsymbol{X}_{j3}\widehat{\beta_3}\right) \end{pmatrix}
$$

Prediction Interval

Consider the following orthogonal decomposition:

$$\begin{pmatrix} \theta_{j4,2} \\ \theta_{j3,1} \end{pmatrix} = \begin{pmatrix} \widehat{\theta_{j4,2}} \\ \widehat{\theta_{j3,1}} \end{pmatrix} + \left\{ \begin{pmatrix} \theta_{j4,2} \\ \theta_{j3,1} \end{pmatrix} - \begin{pmatrix} \widehat{\theta_{j4,2}} \\ \widehat{\theta_{j3,1}} \end{pmatrix} \right\}$$

Such that, consider the estimated predictor

$$\begin{pmatrix} \widehat{\widehat{\theta_{j4,2}}} \\ \widehat{\widehat{\theta_{j3,1}}} \end{pmatrix} = \begin{pmatrix} \phi_{11}\iota_{n_{j4}}^{\top}(\boldsymbol{Y}_{j4} - \boldsymbol{X}_{j4}\widehat{\beta}_4) + \phi_{12}\iota_{n_{j3}}^{\top}(\boldsymbol{Y}_{j3} - \boldsymbol{X}_{j3}\widehat{\beta}_3) \\ \phi_{21}\iota_{n_{j4}}^{\top}(\boldsymbol{Y}_{j4} - \boldsymbol{X}_{j4}\widehat{\beta}_4) + \phi_{22}\iota_{n_{j3}}^{\top}(\boldsymbol{Y}_{j3} - \boldsymbol{X}_{j3}\widehat{\beta}_3) \end{pmatrix},$$

the parameters $\phi_{11}, \phi_{12}, \phi_{21}$ y ϕ_{22} are supposed to be known in the following reasoning.

$$\begin{pmatrix} \widehat{\theta_{j4,2}} \\ \widehat{\theta_{j3,1}} \end{pmatrix} = \boldsymbol{\Phi}_j \begin{pmatrix} \boldsymbol{Y}_{j4} \\ \boldsymbol{Y}_{j3} \end{pmatrix}$$

where $\boldsymbol{\Phi}_j = \begin{pmatrix} \phi_{11}\iota_{n_{j4}}^{\top}\boldsymbol{M}_{j4} & \phi_{12}\iota_{n_{j3}}^{\top}\boldsymbol{M}_{j3} \\ \phi_{21}\iota_{n_{j4}}^{\top}\boldsymbol{M}_{j4} & \phi_{22}\iota_{n_{j3}}^{\top}\boldsymbol{M}_{j3} \end{pmatrix}$, $\boldsymbol{M}_{j3} =$ and similarly, $\boldsymbol{M}_{j4} =$, then

$$\begin{aligned} \mathrm{V}\left\{ \begin{pmatrix} \theta_{j4,2} \\ \theta_{j3,1} \end{pmatrix} - \begin{pmatrix} \widehat{\theta_{j4,2}} \\ \widehat{\theta_{j3,1}} \end{pmatrix} \right\} =& \mathrm{V}\left\{ \begin{pmatrix} \theta_{j4,2} \\ \theta_{j3,1} \end{pmatrix} \right\} + \mathrm{V}\left\{ \begin{pmatrix} \widehat{\theta_{j4,2}} \\ \widehat{\theta_{j3,1}} \end{pmatrix} \right\} \\ & - \mathbf{C}\left\{ \begin{pmatrix} \theta_{j4,2} \\ \theta_{j3,1} \end{pmatrix} ; \begin{pmatrix} \widehat{\theta_{j4,2}} \\ \widehat{\theta_{j3,1}} \end{pmatrix} \right\} - \mathbf{C}\left\{ \begin{pmatrix} \widehat{\theta_{j4,2}} \\ \widehat{\theta_{j3,1}} \end{pmatrix} ; \begin{pmatrix} \theta_{j4,2} \\ \theta_{j3,1} \end{pmatrix} \right\}^{\top} \end{aligned}$$

such that

$$\mathrm{V}\left\{\begin{pmatrix}\theta_{j4,2}\\ \theta_{j3,1}\end{pmatrix}\right\}=\begin{pmatrix}\tau_4^2+\mu^2\tau_2^2 & \mu\tau_2^2\\ \mu\tau_2^2 & \tau_2^2\end{pmatrix}$$

$$\mathrm{V}\left\{\begin{pmatrix}\widehat{\theta_{j4,2}}\\ \widehat{\theta_{j3,1}}\end{pmatrix}\right\}=\boldsymbol{\Phi}_j\begin{pmatrix}\sigma_4^2\boldsymbol{I}_{n_{j4}}+(\tau_4^2+\mu\tau_3^2)\boldsymbol{J}_{n_{j4}} & \iota_{n_{j4}}\iota_{n_{j3}}^\top\mu\tau_3^2\\ \iota_{n_{j3}}\iota_{n_{j4}}^\top\mu\tau_3^2 & \sigma_3^2\boldsymbol{I}_{n_{j3}}+\tau_3^2\boldsymbol{J}_{n_{j3}}\end{pmatrix}\boldsymbol{\Phi}_j^\top$$

$$\mathbf{C}\left\{\begin{pmatrix}\theta_{j4,2}\\ \theta_{j3,1}\end{pmatrix};\begin{pmatrix}\widehat{\theta_{j4,2}}\\ \widehat{\theta_{j3,1}}\end{pmatrix}\right\}=\begin{pmatrix}\iota_{n_{j4}}^\top(\tau_4^2+\mu^2\tau_3^2) & \iota_{n_{j3}}^\top\mu\tau_3^2\\ \iota_{n_{j4}}^\top\mu\tau_3^2 & \iota_{n_{j3}}^\top\tau_3^2\end{pmatrix}\boldsymbol{\Phi}_j^\top$$

This way, if $\mathrm{V}_{e_j}=\mathrm{V}\left\{\begin{pmatrix}\theta_{j4,2}\\ \theta_{j3,1}\end{pmatrix}-\begin{pmatrix}\widehat{\theta_{j4,2}}\\ \widehat{\theta_{j3,1}}\end{pmatrix}\right\}$, then

$$\left\{\begin{pmatrix}\theta_{j4,2}\\ \theta_{j3,1}\end{pmatrix}-\begin{pmatrix}\widehat{\theta_{j4,2}}\\ \widehat{\theta_{j3,1}}\end{pmatrix}\right\}\sim N_2(0,\mathrm{V}_{e_j})$$

Then confidence intervals is given by,

$$\left\{\theta_{j4,2}-\widehat{\theta_{j4,2}}\right\}\sim N_1(0,l_1^\top\mathrm{V}_{e_j}l_1),\quad\text{with}\quad l_1^\top=(0\;1)$$
$$\left\{\theta_{j3,1}-\widehat{\theta_{j3,1}}\right\}\sim N_1(0,l_2^\top\mathrm{V}_{e_j}l_2),\quad\text{with}\quad l_2^\top=(1\;0)$$

Then,

$$IC\{\theta_{j4,2};1-\alpha\}=\left\{\widehat{\theta_{j4,2}}-\sqrt{l_1^\top\widehat{\mathbf{V}_{e_j}}l_1}t_{\alpha/2};\widehat{\theta_{j4,2}}+\sqrt{l_1^\top\widehat{\mathbf{V}_{e_j}}l_1}t_{\alpha/2}\right\}$$
$$IC\{\theta_{j3,1};1-\alpha\}=\left\{\widehat{\theta_{j3,1}}-\sqrt{l_2^\top\widehat{\mathbf{V}_{e_j}}l_2}t_{\alpha/2};\widehat{\theta_{j3,1}}+\sqrt{l_2^\top\widehat{\mathbf{V}_{e_j}}l_2}t_{\alpha/2}\right\}$$

3.3 Estimation Procedure

This section addressed, the estimation procedure of dynamic model of two cohorts, such that for the estimation of the parameters is used the habitual procedure of the hierarchical linear model, so, before performing the calculations of estimation is must consider the equations (3.1a–3.1d), where we can rewritten the equations in matricial form considering all the population, i.e., all the students of all schools, see appendix for matrix notation. Then,

$$\boldsymbol{Y}_3 = \boldsymbol{X}_3\boldsymbol{\beta}_3 + \boldsymbol{L}_{\underline{n_3}}\boldsymbol{\theta}_3 + \boldsymbol{\epsilon}_3 \tag{3.4a}$$

where $\boldsymbol{Y}_3$ are the outcomes scores of all student on the time 3, $\boldsymbol{X}_3$ are the covariates associated at cohort 1, $\boldsymbol{\theta}_3$ are the school effects of all schools in the cohort 1, $\boldsymbol{\epsilon}_3$ are idiosyncratic error of all student and $\boldsymbol{L}_{\underline{n_3}}$ is a block matrix, where each block is a 1's vector.

$$\boldsymbol{Y}_3 = \boldsymbol{X}_3\boldsymbol{\beta}_3 + \boldsymbol{e}_3, \quad \text{where } \boldsymbol{e}_3 = \boldsymbol{L}_{\underline{n_3}}\boldsymbol{\theta}_3 + \boldsymbol{\epsilon}_3 \tag{3.4b}$$

Also, we defined $\mathrm{V}(\boldsymbol{e}_3|\boldsymbol{X}_3, \boldsymbol{X}_4) = \boldsymbol{V}_3$,

where $\boldsymbol{V}_3 = \begin{pmatrix} \boldsymbol{V}_{n_{13}} & & \\ & \ddots & \\ & & \boldsymbol{V}_{n_{J3}} \end{pmatrix}$, such that $\boldsymbol{V}_{n_j3} = \tau_3^2\boldsymbol{J}_{n_{j3}} + \sigma_3^2\boldsymbol{I}_{n_{j3}}$,

for each $j = 1, \dots, J$. But, using the notation annexed in the appendix we can write;

$$\boldsymbol{V}_3 = \tau_3^2\boldsymbol{L}_{\underline{n_3}}\boldsymbol{L}_{\underline{n_3}}^\top + \sigma_3^2\boldsymbol{I}_{N_3} = \sigma_3^2\boldsymbol{Q}_3 + \sigma_3^2\boldsymbol{P}_3\boldsymbol{D}_3\boldsymbol{P}_3 + \tau_3^2\boldsymbol{P}_3\boldsymbol{D}_3^2\boldsymbol{P}_3.$$

Analogously,

$$\boldsymbol{Y}_4 = \boldsymbol{X}_4\boldsymbol{\beta}_4 + \boldsymbol{L}_{\underline{n_4}}\boldsymbol{\theta}_4 + \boldsymbol{\epsilon}_4 \tag{3.5a}$$

such that $\boldsymbol{Y}_4$ are the outcomes scores of all student on the time 4, $\boldsymbol{X}_4$ are the covariates associated at cohort 2, $\boldsymbol{\theta}_4$ are the school effects of all schools in the cohort 2, $\boldsymbol{\epsilon}_4$ are idiosyncratic error of all student and $\boldsymbol{L}_{\underline{n_4}}$ is a block matrix, where each block is a 1's vector.

$$\boldsymbol{Y}_4 = \boldsymbol{X}_4\boldsymbol{\beta}_4 + \boldsymbol{e}_4 \quad \text{Where } \boldsymbol{e}_4 = \boldsymbol{L}_{\underline{n_4}}\boldsymbol{\theta}_4 + \boldsymbol{\epsilon}_4 \tag{3.5b}$$

Also, we defined $\mathrm{V}(\boldsymbol{e}_4|\boldsymbol{X}_3, \boldsymbol{X}_4) = \boldsymbol{V}_4$, where $\boldsymbol{V}_4 = \begin{pmatrix} \boldsymbol{V}_{n_{14}} & & \\ & \ddots & \\ & & \boldsymbol{V}_{n_{J4}} \end{pmatrix}$, such that $\boldsymbol{V}_{n_j4} = (\tau_4^2 + \mu\tau_3^2)\,\boldsymbol{J}_{n_{j4}} + \sigma_4^2 \boldsymbol{I}_{n_{j4}}$,

for each $j = 1, \ldots, J$. But, using the notation annexed in the appendix we can write;

$$\boldsymbol{V}_4 = \tau_4^2 \boldsymbol{L}_{\underline{n_4}} \boldsymbol{L}_{\underline{n_4}}^\top + \sigma_4^2 \boldsymbol{I}_{N_4} = \sigma_4^2 \boldsymbol{Q}_4 + \sigma_4^2 \boldsymbol{P}_4 \boldsymbol{D}_4 \boldsymbol{P}_4 + \left(\tau_4^2 + \mu\tau_3^2\right) \boldsymbol{P}_4 \boldsymbol{D}_4^2 \boldsymbol{P}_4$$

.

3.3.1 Fixed Effects Estimators

The methodology used for the parameters estimation corresponds to the same applied to a standard model and que are used several results demonstrated in Chapter 1.

Estimation of $\boldsymbol{\beta}_3$ and $\boldsymbol{\beta}_4$

a) **Ordinal least squared (OLS)** Using the equation (3.4b) and ordinary least squares (OLS) for the estimation of $\boldsymbol{\beta}_3$ and the equation (3.5b) for $\boldsymbol{\beta}_4$,

$$\hat{\boldsymbol{\beta}}_{3OLS} = \left(\boldsymbol{X}_3^\top \boldsymbol{X}_3\right)^{-1} \left(\boldsymbol{X}_3^\top \boldsymbol{Y}_3\right) \quad \text{and} \quad \hat{\boldsymbol{\beta}}_{4OLS} = \left(\boldsymbol{X}_4^\top \boldsymbol{X}_4\right)^{-1} \left(\boldsymbol{X}_4^\top \boldsymbol{Y}_4\right) \tag{3.6a}$$

Assuming; i)E$(\boldsymbol{X}_3^\top \boldsymbol{e}_3) = 0$, ii) rank$(\boldsymbol{X}_3^\top \boldsymbol{X}_3) = K_3$ is of completed rank iii) E$(\boldsymbol{X}_4^\top \boldsymbol{e}_4) = 0$ and iv) rank$(\boldsymbol{X}_4^\top \boldsymbol{X}_4) = K_4$ is of completed rank.

These estimators are consistent and their asymptotic distribution are,

$$\begin{aligned} \sqrt{N_3}\left(\widehat{\boldsymbol{\beta}}_{3OLS} - \boldsymbol{\beta}_3\right) &\overset{a}{\sim} \mathrm{N}\left(\mathbf{0}\,;\, A_3^{-1}\right), \quad \text{with} \quad A_3 = \left(\frac{1}{N_3}\boldsymbol{X}_3^\top \boldsymbol{X}_3\right) + op(1) \\ \sqrt{N_4}\left(\widehat{\boldsymbol{\beta}}_{4OLS} - \boldsymbol{\beta}_4\right) &\overset{a}{\sim} \mathrm{N}\left(\mathbf{0}\,;\, A_4^{-1}\right), \quad \text{with} \quad A_4 = \left(\frac{1}{N_4}\boldsymbol{X}_4^\top \boldsymbol{X}_4\right) + op(1) \end{aligned}$$

b) **Generalized least squared (GLS)**

$$\widehat{\boldsymbol{\beta}}_{3GLS} = (\boldsymbol{X}_3^\top \boldsymbol{V}_3^{-1} \boldsymbol{X}_3)^{-1} \boldsymbol{X}_3^\top \boldsymbol{V}_3^{-1} \boldsymbol{Y}_3 \quad \text{and} \quad \widehat{\boldsymbol{\beta}}_{4GLS} = (\boldsymbol{X}_4^\top \boldsymbol{V}_4^{-1} \boldsymbol{X}_4)^{-1} \boldsymbol{X}_4^\top \boldsymbol{V}_4^{-1} \boldsymbol{Y}_4 \tag{3.6b}$$

Assuming; i) $\mathrm{E}(\boldsymbol{X}_3^\top \boldsymbol{e}_3) = 0$, ii) $\boldsymbol{V}_3 \equiv \mathrm{E}(\boldsymbol{e}_3^\top \boldsymbol{e}_3)$ is a matrix positive definite and $\mathrm{E}(\boldsymbol{X}_3^\top \boldsymbol{V}_3^{-1} \boldsymbol{X}_3)$ is not singular iii) $\mathrm{E}(\boldsymbol{X}_4^\top \boldsymbol{e}_4) = 0$ and iv) $\boldsymbol{V}_4 \equiv \mathrm{E}(\boldsymbol{e}_4^\top \boldsymbol{e}_4)$ is a matrix positive definite and $\mathrm{E}(\boldsymbol{X}_4^\top \boldsymbol{V}_4^{-1} \boldsymbol{X}_4)$ is not singular.

These estimators are consistent and their asymptotic distribution are,

$$\sqrt{N_3}\left(\widehat{\boldsymbol{\beta}}_{3GLS} - \boldsymbol{\beta}_3\right) \overset{a}{\sim} \mathrm{N}\left(\mathbf{0}\,;\, A_3^{-1}\right), \quad \text{with} \quad A_3 = \left(\frac{1}{N_3} \boldsymbol{X}_3^\top \boldsymbol{V}_3^{-1} \boldsymbol{X}_3\right) + op(1)$$
$$\sqrt{N_4}\left(\widehat{\boldsymbol{\beta}}_{4GLS} - \boldsymbol{\beta}_4\right) \overset{a}{\sim} \mathrm{N}\left(\mathbf{0}\,;\, A_4^{-1}\right), \quad \text{with} \quad A_4 = \left(\frac{1}{N_4} \boldsymbol{X}_4^\top \boldsymbol{V}_4^{-1} \boldsymbol{X}_4\right) + op(1)$$

c) **Feasible Generalized least squared (FGLS)**

$$\widehat{\boldsymbol{\beta}}_{3FGLS} = (\boldsymbol{X}_3^\top \widehat{\boldsymbol{V}}_3^{-1} \boldsymbol{X}_3)^{-1} \boldsymbol{X}_3^\top \widehat{\boldsymbol{V}}_3^{-1} \boldsymbol{Y}_3 \quad \text{and} \quad \widehat{\boldsymbol{\beta}}_{4FGLS} = (\boldsymbol{X}_4^\top \widehat{\boldsymbol{V}}_4^{-1} \boldsymbol{X}_4)^{-1} \boldsymbol{X}_4^\top \widehat{\boldsymbol{V}}_4^{-1} \boldsymbol{Y}_4 \tag{3.6c}$$

Assuming; i) $\mathrm{E}(\boldsymbol{X}_3^\top \boldsymbol{e}_3) = 0$, ii) $\boldsymbol{V}_3 \equiv \mathrm{E}(\boldsymbol{e}_3^\top \boldsymbol{e}_3)$ a matrix positive definite and $\mathrm{E}(\boldsymbol{X}_3^\top \boldsymbol{V}_3 \boldsymbol{X}_3)$ is not singular iii) $\widehat{\boldsymbol{V}}_3 \xrightarrow{P} \boldsymbol{V}_3$, iv) $\mathrm{E}(\boldsymbol{X}_4^\top \boldsymbol{e}_4) = 0$, v) $\boldsymbol{V}_4 \equiv \mathrm{E}(\boldsymbol{e}_4^\top \boldsymbol{e}_4)$ a matrix positive definite and $\mathrm{E}(\boldsymbol{X}_4^\top \boldsymbol{V}_4 \boldsymbol{X}_4)$ is not singular and vi) $\widehat{\boldsymbol{V}}_4 \xrightarrow{P} \boldsymbol{V}_4$.

These estimators are consistent and their asymptotic distribution are,

$$\sqrt{N_3}\left(\widehat{\boldsymbol{\beta}}_{3FGLS} - \boldsymbol{\beta}_3\right) \overset{a}{\sim} \mathrm{N}\left(\mathbf{0}\,;\, A_3^{-1}\right), \quad \text{with} \quad A_3 = \left(\frac{1}{N_3} \boldsymbol{X}_3^\top \boldsymbol{V}_3^{-1} \boldsymbol{X}_3\right) + op(1)$$
$$\sqrt{N_4}\left(\widehat{\boldsymbol{\beta}}_{4FGLS} - \boldsymbol{\beta}_4\right) \overset{a}{\sim} \mathrm{N}\left(\mathbf{0}\,;\, A_4^{-1}\right), \quad \text{with} \quad A_4 = \left(\frac{1}{N_4} \boldsymbol{X}_4^\top \boldsymbol{V}_4^{-1} \boldsymbol{X}_4\right) + op(1)$$

Usually is used as asymptotic variance the following estimators; $\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\beta}}_{3FGLS})} = \widehat{A_3}^{-1}/N_3$ and $\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\beta}}_{4FGLS})} = \widehat{A_4}^{-1}/N_4$, however with heteroskedasticity these are not robust, thus robust estimators of the asymptotic variance are;

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\beta}}_{3FGLS})} = \left(\boldsymbol{X}_3^\top \widehat{\boldsymbol{V}_3}^{-1} \boldsymbol{X}_3\right)^{-1} \left(\boldsymbol{X}_3^\top \widehat{\boldsymbol{V}_3}^{-1} \widehat{\boldsymbol{e}_3} \widehat{\boldsymbol{e}_3}^\top \widehat{\boldsymbol{V}_3}^{-1} \boldsymbol{X}_3\right) \left(\boldsymbol{X}_3^\top \widehat{\boldsymbol{V}_3}^{-1} \boldsymbol{X}_3\right)^{-1}$$

and

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\beta}}_{4FGLS})} = \left(\boldsymbol{X}_4^\top \widehat{\boldsymbol{V}_4}^{-1} \boldsymbol{X}_4\right)^{-1} \left(\boldsymbol{X}_4^\top \widehat{\boldsymbol{V}_4}^{-1} \widehat{\boldsymbol{e}_4} \widehat{\boldsymbol{e}_4}^\top \widehat{\boldsymbol{V}_4}^{-1} \boldsymbol{X}_4\right) \left(\boldsymbol{X}_4^\top \widehat{\boldsymbol{V}_4}^{-1} \boldsymbol{X}_4\right)^{-1}$$

3.3.2 Estimators of variance components

The estimation of variance components is through the method of moments. Then,

Estimation of σ_3^2, and σ_4^2

Apply the W-operator to in the equations (3.4b) and (3.5b), we obtain the following (see Appendix C.1.1 for matricial notation).

The within regressions of both cohorts are formulated of the following way,

$$\begin{aligned} \boldsymbol{Q}_4\boldsymbol{Y}_3 &= \boldsymbol{Q}_3\boldsymbol{X}_3\boldsymbol{\beta}_3 + \boldsymbol{Q}_3\boldsymbol{e}_3 \\ \boldsymbol{Q}_4\boldsymbol{Y}_4 &= \boldsymbol{Q}_4\boldsymbol{X}_4\boldsymbol{\beta}_4 + \boldsymbol{Q}_4\boldsymbol{e}_4 \end{aligned}$$

where,

$$\begin{array}{ll} \text{(a.1)}\ \ \mathrm{V}\left(\boldsymbol{Q}_3\boldsymbol{e}_3|\boldsymbol{X}_4,\boldsymbol{X}_3\right) = \boldsymbol{Q}_3\boldsymbol{V}_3\boldsymbol{Q}_3^\top = \sigma_3^2\boldsymbol{Q}_3 & \text{(a.2)}\ \ \widehat{\boldsymbol{\beta}_3}^w = (\boldsymbol{X}_3^\top\boldsymbol{Q}_3\boldsymbol{X}_3)^{-1}\boldsymbol{X}_3^\top\boldsymbol{Q}_3\boldsymbol{Y}_3 \\ \text{(a.3)}\ \ \mathrm{V}\left(\boldsymbol{Q}_4\boldsymbol{e}_4|\boldsymbol{X}_4,\boldsymbol{X}_3\right) = \boldsymbol{Q}_4\boldsymbol{V}_4\boldsymbol{Q}_4^\top = \sigma_4^2\boldsymbol{Q}_4 & \text{(a.4)}\ \ \widehat{\boldsymbol{\beta}_4}^w = (\boldsymbol{X}_4^\top\boldsymbol{Q}_4\boldsymbol{X}_4)^{-1}\boldsymbol{X}_4^\top\boldsymbol{Q}_4\boldsymbol{Y}_4 \end{array}$$

Now, if we are defined $\widehat{\boldsymbol{e}}_3^w = (\boldsymbol{Y}_3 - \boldsymbol{X}_3\widehat{\boldsymbol{\beta}_3}^w)$ and $\widehat{\boldsymbol{e}}_4^w = (\boldsymbol{Y}_4 - \boldsymbol{X}_4\widehat{\boldsymbol{\beta}_4}^w)$ then

$$\begin{aligned} \boldsymbol{Q}_3\widehat{\boldsymbol{e}}_3^w &= \boldsymbol{Q}_3(\boldsymbol{Y}_3 - \boldsymbol{X}_3\widehat{\boldsymbol{\beta}_3}^w) \qquad & \boldsymbol{Q}_4\widehat{\boldsymbol{e}}_4^w &= \boldsymbol{Q}_4(\boldsymbol{Y}_4 - \boldsymbol{X}_4\widehat{\boldsymbol{\beta}_4}^w) \\ &= \mathrm{M}_3\boldsymbol{Y}_3 & &= \mathrm{M}_4\boldsymbol{Y}_4 \end{aligned}$$

where $\mathrm{M}_3 = (\boldsymbol{Q}_3 - \boldsymbol{Q}_3\boldsymbol{X}_3(\boldsymbol{X}_3^\top\boldsymbol{Q}_3\boldsymbol{X}_3)^{-1}\boldsymbol{X}_3^\top\boldsymbol{Q}_3)$ and

$\mathrm{M}_4 = (\boldsymbol{Q}_4 - \boldsymbol{Q}_4\boldsymbol{X}_4(\boldsymbol{X}_4^\top\boldsymbol{Q}_4\boldsymbol{X}_4)^{-1}\boldsymbol{X}_4^\top\boldsymbol{Q}_4)$,

such that $\mathrm{M}_3\mathrm{M}_3 = \mathrm{M}_3$, $\mathrm{M}_3^\top\mathrm{M}_3 = \mathrm{M}_3$,
$\mathrm{M}_4\mathrm{M}_4 = \mathrm{M}_4$, $\mathrm{M}_4^\top\mathrm{M}_4 = \mathrm{M}_4$.

Then, it is easy to check that $\widehat{\boldsymbol{e}}_3^{w\top}\boldsymbol{Q}_3\widehat{\boldsymbol{e}}^w = \boldsymbol{e}_3^\top\mathrm{M}_3\boldsymbol{e}_3$ and $\widehat{\boldsymbol{e}}_4^{w\top}\boldsymbol{Q}_4\widehat{\boldsymbol{e}}_4^w = \boldsymbol{e}_4^\top\mathrm{M}_4\boldsymbol{e}_4$. This way,

$$
\begin{aligned}
\text{(b.1)}\ \ \mathrm{E}\left(\widehat{\boldsymbol{e}_3}^{\top}\boldsymbol{M}_3\widehat{\boldsymbol{e}_3}\right) &= \operatorname{tr}\left(\boldsymbol{M}3\mathrm{V}(\boldsymbol{e}_3)\right) \\
&= \sigma_3^2\left(N_3 - J\right) - \sigma_3^2\left(K_3^*\right)
\end{aligned}
$$

$$
\begin{aligned}
\text{(b.2)}\ \ \mathrm{E}\left(\widehat{\boldsymbol{e}_4}^{\top}\boldsymbol{M}_4\widehat{\boldsymbol{e}_4}\right) &= \operatorname{tr}\left(\boldsymbol{M}4\mathrm{V}(\boldsymbol{e}_4)\right) \\
&= \sigma_4^2\left(N_4 - J\right) - \sigma_4^2\left(K_4^*\right)
\end{aligned}
$$

Thus be have,

$$
\widehat{\sigma_3^2} = \frac{\left(\boldsymbol{Y}_3 - \boldsymbol{X}_3\widehat{\boldsymbol{\beta}_3}^{\boldsymbol{Q}_3}\right)^{\top}\boldsymbol{Q}_3\left(\boldsymbol{Y}_3 - \boldsymbol{X}_3\widehat{\boldsymbol{\beta}_3}^{\boldsymbol{Q}_3}\right)}{N_3 - J - K_3^*} \tag{3.7}
$$

$$
\widehat{\sigma_4^2} = \frac{\left(\boldsymbol{Y}_3 - \boldsymbol{X}_3\widehat{\boldsymbol{\beta}_3}^{\boldsymbol{Q}_3}\right)^{\top}\boldsymbol{Q}_3\left(\boldsymbol{Y}_3 - \boldsymbol{X}_3\widehat{\boldsymbol{\beta}_3}^{\boldsymbol{Q}_3}\right)}{N_3 - J - K_3^*} \tag{3.8}
$$

where $K_4^* \leq K_4$ $K_3^* \leq K_3$ are the number of covariates non-zero in the within regression of second and first cohort respectively.

Estimation of τ_3^2, τ_4^2

Apply the Between-operator to in the equations (3.4b) and (3.5b), we the following regression, The within regressions of both cohorts are formulated of the following way,

$$
\begin{aligned}
\boldsymbol{P}_4\boldsymbol{Y}_3 &= \boldsymbol{P}_3\boldsymbol{X}_3\boldsymbol{\beta}_3 + \boldsymbol{P}_3\boldsymbol{e}_3 \\
\boldsymbol{P}_4\boldsymbol{Y}_4 &= \boldsymbol{Q}_4\boldsymbol{X}_4\boldsymbol{\beta}_4 + \boldsymbol{P}_4\boldsymbol{e}_4
\end{aligned}
$$

where,

(c.1) $\mathrm{V}\left(\boldsymbol{P}_3\boldsymbol{e}_3|\boldsymbol{X}_4,\boldsymbol{X}_3\right) = \boldsymbol{P}_3\boldsymbol{V}_3\boldsymbol{P}_3^{\top} = \sigma_3^2\boldsymbol{P}_3\boldsymbol{D}_3\boldsymbol{P}_3^{\top} + \tau_3^2\boldsymbol{P}_3\boldsymbol{D}_3^2\boldsymbol{P}_3^{\top}$,

(c.2) $\widehat{\boldsymbol{\beta}_3}^{b} = (\boldsymbol{X}_3^{\top}\boldsymbol{P}_3^{\top}\boldsymbol{P}_3\boldsymbol{X}_3)^{-1}\boldsymbol{X}_3^{\top}\boldsymbol{P}_3^{\top}\boldsymbol{P}_3\boldsymbol{Y}_3$,

(c.3) $\mathrm{V}\left(\boldsymbol{P}_4\boldsymbol{e}_4|\boldsymbol{X}_4,\boldsymbol{X}_3\right) = \boldsymbol{P}_4\boldsymbol{V}_4\boldsymbol{P}_4^{\top} = \sigma_4^2\boldsymbol{P}_4\boldsymbol{D}_4\boldsymbol{P}_3^{\top} + (\tau_4^2 + \mu\tau_3^2)\boldsymbol{P}_4\boldsymbol{D}_4^2\boldsymbol{P}_4^{\top}$,

(c.4) $\widehat{\boldsymbol{\beta}_4}^{b} = (\boldsymbol{X}_4^{\top}\boldsymbol{P}_4^{\top}\boldsymbol{P}_4\boldsymbol{X}_4)^{-1}\boldsymbol{X}_4^{\top}\boldsymbol{P}_4^{\top}\boldsymbol{P}_4\boldsymbol{Y}_4$.

Note that $\boldsymbol{P}_3\boldsymbol{D}_3^2\boldsymbol{P}_3^{\top} = \boldsymbol{P}_4\boldsymbol{D}_4^2\boldsymbol{P}_4^{\top} = \boldsymbol{I}_J$, $\boldsymbol{P}_3\boldsymbol{D}_3\boldsymbol{P}_3^{\top} = \boldsymbol{D}_3^{-1}$ and $\boldsymbol{P}_4\boldsymbol{D}_4\boldsymbol{P}_4^{\top} = \boldsymbol{D}_4^{-1}$.

Now, we are defined

$$\widehat{\boldsymbol{e}}_3^b = (\boldsymbol{Y}_3 - \boldsymbol{X}_3\widehat{\boldsymbol{\beta}_3}^b), \quad \text{and} \quad \widehat{\boldsymbol{e}}_4^b = (\boldsymbol{Y}_4 - \boldsymbol{X}_4\widehat{\boldsymbol{\beta}_4}^b)$$

therefore, $\boldsymbol{P}_3\widehat{\boldsymbol{e}}_3^b = \boldsymbol{P}_3(\boldsymbol{Y}_3 - \boldsymbol{X}_3\widehat{\boldsymbol{\beta}_3}^b) = \boldsymbol{T}_3\boldsymbol{P}_3(\boldsymbol{L}_{\underline{n_3}}\boldsymbol{\theta}_3 + \boldsymbol{\epsilon}_3)$, and
$\boldsymbol{P}_4\widehat{\boldsymbol{e}}_4^b = \boldsymbol{P}_4(\boldsymbol{Y}_4 - \boldsymbol{X}_4\widehat{\boldsymbol{\beta}_4}^b) = \boldsymbol{T}_4\boldsymbol{P}_4(\boldsymbol{L}_{\underline{n_4}}\boldsymbol{\theta}_4 + \boldsymbol{\epsilon}_4)$

where $\boldsymbol{T}_3 = (\boldsymbol{I}_J - \boldsymbol{P}_3\boldsymbol{X}_3(\boldsymbol{X}_3^\top\boldsymbol{P}_3^\top\boldsymbol{P}_3\boldsymbol{X}_3)^{-1}\boldsymbol{X}_3^\top\boldsymbol{P}_3^\top)$, and
$\boldsymbol{T}_4 = (\boldsymbol{I}_J - \boldsymbol{P}_4\boldsymbol{X}_4(\boldsymbol{X}_4^\top\boldsymbol{P}_4^\top\boldsymbol{P}_4\boldsymbol{X}_4)^{-1}\boldsymbol{X}_4^\top\boldsymbol{P}_4^\top)$,

such that $\boldsymbol{T}_3\boldsymbol{T}_3 = \boldsymbol{T}_3$, $\boldsymbol{T}_3^\top\boldsymbol{T}_3 = \boldsymbol{T}_3$.
$\boldsymbol{T}_4\boldsymbol{T}_4 = \boldsymbol{T}_4$ and $\boldsymbol{T}_4^\top\boldsymbol{T}_4 = \boldsymbol{T}_4$.

Then, as $\widehat{\boldsymbol{e}_3}^{b\top}\boldsymbol{P}_3^\top\boldsymbol{P}_3\widehat{\boldsymbol{e}}_3^b = (\boldsymbol{L}_{\underline{n_3}}\boldsymbol{\theta}_3 + \boldsymbol{\epsilon}_3)^\top\boldsymbol{P}_3^\top\boldsymbol{T}_3\boldsymbol{P}_3(\boldsymbol{L}_{\underline{n_3}}\boldsymbol{\theta}_3 + \boldsymbol{\epsilon}_3)$, and
$\widehat{\boldsymbol{e}_4}^{b\top}\boldsymbol{P}_4^\top\boldsymbol{P}_4\widehat{\boldsymbol{e}}_4^b = (\boldsymbol{L}_{\underline{n_4}}\boldsymbol{\theta}_4 + \boldsymbol{\epsilon}_4)^\top\boldsymbol{P}_4^\top\boldsymbol{T}_4\boldsymbol{P}_4(\boldsymbol{L}_{\underline{n_4}}\boldsymbol{\theta}_4 + \boldsymbol{\epsilon}_4)$.

This way,

$$\begin{aligned}
\text{(d.1)}\ \ \mathrm{E}\left(\widehat{\boldsymbol{e}_3}^{b\top}\boldsymbol{T}_3\widehat{\boldsymbol{e}_3^b}\right) &= \tau_3^2\mathrm{tr}\,(\boldsymbol{T}_3) + \sigma_3^2\mathrm{tr}\left(\boldsymbol{T}_3\boldsymbol{D}_3^{-1}\right)\\
&= \tau_3^2(J - K_3^{**}) + \sigma_3^2\left(\textstyle\sum_{j=1}^J(1/n_{j3}) - K_3^{***}\right)
\end{aligned}$$

$$\begin{aligned}
\text{(d.2)}\ \ \mathrm{E}\left(\widehat{\boldsymbol{e}_4}^{b\top}\boldsymbol{T}_4\widehat{\boldsymbol{e}_4^b}\right) &= \tau_4^2\mathrm{tr}\,(\boldsymbol{T}_4) + \sigma_4^2\mathrm{tr}\left(\boldsymbol{T}_4\boldsymbol{D}_4^{-1}\right)\\
&= (\tau_4^2 + \mu\tau_3^2)(J - K_4^{**}) + \sigma_4^2\left(\textstyle\sum_{j=1}^J(1/n_{j4}) - K_4^{***}\right)
\end{aligned}$$

Thus be have,

$$\widehat{\tau_3^2} = \frac{\left(\boldsymbol{Y}_3 - \boldsymbol{X}_3\widehat{\boldsymbol{\beta}_3}^b\right)^\top\boldsymbol{P}_3^\top\boldsymbol{P}_3\left(\boldsymbol{Y}_3 - \boldsymbol{X}_3\widehat{\boldsymbol{\beta}_3}^b\right) - \widehat{\sigma_3^2}\left(\sum_{j=1}^J(1/n_{j3}) - K_3^{***}\right)}{J - K_3^{**}} \tag{3.9}$$

$$\widehat{\tau_4^2} = \frac{\left(\boldsymbol{Y}_4 - \boldsymbol{X}_4\widehat{\boldsymbol{\beta}_4}^b\right)^\top\boldsymbol{P}_4^\top\boldsymbol{P}_4\left(\boldsymbol{Y}_4 - \boldsymbol{X}_4\widehat{\boldsymbol{\beta}_4}^b\right) - \sigma_4^2\left(\sum_{j=1}^J(1/n_{j3}) - K_4^{***}\right)}{J - K_4^{**}} - \mu^2\tau_3^2 \tag{3.10}$$

where, $K_3^{***} = \mathrm{tr}\left\{(\boldsymbol{X}_3^\top\boldsymbol{P}_3^\top\boldsymbol{P}_3\boldsymbol{P}_3)^{-1}\boldsymbol{X}_3^\top\boldsymbol{P}_3^\top\boldsymbol{D}_3^{-1}\boldsymbol{P}_3\boldsymbol{X}_3\right\}$,
$K_3^{**} = \mathrm{tr}\left\{(\boldsymbol{X}_3^\top\boldsymbol{P}_3^\top\boldsymbol{P}_3\boldsymbol{P}_3)^{-1}(\boldsymbol{X}_3^\top\boldsymbol{P}_3^\top\boldsymbol{P}_3\boldsymbol{X}_3)\right\}$,
$K_4^{***} = \mathrm{tr}\left\{(\boldsymbol{X}_4^\top\boldsymbol{P}_4^\top\boldsymbol{P}_4\boldsymbol{X}_4)^{-1}\boldsymbol{X}_4^\top\boldsymbol{P}_4^\top\boldsymbol{D}_4^{-1}\boldsymbol{P}_4\boldsymbol{X}_4\right\}$,
$K_4^{**} = \mathrm{tr}\left\{(\boldsymbol{X}_4^\top\boldsymbol{P}_4^\top\boldsymbol{P}_4\boldsymbol{X}_4)^{-1}(\boldsymbol{X}_4^\top\boldsymbol{P}_4^\top\boldsymbol{P}_4\boldsymbol{X}_4)\right\}$

Estimation of μ

Using the between-residues seen in the previous section

$$
\begin{aligned}
\widehat{\boldsymbol{e}}_4^b &= (\boldsymbol{Y}_4 - \boldsymbol{X}_4\widehat{\boldsymbol{\beta}}_4^b) \\
\widehat{\boldsymbol{e}}_3^b &= (\boldsymbol{Y}_3 - \boldsymbol{X}_3\widehat{\boldsymbol{\beta}}_3^b)
\end{aligned}
$$

we obtain

$$
\mathrm{E}\left(\widehat{\boldsymbol{e}}_4^{b\top}\boldsymbol{P}_4^\top\boldsymbol{P}_3\widehat{\boldsymbol{e}}_3^b\right) = \mu\tau_3^2\mathrm{tr}\left\{\boldsymbol{T}_4^\top\boldsymbol{T}_3\right\}
$$

This way,

$$
\mu = \frac{\widehat{\boldsymbol{e}_4^b}^\top\boldsymbol{P}_4^\top\boldsymbol{P}_3\widehat{\boldsymbol{e}_3^b}}{\tau_3^2\left(J - m^* - m^{**} + m^{***}\right)} \tag{3.11}
$$

where

$$
\begin{aligned}
m^* &= \mathrm{r}\left\{\boldsymbol{X}_4^\top\boldsymbol{P}_4^\top\boldsymbol{P}_4\boldsymbol{X}_4\right\} \\
m^{**} &= \mathrm{r}\left\{\boldsymbol{X}_3^\top\boldsymbol{P}_3^\top\boldsymbol{P}_3\boldsymbol{X}_3\right\} \\
m^{***} &= \mathrm{tr}\left\{\boldsymbol{P}_4\boldsymbol{X}_4(\boldsymbol{X}_4^\top\boldsymbol{P}_4^\top\boldsymbol{P}_4\boldsymbol{X}_4)^{-1}\boldsymbol{X}_4^\top\boldsymbol{P}_4^\top\boldsymbol{P}_3\boldsymbol{X}_3(\boldsymbol{X}_3^\top\boldsymbol{P}_3^\top\boldsymbol{P}_3\boldsymbol{X}_3)^{-1}\boldsymbol{X}_3^\top\boldsymbol{P}_3^\top\right\}
\end{aligned}
$$

3.3.3 Summary of the Estimation Process

1. The process starts using the model's equations (3.4b)–(3.5b), such that $\boldsymbol{\beta}_3$ and $\boldsymbol{\beta}_4$ are estimated through of equations (3.6a) and (3.6b) respectively.

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{3OLS} &= \left(\boldsymbol{X}_3^\top \boldsymbol{X}_3\right)^{-1}\left(\boldsymbol{X}_3^\top \boldsymbol{Y}_3\right)\\ \hat{\boldsymbol{\beta}}_{4OLS} &= \left(\boldsymbol{X}_4^\top \boldsymbol{X}_4\right)^{-1}\left(\boldsymbol{X}_4^\top \boldsymbol{Y}_4\right)\end{aligned}$$

2. Completed steps 1, be must estimate the Within Residual of the regressions (3.4b)–(3.5b). Then, they are estimated σ_3^2 and σ_4^2 by the equation (3.7) and (3.8) respectively.

$$\widehat{\sigma_3^2} = \frac{\left(\boldsymbol{Y}_3 - \boldsymbol{X}_3\widehat{\boldsymbol{\beta}_3}^{\boldsymbol{Q}_3}\right)^\top \boldsymbol{Q}_3\left(\boldsymbol{Y}_3 - \boldsymbol{X}_3\widehat{\boldsymbol{\beta}_3}^{\boldsymbol{Q}_3}\right)}{N_3 - J - K_3^*}$$

$$\widehat{\sigma_4^2} = \frac{\left(\boldsymbol{Y}_3 - \boldsymbol{X}_3\widehat{\boldsymbol{\beta}_3}^{\boldsymbol{Q}_3}\right)^\top \boldsymbol{Q}_3\left(\boldsymbol{Y}_3 - \boldsymbol{X}_3\widehat{\boldsymbol{\beta}_3}^{\boldsymbol{Q}_3}\right)}{N_3 - J - K_3^*}$$

3. Completed steps 1–2, be must estimate the Between Residual of the regressions (3.4b). Then, $\widehat{\tau_3^2}$ is estimated by the equation (3.9),

$$\widehat{\tau_3^2} = \frac{\left(\boldsymbol{Y}_3 - \boldsymbol{X}_3\widehat{\boldsymbol{\beta}_3}^{b}\right)^\top \boldsymbol{P}_3^\top \boldsymbol{P}_3\left(\boldsymbol{Y}_3 - \boldsymbol{X}_3\widehat{\boldsymbol{\beta}_3}^{b}\right) - \widehat{\sigma_3^2}\left(\sum_{j=1}^{J}(1/n_{j3}) - K_3^{***}\right)}{J - K_3^{**}}$$

4. Completed steps 1–3, be must estimate the Between Residual of the regressions (3.4b) and (3.5b). Then, $\widehat{\mu}$ is estimated by the equation (3.11),

$$\widehat{\mu} = \frac{\widehat{\boldsymbol{e}_4^b}^\top \boldsymbol{P}_4^\top \boldsymbol{P}_3 \widehat{\boldsymbol{e}_3^b}}{\widehat{\tau_3}^2\left(J - m^* - m^{**} + m^{***}\right)}$$

5. Completed steps 1–4, be must estimate the Between Residual of the regression (3.5b). Then, and $\widehat{\tau_4^2}$ is estimated by the equation (3.10),

$$\widehat{\tau_4^2} = \frac{\left(\boldsymbol{Y}_4 - \boldsymbol{X}_4\widehat{\boldsymbol{\beta}_4}^b\right)^\top \boldsymbol{P}_4^\top \boldsymbol{P}_4 \left(\boldsymbol{Y}_4 - \boldsymbol{X}_4\widehat{\boldsymbol{\beta}_4}^b\right) - \sigma_4^2 \left(\sum_{j=1}^J (1/n_{j3}) - K_4^{***}\right)}{J - K_4^{**}} - \mu^2\tau_3^2$$

6. Completed the steps 1–5, it is possible calculate the estimation of $\widehat{\mathbf{V}}_3$ and $\widehat{\mathbf{V}}_4$.

$$\begin{aligned}
\widehat{\boldsymbol{V}_3} &= \boldsymbol{Q}_3\widehat{\sigma_3^2} + \boldsymbol{P}_3\boldsymbol{D}_3^2\boldsymbol{P}_3^\top\widehat{\tau_3^2} + \boldsymbol{P}_3\boldsymbol{D}_3\boldsymbol{P}_3^\top\widehat{\sigma_3^2} \\
\widehat{\boldsymbol{V}_4} &= \boldsymbol{Q}_4\widehat{\sigma_4^2} + \boldsymbol{P}_4\boldsymbol{D}_4^2\boldsymbol{P}_4^\top(\widehat{\tau_4^2} + \widehat{\mu}\widehat{\tau_3^2}) + \boldsymbol{P}_4\boldsymbol{D}_4\boldsymbol{P}_4^\top\widehat{\sigma_4^2}
\end{aligned}$$

7. Using the results of step 6, the estimation of $\boldsymbol{\beta}_3$ and $\boldsymbol{\beta}_4$ are recalculated by equation (3.6c), i.e. by feasible generalized least squared,

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{3FGLS} &= (\boldsymbol{X}_3^\top\widehat{\boldsymbol{V}}_3^{-1}\boldsymbol{X}_3)^{-1}\boldsymbol{X}_3^\top\widehat{\boldsymbol{V}}_3^{-1}\boldsymbol{Y}_3 \\
\widehat{\boldsymbol{\beta}}_{4FGLS} &= (\boldsymbol{X}_4^\top\widehat{\boldsymbol{V}}_4^{-1}\boldsymbol{X}_4)^{-1}\boldsymbol{X}_4^\top\widehat{\boldsymbol{V}}_4^{-1}\boldsymbol{Y}_4
\end{aligned}$$

3.4 Aplication of HLM across the time to educational data

3.4.1 Simulation Study

In this simulation study we will generate data from a hierarchical lineal model with dependence across the time defined in equations (3.1a), (3.1b), (3.1c) and (3.1d). To generate a simulation of a educational database, we consider that the population of 300 schools, and in each school there are 20 to 30 students. The simulation procedure is detailed below,

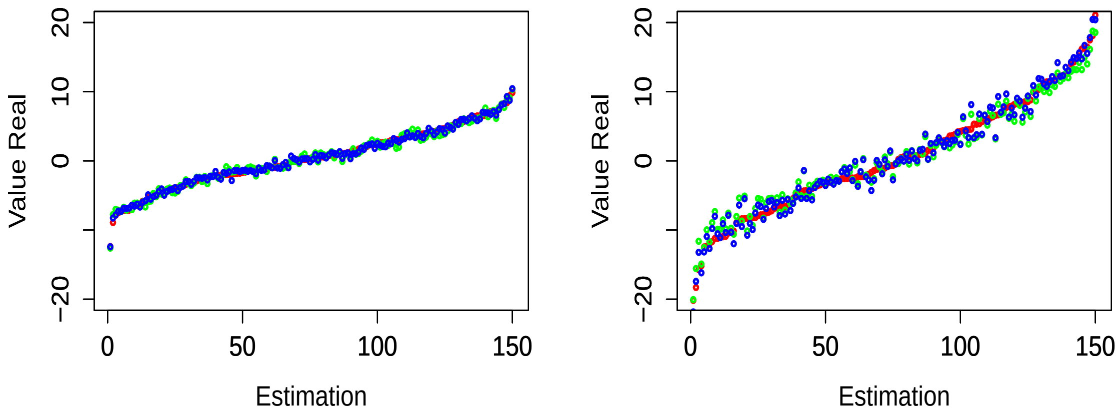
1. Do $i = 1$ to J, with J is the school total,
   - Generate n_{j3} for each j, where $20 \leq n_{j3} \leq 30$
   - Generate n_{j4} for each j, where $20 \leq n_{j4} \leq 30$
   - $\theta_{j31} \sim \mathrm{N}(0, 25)$ for each j.
   - $\theta_{j42} \sim \mathrm{N}(\mu\theta_{31}, 64)$, μ is considered in two sceneries different.
2. Generate 4 covariables, these may be generated from any distribution. In this particular case, these are assumed as;$\boldsymbol{X}_{1j3} = \imath_{n_{j3}}$, $\boldsymbol{X}_{2j3} \sim \mathrm{N}(100, 20)$, $\boldsymbol{X}_{3j3} \sim \mathrm{N}(-1, 10)$ and $\boldsymbol{X}_{4j3}$ correspond to contemporaneous score of $\boldsymbol{X}_{2j3}$.
3. Generate 4 covariables, these may be generated from any distribution. In this particular case, these are assumed as;$\boldsymbol{X}_{1j4} = \imath_{n_{j4}}$, $\boldsymbol{X}_{2j4} \sim \mathrm{N}(200, 20)$, $\boldsymbol{X}_{3j4} \sim \mathrm{N}(0, 10)$ and $\boldsymbol{X}_{4j4}$ correspond to contemporaneous score of $\boldsymbol{X}_{2j4}$.
4. Generate a error associated to the equation (3.1a). So, $\epsilon_{ij3} \sim \mathrm{N}(0; 289)$
5. Generate a error associated to the equation (3.1c). So, $\epsilon_{ij3} \sim \mathrm{N}(0; 324)$
6. Once the steps 5 – 6 were performed, must be simulated the dependent variables by,
   - $Y_{ij3} = \boldsymbol{X}_{ij3}^{\top}\boldsymbol{\beta}_3 + \theta_{j31} + \epsilon_{ij3}$, Where $\boldsymbol{\beta}_3^{\top} = \begin{pmatrix} 2 & 0.5 & 20 & 5 \end{pmatrix}$
   - $Y_{ij4} = \boldsymbol{X}_{ij4}^{\top}\boldsymbol{\beta}_4 + \theta_{j42} + \epsilon_{ij4}$, Where $\boldsymbol{\beta}_4^{\top} = \begin{pmatrix} 7 & 0.9 & 15 & 0.5 \end{pmatrix}$

Simulation Results

This simulation study aims to analyse the computational performance of our model proposed in this chapter. We consider two scenarios: (I) Small parameter of association on the time between two cohorts different, $\mu = 0.5$, and (II) Big parameter of association on the time, $\mu = 5$.

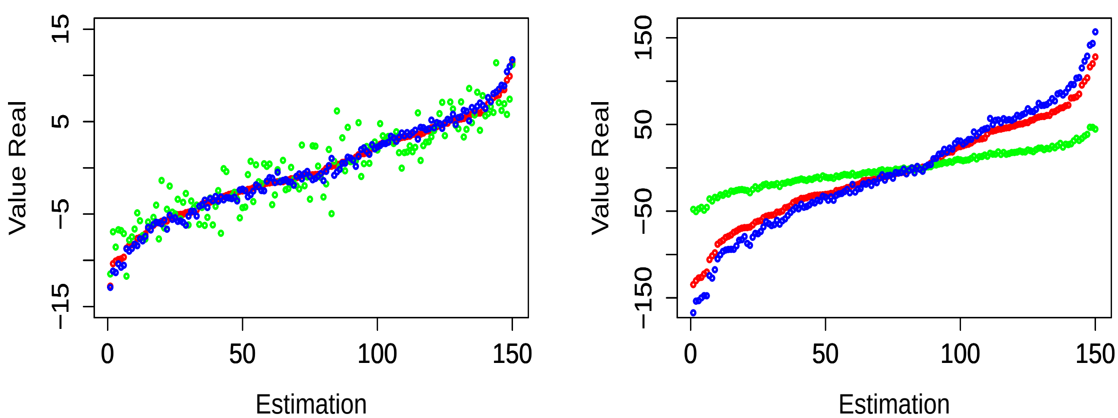
The study of simulation consider Monte Carlo samples of size $10, 50, 100$ and 500. Each sample was adjusted by both models standard HLM and dynamical HLM. Then, the means a standard deviation of estimation all parameter are registered in the appendix C.2. As a general conclusion, Independently of scenery in both models are observed that the estimation of parameters are close to Their actual values, with the exception of tau^2 4, variance associated to school random effect in the second cohort, its real value is 144 but the standard model overestimate this parameter, although when $mu = 0.5$ the differences are slight, when $\mu = 5$ the differences are high, it is 18 times higher that value real.

Moreover the value added estimation is represented in the following graphics, so that it all the graphics in axis X is the order ascendant of real value-added (Colour red), the blue points are the estimates through our proposed model and points green the estimation of standard model. When μ is 0.5 the estimations of both models are close to real value, although variability of estimation of value added is greater in the second cohort, see Figure 3.4. However this similarity between the estimates of both models begin to disappear with value of μ at least 1, see Figures 3.5 to 3.9.



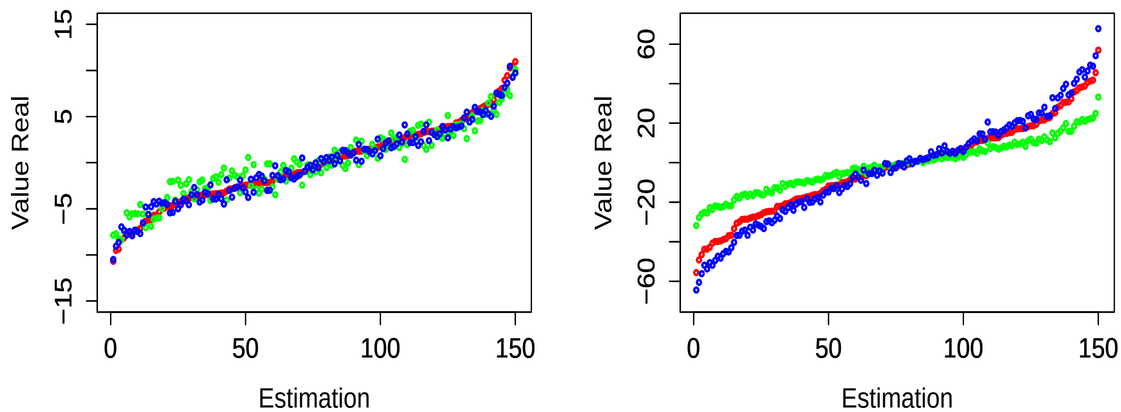
(a) First cohort with $\mu = 0.5$ (b) Second cohort with $\mu = 0.5$

Figure 3.4: Comparation between of value-added estimates on both cohorts with $\mu = 0.5$



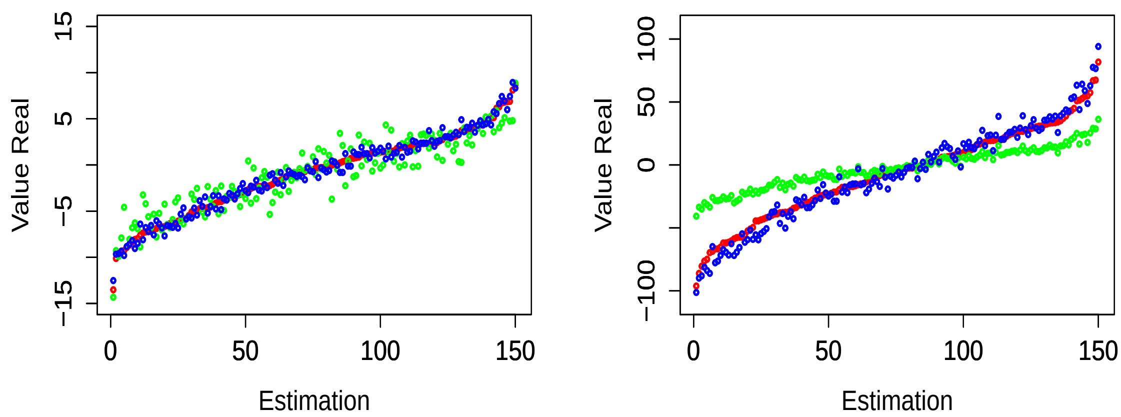
(a) First cohort with $\mu = 1$ (b) Second cohort with $\mu = 1$

Figure 3.5: Comparation between of value-added estimates on both cohorts with $\mu = 1$
.



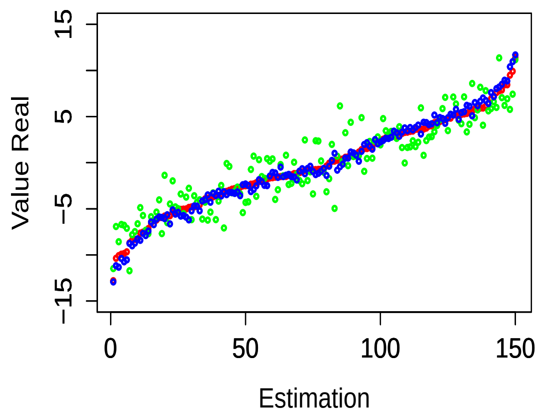
(a) First cohort with $\mu = 2$ (b) Second cohort with $\mu = 2$

Figure 3.6: Comparation between of value-added estimates on both cohorts with $\mu = 2$
.

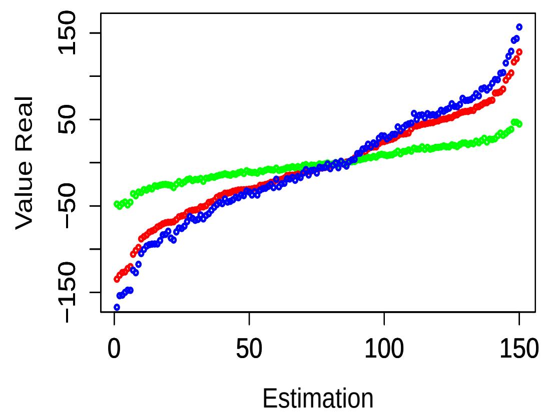


(a) First cohort with $\mu = 3$ (b) Second cohort with $\mu = 3$

Figure 3.7: Comparation between of value-added estimates on both cohorts with $\mu = 3$
.

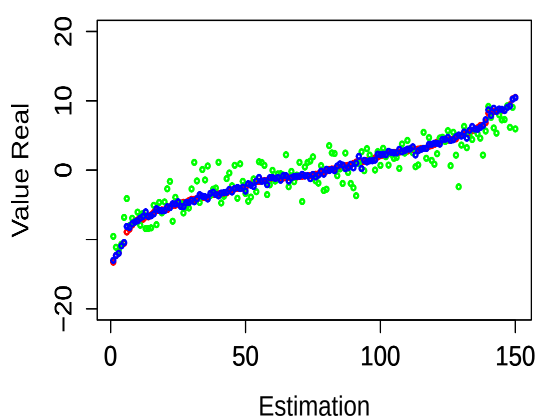


(a) First cohort with $\mu = 4$

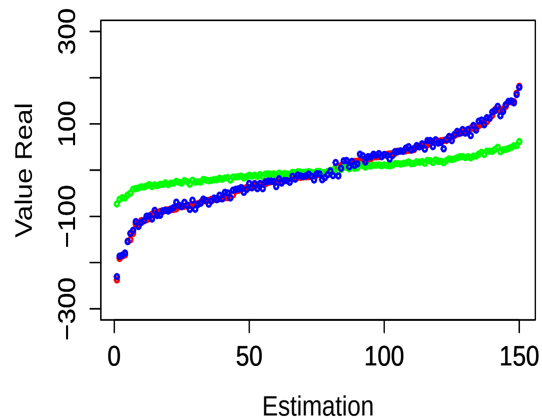


(b) Second cohort with $\mu = 4$

Figure 3.8: Comparation between of value-added estimates on both cohorts with $\mu = 4$
.



(a) First cohort with $\mu = 5$

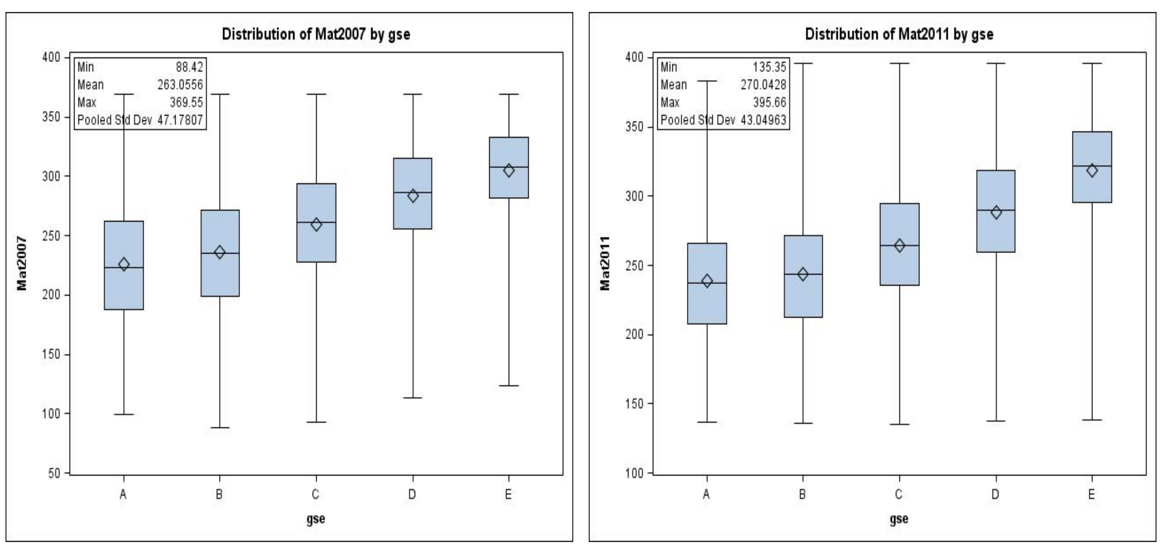


(b) Second cohort with $\mu = 5$

Figure 3.9: Comparation between of value-added estimates on both cohorts with $\mu = 5$
.

3.4.2 Data application

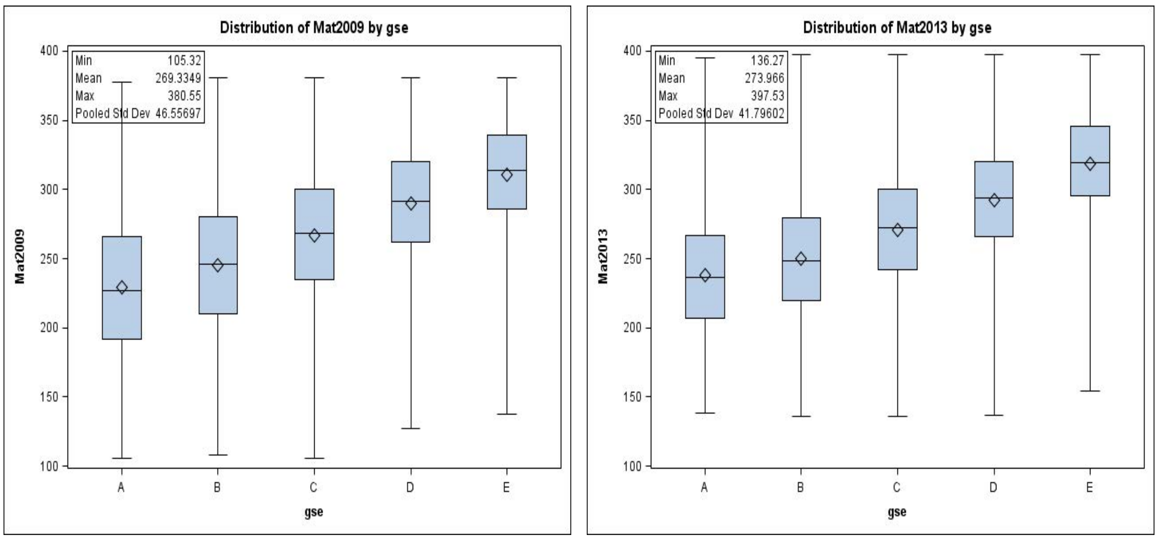
In this section we apply our value-added model proposed in this chapter on SIMCE. Specifically we use the score of mathematics in the cohorts 2007-2011 and 2009-2013. This application only consider schools of great cities; Región de Valparaíso, Región del Bío-Bío and Región Metropolitana. Moreover, the Figures 3.10 and 3.11 present the distribution of mathematics SIMCE by group social economic in a sub-population of Chile, in them not only appreciate the difference in scores by socioeconomic level, also it is observed a decrease in the variability within groups of a measurement to compare the previous and post score of the cohorts.



(a) Mathematic SIMCE 2007 (b) Mathematic SIMCE 2011

Figure 3.10: Distribution of first cohort 2007-2011 by GSE

.



(a) Mathematic SIMCE 2009 (b) Mathematic SIMCE 2013

Figure 3.11: Distribution of second cohort 2009-2013 by GSE

.

Specifically, the analysis of school improvement considered 1,315 schools, such that in the cohort 2007-2011 there are 60,889 students while that the cohort 2009-2013 has 50,364 students. For each cohort is a school effect. Then we say that the analysis of school improvement explicit the assumption that the past determines the future if school effect for the 2007-2011 dependent of school effect for 2009-2013.

The estimators of parameters obtained through dynamic value-added model with the parameter of associated of school effects correspond to; $\sigma_3^2 = 922.638$, $\sigma_4^2 = 868.042$, $\tau_3^2 = 155.425$, $\tau_4^2 = 1158.660$, while that the estimation of fixed effects are $\beta_3^\top = (66.720, 0.581, 7.442, 0.572)$, $\beta_4^\top = (80.305, 0.567, 8.144, 0.055)$ considering as covariates; a intercept, the prior score, IVE and contemporaneous score. Moreover, estimation of parameter μ is 0.48.

If the parameter of dependence is less than 1, we can say that, for the group of schools under analysis, the past determines the future in a very weak way; If this parameter is greater than 1, we say that the past determines the future strong way; bigger is μ the dependence is stronger. Moreover, if the parameter is equal to 0, then past and future are not at all related. Although the parameter is low $\mu = 0.48$ Moreover, we classified the schools according the quartile of value added estimation obtained through our model proposed, then we graph the distribution of estimates of both cohorts (boxplot), such that the green colour are the estimation of standard value-added model and blue is the same model but considering the association parameter between school effects.See Figure 3.12, when we observed the difference in the estimation of value added those are not non-existent, despite a small value of μ.

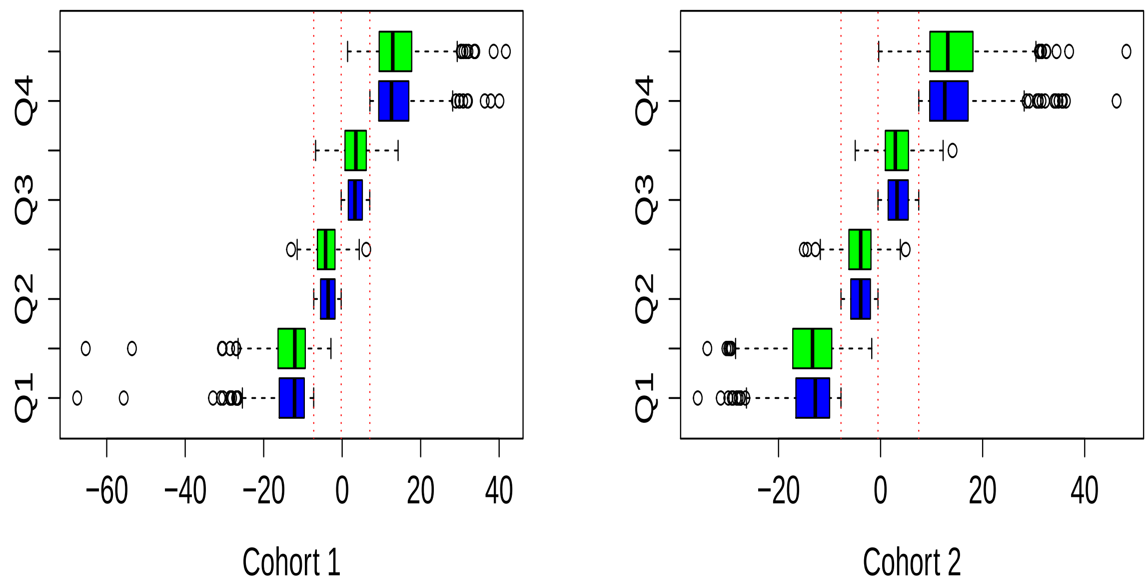


Figure 3.12: Comparison by quantile for estimation of value-added through HLM and dynamic HLM.
.

3.5 Final remarks

From a methodological point of view, the novelty of this proposed is the specification and estimation of a value added model under school effects dependent across the time. Specifically,

- The model estimates the fixed effects as well as the component variances in an unbiased way.
- The model estimates the correlation between the school effects across the time.
- The model estimates the values added by the schools effect, but the predictions dependent on the whole trajectory of scores within the school.

From an empirical point of view, the novelties are the following:

- To perform a value added analysis of the Chilean educational system considering the dependent school effects,
- To make a comparison between the standard value added analysis and this new proposed.

As a final comment, our model does not have a computational cost, its convergence is quick, even when is used on SIMCE educational data that it has a large volume of students.

Chapter 4:

Conclusions and Future work

Chapter 4: Conclusions and Future Work

This thesis was developed in the context of school effectiveness. We have addressed two different frameworks. In the chapter 2 we developed an extension to the hierarchical linear mixed model that include instrumental variables for to resolve of endogeneity problem, extention to Manzi et al. (2014). Whereas in chapter 3 we worked in a new hierarchical lineal mixed model that include a parameter of association between school effect across the time.

To conclude this work, in this last chapter we include; i) main conclusions, ii) some final comments in the context of value added models, and iii) future works.

4.1 Conclusion

> *... An effective school is one in which students progress further than might be expected from consideration of its intake ......*
>
> Mortimore (1991)

School effectiveness seeks to identify the 'Value Added' schools through outcomes of their student. Typically the value-added models are used to estimate the contribution of teachers, educational programs or schools to student achievement. From a methodological point of view, this can be achieved by modelling student's scores taking into account the differences in prior achievements and possibly other measured characteristics in the form of covariates at both the school level and the student level; see Braun et al. (2010); OECD

(2008); Raudenbush (2004) and Timmermans et al. (2011). The role of the covariates is to characterize a school of reference with respect to which the Value-Added is substantively interpreted.

As a matter of fact, the Value-Added of a school is a comparison between the conditional expected scores in a given school and the conditional expected score in the school of reference: if the covariates are modified, the school of reference is also modified and the meaning of Value-Added changes. Because the interest is to know the net contribution of a school, and the covariates have an influence on the student performance, an important requirement for the covariates is that they have to be unrelated to the internal pedagogical processes performed by a school. Using the econometric jargon, it is said that the covariates are *exogenous* with respect to the school. A standard approach to model the school Value-Added is the use of hierarchical linear models (HLM), or multilevel models Goldstein (2002); Snijders and Bosker (1999), due to the hierarchical structure of the data where students are nested into schools (see chapter 1). Under this approach, students scores are explained by their previous achievement, some covariates and a random effect representing the school effect. A measure of Value-Added has been typically obtained as the prediction of the random school effect Aitkin and Longford (1986); Longford (2012); Raudenbush and Willms (1995); Tekwe et al. (2004).

Throughout this thesis we mention some problems associated with the methodology used to estimate value added, that these are our motivation, the beginning of this dissertation. A little update of this issues are mentioned below,

(i) **Homogeneity of educational data sets:**

In practice often be found in the educational data sets, more of a group of students or of schools with similar social features in common, such that their yields are similar within them, but different between groups. Examples of those features are social mix, balance of intake, school mix and school composition Rutter, Maughan, Mortimore, and Ouston (1979); Thrupp (1995). The solution is use heteroscedastic value added model. Thus, for instance, if schools can be grouped in five socio-economics groups, then the variance of the the distribution generating the school effects is specified as a function of the socio-economic status (SES):

(ii) **Independence between school effect and covariates:**

This was discussed in the chapter 2, the endogeneity is a problem present in some educational data, SIMCE is a clear example. **?** solved this difficult using to instrumental variables in the structure of value added model. Also, we repeatedly mentioned that the value added is different to school effect.

(iii) **Independence across the time:**

This consideration was addressed in chapter 3. While there are studies that have evaluated the persistence of value added over time, the methodologies assume independent between the cohorts school effect, because in the modelling of value added no considered the past of schools.

The methodologies used and developed in this dissertation are contributions in the context of school effectiveness. We have seen how sensitive are the estimates of these models and the consequences which a unfair classification could cause under the accountability system.

To finalized this section just mention that study of Monte Carlo suggest that both models are defectives computationally, their estimation are quick , under time cost. Also, they an adequate fit of error low. Then, the question is Why? and When? use these models, as any statistical model the answer is depend, it depends the context and objectives.

4.2 Future Work

While the methodologies developed in this thesis are based on the educational context, these are not restricted only educational data may well be applied in other contexts. The future works derived of this dissertation are detailed below.

(a) **Applications problem**: What happens if we use a representative sample of population for estimate and classify each school system school?, What are the problems associated to the sample?, What must be the characteristics of sample design and sample size?, How the value added can be stable on the time without it is affect by change subtitle on the population.

(b) **Global Measures**;

(i) A natural extension of this thesis is the developed of a global measure that consider the endogeneity and dependence on the time, until now are two model that compete for its capacity of fit.

(ii) Other extension but maybe no so natural it is to develop a robust measure of VA when information on various levels within a school is available. The schools commonly have scores of more than two educational levels. For example, two group of students on level educational different that form two cohorts on the same time. Illustrating this idea, the student of 4th and 6th degree have their score on the time t (pre-score), the same student on the time t+1 have their post-score on the levels 5th and 7th degree respectively. Of course, we think and assumption that calculating measures of value-added using only part of the score information could lead to erroneous conclusions about the effectiveness of a school. For example, it could happen that the value added calculated for a particular school using score information for levels 4rd and 5th is very different from that calculated using information from the 6th and 7th grade in the same school.

(iii) The following step will be monitoring student achievement every year. The fact that a particular school assess several educational level every year, inevitably leads to a complex data design.

References

Aitkin, M., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society. Series A*, *149*, 1-43.

Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: from supply and demand to natural experiments. *The Journal of Economic Perspectives*, *15*, 69-85.

Baker, E., Barton, P., Darling-Hammond, L., Haertel, E., Ladd, H., Linn, R., … Shepard, L. (2010). Problems with the use of student test scores to evaluate teachers (epi briefing paper 278). *Washington D. C.: Economic Policy Institute.*.

Barndorff-Nielsen, O. ("1978). *Information and exponential families*. Wiley.

Bauer, Gottfredson, Dean, & Zucker. (2013). Analyzing repeated measures data on individuals nested within groups: accounting for dynamic group effects. *Psychological Methods*, *18 (1)*.

Bollen, K. A. (1989). *Structural equations with latent variables*. London: Wiley.

Braun, H., Chudowsky, N., & Koening, J. (2010). *Getting value out of value-added: Report of a workshop*. The National Academies Press.

Briggs, D. C., & Weeks, J. P. (2011). The persistence of value-added school effects. *Journal of Educational and Behavioral Statistics*, *36 (5)*.

Carrasco, A., & San Martín, E. (2012). Voucher system and school effectiveness: Reassessing school performance difference and parental choice decision-making. *Estudios de Economia*, *39*, 123-141.

Cohen, Bloom, & Malin. (1996). Educating all children: A global agenda. *Cambridge, MA: The MIT Press*.

Del Pino, G., González, J., Manzi, J., & San Martín, E. (2009). Estudio de valor agregado y progreso 3 básico 2006-4 básico 2007. *Technical Report, Ministry of Education of the Chilean Government*.

Del Pino, G., San Martín, E., de la Cruz, R., Manzi, J., González, J., & Taut, S. (2008). Estudio de valor agregado y progreso 3 básico 2006-4 básico 2007. *Technical Report, Ministry of Education of the Chilean Government*.

Del Pino, G., San Martín, E., Manzi, J., González, J., & Taut, S. (2008). Estudio de valor agregado y progreso 8 básico 2004-2 medio 2006. *Technical Report, Ministry of Education of the Chilean Government*.

Engle, R., Hendry, D., & Richard, J. (1983). Exogeneity. *Econometrica*, *51*, 277-304.

Florens, J. P., Marimoutou, V., & Péguin-Feissolle, A. (2007). *Econometric modeling and inference*. Cambridge: Cambridge University Press.

Florens, J. P., & Mouchart, M. (1985). Conditioning in dynamic models. *Journal of Time Series Analysis*, *6*, 15-34.

Gansle, K., Noell, G., & Burns, J. (2012). Do student achievement outcomes differ across teacher preparation programs? an analysis of teacher education in lousiana. *Journal of Teacher Education*, *63*, 304-317.

Goldstein, H. (2002). *Multilevel statistical models*. London: Edward Arnold.

Goldstein, H. (8). Methods in school effectiveness resarch. *School Effectiveness and School Improvement*, *369-395*.

González, J., San Martín, E., Manzi, J., & Del Pino, G. (2010, Agosto). Mixed modeling approach to value-added analysis: Features, problems, and recent developments in the chilean case. *Paper presented at the Second Biennial Meeting of the EARLI Special Interest Group 18 "Educational Effectiveness: Models, Methods and Applications", Leuven, Belgium*.

Gray, Goldstein, & Jesson. (1996). Changes and improvements in schools' effectiveness: trends over five years. *Research Papers in Education*, *11(11)*, 35–51.

Gray, Goldstein, & Thomas. (2001). Predicting the future: the role of past performance in determining trends in institutional effectiveness at a level. *British Educational Research Journal*, *27 (4)*, 391–405.

Gray, Goldstein, & Thomas. (2003). Of trends and trajectories: Searching for patterns in school improvement. *British Educational Research Journa*, *29*.

Gray, J., Goldstein, H., & Jesson, D. (2012). Changes and improvements in schools' effectiveness: trends over five years. *Research papers in education*, *11*.

Gray, J., Hopkins, D., Reynolds, D., Wilcox, B., & Farrell, D., S. Jesson. (1999). Improving school: performance and potential. *Buckingham: Open University Press*.

Hendry, D. H., & Richard, J. F. (1983). The econometric analysis of economic time series (with discussion). *International Statistical Review*, *51*, 111-148.

Kim, J. S., & Frees, E. W. (2007). Multilevel modelling with correlated effects. *Psychometrika*, *72*, 505-533.

Kinsler, J. (2012). Beyond levels and growth estimating teacher value-added and its persistence. *Journal of Human Resources*, *47 (3)*, 722-753.

Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structural analysis. In *Measurement and prediction.*

Longford, N. T. (2012). A revision of school effectiveness analysis. *Journal of Educational and Behavioral Statistics*, *37*, 157–179.

Manzi, J., & Preiss, D. (2013). Educational assessment and educational achievement in south america. *In: International Guide to Student Achievement, J. Hattie and E. M. Anderman, chapter 9. Taylor and Friends: New York*.

Manzi, J., San Martín, E., & Van Bellegem, S. (2014). School systme evaluation by value added analysis under endogeneity. *Psychometrika*, *79*, 130–153.

Meckes, L., & Carrasco, R. (2010). Two decades of simce: An overview of the national assessment system in chile. *Assessment in Education: Principles, Policy & Practice*, *17 (2)*, 233-248.

Milla, J., San Martín, E., & Van Bellegem, S. (2014). Value added analysis of tertiary education in colombia. *First year report to the Instituto Colombiano para la Evaluacin de la Educacin ICFES, Bogotá, Colombia*.

Mortimore. (1991). Effective schools from a british perspective. In *Rethinking effective schools: Research and practice.*

Mouchart, M., Russo, F., & Wunsch, G. (2010). Inferring causal relations by modelling structures. *Statistica*, *70*, 411-432.

OECD. (2008). *Measuring improvements in learning outcomes. best practices to assess the value-added of schools*. OECD publications.

Page, G., Orellana, J., San Martín, E., & González, J. (2015). Exploring complete school effectiveness via quantile value-added. *submitted*.

Pugh, G., & Mangan, J. (2003). What's in a trend? a comment on gray, goldstein and thomas (2001). *British Educational Research Journal*, *29 (1)*.

Randolph, W. C. (1988). A transformation for heterocedastic error components regression models. *Economics Letters*, *27*, 349-354.

Raudenbush, S. W. (2004). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics*, *29*, 121-129.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis method*. London: SAGE Publication.

Raudenbush, S. W., & Willms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, *20*(4), 307-335.

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay and student achievement. *Quarterly Journal of Economics*, *125(1)*.

Rutter, M., Maughan, B., Mortimore, P., & Ouston, J. (1979). *Fifteen thousand hours: Secondary schools and their effect on children*. Harvard University Press.

San Martín, E., & Carrasco, A. (2012). Clasificación de escuelas en la educativa: Contribucióon de modelos de valor-agregado para una responsabilización justa. *Serie Temas de la Agenda Pública, Centro de Políticas Públicas UC 7*, *53*.

Santelices, M. V., Galleguillos, P., González, J., & Taut, S. (2015). A study about teacher quality in chile: the role of context where teacher works and value added measures. *Psyhke*, *24 (1)*.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage Publication.

Spanos, A. (1994). On modeling heteroskedasticity: The student's t and elliptica linear regression models. *Econometric Theory*, *10*, 286-315.

Spencer, G. H., & Fielding, A. (2002). A comparison of modelling strategies for value-added analyses of educational data. *Computational Statistics*, *17*, 103-116.

Steele, F., Rasbash, J., & Jenkins, J. (2013). A multilevel simultaneous equations model for within-cluster dynamic effects, with an application to reciprocal parent-child and sibling effects. *Psychological Methods*, *18 (1)*.

Teddlie, C., & Reynolds, D. (2000). The international handbook of school effectiveness research.

Tekwe, C., Carter, R., Changing, M., Algina, J., Lucas, M., Roth, J., … Resnick, M. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics*, *29*, 11–36.

Thieme, C., Prior, D., Tortosa-Ausina, E., & Gempp, R. (2013). Value added and contextual factors in education: Evidence from chilean chools. *Working Papers 2013129, Fundacion BBVA / BBVA Foundation*.

Thomas. (2001). Dimensions of secondary school effectiveness: comparisons across regions, school. *Effectiveness and School Improvement*, *12 (4)*.

Thrupp, M. (1995). The school mix effect: the history f an enduring problem in educational researc, policy and practice. *School Effectiveness and School Improvement*, *12*, 41-84.

Timmermans, Doolaard, & Wolf, D. (2011). Conceptual and empirical differences among various value-added models for accountability. *School Effectiveness & School Improvement*, *22 (4)*, 393-413.

Wooldridge, J. (2002). *Econometric analysis of cross section and panel data*. Mason: South-Western College Pub.

Wooldridge, J. (2008). *Introductory econometrics: a modern approach*. Mason: South-Western College Pub.

Wunsh, G., Mouchart, M., & Russo, F. (2014). Functions and mechanism in structural-modelling explanations. *Journal for General Philosophy of Science*, *45*, 187-208.

A Appendix of chapter I: School Effectiveness

A.1 Technical Appendix

A.1.1 Notation

First : For all dataset is defined the following expressions;

$\mathcal{J}$: It represents the set of all school.

J : It represents the total school, card $(\mathcal{J})$.

n_j : It represents total students in school j, $j \in \mathcal{J}$.

Second : A student level the following expressions are defined;

Y_{ij} : It is the score (scalar), for the student i that belong school j, with $j = 1, \ldots, J$ and $i = 1, \ldots, n_j$

$\boldsymbol{X}_{ij}$: It is a K-dimensional vector of K-explanatory variables for student i, in the school j.

Third : A school level the following expressions are defined;

$\boldsymbol{Y}_j = \begin{pmatrix} Y_{1j} \\ \vdots \\ Y_{n_j j} \end{pmatrix}$: It is a vector of all scores in school j, such that $\boldsymbol{Y}_j \in \mathbb{R}^{n_j}$.

$\boldsymbol{X}_j = \begin{pmatrix} \boldsymbol{X}_{1j}^\top \\ \vdots \\ \boldsymbol{X}_{n_j j}^\top \end{pmatrix}$: It is a matrix of dimension $n_j \times K$, where K is the number of covariates.

θ_j : It represents the school effect of school j.

Also, we define others matrices very useful, (to school level).

$\imath_{n_j} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$: It is a vector 1's of dimension $n_j \times 1$, for $j = 1, 2 \ldots, J$.

$\boldsymbol{J}_{n_j} = \imath_{n_j} \imath_{n_j}^\top$: It is a matrix of dimension $n_j \times n_j$ of 1's.

$\overline{\boldsymbol{J}}_{n_j} = \boldsymbol{J}_{n_j}/n_j$: It is a matrix of dimension $n_j \times n_j$ of $1/n_j'$s.

$\boldsymbol{E}_{n_j} = \boldsymbol{I}_{n_j} - \overline{\boldsymbol{J}}_{n_j}$: It is matrix of dimension $n_j \times n_j$.

Fourth : A dataset level the following expressions are defined;

$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \vdots \\ \boldsymbol{Y}_J \end{pmatrix}$: It is a vector of dimension $N \times 1$, where $N = \sum_{j=1}^{J} n_j$.

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_J \end{pmatrix} \quad : \quad \text{It is a matrix of dimension } N \times K.$$

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_J \end{pmatrix} \quad : \quad \text{It is a vector of dimension } J.$$

$$\boldsymbol{L}_{\underline{n}} = \begin{bmatrix} \imath_{n_1} & & \\ & \ddots & \\ & & \imath_{n_J} \end{bmatrix} \quad : \quad \text{It is a } N \times J \text{ matrix.}$$

$$\boldsymbol{D} = \begin{bmatrix} n_1 & & \\ & \ddots & \\ & & n_J \end{bmatrix} \quad : \quad \text{It is a diagonal matrix of dimension } J \times J$$

$$\boldsymbol{P} = \begin{bmatrix} \imath_{n_1}^\top / n_1 & & \\ & \ddots & \\ & & \imath_{n_J}^\top / n_J \end{bmatrix}$$

: It is a matrix of dimension $J \times N$. Also It is called Between-matrix, and It can be obtained by $\boldsymbol{P} = \boldsymbol{D}^{-1}\boldsymbol{L}_{\underline{n}}^\top$

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{E}_{n_1} & & \\ & \ddots & \\ & & \boldsymbol{E}_{n_J} \end{bmatrix}$$

: It is a idempotent matrix of dimension $N \times N$. Also, It is called Within-matrix .

A.1.2 The semantic of the conditional expectation

The structure underlying most of the statistical models used in educational measurement, and particularly in SER, can be depicted through conditional expectations. Broadly speaking, the target of quantitative analysis is to relate a dependent variable $\boldsymbol{Y}$ to a set of covariates or explanatory factors, denoted by $\boldsymbol{X}$. This relationship has two components: the part of $\boldsymbol{Y}$ that can be explained by $\boldsymbol{X}$, and the part of $\boldsymbol{Y}$ that cannot be explained by $\boldsymbol{X}$. The first component corresponds to the conditional expectation of $\boldsymbol{Y}$ given $\boldsymbol{X}$, and it is typically denoted as $\mathrm{E}(\boldsymbol{Y}|\boldsymbol{X})$, also called the regression of $\boldsymbol{Y}$ on $\boldsymbol{X}$. Accordingly, the dependent variable $\boldsymbol{Y}$ is decomposed as

$$\boldsymbol{Y} = \mathrm{E}(\boldsymbol{Y}|\boldsymbol{X}) + \{\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y}|\boldsymbol{X})\}$$

where the error or residual term $\{\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y}|\boldsymbol{X})\}$, the part of Y that cannot be explained by the covariates $\boldsymbol{X}$, is by construction uncorrelated with or orthogonal to $\mathrm{E}(Y|\boldsymbol{X})$.

Technically speaking, the conditional expectation $E(\boldsymbol{Y}|\boldsymbol{X})$ is computed on the basis of the joint distribution of $(\boldsymbol{Y}, \boldsymbol{X})$. However, in practice, this conditional expectation is specified in a particular form. Thus, for instance, it can be assumed that the relationships between Y and $\boldsymbol{X}$ is linear in $\boldsymbol{X}$. As the reader can recognize, this corresponds to a linear regression. Other specifications are possible, as for instance a nonlinear function of one or more explanatory factors.

B Appendix of chapter II: Endogenous Value-Added Models for Subgroups of Schools

B.1 Technical Appendix

B.1.1 Notation

First : For all dataset is defined the following expressions;

$\mathcal{S}$: It represents the set of groups' labels

S : It represents the total groups, card($\mathcal{S}$)

$\mathcal{J}$: It represents the set of all school.

J : It represents the total school, card $(\mathcal{J})$.

$\rho(\cdot)$: It is a function induces a partition of $\mathcal{J}$, $\{\mathcal{J}_1, \ldots, \mathcal{J}_S\}$ on $\{1, \ldots, J\}$. This way, $\rho(.)$ Is the grouping school function is defined from $\{1, \ldots, J\}$ to $\{1, \ldots, S\}$ as $j \in \{1, \ldots, J\} \longmapsto \rho(j) = s \in \{1, \ldots, S\}$.

J_s : It is a number of school in the group s, ie card($\mathcal{J}_s$),

$n_j^{(s)}$: It represents total students in school j, $j \in \mathcal{J}_s$.

Second : A student level the following expressions are defined;

$Y_{ij}^{(s)}$: It is the contemporaneous test score of student i belonging to school $j \in \mathcal{J}_s$.

$\boldsymbol{X}_{ij}^{(s)}$: It is a vector K_1-dimensional, the exogenous variables associated to pupil i belonging to school $j \in \mathcal{J}_s$.

$\boldsymbol{W}_{ij}^{(s)}$: It is a L-dimensional vector, the instrumental variables associated to pupil i belonging to school $j \in \mathcal{J}_s$.

$\boldsymbol{Z}_{ij}^{(s)}$: It is a K_2-dimensional vector, the endogenous variables associated to pupil i belonging to school $j \in \mathcal{J}_s$.

$\theta_j^{(s)}$: It represents the school effect of school $j \in \mathcal{J}_s$.

Third : For each school $j \in \mathcal{J}_s$ with $s = 1, \ldots, S$ following expressions are defined;

$$\boldsymbol{Y}_j^{(s)} = \begin{pmatrix} Y_{1j}^{(s)} \\ \vdots \\ Y_{n_j^{(s)}j}^{(s)} \end{pmatrix}$$

: It is a vector of all scores in school j, such that $\boldsymbol{Y}_j^{(s)} \in \mathbb{R}^{n_j^{(s)}}$.

$$\boldsymbol{X}_j^{(s)} = \begin{pmatrix} \boldsymbol{X}_{1j}^{(s)\top} \\ \vdots \\ \boldsymbol{X}_{n_j^{(s)}j}^{(s)\top} \end{pmatrix}$$

: It is a matrix of dimension $n_j^{(s)} \times K_1$, where K_1 is the number of exogenous variables.

$$\boldsymbol{W}_j^{(s)} = \begin{pmatrix} \boldsymbol{W}_{1j}^{(s)\top} \\ \vdots \\ \boldsymbol{W}_{n_j^{(s)}j}^{(s)\top} \end{pmatrix}$$

: It is a matrix of dimension $n_j^{(s)} \times L$, where L is the number of instrumental variables.

$$\boldsymbol{Z}_j^{(s)} = \begin{pmatrix} \boldsymbol{Z}_{1j}^{(s)\top} \\ \vdots \\ \boldsymbol{Z}_{n_j^{(s)}j}^{(s)\top} \end{pmatrix}$$: It is a matrix of dimension $n_j^{(s)} \times K_2$, where K_2 is the number of endogenous variables.

$\theta_j^{(s)}$: It represents the school effect of school $j \in \mathcal{J}_s$.

Also, we define others matrices very useful, (to school level).

$$\imath_{n_j^{(s)}} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$: It is a vector 1's of dimension $n_j^{(s)} \times 1$, for $j = 1, 2 \ldots, J_s$.

$\boldsymbol{J}_{n_j^{(s)}} = \imath_{n_j^{(s)}} \imath_{n_j^{(s)}}^\top$: It is a matrix of dimension $n_j^{(s)} \times n_j^{(s)}$ of 1's.

$\overline{\boldsymbol{J}}_{n_j^{(s)}} = \boldsymbol{J}_{n_j^{(s)}} / n_j^{(s)}$: It is a matrix of dimension $n_j^{(s)} \times n_j^{(s)}$ of $1/n_j^{(s)}$'s.

$\boldsymbol{E}_{n_j^{(s)}} = \boldsymbol{I}_{n_j^{(s)}} - \overline{\boldsymbol{J}}_{n_j^{(s)}}$: It is matrix of dimension $n_j^{(s)} \times n_j^{(s)}$.

Fourth : For each $s \in \mathcal{S}$, the following expressions are defined;

$N_s = \sum_{j \in \mathcal{J}_s} n_j^{(s)}$: Total of students in the group s.

$$\boldsymbol{Y}^{(s)} = \begin{pmatrix} \boldsymbol{Y}_1^{(s)} \\ \vdots \\ \boldsymbol{Y}_{J_s}^{(s)} \end{pmatrix}$$: It is a vector of dimension $N_s \times 1$.

$$\boldsymbol{X}^{(s)} = \begin{pmatrix} \boldsymbol{X}_1^{(s)} \\ \vdots \\ \boldsymbol{X}_{J_s}^{(s)} \end{pmatrix}$$: It is a matrix of dimension $N_s \times K_1$.

$$\boldsymbol{W}^{(s)} = \begin{pmatrix} \boldsymbol{W}_1^{(s)} \\ \vdots \\ \boldsymbol{W}_{J_s}^{(s)} \end{pmatrix}$$: It is a matrix of dimension $N_s \times L$.

$$\boldsymbol{Z}^{(s)} = \begin{pmatrix} \boldsymbol{Z}_1^{(s)} \\ \vdots \\ \boldsymbol{Z}_{J_s}^{(s)} \end{pmatrix}$$: It is a matrix of dimension $N_s \times K_2$.

$$\boldsymbol{\theta}^{(s)} = \begin{pmatrix} \theta_1^{(s)} \\ \vdots \\ \theta_{J_s}^{(s)} \end{pmatrix}$$: It is a vector of dimension J_s.

$$\boldsymbol{L}_{\underline{n}_s} = \begin{bmatrix} \imath_{n_1^{(s)}} & & \\ & \ddots & \\ & & \imath_{n_{J_s}^{(s)}} \end{bmatrix}$$: It is a $N_s \times J_s$ matrix.

$$\boldsymbol{D}_s = \begin{bmatrix} n_1^{(s)} & & \\ & \ddots & \\ & & n_{J_s}^{(s)} \end{bmatrix}$$: It is a diagonal matrix of dimension $J_s \times J_s$

$$\boldsymbol{P}_s = \begin{bmatrix} \imath^\top_{n_1^{(s)}}/n_1^{(s)} & & \\ & \ddots & \\ & & \imath^\top_{n_{J_s}^{(s)}}/n_{J_s}^{(s)} \end{bmatrix}$$: It is a matrix of dimension $J_s \times N_s$. Also It is called Between-matrix or Between-operator, and It can be obtained by $\boldsymbol{P}_s = \boldsymbol{D}_s^{-1}\boldsymbol{L}_{\underline{n}_s}^\top$

$$\boldsymbol{Q}_s = \begin{bmatrix} \boldsymbol{E}_{n_1^{(s)}} & & \\ & \ddots & \\ & & \boldsymbol{E}_{n_{J_s}^{(s)}} \end{bmatrix}$$: It is a idempotent matrix of dimension $N_s \times N_s$. Also, It is called Within-matrix or Within-operator.

Some useful matrix multiplications,;

$$\boldsymbol{D}_s^{-1} = \boldsymbol{P}_s \boldsymbol{P}_s^\top$$

$$\boldsymbol{P}_s^\top \boldsymbol{D}_s^2 \boldsymbol{P}_s = \begin{bmatrix} J_{n_1^{(s)}} & & \\ & \ddots & \\ & & J_{n_{J_s}^{(s)}} \end{bmatrix}$$

$$\boldsymbol{P}_s^\top \boldsymbol{D}_s \boldsymbol{P}_s = \begin{bmatrix} \overline{\boldsymbol{J}}_{n_1^{(s)}} & & \\ & \ddots & \\ & & \overline{\boldsymbol{J}}_{n_{J_s}^{(s)}} \end{bmatrix}, \text{ is a } N_s \times N_s \text{ matrix.}$$

Fifth : A dataset level the following expressions are defined;

$N = \sum_{s=1}^{S} N_s$: It is the number of observations in the database

$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}^{(1)} \\ \vdots \\ \boldsymbol{Y}^{(S)} \end{pmatrix}$: It is a $N \times 1$ vector.

$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}^{(1)} \\ \vdots \\ \boldsymbol{X}^{(S)} \end{pmatrix}$: It is a $N \times K_1$ matrix.

$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{Z}^{(1)} \\ \vdots \\ \boldsymbol{Z}^{(S)} \end{pmatrix}$: It is a matrix of dimension $N \times K_2$

$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{W}^{(1)} \\ \vdots \\ \boldsymbol{W}^{(S)} \end{pmatrix}$: It is a matrix of dimension $N \times L$ matrix.

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}^{(1)} \\ \vdots \\ \boldsymbol{\theta}^{(S)} \end{pmatrix}$$: It is a $J_s \times 1$ vector.

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{Q}_1 & & \\ & \ddots & \\ & & \boldsymbol{Q}_S \end{bmatrix}$$: It is a matrix of dimension $N \times N$.

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{P}_1 & & \\ & \ddots & \\ & & \boldsymbol{P}_S \end{bmatrix}$$: It is a matrix of dimension $J \times J$.

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{D}_1 & & \\ & \ddots & \\ & & \boldsymbol{D}_S \end{bmatrix}$$: It is a matrix of dimension $J \times J$.

$$\boldsymbol{D}^{-1} = \begin{bmatrix} \boldsymbol{D}_1^{-1} & & \\ & \ddots & \\ & & \boldsymbol{D}_S^{-1} \end{bmatrix}$$: It is a matrix of dimension $J \times J$. It can be written as $\boldsymbol{D}^{-1} = \boldsymbol{P}\boldsymbol{P}^\top$

B.2 Study of Simulation

B.2.1 Results in scenario I, different parameter of association by endoguenous variables for each group

Table B.1: Results of Monte Carlo simulations , Scenario I

Monte Carlo estimations			
Mean VA	Mean θ	standard deviation VA	standard deviation θ
0.003	0.258	0.007	0.576
0.003	0.225	0.009	0.551
0.003	0.256	0.009	0.600
0.004	0.303	0.010	0.684

Table B.2: Results of Monte Carlo simulations for fixed effects estimation on first level ($\boldsymbol{\beta}$,γ), Scenario I

		Monte Carlo estimations							
	Real	Sample= 50		Sample= 100		Sample= 500		Sample= 1000	
$Vec(\boldsymbol{\beta})$	$\begin{pmatrix} 1.0 \\ 3.0 \end{pmatrix}$	0.980 2.951	(0.053) (0.093)	1.000 2.938	(0.069) (0.102)	0.987 2.947	(0.066) (0.098)	0.990 2.941	(0.063) (0.103)
$Vec(\gamma)$	$\begin{pmatrix} 7.0 \\ 4.0 \end{pmatrix}$	6.995 4.008	(0.010) (0.012)	6.994 4.009	(0.011) (0.015)	6.996 4.008	(0.010) (0.014)	6.995 4.009	(0.011) (0.014)

Table B.3: Results of Monte Carlo simulations for fixed effects's estimation ($\boldsymbol{A}$, $\boldsymbol{H}$) on second level, Scenario I

	Real	Monte Carlo estimations: Sample= 50		Sample= 100		Sample= 500		Sample= 1000	
Vec(A)	6.8	6.801	(0.021)	6.800	(0.018)	6.800	(0.017)	6.800	(0.017)
	5.2	5.204	(0.012)	5.198	(0.018)	5.200	(0.016)	5.200	(0.016)
	−0.9	− 0.897	(0.019)	− 0.901	(0.020)	− 0.900	(0.019)	− 0.900	(0.019)
	6.1	6.101	(0.015)	6.098	(0.017)	6.100	(0.016)	6.101	(0.015)
Vec(H)	3.6	3.599	(0.017)	3.597	(0.020)	3.598	(0.019)	3.600	(0.018)
	2.1	2.101	(0.019)	2.101	(0.017)	2.099	(0.017)	2.100	(0.017)
	0.5	0.500	(0.016)	0.500	(0.017)	0.499	(0.017)	0.501	(0.018)
	2.3	2.298	(0.015)	2.302	(0.016)	2.300	(0.015)	2.301	(0.015)
	4.8	4.799	(0.017)	4.803	(0.014)	4.799	(0.019)	4.800	(0.019)
	2.9	2.904	(0.017)	2.902	(0.014)	2.900	(0.016)	2.900	(0.016)

Table B.4: Results of Monte Carlo simulations for variance estimation (σ_s^2), Scenario I

	Real	Monte Carlo estimations: Sample= 50		Sample= 100		Sample= 500		Sample= 1000	
σ_1^2	38	37.995	(4.828)	38.015	(4.882)	37.739	(4.966)	38.087	(4.732)
σ_2^2	25	24.358	(3.086)	24.536	(3.378)	24.813	(2.978)	24.529	(3.028)
σ_3^2	42	43.117	(5.190)	42.384	(5.757)	41.885	(6.516)	41.795	(6.025)
σ_4^2	32	31.273	(4.157)	31.844	(4.041)	31.675	(3.951)	31.857	(4.088)
σ_5^2	21	19.969	(3.350)	21.556	(3.279)	20.842	(3.323)	20.974	(3.414)

Table B.5: Results of Monte Carlo simulations for variance estimation (τ_s^2), Scenario I

	Real	Monte Carlo estimations: Sample= 50		Sample= 100		Sample= 500		Sample= 1000	
τ_1^2	28	24.203	(4.312)	25.976	(5.759)	25.316	(5.000)	25.034	(5.508)
τ_2^2	15	12.835	(2.976)	12.838	(3.250)	13.002	(3.341)	12.807	(3.392)
τ_3^2	32	30.087	(7.612)	31.948	(8.528)	31.291	(8.061)	31.552	(8.569)
τ_4^2	22	20.788	(4.284)	22.102	(4.490)	21.238	(4.207)	21.422	(4.825)
τ_5^2	11	10.470	(2.824)	10.567	(3.073)	10.340	(2.509)	10.502	(3.102)

Table B.6: Results of Monte Carlo simulations for estimation of variance covariance matrix ($\mathbf{\Phi}_s$), Scenario I

		Monte Carlo estimations							
	Real	Sample= 50		Sample= 100		Sample= 500		Sample= 1000	
$Vec(\mathbf{\Phi_1})$	7.0	7.005	(0.174)	7.001	(0.165)	7.015	(0.168)	6.993	(0.162)
	3.0	2.990	(0.111)	3.001	(0.111)	3.008	(0.100)	2.996	(0.105)
	3.0	2.990	(0.111)	3.001	(0.111)	3.008	(0.100)	2.996	(0.105)
	4.0	3.994	(0.099)	4.002	(0.106)	4.004	(0.093)	3.994	(0.095)
$Vec(\mathbf{\Phi_2})$	3.5	3.511	(0.083)	3.505	(0.084)	3.497	(0.087)	3.500	(0.083)
	1.3	1.306	(0.072)	1.300	(0.090)	1.300	(0.081)	1.300	(0.083)
	1.3	1.306	(0.072)	1.300	(0.090)	1.300	(0.081)	1.300	(0.083)
	6.0	6.028	(0.120)	6.011	(0.145)	6.001	(0.139)	5.993	(0.143)
$Vec(\mathbf{\Phi_3})$	9.0	8.995	(0.247)	9.008	(0.199)	9.003	(0.205)	9.005	(0.222)
	4.1	4.127	(0.162)	4.116	(0.124)	4.099	(0.140)	4.099	(0.150)
	4.1	4.127	(0.162)	4.116	(0.124)	4.099	(0.140)	4.099	(0.150)
	6.5	6.527	(0.156)	6.518	(0.154)	6.497	(0.150)	6.499	(0.156)
$Vec(\mathbf{\Phi_4})$	6.0	5.963	(0.136)	6.024	(0.140)	6.012	(0.153)	5.993	(0.145)
	1.5	1.506	(0.075)	1.509	(0.088)	1.506	(0.087)	1.498	(0.084)
	1.5	1.506	(0.075)	1.509	(0.088)	1.506	(0.087)	1.498	(0.084)
	4.0	4.010	(0.064)	4.002	(0.099)	4.000	(0.093)	3.999	(0.098)
$Vec(\mathbf{\Phi_5})$	5.3	5.308	(0.128)	5.304	(0.130)	5.299	(0.134)	5.296	(0.127)
	2.9	2.911	(0.105)	2.896	(0.091)	2.897	(0.093)	2.896	(0.092)
	2.9	2.911	(0.105)	2.896	(0.091)	2.897	(0.093)	2.896	(0.092)
	3.8	3.820	(0.109)	3.797	(0.088)	3.799	(0.092)	3.798	(0.093)

Table B.7: Results of Monte Carlo simulations for the parameter of marginal effect of te school effect and endogenous covariates ($\boldsymbol{\delta}_s$), Scenario I

	Real	Monte Carlo estimations							
		Sample= 50		Sample= 100		Sample= 500		Sample= 1000	
$\boldsymbol{\delta}_1$	$\begin{pmatrix} 1.57 \\ 3.38 \end{pmatrix}$	1.770 3.453	(0.209) (0.235)	1.735 3.473	(0.278) (0.319)	1.822 3.469	(1.636) (0.245)	1.778 3.477	(0.265) (0.268)
$\boldsymbol{\delta}_2$	$\begin{pmatrix} 2.39 \\ 1.13 \end{pmatrix}$	2.708 1.245	(0.450) (0.447)	2.657 1.285	(0.439) (0.376)	2.730 1.276	(1.155) (0.405)	2.687 1.258	(1.043) (0.426)
$\boldsymbol{\delta}_3$	$\begin{pmatrix} -4.70 \\ 3.10 \end{pmatrix}$	− 5.014 3.377	(0.541) (0.446)	− 4.924 3.326	(0.518) (0.386)	− 4.924 3.326	(0.545) (0.419)	− 4.966 3.373	(0.612) (0.502)
$\boldsymbol{\delta}_4$	$\begin{pmatrix} 4.1 \\ 1.08 \end{pmatrix}$	4.275 1.130	(0.372) (0.099)	4.117 1.136	(0.293) (0.129)	4.206 1.147	(0.354) (0.148)	4.200 1.156	(0.418) (0.329)
$\boldsymbol{\delta}_5$	$\begin{pmatrix} 5.13 \\ 2.4 \end{pmatrix}$	5.437 2.464	(0.527) (0.310)	5.357 2.499	(0.568) (0.272)	5.391 2.485	(0.641) (0.287)	5.406 2.472	(0.771) (0.329)

B.2.2 Results in scenario II, different parameter of association by endoguenous variables for each group

Table B.8: Results of Monte Carlo simulations , Scenario II

Monte Carlo estimations			
Mean VA	Mean θ	standard deviation VA	standard deviation θ
0.006	0.147	0.014	0.278
0.007	0.158	0.015	0.303
0.007	0.156	0.017	0.308
0.007	0.157	0.016	0.302

Table B.9: Results of Monte Carlo simulations for fixed effects's estimation on first level ($\boldsymbol{\beta}$,$\boldsymbol{\gamma}$), Scenario II

	Real	Monte Carlo estimations							
		Sample= 50		Sample= 100		Sample= 500		Sample= 1000	
$Vec(\boldsymbol{\beta})$	1.0	0.991	(0.061)	0.985	(0.060)	0.990	(0.065)	0.991	(0.064)
	3.0	3.151	(0.100)	3.180	(0.098)	3.177	(0.101)	3.168	(0.105)
$Vec(\boldsymbol{\gamma})$	7.0	7.020	(0.011)	7.022	(0.009)	7.021	(0.011)	7.020	(0.011)
	4.0	3.977	(0.014)	3.974	(0.014)	3.975	(0.014)	3.975	(0.014)

Table B.10: Results of Monte Carlo simulations for fixed effects's estimation on second level ($\boldsymbol{A}$, $\boldsymbol{H}$), Scenario II

	Real	Monte Carlo estimations							
		Sample= 50		Sample= 100		Sample= 500		Sample= 1000	
Vec(A)	6.8	6.800	(0.016)	6.801	(0.016)	6.802	(0.018)	6.801	(0.017)
	5.2	5.198	(0.015)	5.201	(0.013)	5.200	(0.016)	5.201	(0.015)
	−0.9	− 0.905	(0.014)	− 0.899	(0.015)	− 0.901	(0.018)	− 0.899	(0.018)
	6.1	6.100	(0.013)	6.103	(0.016)	6.100	(0.015)	6.101	(0.015)
Vec(H)	3.6	3.601	(0.017)	3.599	(0.020)	3.601	(0.018)	3.600	(0.017)
	2.1	2.099	(0.014)	2.100	(0.015)	2.100	(0.016)	2.100	(0.015)
	0.5	0.501	(0.017)	0.497	(0.018)	0.501	(0.018)	0.500	(0.018)
	2.3	2.302	(0.017)	2.297	(0.017)	2.300	(0.015)	2.300	(0.016)
	4.8	4.801	(0.015)	4.802	(0.017)	4.800	(0.019)	4.800	(0.017)
	2.9	2.899	(0.012)	2.903	(0.016)	2.902	(0.016)	2.900	(0.015)

Table B.11: Results of Monte Carlo simulations for estimation of variance (σ_s^2), Scenario II

		Monte Carlo estimations							
	Real	Sample= 50		Sample= 100		Sample= 500		Sample= 1000	
σ_1^2	38	37.634	(4.524)	37.166	(5.274)	37.408	(5.124)	37.335	(5.013)
σ_2^2	25	26.342	(2.602)	25.808	(3.474)	25.417	(3.196)	25.351	(3.094)
σ_3^2	42	41.967	(5.781)	41.511	(5.822)	42.077	(6.036)	41.181	(5.944)
σ_4^2	32	30.763	(4.062)	31.375	(4.289)	31.336	(3.976)	31.558	(4.029)
σ_5^2	21	20.541	(3.483)	20.262	(3.228)	21.091	(3.437)	20.818	(3.402)

Table B.12: Results of Monte Carlo simulations for estimation of variance school effect (τ_s^2), Scenario II

		Monte Carlo estimations							
	Real	Sample= 50		Sample= 100		Sample= 500		Sample= 1000	
τ_1^2	28	24.350	(3.307)	24.558	(3.678)	24.502	(3.677)	24.643	(3.809)
τ_2^2	15	13.386	(2.348)	12.724	(1.995)	13.078	(2.207)	13.020	(2.164)
τ_3^2	32	27.888	(4.116)	28.001	(4.261)	28.448	(4.455)	28.387	(4.100)
τ_4^2	22	19.513	(3.126)	19.348	(2.966)	19.336	(2.954)	19.320	(3.066)
τ_5^2	11	9.533	(1.592)	9.297	(1.632)	9.313	(1.702)	9.407	(1.712)

Table B.13: Results of Monte Carlo simulations for estimation of variance covariance matrix ($\mathbf{\Phi}_s$), Scenario II

		Monte Carlo estimations							
	Real	Sample= 50		Sample= 100		Sample= 500		Sample= 1000	
$Vec(\mathbf{\Phi_1})$	$\begin{pmatrix} 7.0 \\ 3.0 \\ 3.0 \\ 4.0 \end{pmatrix}$	6.982 2.991 2.991 4.007	(0.168) (0.109) (0.109) (0.102)	7.003 2.998 2.998 4.004	(0.157) (0.097) (0.097) (0.092)	6.994 2.998 2.998 4.001	(0.166) (0.106) (0.106) (0.097)	6.992 2.997 2.997 3.996	(0.166) (0.101) (0.101) (0.094)
$Vec(\mathbf{\Phi_2})$	$\begin{pmatrix} 3.5 \\ 1.3 \\ 1.3 \\ 6.0 \end{pmatrix}$	3.472 1.286 1.286 5.981	(0.085) (0.081) (0.081) (0.121)	3.491 1.307 1.307 5.999	(0.085) (0.081) (0.081) (0.139)	3.500 1.302 1.302 6.000	(0.085) (0.079) (0.079) (0.148)	3.498 1.296 1.296 5.999	(0.087) (0.080) (0.080) (0.149)

Continued on the next page

$Vec(\mathbf{\Phi_3})$	$\begin{pmatrix} 9.0 \\ 4.1 \\ 4.1 \\ 6.5 \end{pmatrix}$	8.974 (0.223) 4.114 (0.167) 4.114 (0.167) 6.525 (0.149)	9.020 (0.199) 4.114 (0.140) 4.114 (0.140) 6.528 (0.154)	8.979 (0.211) 4.090 (0.146) 4.090 (0.146) 6.507 (0.159)	8.994 (0.209) 4.102 (0.146) 4.102 (0.146) 6.506 (0.159)
$Vec(\mathbf{\Phi_4})$	$\begin{pmatrix} 6.0 \\ 1.5 \\ 1.5 \\ 4.0 \end{pmatrix}$	5.975 (0.132) 1.501 (0.075) 1.501 (0.075) 4.005 (0.062)	5.984 (0.129) 1.514 (0.079) 1.514 (0.079) 4.001 (0.098)	5.998 (0.147) 1.499 (0.086) 1.499 (0.086) 4.000 (0.092)	6.001 (0.140) 1.505 (0.087) 1.505 (0.087) 4.004 (0.098)
$Vec(\mathbf{\Phi_5})$	$\begin{pmatrix} 5.3 \\ 2.9 \\ 2.9 \\ 3.8 \end{pmatrix}$	5.298 (0.125) 2.889 (0.094) 2.889 (0.094) 3.791 (0.096)	5.290 (0.107) 2.887 (0.081) 2.887 (0.081) 3.786 (0.083)	5.293 (0.124) 2.898 (0.090) 2.898 (0.090) 3.797 (0.093)	5.298 (0.133) 2.897 (0.091) 2.897 (0.091) 3.798 (0.090)

Table B.14: Results of Monte Carlo simulations for the parameter of marginal effect of school effect and endogenous covariates ($\boldsymbol{\delta}_s$), Scenario II

		Monte Carlo estimations			
	Real	Sample= 50	Sample= 100	Sample= 500	Sample= 1000
$\boldsymbol{\delta}_1$	$\begin{pmatrix} 2.0 \\ -0.5 \end{pmatrix}$	2.126 (0.098) − 0.531 (0.022)	2.149 (0.087) − 0.535 (0.024)	2.146 (0.103) − 0.536 (0.027)	2.146 (0.110) − 0.536 (0.029)
$\boldsymbol{\delta}_2$	$\begin{pmatrix} 2.0 \\ -0.5 \end{pmatrix}$	2.153 (0.128) − 0.542 (0.033)	2.197 (0.137) − 0.553 (0.036)	2.165 (0.154) − 0.543 (0.043)	2.164 (0.141) − 0.541 (0.038)
$\boldsymbol{\delta}_3$	$\begin{pmatrix} 2.0 \\ -0.5 \end{pmatrix}$	2.134 (0.096) − 0.533 (0.024)	2.144 (0.101) − 0.538 (0.028)	2.135 (0.102) − 0.534 (0.028)	2.132 (0.097) − 0.533 (0.026)
$\boldsymbol{\delta}_4$	$\begin{pmatrix} 2.0 \\ -0.5 \end{pmatrix}$	2.164 (0.148) − 0.541 (0.040)	2.145 (0.104) − 0.535 (0.029)	2.150 (0.106) − 0.537 (0.028)	2.152 (0.112) − 0.538 (0.030)
$\boldsymbol{\delta}_5$	$\begin{pmatrix} 2.0 \\ -0.5 \end{pmatrix}$	2.176 (0.157) − 0.539 (0.039)	2.195 (0.143) − 0.546 (0.037)	2.203 (0.205) − 0.551 (0.054)	2.198 (0.186) − 0.550 (0.047)

B.2.3 Results in scenario III, same parameter of association by endoguenous variables for each group

Table B.15: Results of Monte Carlo simulations , Scenario III

Monte Carlo estimations			
Mean VA	Mean θ	standard deviation VA	standard deviation θ
0.057	0.054	0.134	0.129
0.061	0.053	0.144	0.130
0.056	0.049	0.126	0.115
0.057	0.050	0.130	0.119

Table B.16: Results of Monte Carlo simulations for fixed effects's estimation on first level ($\boldsymbol{\beta}$,$\boldsymbol{\gamma}$), Scenario III

	Real	Monte Carlo estimations Sample= 50		Sample= 100		Sample= 500		Sample= 1000	
β	1.0	1.103	(0.071)	1.100	(0.062)	1.112	(0.060)	1.112	(0.063)
	3.0	3.170	(0.100)	3.159	(0.121)	3.152	(0.111)	3.164	(0.110)
γ	7.0	7.006	(0.010)	7.005	(0.013)	7.003	(0.011)	7.004	(0.011)
	4.0	3.973	(0.013)	3.975	(0.016)	3.975	(0.015)	3.974	(0.015)

Table B.17: Results of Monte Carlo simulations for fixed effects's estimation on second level ($\boldsymbol{A}$, $\boldsymbol{H}$), Scenario III

	Real	Monte Carlo estimations Sample= 50		Sample= 100		Sample= 500		Sample= 1000	
$Vec(\boldsymbol{A})$	6.8	6.800	(0.015)	6.801	(0.018)	6.801	(0.018)	6.799	(0.017)
	5.2	5.199	(0.016)	5.202	(0.018)	5.200	(0.015)	5.200	(0.015)
	-0.9	-0.897	(0.017)	-0.900	(0.019)	-0.902	(0.017)	-0.901	(0.018)
	6.1	6.105	(0.015)	6.101	(0.017)	6.099	(0.015)	6.100	(0.015)
$Vec(\boldsymbol{H})$	3.6	3.600	(0.018)	3.598	(0.017)	3.601	(0.017)	3.600	(0.017)
	2.1	2.102	(0.015)	2.099	(0.015)	2.101	(0.016)	2.100	(0.015)
	0.5	0.497	(0.017)	0.500	(0.018)	0.499	(0.018)	0.500	(0.018)
	2.3	2.297	(0.017)	2.298	(0.016)	2.300	(0.017)	2.300	(0.015)
	4.8	4.799	(0.016)	4.798	(0.017)	4.802	(0.017)	4.800	(0.018)
	2.9	2.900	(0.015)	2.899	(0.015)	2.901	(0.015)	2.900	(0.016)

Table B.18: Results of Monte Carlo simulations for variance estimation (σ_s^2), Scenario III

	Real	Monte Carlo estimations Sample= 50		Sample= 100		Sample= 500		Sample= 1000	
σ_1^2	38	38.820	(4.802)	39.175	(5.241)	39.302	(4.999)	39.542	(5.012)
σ_2^2	25	26.202	(3.728)	26.241	(3.453)	26.826	(3.103)	26.378	(3.234)
σ_3^2	42	44.057	(6.064)	43.120	(6.231)	44.200	(5.776)	44.086	(6.090)
σ_4^2	32	33.211	(4.250)	33.237	(3.689)	32.898	(4.047)	33.024	(3.847)
σ_5^2	21	22.670	(3.976)	21.973	(3.370)	22.589	(3.224)	22.458	(3.363)

Table B.19: Results of Monte Carlo simulations for variance estimation (τ_s^2), Scenario III

	Real	Monte Carlo estimations Sample= 50		Sample= 100		Sample= 500		Sample= 1000	
τ_1^2	28	27.346	(3.078)	27.609	(3.165)	27.779	(3.448)	27.897	(3.612)
τ_2^2	15	15.254	(1.870)	15.060	(1.733)	15.036	(1.979)	15.025	(1.938)
τ_3^2	32	31.414	(3.592)	31.689	(4.242)	31.719	(3.945)	31.666	(3.859)
τ_4^2	22	21.825	(2.471)	21.975	(2.977)	21.769	(2.619)	21.963	(2.817)
τ_5^2	11	10.752	(1.401)	11.110	(1.640)	11.185	(1.510)	11.100	(1.475)

Table B.20: Results of Monte Carlo simulations for estimation of variance covariance matrix ($\mathbf{\Phi}_s$), Scenario III

	Real	Monte Carlo estimations Sample= 50		Sample= 100		Sample= 500		Sample= 1000	
$Vec(\mathbf{\Phi_1})$	7.0	6.979	(0.171)	6.990	(0.153)	6.999	(0.170)	7.004	(0.166)
	3.0	2.993	(0.111)	2.993	(0.096)	2.997	(0.107)	3.000	(0.102)
	3.0	2.993	(0.111)	2.993	(0.096)	2.997	(0.107)	3.000	(0.102)
	4.0	4.003	(0.094)	3.990	(0.099)	3.997	(0.096)	4.000	(0.095)
$Vec(\mathbf{\Phi_2})$	3.5	3.484	(0.083)	3.498	(0.093)	3.503	(0.088)	3.500	(0.082)
	1.3	1.299	(0.077)	1.300	(0.073)	1.303	(0.078)	1.300	(0.081)
	1.3	1.299	(0.077)	1.300	(0.073)	1.303	(0.078)	1.300	(0.081)
	6.0	5.983	(0.146)	5.989	(0.127)	6.001	(0.140)	6.001	(0.147)

Continued on the next page

$Vec(\mathbf{\Phi_3})$	9.0	9.001 (0.220)	8.993 (0.233)	8.999 (0.214)	9.008 (0.214)
	4.1	4.099 (0.139)	4.100 (0.150)	4.105 (0.148)	4.106 (0.146)
	4.1	4.099 (0.139)	4.100 (0.150)	4.105 (0.148)	4.106 (0.146)
	6.5	6.481 (0.170)	6.522 (0.146)	6.503 (0.158)	6.498 (0.151)
$Vec(\mathbf{\Phi_4})$	6.0	5.978 (0.138)	6.009 (0.146)	6.000 (0.140)	6.000 (0.139)
	1.5	1.489 (0.074)	1.496 (0.088)	1.503 (0.085)	1.504 (0.086)
	1.5	1.489 (0.074)	1.496 (0.088)	1.503 (0.085)	1.504 (0.086)
	4.0	4.013 (0.079)	3.997 (0.087)	3.999 (0.095)	4.006 (0.095)
$Vec(\mathbf{\Phi_5})$	5.3	5.308 (0.135)	5.314 (0.142)	5.298 (0.127)	5.302 (0.123)
	2.9	2.891 (0.101)	2.907 (0.103)	2.899 (0.091)	2.900 (0.092)
	2.9	2.891 (0.101)	2.907 (0.103)	2.899 (0.091)	2.900 (0.092)
	3.8	3.790 (0.103)	3.800 (0.098)	3.797 (0.092)	3.796 (0.093)

Table B.21: Results of Monte Carlo simulations for the parameter of marginal effect of school effect and endogenous covariates ($\boldsymbol{\delta}_s$), Scenario III

		Monte Carlo estimations			
	Real	Sample= 50	Sample= 100	Sample= 500	Sample= 1000
$\boldsymbol{\delta}_1$	-0.5	-0.502 (0.015)	-0.504 (0.015)	-0.504 (0.015)	-0.504 (0.014)
	-0.5	-0.505 (0.015)	-0.504 (0.014)	-0.504 (0.013)	-0.504 (0.013)
$\boldsymbol{\delta}_2$	-0.5	-0.498 (0.017)	-0.500 (0.016)	-0.502 (0.017)	-0.502 (0.016)
	-0.5	-0.501 (0.019)	-0.502 (0.019)	-0.503 (0.019)	-0.503 (0.018)
$\boldsymbol{\delta}_3$	-0.5	-0.505 (0.014)	-0.501 (0.016)	-0.504 (0.015)	-0.504 (0.015)
	-0.5	-0.505 (0.015)	-0.502 (0.016)	-0.504 (0.013)	-0.505 (0.014)
$\boldsymbol{\delta}_4$	-0.5	-0.506 (0.016)	-0.504 (0.012)	-0.504 (0.015)	-0.505 (0.016)
	-0.5	-0.507 (0.014)	-0.506 (0.013)	-0.504 (0.014)	-0.504 (0.015)
$\boldsymbol{\delta}_5$	-0.5	-0.503 (0.022)	-0.501 (0.021)	-0.501 (0.021)	-0.502 (0.020)
	-0.5	-0.503 (0.019)	-0.500 (0.020)	-0.501 (0.020)	-0.501 (0.019)

C Appendix of chapter III: On the modeling of school improvement through a time dependent value-added model

C.1 Technical Appendix

C.1.1 Notation

First : For all dataset is defined the following expressions;

$\mathcal{J}$: It represents the set of all school.

J : It represents the total school, card $(\mathcal{J})$.

n_{jt} : It represents total students in school j on the time t, $j \in \mathcal{J}_s$.

Second : A student level the following expressions are defined;

Y_{ijt} : It is the score (scalar), for student i belonging to school j . Such that $i = 1, \ldots, n_{jt}$, with n_{jt} total students in school j on the time t and $j \in \mathcal{J}$.

$\boldsymbol{X}_{ijt}$: It is a K_t-dimensional vector of K_t-explanatory variables for student i, in the school j on the time t. The vector of explanatory variables included as variable to the prior score of student i.

Third : A school level the following expressions are defined;

$\boldsymbol{Y}_{jt} = \begin{pmatrix} Y_{1jt} \\ \vdots \\ Y_{n_{jt}jt} \end{pmatrix}$: It is a vector of all scores in school j on the time t, such that $\boldsymbol{Y}_{jt} \in \mathbb{R}^{n_{jt}}$.

$\boldsymbol{X}_{jt} = \begin{pmatrix} \boldsymbol{X}_{1jt}^{\top} \\ \vdots \\ \boldsymbol{X}_{n_{jt}jt}^{\top} \end{pmatrix}$: It is a matrix of dimension $n_{jt} \times K_t$, where K_t is the number of covariates.

$\theta_{jt,t-1}$: It represents the school effect of school j on the cohort t–$t-1$, $j \in \mathcal{J}$.

Also, we define others matrices very useful, (to school level).

$\imath_{n_{jt}} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$: It is a vector 1's of dimension $n_{jt} \times 1$, for $j = 1, 2 \ldots, J$.

$\boldsymbol{J}_{n_{jt}} = \imath_{n_{jt}} \imath_{n_{jt}}^{\top}$: It is a matrix of dimension $n_{jt} \times n_{jt}$ of 1's.

$\overline{\boldsymbol{J}}_{n_{jt}} = \boldsymbol{J}_{n_{jt}} / n_{jt}$: It is a matrix of dimension $n_{jt} \times n_{jt}$ of $1/n_{jt}$'s.

$\boldsymbol{E}_{n_{jt}} = \boldsymbol{I}_{n_{jt}} - \overline{\boldsymbol{J}}_{n_{jt}}$: It is matrix of dimension $n_{jt} \times n_{jt}$.

Fourth : For each t the following expressions are defined;

$N_t = \sum_{j \in \mathcal{J}} n_{jt}$: Total of students on the time t.

$\boldsymbol{Y}_{t} = \begin{pmatrix} \boldsymbol{Y}_{1t} \\ \vdots \\ \boldsymbol{Y}_{Jt} \end{pmatrix}$: It is a vector of dimension $N_t \times 1$.

$$\boldsymbol{X}_t = \begin{pmatrix} \boldsymbol{X}_{1t} \\ \vdots \\ \boldsymbol{X}_{Jt} \end{pmatrix} \quad : \quad \text{It is a matrix of dimension } N_t \times K_t.$$

$$\boldsymbol{\theta}_t = \begin{pmatrix} \theta_{1t,t-1} \\ \vdots \\ \theta_{Jt,t-1} \end{pmatrix} \quad : \quad \text{It is a vector of dimension } J.$$

$$\boldsymbol{L}_{\underline{n_t}} = \begin{bmatrix} \imath_{n_{1t}} & & \\ & \ddots & \\ & & \imath_{n_{Jt}} \end{bmatrix} \quad : \quad \text{It is a } N_t \times J \text{ matrix.}$$

$$\boldsymbol{D}_t = \begin{bmatrix} n_{1t} & & \\ & \ddots & \\ & & n_{Jt} \end{bmatrix} \quad : \quad \text{It is a diagonal matrix of dimension } J \times J$$

$$\boldsymbol{P}_t = \begin{bmatrix} \imath_{n_{1t}}^\top / n_{1t} & & \\ & \ddots & \\ & & \imath_{n_{Jt}}^\top / n_{Jt} \end{bmatrix}$$

: It is a matrix of dimension $J \times N_t$. Also It is called Between-matrix, and It can be obtained by $\boldsymbol{P}_t = \boldsymbol{D}_t^{-1} \boldsymbol{L}_{\underline{n_t}}^\top$

$$\boldsymbol{Q}_t = \begin{bmatrix} \boldsymbol{E}_{n_{1t}} & & \\ & \ddots & \\ & & \boldsymbol{E}_{n_{Jt}} \end{bmatrix}$$

: It is a idempotent matrix of dimension $N_t \times N_t$. Also, It is called Within-matrix .

C.1.2 Joint Distribution

From the section 3.2.1, one can obtain the following calculations of averages.

1. $\mu_{\theta_3} = \mathbf{E}(\theta_{j3,1}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}) = 0$

   $\Rightarrow \mu_{\theta_3} = 0$

2. $\mu_{\theta_4} = \mathbf{E}(\theta_{j4,2}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}) = \mathbf{E}(\mathbf{E}(\theta_{j4,2}|\theta_{j3,1}, , \mathbf{X}_{j4}, \mathbf{X}_{j3})|, \mathbf{X}_{j4}, \mathbf{X}_{j3})$
   $= \mathbf{E}(\mu\theta_{j3,1}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}) = \mu\mathbf{E}(\theta_{j3,1}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}) =$

   $\Rightarrow \mu_{\theta_4} = 0$

3. $\mu_{j3} = \mathbf{E}(\mathbf{Y}_{j3}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}) = \mathbf{E}(\mathbf{E}(\mathbf{Y}_{j3}|\theta_{j3,1}, , \mathbf{X}_{j4}, \mathbf{X}_{j3})|, \mathbf{X}_{j4}, \mathbf{X}_{j3})$
   $= \mathbf{E}(\mathbf{X}_{j3}\beta_{\mathbf{3}} + \iota_{n_{j3}}\theta_{j3,1}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}) = \mathbf{X}_{j3}\beta_{\mathbf{3}} + \mathbf{E}(\iota_{n_{j3}}\theta_{j3,1}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}) = \mathbf{X}_{j3}\beta_{\mathbf{3}}$
   $= \mathbf{X}_{j3}\beta_{\mathbf{3}}$

   $\Rightarrow \mu_{j3} = \mathbf{X}_{j3}\beta_{\mathbf{3}}$

4. $\mu_{j4} = \mathbf{E}(\mathbf{Y}_{j4}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}) = \mathbf{E}(\mathbf{E}(\mathbf{Y}_{j4}|\mathbf{Y}_{j3}, \theta_{j4,2}, , \mathbf{X}_{j4}, \mathbf{X}_{j3})|, \mathbf{X}_{j4}, \mathbf{X}_{j3})$
   $= \mathbf{E}(\mathbf{X}_{j4}\beta_{\mathbf{4}} + \iota_{n_{j4}}\theta_{j4,2}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}) = \mathbf{X}_{j4}\beta_{\mathbf{4}} + \mathbf{E}(\iota_{n_{j4}}\theta_{j4,2}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}) = \mathbf{X}_{j4}\beta_{\mathbf{4}}$

   $\Rightarrow \mu_{j4} = \mathbf{X}_{j4}\beta_{\mathbf{4}}$

From the section 3.2.1, one can obtain the following calculations of variances,

1. $\Sigma_{44} = \mathbf{V}(\theta_{j3,1}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}) = \tau_3^2$
$\Rightarrow \Sigma_{44} = \tau_3^2$

2. $\Sigma_{33} = \mathbf{V}(\theta_{j4,2}|, \mathbf{X}_{j4}, \mathbf{X}_{j3})$
$= \mathbf{E}(\mathbf{V}(\theta_{j4,2}|\mathbf{Y}_{j3}, \theta_{j3,1}, , \mathbf{X}_{j4}, \mathbf{X}_{j3})|, \mathbf{X}_{j4}, \mathbf{X}_{j3})$
$+ \mathbf{V}(\mathbf{E}(\theta_{j4,2}|\mathbf{Y}_{j3}, \theta_{j3,1}, , \mathbf{X}_{j4}, \mathbf{X}_{j3})|, \mathbf{X}_{j4}, \mathbf{X}_{j3})$
$= \tau_4^2 + \mu^2 \mathbf{V}(\theta_{3j}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}) = \tau_4^2 + \mu^2 \tau_3^2$
$\Rightarrow \Sigma_{33} = \tau_4^2 + \mu^2 \tau_3^2$

3. $\Sigma_{22} = \mathbf{V}(\mathbf{Y}_{j3}|, \mathbf{X}_{j4}, \mathbf{X}_{j3})$
$= \mathbf{E}(\sigma_3^2 \mathbf{I}_{n_{j3}}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}) + \mathbf{V}(\mathbf{X}_{j3}\beta_2 + \iota_{n_{j3}}\theta_{j3,1}|, \mathbf{X}_{j4}, \mathbf{X}_{j3})$
$= \sigma_3^2 \mathbf{I}_{n_{j3}} + \iota_{n_{j3}} \tau_3^2 \iota_{n_{j3}}^\top = \sigma_3^2 \mathbf{I}_{n_{j3}} + \tau_3^2 n_{j3} \frac{1}{n_{j3}} \iota_{n_{j3}} \iota_{n_{j3}}^\top = \sigma_3^2 \mathbf{I}_{n_{j3}} + \tau_3^2 n_{j3} \bar{\mathbf{J}}_{n_{j3}}$
$= \sigma_3^2 \mathbf{E}_{n_{j3}} + (\sigma_3^2 + \tau_3^2 n_{j3}) \bar{\mathbf{J}}_{n_{j3}}$
$\Rightarrow \Sigma_{22} = \sigma_3^2 \mathbf{E}_{n_{j3}} + (\sigma_3^2 + \tau_3^2 n_{j3}) \bar{\mathbf{J}}_{n_{j3}}$

4. $\Sigma_{11} = \mathbf{V}(\mathbf{Y}_{j4}|, \mathbf{X}_{j4}, \mathbf{X}_{j3})$
$= \mathbf{E}(\sigma_4^2 \mathbf{I}_{n_{j4}}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}) + \mathbf{V}(\mathbf{X}_{j4}\beta_4 + \iota_{n_{j4}}\theta_{j4,2}|, \mathbf{X}_{j4}, \mathbf{X}_{j3})$
$= \sigma_4^2 \mathbf{I}_{n_{j4}} + \iota_{n_{j4}} (\tau_4^2 + \mu^2 \tau_3^2) \iota_{n_{j4}}^\top = \sigma_4^2 \mathbf{I}_{n_{j4}} + (\tau_4^2 + \mu^2 \tau_3^2) n_{j4} \bar{\mathbf{J}}_{n_{j4}}$
$= \sigma_4^2 \mathbf{E}_{n_{j3}} + (\sigma_4^2 + (\tau_4^2 + \mu^2 \tau_3^2) n_{j4}) \bar{\mathbf{J}}_{n_{j4}}$
$\Rightarrow \Sigma_{11} = \sigma_4^2 \mathbf{E}_{n_{j4}} + (\sigma_4^2 + (\tau_4^2 + \mu^2 \tau_3^2) n_{j4}) \bar{\mathbf{J}}_{n_{j4}}$

From the section 3.2.1, one can obtain the following calculations of variances,

1. $\Sigma_{34} = \mathbf{Cov}\left(\theta_{j4,2}, \theta_{j3,1}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}\right)$
$= \mathbf{Cov}\left(\mu\theta_{j3,1}, \theta_{j3,1}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}\right) = \mu\mathbf{V}\left(\theta_{j3,1}, \theta_{j3,1}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}\right) = \mu\tau_3^2$
$\Rightarrow \Sigma_{34} = \mu\tau_3^2$

2. $\Sigma_{24} = \mathbf{Cov}\left(\mathbf{Y}_{j3}, \theta_{j3,1}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}\right)$
$=\mathbf{Cov}\left(\mathbf{X}_{j3}\beta_2 + \iota_{n_{j3}}\theta_{j3,1}, \theta_{j3,1}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}\right) = \iota_{n_{j3}}\mathbf{V}\left(\theta_{j3,1}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}\right)$
$= \iota_{n_{j3}}\tau_3^2$
$\Rightarrow \Sigma_{24} = \iota_{n_{j3}}\tau_3^2$

3. $\Sigma_{23} = \mathbf{Cov}\left(\mathbf{Y}_{j3}, \theta_{j4,2}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}\right)$
$= \mathbf{Cov}\left(\mathbf{X}_{j3}\beta_2 + \theta_{j3,1}\iota_{n_{j3}}, \theta_{j4,2}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}\right) = \mathbf{Cov}\left(\theta_{j3,1}, \theta_{j4,2}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}\right)\iota_{n_{j3}} = \mu\tau_3^2\iota_{n_{j3}}$
$\Rightarrow \Sigma_{23} = \mu\tau_3^2\iota_{n_{j3}}$

4. $\Sigma_{14} = \mathbf{Cov}\left(\mathbf{Y}_{j4}, \theta_{j3,1}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}\right)$
$= \mathbf{Cov}\left(\mathbf{X}_{j4}\beta_4 + \theta_{j4,2}\iota_{n_{j4}}, \theta_{j3,1}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}\right) = \mathbf{Cov}\left(\theta_{j3,1}, \theta_{j4,2}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}\right)\iota_{n_{j4}}$
$= \mu\tau_3^2\iota_{n_{j4}}$
$\Rightarrow \Sigma_{14} = \mu\tau_3^2\iota_{n_{j4}}$

5. $\Sigma_{13} = \mathbf{Cov}\left(\mathbf{Y}_{j4}, \theta_{j4,2}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}\right)$
$= \mathbf{Cov}\left(\mathbf{X}_{j4}\beta_4 + \theta_{j4,2}\iota_{n_{j4}}, \theta_{j4,2}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}\right) = \mathbf{V}\left(\theta_{j4,2}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}\right)\iota_{n_{j4}}$
$= (\tau_4^2 + \mu^2\tau_3^2)\iota_{n_{j4}}$
$\Rightarrow \Sigma_{14} = (\tau_4^2 + \mu^2\tau_3^2)\iota_{n_{j4}}$

6. $\Sigma_{12} = \mathbf{Cov}\left(\mathbf{Y}_{j4}, \mathbf{Y}_{j3}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}\right)$
$= \mathbf{Cov}\left(\mathbf{X}_{j4}\beta_4 + \theta_{j4,2}\iota_{n_{j4}}, \mathbf{X}_{j3}\beta_2 + \theta_{j3,1}\iota_{n_{j3}}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}\right)$
$= \mathbf{Cov}\left(\theta_{j4,2}\iota_{n_{j4}}, \theta_{j3,1}\iota_{n_{j3}}|, \mathbf{X}_{j4}, \mathbf{X}_{j3}\right) = \mu\tau_3^2\iota_{n_{j4}}\iota_{n_{j3}}^\top$
$\Rightarrow \Sigma_{12} = \mu\tau_3^2\iota_{n_{j4}}\iota_{n_{j3}}^\top$

C.1.3 Inverse of a matrix

Let, $\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ & \mathbf{A}_{22} \end{pmatrix}$ a symmetric matrix, such that; $\mathbf{AB} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix, ie, it is diagonal matrix of $1^\top$s and $\mathbf{B}$ is the inverse matrix of $\mathbf{A}$. Then,

$$\begin{aligned} \mathbf{AB} &= \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ & \mathbf{A}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ & \mathbf{B}_{22} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{12}^\top & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ & \mathbf{A}_{12}^\top\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ & \mathbf{I} \end{pmatrix} \end{aligned}$$

This way,

$$\mathbf{I} = \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{12}^\top \tag{C.1}$$

$$\mathbf{0} = \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \tag{C.2}$$

$$\mathbf{I} = \mathbf{A}_{12}^\top\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \tag{C.3}$$

Therefore, from (C.2) $\mathbf{A}_{11}\mathbf{B}_{12} = -\mathbf{A}_{12}\mathbf{B}_{22} \Rightarrow \mathbf{B}_{12} = -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22}$, substituting this equality into equation (C.3), it is obtained the following;

$$\begin{aligned} \mathbf{I} &= -\mathbf{A}_{12}^\top\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22} + \mathbf{A}_{22}\mathbf{B}_{22} \\ \mathbf{I} &= (\mathbf{A}_{22} - \mathbf{A}_{12}^\top\mathbf{A}_{11}^{-1}\mathbf{A}_{12})\mathbf{B}_{22} \end{aligned}$$

$$\Rightarrow \quad \mathbf{B}_{22} = (\mathbf{A}_{22} - \mathbf{A}_{12}^\top\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}.$$

Otherwise, substituting $\mathbf{B}_{12}$ into equation (C.1), it is obtained the following;

$$\begin{aligned}
\mathbf{I} \quad &= \quad \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}(-\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22})^{\top} \\
&= \quad \mathbf{A}_{11}\mathbf{B}_{11} - \mathbf{A}_{12}\mathbf{B}_{22}\mathbf{A}_{12}^{\top}\mathbf{A}_{11}^{-1} \qquad /\mathbf{A}_{11}^{-1} \\
\Rightarrow \mathbf{A}_{11}^{-1} &= \quad \mathbf{B}_{11} - \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22}\mathbf{A}_{12}^{\top}\mathbf{A}_{11}^{-1} \\
\Rightarrow \mathbf{B}_{11} &= \quad \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22}\mathbf{A}_{12}^{\top}\mathbf{A}_{11}^{-1} \\
\\
\Rightarrow \mathbf{B}_{11} &= \quad \mathbf{A}_{11}^{-1}(\mathbf{I} + \mathbf{A}_{12}\mathbf{B}_{22}\mathbf{A}_{12}^{\top}\mathbf{A}_{11}^{-1})
\end{aligned}$$

Therefore,

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ & \mathbf{B}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11}^{-1}(\mathbf{I} + \mathbf{A}_{12}\mathbf{B}_{22}\mathbf{A}_{12}^{\top}\mathbf{A}_{11}^{-1}) & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22} \\ & \mathbf{B}_{22} \end{pmatrix}$$

where $\mathbf{B}_{22} = (\mathbf{A}_{22} - \mathbf{A}_{12}^{\top}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}$.

$$
\begin{aligned}
\mathbf{AB} &= \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ & \mathbf{A}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ & \mathbf{B}_{22} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{12}^{\top} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ & \mathbf{A}_{12}^{\top}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{A}_{11}\mathbf{A}_{11}^{-1}(\mathbf{I} + \mathbf{A}_{12}\mathbf{B}_{22}\mathbf{A}_{12}^{\top}\mathbf{A}_{11}^{-1}) + \mathbf{A}_{12}(-\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22})^{\top} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ & \mathbf{A}_{12}^{\top}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{pmatrix} \\
&= \begin{pmatrix} (\mathbf{I} + \mathbf{A}_{12}\mathbf{B}_{22}\mathbf{A}_{12}^{\top}\mathbf{A}_{11}^{-1}) - \mathbf{A}_{12}\mathbf{B}_{22}\mathbf{A}_{12}^{\top}\mathbf{A}_{11}^{-1} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ & \mathbf{A}_{12}^{\top}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{I} & -\mathbf{A}_{11}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22} + \mathbf{A}_{12}\mathbf{B}_{22} \\ & -\mathbf{A}_{12}^{\top}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22} + \mathbf{A}_{22}\mathbf{B}_{22} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{I} & \mathbf{A}_{12}\mathbf{B}_{22} + \mathbf{A}_{12}\mathbf{B}_{22} \\ & (-\mathbf{A}_{12}^{\top}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} + \mathbf{A}_{22})\mathbf{B}_{22} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ & (\mathbf{A}_{22} - \mathbf{A}_{12}^{\top}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})\mathbf{B}_{22} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ & \mathbf{B}_{22}^{-1}\mathbf{B}_{22} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ & \mathbf{I} \end{pmatrix}
\end{aligned}
$$

It demonstrated that $\mathbf{B}$ is the inverse of the symmetric matrix $\mathbf{A}$

Inverse of variance-covariance matrix

The variance-covariance matrix correspond to,

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ & \Sigma_{22} \end{pmatrix}$$

Therefore, using appendix C.1.3 be have to,

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_{11}^{-1}(\mathbf{I} + \Sigma_{12}\mathbf{G}\Sigma_{12}^{\top}\Sigma_{11}^{-1}) & -\Sigma_{11}^{-1}\Sigma_{12}\mathbf{G} \\ & \mathbf{G} \end{pmatrix}$$

where $\mathbf{G} = (\Sigma_{22} - \Sigma_{12}^{\top}\Sigma_{11}^{-1}\Sigma_{12})^{-1}$.

Now, in this case, model of two cohorts,

$$\left(\begin{array}{c|c} \mathbf{Y}_{j4} & \\ & \mathbf{X}_{j4}, \mathbf{X}_{j3} \\ \mathbf{Y}_{j3} & \end{array} \right) \sim N\left(\begin{pmatrix} \mu_{j4} \\ \mu_{j3} \end{pmatrix}; \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ & \Sigma_{22} \end{pmatrix} \right)$$

the elements of variance covariance matrix correspond to,

$$\begin{aligned} \Sigma_{11} &= \sigma_4^2\mathbf{E}_{n_{j4}} + (\sigma_4^2 + (\tau_4^2 + \mu^2\tau_3^2)n_{j4})\bar{\mathbf{J}}_{n_{j4}} \\ \Sigma_{11} &= \sigma_4^2\mathbf{E}_{n_{j4}} + l_4\bar{\mathbf{J}}_{n_{j4}}, \text{ where } l_4 = \sigma_4^2 + (\tau_4^2 + \mu^2\tau_3^2)n_{j4} \end{aligned}$$

$$\Sigma_{12} = \iota_{n_{j4}}\iota_{n_{j3}}^{\top}\mu\tau_3^2$$

$$\begin{aligned} \Sigma_{22} &= \sigma_3^2\mathbf{E}_{n_{j3}} + (\sigma_3^2 + \tau_3^2 n_{j3})\bar{\mathbf{J}}_{n_{j3}} \\ \Sigma_{22} &= \sigma_3^2\mathbf{E}_{n_{j3}} + l_2\bar{\mathbf{J}}_{n_{j3}}, \text{ where } l_2 = \sigma_3^2 + \tau_3^2 n_{j3} \end{aligned}$$

therefore

$$\begin{aligned}\Sigma_{11}^{-1} &= (1/\sigma_4^2)\mathbf{E}_{n_{j4}} + (1/l_4)\bar{\mathbf{J}}_{n_{j4}} \\ \Sigma_{22}^{-1} &= (1/\sigma_3^2)\mathbf{E}_{n_{j4}} + (1/l_2)\bar{\mathbf{J}}_{n_{j3}}\end{aligned}$$

continue with the calculations of the matrix elements Σ^{-1}, then

$$\begin{aligned}
\mathbf{G} &= \Big(\sigma_3^2\mathbf{E}_{n_{j3}} + l_2\bar{\mathbf{J}}_{n_{j3}} - \\
&\quad - \iota_{n_{j3}}\iota_{n_{j4}}^\top\mu\tau_3^2\left((1/\sigma_4^2)\mathbf{E}_{n_{j4}} + (1/l_4)\bar{\mathbf{J}}_{n_{j4}}\right)\iota_{n_{j4}}\iota_{n_{j3}}^\top\mu\tau_3^2\Big)^{-1} \\
&= \left(\sigma_3^2\mathbf{E}_{n_{j3}} + l_2\bar{\mathbf{J}}_{n_{j3}} - n_{j4}\mu\tau_3^2(1/l_4)\iota_{n_{j3}}\iota_{n_{j3}}^\top\mu\tau_3^2\right)^{-1} \\
&= \left(\sigma_3^2\mathbf{E}_{n_{j3}} + \left(l_2 - n_{j3}n_{j4}\mu^2\tau_3^4(1/l_4)\right)\bar{\mathbf{J}}_{n_{j3}}\right)^{-1} \\
&= \left(\sigma_3^2\mathbf{E}_{n_{j3}} + r_2\bar{\mathbf{J}}_{n_{j3}}\right)^{-1}, \text{ where } r_2 = \left(l_2 - n_{j3}n_{j4}\mu^2\tau_3^4(1/l_4)\right) \\
&= (1/\sigma_3^2)\mathbf{E}_{n_{j3}} + (1/r_2)\bar{\mathbf{J}}_{n_{j3}}
\end{aligned}$$

$$\Rightarrow \mathbf{G} = (1/\sigma_3^2)\mathbf{E}_{n_{j3}} + (1/r_2)\bar{\mathbf{J}}_{n_{j3}}$$

$$\begin{aligned}
\Sigma_{11}^{-1}\Sigma_{12}\mathbf{G} &= \left((1/\sigma_4^2)\mathbf{E}_{n_{j4}} + (1/l_4)\bar{\mathbf{J}}_{n_{j4}}\right)\left(\iota_{n_{j4}}\iota_{n_{j3}}^\top\mu\tau_3^2\right) \\
&\quad \left((1/\sigma_3^2)\mathbf{E}_{n_{j3}} + (1/r_2)\bar{\mathbf{J}}_{n_{j3}}\right) \\
&= \left((1/\sigma_4^2)\mathbf{E}_{n_{j4}} + (1/l_4)\bar{\mathbf{J}}_{n_{j4}}\right)\left(\iota_{n_{j4}}\iota_{n_{j3}}^\top\mu\tau_3^2/r_2\right) \\
&= \left(\mu\tau_3^2/(r_2l_4)\right)\iota_{n_{j4}}\iota_{n_{j3}}^\top
\end{aligned}$$

$$\Rightarrow \Sigma_{11}^{-1}\Sigma_{12}\mathbf{G} = \left(\mu\tau_3^2/(r_2l_4)\right)\iota_{n_{j4}}\iota_{n_{j3}}^\top$$

$$\begin{aligned}
\Sigma_{11}^{-1}(\mathbf{I} + \Sigma_{12}\mathbf{G}\Sigma_{12}^\top\Sigma_{11}^{-1}) &= \left((1/\sigma_4^2)\mathbf{E}_{n_{j4}} + (1/l_4)\bar{\mathbf{J}}_{n_{j4}}\right) \\
&\quad \left(\mathbf{I} + \iota_{n_{j4}}\iota_{n_{j3}}^\top\mu\tau_3^2\left(\mu\tau_3^2/(r_2l_4)\right)\iota_{n_{j3}}\iota_{n_{j4}}^\top\right) \\
&= \left((1/\sigma_4^2)\mathbf{E}_{n_{j4}} + (1/l_4)\bar{\mathbf{J}}_{n_{j4}}\right) \\
&\quad \left(\mathbf{I} + n_{j3}\mu^2\tau_3^4/(r_2l_4)\iota_{n_{j4}}\iota_{n_{j4}}^\top\right) \\
&= \left((1/\sigma_4^2)\mathbf{E}_{n_{j4}} + (1/l_4)\bar{\mathbf{J}}_{n_{j4}}\right) \\
&\quad \left(\mathbf{I} + n_{j3}n_{j4}\mu^2\tau_3^4/(r_2l_4)\bar{\mathbf{J}}_{n_{j4}}\right) \\
&= (1/\sigma_4^2)\mathbf{E}_{n_{j4}} + (1/l_4)\bar{\mathbf{J}}_{n_{j4}} + n_{j3}n_{j4}\mu^2\tau_3^4/(r_2l_4^2)\bar{\mathbf{J}}_{n_{j4}} \\
&= (1/\sigma_4^2)\mathbf{E}_{n_{j4}} + \left(1/l_4 + n_{j3}n_{j4}\mu^2\tau_3^4/(r_2l_4^2)\right)\bar{\mathbf{J}}_{n_{j4}}
\end{aligned}$$

$$\Rightarrow \Sigma_{11}^{-1}(\mathbf{I} + \Sigma_{12}\mathbf{G}\Sigma_{12}^\top\Sigma_{11}^{-1}) = (1/\sigma_4^2)\mathbf{E}_{n_{j4}} + \left(1/l_4 + n_{j3}n_{j4}\mu^2\tau_3^4/(r_2l_4^2)\right)\bar{\mathbf{J}}_{n_{j4}}$$

this way,

$$\Sigma^{-1} = \begin{pmatrix} (1/\sigma_4^2)\mathbf{E}_{n_{j4}} + (1/l_4 + n_{j3}n_{j4}\mu^2\tau_3^4/(r_2 l_4^2))\,\bar{\mathbf{J}}_{n_{j4}} & -(\mu\tau_3^2/(r_2 l_4))\,\iota_{n_{j4}}\iota_{n_{j3}}^\top \\ & (1/\sigma_3^2)\mathbf{E}_{n_{j3}} + (1/r_2)\bar{\mathbf{J}}_{n_{j3}} \end{pmatrix}$$

However, note that

$$\begin{aligned}
\left(1/l_4 + n_{j3}n_{j4}\mu^2\tau_3^4/(r_2 l_4^2)\right) =& \frac{1}{l_4^2}\left(l_4 + \frac{1}{r_2}n_{j3}n_{j4}\mu^2\tau_3^4\right) \\
=& \frac{1}{r_2 l_4^2}\left(r_2 l_4 + n_{j3}n_{j4}\mu^2\tau_3^4\right) \\
=& \frac{1}{r_2 l_4^2}\left((l_2 - n_{j3}n_{j4}\mu^2\tau_3^4/l_4)l_4 + n_{j3}n_{j4}\mu^2\tau_3^4\right) \\
=& \frac{1}{r_2 l_4^2}\left((l_2 l_4 - n_{j3}n_{j4}\mu^2\tau_3^4) + n_{j3}n_{j4}\mu^2\tau_3^4\right) \\
=& \frac{1}{r_2 l_4^2}\left(l_2 l_4\right) \\
=& \frac{l_2}{r_2 l_4}
\end{aligned}$$

This way,

$$\Sigma^{-1} = \begin{pmatrix} (1/\sigma_4^2)\mathbf{E}_{n_{j4}} + (l_2/(r_2 l_4))\,\bar{\mathbf{J}}_{n_{j4}} & -(\mu\tau_3^2/(r_2 l_4))\,\iota_{n_{j4}}\iota_{n_{j3}}^\top \\ & (1/\sigma_3^2)\mathbf{E}_{n_{j3}} + (1/r_2)\bar{\mathbf{J}}_{n_{j3}} \end{pmatrix}$$

C.1.4 Determinant of block matrices

Suppose A, B, C, and D are matrices of dimension $n \times n$, $n \times m$, $m \times n$, and $m \times m$, respectively. Then, from the Leibniz formula, or from a decomposition like one has

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A & 0 \\ C & I_m \end{pmatrix} \begin{pmatrix} I_n & A^{-1}B \\ 0 & D - CA^{-1}B \end{pmatrix}$$

This way, when A is invertible, one has

$$\left| \begin{pmatrix} A & B \\ C & D \end{pmatrix} \right| = |A||D - CA^{-1}B|$$

Determinant of variance-covariance matrix

The variance-covariance matrix correspond to,

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ & \Sigma_{22} \end{pmatrix}$$

Therefore, using appendix C.1.4 be have to,

$$|\Sigma| = |\Sigma_{11}||\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}|$$

Now, in this case, model of two cohorts, the elements of variance covariance matrix correspond to,

$$
\begin{aligned}
\Sigma_{11} &= \sigma_4^2 \mathbf{E}_{n_{j4}} + (\sigma_4^2 + (\tau_4^2 + \mu^2\tau_3^2) n_{j4}) \bar{\mathbf{J}}_{n_{j4}} \\
\Sigma_{11} &= \sigma_4^2 \mathbf{E}_{n_{j4}} + l_4 \bar{\mathbf{J}}_{n_{j4}}, \quad \text{where } \; l_4 = \sigma_4^2 + (\tau_4^2 + \mu^2\tau_3^2) n_{j4}
\end{aligned}
$$

$$
\Sigma_{12} = \iota_{n_{j4}} \iota_{n_{j3}}^\top \mu \tau_3^2
$$

$$
\begin{aligned}
\Sigma_{22} &= \sigma_3^2 \mathbf{E}_{n_{j3}} + (\sigma_3^2 + \tau_3^2 n_{j3}) \bar{\mathbf{J}}_{n_{j3}} \\
\Sigma_{22} &= \sigma_3^2 \mathbf{E}_{n_{j3}} + l_2 \bar{\mathbf{J}}_{n_{j3}}, \quad \text{where } \; l_2 \sigma_3^2 + \tau_3^2 n_{j3}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
|\Sigma_{11}| &= \left| \sigma_4^2 \mathbf{E}_{n_{j4}} + l_4 \bar{\mathbf{J}}_{n_{j4}} \right| \\
&= \prod_{i=1}^{n_{j4}} \lambda_i \quad \text{where} \lambda_i \text{are the eigen values} \\
&= \sigma_4^{2(n_{j4}-1)} l_4
\end{aligned}
$$

$$
\begin{aligned}
\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} &= \mu^2 \tau_3^4 \iota_{n_{j3}} \iota_{n_{j4}}^\top \left(\frac{1}{\sigma_4^2} \mathbf{E}_{n_{j4}} + \frac{1}{l_4} \bar{\mathbf{J}}_{n_{j4}} \right) \iota_{n_{j4}} \iota_{n_{j3}}^\top \\
&= n_{j4} \mu^2 \tau_3^4 \iota_{n_{j3}} \frac{1}{l_4} \iota_{n_{j3}}^\top \\
&= n_{j3} n_{j4} \mu^2 \tau_3^4 \bar{\mathbf{J}}_{n_{j3}}
\end{aligned}
$$

$$
\begin{aligned}
\left| \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \right| &= \left| \sigma_3^2 \mathbf{E}_{n_{j3}} + l_2 \bar{\mathbf{J}}_{n_{j3}} - n_{j3} n_{j4} \mu^2 \tau_3^4 \bar{\mathbf{J}}_{n_{j3}} \right| \\
&= \left| \sigma_3^2 \mathbf{E}_{n_{j3}} + (l_2 - n_{j3} n_{j4} \mu^2 \tau_3^4) \bar{\mathbf{J}}_{n_{j3}} \right| \\
&= \sigma_3^{2(n_{j3}-1)} \left(l_2 - n_{j3} n_{j4} \mu^2 \tau_3^4 \right)
\end{aligned}
$$

This way,

$$
|\Sigma_j| = \sigma_4^{2(n_{j4}-1)} \sigma_3^{2(n_{j3}-1)} l_4 \left(l_2 - n_{j3} n_{j4} \mu^2 \tau_3^4 \right)
$$

C.1.5 School effects

Expected of school effects

$$
\begin{aligned}
\alpha_j =& \mathbb{E}\left(\begin{matrix}\theta_{j4,2}\\ \theta_{j3,1}\end{matrix}\;\middle|\; \mathbf{X}_{j4},\mathbf{X}_{j2}\right) + \mathrm{Cov}\left(\begin{matrix}\theta_{j4,2}\\ \theta_{j3,1}\end{matrix}\;;\;\begin{matrix}\mathbf{Y}_{j4}\\ \mathbf{Y}_{j2}\end{matrix}\;\middle|\; \mathbf{X}_{j4},\mathbf{X}_{j2}\right)\mathbf{V}^{-1}\left(\begin{matrix}\mathbf{Y}_{j4}\\ \mathbf{Y}_{j2}\end{matrix}\;\middle|\; \mathbf{X}_{j4},\mathbf{X}_{j2}\right)\\
&\left(\begin{pmatrix}\mathbf{Y}_{j4}\\ \mathbf{Y}_{j2}\end{pmatrix} - \mathbb{E}\left(\begin{matrix}\mathbf{Y}_{j4}\\ \mathbf{Y}_{j2}\end{matrix}\;\middle|\; \mathbf{X}_{j4},\mathbf{X}_{j2}\right)\right)\\
=& \begin{pmatrix}\iota_{n_{4j}}^\top(\tau_4^2+\mu^2\tau_2^2) & \iota_{n_{2j}}^\top\mu\tau_2^2\\ \iota_{n_{4j}}^\top\mu\tau_2^2 & \iota_{n_{2j}}^\top\tau_2^2\end{pmatrix}\begin{pmatrix}\sigma_4^2\mathbf{I}_{n_{4j}}+(\tau_4^2+\mu^2\tau_2^2)\mathbf{J}_{n_{4j}} & \iota_{n_{4j}}\iota_{n_{2j}}^\top\mu\tau_2^2\\ \iota_{n_{2j}}\iota_{n_{4j}}^\top\mu\tau_2^2 & \sigma_2^2\mathbf{I}_{n_{2j}}+\tau_2^2\mathbf{J}_{n_{2j}}\end{pmatrix}^{-1}\\
&\begin{pmatrix}\mathbf{Y}_{j4}-\mathbf{X}_{j4}\beta_4\\ \mathbf{Y}_{j2}-\mathbf{X}_{j2}\beta_2\end{pmatrix}\\
=& \begin{pmatrix}\iota_{n_{4j}}^\top(\tau_4^2+\mu^2\tau_2^2) & \iota_{n_{2j}}^\top\mu\tau_2^2\\ \iota_{n_{4j}}^\top\mu\tau_2^2 & \iota_{n_{2j}}^\top\tau_2^2\end{pmatrix}\begin{pmatrix}\sigma_4^2\mathbf{I}_{n_{4j}}+(\tau_4^2+\mu^2\tau_2^2)\mathbf{J}_{n_{4j}} & \iota_{n_{4j}}\iota_{n_{2j}}^\top\mu\tau_2^2\\ \iota_{n_{2j}}\iota_{n_{4j}}^\top\mu\tau_2^2 & \sigma_2^2\mathbf{I}_{n_{2j}}+\tau_2^2\mathbf{J}_{n_{2j}}\end{pmatrix}^{-1}\\
&\begin{pmatrix}\mathbf{Y}_{j4}-\mathbf{X}_{j4}\beta_4\\ \mathbf{Y}_{j2}-\mathbf{X}_{j2}\beta_2\end{pmatrix}\\
=& \begin{pmatrix}\iota_{n_{4j}}^\top(\tau_4^2+\mu^2\tau_2^2) & \iota_{n_{2j}}^\top\mu\tau_2^2\\ \iota_{n_{4j}}^\top\mu\tau_2^2 & \iota_{n_{2j}}^\top\tau_2^2\end{pmatrix}\\
&\begin{pmatrix}(1/\sigma_4^2)\mathbf{E}_{n_{4j}}+(l_2/(r_2l_4))\,\bar{\mathbf{J}}_{n_{4j}} & -(\mu\tau_2^2/(r_2l_4))\,\iota_{n_{4j}}\iota_{n_{2j}}^\top\\ & (1/\sigma_2^2)\mathbf{E}_{n_{2j}}+(1/r_2)\bar{\mathbf{J}}_{n_{2j}}\end{pmatrix}\\
&\begin{pmatrix}\mathbf{Y}_{j4}-\mathbf{X}_{j4}\beta_4\\ \mathbf{Y}_{j2}-\mathbf{X}_{j2}\beta_2\end{pmatrix}\\
=& \begin{pmatrix}\phi_{11}\iota_{n_{4j}}^\top & \phi_{12}\iota_{n_{2j}}^\top\\ \phi_{21}\iota_{n_{4j}}^\top & \phi_{22}\iota_{n_{2j}}^\top\end{pmatrix}\begin{pmatrix}\mathbf{Y}_{j4}-\mathbf{X}_{j4}\beta_4\\ \mathbf{Y}_{j2}-\mathbf{X}_{j2}\beta_2\end{pmatrix}
\end{aligned}
$$

where

$$
\begin{aligned}
\phi_{11} =& (1/(r_2 l_4)) \left\{ (\tau_4^2 + \mu^2 \tau_2^2) l_2 - n_{2j} \mu^2 \tau_2^4) \right\} \\
\phi_{12} =& - n_{4j} (\tau_4^2 + \mu^2 \tau_2^2) \left(\mu \tau_2^2 / (r_2 l_4) \right) + \mu \tau_2^2 / r2 \\
=& (1/(r_2 l_4)) \left(-n_{4j} (\tau_4^2 + \mu^2 \tau_2^2) \mu \tau_2^2 + l_4 \mu \tau_2^2 \right) \\
=& (1/(r_2 l_4)) \left(-n_{4j} (\tau_4^2 + \mu^2 \tau_2^2) \mu \tau_2^2 + (\sigma_4^2 + (\tau_4^2 + \mu^2 \tau_2^2) n_{4j}) \mu \tau_2^2 \right) \\
=& (\sigma_4^2 \mu \tau_2^2) / (r_2 l_4) \\
\phi_{21} =& \mu \tau_2^2 (l_2 / (r_2 l_4)) - \mu \tau_2^4 / (r_2 l_4) n_{2j} \\
=& (1/(r_2 l_4)) \left(\mu \tau_2^2 (\sigma_2^2 + \tau_2^2 n_{2j}) - \mu \tau_2^4 n_{2j} \right) \\
=& (\mu \tau_2^2 \sigma_2^2) / (r_2 l_4) \\
\phi_{22} =& - n_{4j} \mu^2 \tau_2^4 / (r_2 l_4) + \tau_2^2 / r_2 \\
=& \tau_2^2 (\sigma_4^2 + \tau_4^2 n_{4j}) / (r_2 l_4)
\end{aligned}
$$

This way,

$$
\alpha_j = \begin{pmatrix} \phi_{11} \iota_{n_{4j}}^\top (\mathbf{Y}_{j4} - \mathbf{X}_{j4} \beta_4) + \phi_{12} \iota_{n_{2j}}^\top (\mathbf{Y}_{j2} - \mathbf{X}_{j2} \beta_2) \\ \phi_{21} \iota_{n_{4j}}^\top (\mathbf{Y}_{j4} - \mathbf{X}_{j4} \beta_4) + \phi_{22} \iota_{n_{2j}}^\top (\mathbf{Y}_{j2} - \mathbf{X}_{j2} \beta_2) \end{pmatrix}
$$

Variance of school effects

$$
\Lambda_j = \mathbf{V}\left(\begin{array}{c}\theta_{j4,2}\\ \theta_{j3,1}\end{array}\middle| \mathbf{X}_{j4},\mathbf{X}_{j2}\right) - \mathrm{Cov}\left(\begin{array}{c}\theta_{j4,2}\\ \theta_{j3,1}\end{array};\begin{array}{c}\mathbf{Y}_{j4}\\ \mathbf{Y}_{j2}\end{array}\middle| \mathbf{X}_{j4},\mathbf{X}_{j2}\right)\mathbf{V}^{-1}\left(\begin{array}{c}\mathbf{Y}_{j4}\\ \mathbf{Y}_{j2}\end{array}\middle| \mathbf{X}_{j4},\mathbf{X}_{j2}\right)
$$

$$
\mathrm{Cov}\left(\begin{array}{c}\theta_{j4,2}\\ \theta_{j3,1}\end{array};\begin{array}{c}\mathbf{Y}_{j4}\\ \mathbf{Y}_{j2}\end{array}\middle| \mathbf{X}_{j4},\mathbf{X}_{j2}\right)^{\top}
$$

$$
= \begin{pmatrix}\tau_4^2+\mu^2\tau_2^2 & \mu\tau_2^2\\ \mu\tau_2^2 & \tau_2^2\end{pmatrix} - \begin{pmatrix}\iota_{n_{4j}}^{\top}(\tau_4^2+\mu^2\tau_2^2) & \iota_{n_{2j}}^{\top}\mu\tau_2^2\\ \iota_{n_{4j}}^{\top}\mu\tau_2^2 & \iota_{n_{2j}}^{\top}\tau_2^2\end{pmatrix}
$$

$$
\begin{pmatrix}\sigma_4^2\mathbf{I}_{n_{4j}}+(\tau_4^2+\mu\tau_2^2)\mathbf{J}_{n_{4j}} & \iota_{n_{4j}}\iota_{n_{2j}}^{\top}\mu\tau_2^2\\ \iota_{n_{2j}}\iota_{n_{4j}}^{\top}\mu\tau_2^2 & \sigma_2^2\mathbf{I}_{n_{2j}}+\tau_2^2\mathbf{J}_{n_{2j}}\end{pmatrix}^{-1}\begin{pmatrix}\iota_{n_{4j}}(\tau_4^2+\mu^2\tau_2^2) & \iota_{n_{2j}}\mu\tau_2^2\\ \iota_{n_{4j}}\mu\tau_2^2 & \iota_{n_{2j}}\tau_2^2\end{pmatrix}
$$

$$
= \begin{pmatrix}\tau_4^2+\mu^2\tau_2^2 & \mu\tau_2^2\\ \mu\tau_2^2 & \tau_2^2\end{pmatrix} - \begin{pmatrix}\phi_{11}\iota_{n_{4j}}^{\top} & \phi_{12}\iota_{n_{2j}}^{\top}\\ \phi_{21}\iota_{n_{4j}}^{\top} & \phi_{22}\iota_{n_{2j}}^{\top}\end{pmatrix}\begin{pmatrix}\iota_{n_{4j}}(\tau_4^2+\mu^2\tau_2^2) & \iota_{n_{4j}}\mu\tau_2^2\\ \iota_{n_{2j}}\mu\tau_2^2 & \iota_{n_{2j}}\tau_2^2\end{pmatrix}
$$

$$
= \begin{pmatrix}\tau_4^2+\mu^2\tau_2^2 & \mu\tau_2^2\\ \mu\tau_2^2 & \tau_2^2\end{pmatrix} - \begin{pmatrix}\phi_{11}n_{4j}(\tau_4^2+\mu^2\tau_2^2)+n_{2j}\phi_{12}\mu\tau_2^2 & \phi_{11}n_{4j}\mu\tau_2^2+\phi_{12}n_{2j}\tau_2^2\\ \phi_{21}n_{4j}(\tau_4^2+\mu^2\tau_2^2)+n_{2j}\phi_{22}\mu\tau_2^2 & \phi_{21}n_{4j}\mu\tau_2^2+\phi_{22}n_{2j}\tau_2^2\end{pmatrix}
$$

$$
= \begin{pmatrix}\tau_4^2+\mu^2\tau_2^2-(\phi_{11}n_{4j}(\tau_4^2+\mu^2\tau_2^2)+n_{2j}\phi_{12}\mu\tau_2^2) & \mu\tau_2^2-(\phi_{11}n_{4j}\mu\tau_2^2+\phi_{12}n_{2j}\tau_2^2)\\ \mu\tau_2^2-(\phi_{21}n_{4j}(\tau_4^2+\mu^2\tau_2^2)+n_{2j}\phi_{22}\mu\tau_2^2) & \tau_2^2-(\phi_{21}n_{4j}\mu\tau_2^2+\phi_{22}n_{2j}\tau_2^2)\end{pmatrix}
$$

$$
= \begin{pmatrix}\sigma_4^2(\tau_4^2+\mu\tau_2^2)(l_2-n_{2j}\mu^2\tau_2^4)/(r_2l_4) & \sigma_2^2\sigma_4^2\tau_2^2\mu/(r_2l_4)\\ \sigma_2^2\sigma_4^2\tau_2^2\mu/(r_2l_4) & \tau_2^2\sigma_2^2\left(\sigma_4^2+\tau_4^2n_{4j}\right)/(r_2l_4)\end{pmatrix}
$$

C.2 Study of Simulation

C.2.1 Results in scenario I, $\mu = 5$

Table C.1: Results of Monte Carlo simulations of components variance in 2HLM model, with $\mu = 5$

	Real	Monte Carlo estimations							
		Sample= 10		Sample= 50		Sample= 100		Sample= 500	
μ	5	4.998	(0.113)	4.957	(0.238)	5.047	(0.231)	5.006	(0.226)
σ_3^2	729	724.633	(13.905)	730.450	(11.869)	728.943	(13.575)	728.852	(11.519)
σ_4^2	900	895.184	(12.494)	900.134	(17.582)	902.952	(13.473)	899.340	(16.298)
τ_3^2	100	97.939	(7.281)	99.349	(9.366)	97.615	(9.354)	99.793	(10.768)
τ_4^2	144	160.096	(61.952)	143.173	(81.978)	139.715	(87.748)	140.902	(90.239)

Table C.2: Results of Monte Carlo simulations of fixed effects on 2HLM model, with $\mu = 5$

	Real	Monte Carlo estimations							
		Sample= 10		Sample= 50		Sample= 100		Sample= 500	
β_4	7.0	6.986	(0.179)	6.999	(0.187)	6.987	(0.187)	6.992	(0.176)
	0.9	0.905	(0.022)	0.901	(0.016)	0.899	(0.017)	0.899	(0.017)
	15.0	14.998	(0.032)	15.001	(0.035)	14.997	(0.035)	15.001	(0.035)
	5.0	4.997	(0.021)	5.000	(0.020)	5.002	(0.021)	5.000	(0.022)
β_3	2.0	2.018	(0.141)	2.007	(0.320)	2.039	(0.304)	1.994	(0.338)
	0.5	0.480	(0.146)	0.492	(0.321)	0.463	(0.303)	0.505	(0.337)
	20.0	20.011	(0.025)	20.001	(0.029)	19.999	(0.035)	19.999	(0.033)
	0.5	0.499	(0.014)	0.502	(0.017)	0.497	(0.018)	0.499	(0.017)

Table C.3: Results of Monte Carlo simulations of variance components in 1HLM model, with $\mu = 5$

	Real	Monte Carlo estimations							
		Sample= 10		Sample= 50		Sample= 100		Sample= 500	
σ_3^2	729	724.63	(13.906)	730.45	(11.870)	728.94	(13.576)	728.85	(11.520)
σ_4^2	900	895.18	(12.495)	900.13	(17.583)	902.95	(13.473)	899.34	(16.298)
τ_3^2	100	97.94	(7.282)	99.35	(9.366)	97.61	(9.354)	99.79	(10.768)
τ_4^2	144	2606.82	(178.095)	2580.02	(201.911)	2620.50	(212.442)	2634.00	(232.330)

Table C.4: Results of Monte Carlo simulations of fixed effects estimation in 1HLM model, with $\mu = 5$

	Real	Monte Carlo estimations							
		Sample= 10		Sample= 50		Sample= 100		Sample= 500	
β_4	7.0	6.981	(0.181)	7.001	(0.188)	6.987	(0.186)	6.992	(0.176)
	0.9	0.905	(0.022)	0.901	(0.016)	0.899	(0.017)	0.899	(0.017)
	15.0	14.999	(0.032)	15.001	(0.036)	14.997	(0.036)	15.001	(0.035)
	5.0	4.997	(0.021)	5.000	(0.020)	5.002	(0.021)	5.000	(0.022)
β_3	2.0	2.018	(0.141)	2.007	(0.322)	2.039	(0.304)	1.994	(0.338)
	0.5	0.480	(0.146)	0.492	(0.321)	0.463	(0.303)	0.505	(0.337)
	20.0	20.011	(0.025)	20.001	(0.029)	19.999	(0.035)	19.999	(0.033)
	0.5	0.499	(0.014)	0.502	(0.017)	0.497	(0.018)	0.499	(0.017)

C.2.2 Results in scenario II, $\mu = 0.5$

Table C.5: Results of Monte Carlo simulations of components variance in 2HLM model, with $\mu = 0.5$

	Real	Monte Carlo estimations							
		Sample= 10		Sample= 50		Sample= 100		Sample= 500	
μ	0.5	0.542	(0.0833)	0.498	(0.084)	0.494	(0.0811)	0.5	(0.0932)
σ_3^2	729	731.831	(15.5124)	730.304	(12.8073)	729.225	(10.9322)	729.457	(11.6532)
σ_4^2	900	905.915	(11.5656)	893.665	(14.7667)	901.438	(14.0915)	899.972	(15.5468)
τ_3^2	100	100.422	(14.7331)	101.656	(10.6117)	100.505	(9.7297)	99.597	(10.3115)
τ_4^2	144	140.523	(18.1887)	143.08	(15.28)	146.446	(13.4957)	142.284	(16.1153)

Table C.6: Results of Monte Carlo simulations of fixed effects estimation in 2HLM model, with $\mu = 0.5$

	Real	Monte Carlo estimations							
		Sample= 10		Sample= 50		Sample= 100		Sample= 500	
β_4	7.0	7.046	(0.102)	6.998	(0.106)	7.008	(0.111)	7.004	(0.106)
	0.9	0.907	(0.012)	0.900	(0.009)	0.899	(0.010)	0.900	(0.009)
	15.0	14.990	(0.018)	15.001	(0.019)	15.002	(0.020)	14.998	(0.019)
	5.0	4.992	(0.011)	4.999	(0.009)	5.000	(0.011)	5.000	(0.010)
β_2	2.0	2.036	(0.148)	2.022	(0.218)	2.013	(0.208)	1.999	(0.216)
	0.5	0.461	(0.147)	0.477	(0.222)	0.4863	(0.209)	0.5009	(0.215)
	20.0	19.993	(0.019)	20.001	(0.019)	20.001	(0.018)	20.000	(0.018)
	0.5	0.503	(0.011)	0.500	(0.010)	0.500	(0.009)	0.500	(0.010)

Table C.7: Results of Monte Carlo simulations of variance components estimation in 1HLM model, with $\mu = 0.5$

	Real	Monte Carlo estimations							
		Sample= 10		Sample= 50		Sample= 100		Sample= 500	
σ_3^2	729	731.831	(15.5124)	730.304	(12.8073)	729.225	(10.9322)	729.457	(11.6532)
σ_4^2	900	905.915	(11.5656)	893.665	(14.7667)	901.438	(14.0915)	899.972	(15.5468)
τ_3^2	100	100.422	(14.7331)	101.656	(10.6117)	100.505	(9.7297)	99.597	(10.3115)
τ_4^2	144	169.989	(14.3712)	168.783	(16.2507)	171.643	(15.1116)	167.929	(17.4677)

Table C.8: Results of Monte Carlo simulations of fixed effects estimation in 1HLM model, with $\mu = 0.5$

	Real	Monte Carlo estimations							
		Sample= 10		Sample= 50		Sample= 100		Sample= 500	
β_4	7.0	7.046	(0.101)	6.998	(0.106)	7.008	(0.111)	7.004	(0.106)
	0.9	0.907	(0.012)	0.900	(0.009)	0.899	(0.010)	0.900	(0.009)
	15.0	14.990	(0.018)	15.001	(0.019)	15.002	(0.020)	14.998	(0.019)
	5.0	4.992	(0.011)	4.999	(0.009)	5.000	(0.011)	5.000	(0.010)
β_3	2.0	2.036	(0.148)	2.022	(0.218)	2.013	(0.208)	1.999	(0.216)
	0.5	0.461	(0.147)	0.477	(0.222)	0.486	(0.208)	0.500	(0.215)
	20.0	19.993	(0.019)	20.001	(0.019)	20.001	(0.018)	20.000	(0.018)
	0.5	0.503	(0.011)	0.500	(0.010)	0.500	(0.009)	0.500	(0.010)