



PONTIFICIA  
UNIVERSIDAD  
CATÓLICA  
DE CHILE

FACULTAD DE EDUCACIÓN

**Neo-Classical Test Theory: On the modelling of individual learning.  
Theory , Statistical Estimation and Educational Applications**

**Neo-Teoría Clásica de Test: Modelando el aprendizaje de las personas.  
Teoría, estimación estadística y aplicaciones en educación**

por

DANNY ANDRES AVELLO FERNÁNDEZ

Tesis Presentada a la Facultad de Educación  
Pontificia Universidad Católica de Chile,  
Para optar al grado de Doctor en Ciencias de la Educación

Profesor Guía: Dr. Ernesto San Martín

July 29, 2019

Santiago, Chile

©2019, Danny Andrés Avello Fernández

©2019, Danny Andrés Avello Fernández Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica que acredita al trabajo y a su autor.

Fecha: Julio, 2019

Firma:

*Qué valor tendría existir sin al menos un pensamiento sobre estar...  
tu conciencia abandonó nuestro tiempo,  
pero tu tiempo se congeló en mi conciencia.*

Para mi Abuela Iris



# Agradecimientos

Agradezco a mi tutor Ernesto San Martín, no sólo por su imprescindible apoyo en la elaboración y desarrollo de esta tesis, sino además por haberme presentado tantos problemas interesantes durante todo el doctorado. Agradezco a toda mi comisión, Jorge González, Verónica Santalices y Paul De Boeck, cada una de sus preguntas fueron un reto que siempre terminaron mejorando esta investigación; tanto en los aspectos teóricos como aplicados. En particular agradezco al Dr. Paul De Boeck por haber compartido conmigo su admirable experiencia como investigador, las discusiones que mantuve con usted fueron fundamentales para interpretar sicológicamente los resultados matemáticos encontrados.

Agradezco a mis padres Olga y Luis, no es exagerado decir que si no fuera por ustedes no estaría acá, muchas gracias por todo. A mis hermanos Luis y Paula, su apoyo, ayuda y ejemplo han sido vital en todo mi desarrollo y trayectoria académica, sino fuera por ustedes probablemente no habría seguido este camino.

Agradezco a Ana María, has convivido conmigo los aspectos más personales y difíciles de hacer un postgrado, pero de alguna manera has hecho posible que todo parezca más fácil y agradable. Gracias por alegar, reir, discutir y disfrutar conmigo. No es casualidad que con éste llevemos cinco procesos de graduación juntos.

Agradezco a toda la “Fuerza” por las salidas y risas compartidas. Agradezco también a Sabina, aunque estuvimos en programas separados fuiste lo más cercana a una compañera de curso que he tenido, fue muy bueno haber podido discutir aspectos de programación, matemáticos y de presentación contigo.

Agradezco a Charlie porque de alguna manera logró entender durante los tiempos más laboriosos que no podría salir con él, y porque incluso cuando el sueño lo venció se mantuvo cerca trabajando conmigo.

Finalmente agradezco a CONICYT del Gobierno de Chile, por haber financiado esta investigación a través de la beca Doctorado Nacional año 2014, folio 21141210.



# Contents

Contents . . . . .	iii
List of Tables . . . . .	v
List of Figures . . . . .	vi
Abstract . . . . .	vii
Resumen . . . . .	ix
<b>I Introduction</b>	<b>1</b>
<b>II Neo-Classical Test Theory: A Geometrical Approach</b>	<b>5</b>
2.1 Observed score, Latent variables and Hilbert space . . . . .	5
2.2 Neo-Classical Test Theory . . . . .	13
2.2.1 Classical and Dual reliabilities . . . . .	15
2.2.2 Relation between Reliabilities . . . . .	18
2.2.3 Splitting Subspaces and Axiom of Local Independence . . . . .	24
2.2.4 Multidimensional Latent Variables . . . . .	32
2.2.5 Final Comments . . . . .	36
<b>III Marginal Distribution Estimation of a Latent Variable</b>	<b>38</b>
3.1 Model Specifications . . . . .	39
3.2 Method of Estimation . . . . .	41
3.3 Simulation Study . . . . .	43
3.4 Application with real datas . . . . .	51

<b>IV Conclusion</b>	<b>56</b>
<b>Bibliography</b>	<b>59</b>
<b>References</b>	<b>59</b>
<b>Appendix</b>	<b>67</b>

# List of Tables

2.1 Examples of Observed scores . . . . .	7
3.1 MISE . . . . .	47

# List of Figures

2.1	Orthogonal Complement . . . . .	11
2.2	Orthogonal Projection Operators . . . . .	12
2.3	Orthogonal Projection from the Observed subspace onto the Latent Subspace . . . . .	14
2.4	Conditional Orthogonality . . . . .	26
3.1	Latent variable from gamma and normal distribution equal errors . .	44
3.2	Latent variable from gamma and normal distribution different errors .	45
3.3	Latent variable from gamma and normal distributions different errors best reliability . . . . .	46
3.4	MISE and variance of the errors . . . . .	48
3.5	Error distributions for binaries observed scores . . . . .	49
3.6	Discrete latent variable . . . . .	50
3.7	Estimation of $\theta$ 's density from real datas . . . . .	51
3.8	Densities of errors in real datas Language . . . . .	52
3.9	Densities of errors in real datas Mathematic . . . . .	54

# Abstract

Latent variables used in Item Response Theory (IRT) models are specified by the axiom of local independence (ALI). By doing this, latent variables explain the non random variations in the observed scores, thus, latent variable represent what is meaningful to the researcher. High impact decisions are made based on the estimates of them. Selection in some universities and grants are just some examples. However, the marginal distribution of latent variables is not identified in IRT models, hence, it is not possible to infer from the observed information an interpretation of a latent variable in this context. Usually their distributions are assumed as normal, even if they are meant to represent the substantive behind the data. Our main goal was to identify the marginal distribution of latent variables in a set of IRT models, and estimate this distribution.

We frame all IRT models that can be formulated as generalized linear mixed models in a Hilbert space. In this framework the ALI is replaced by conditional orthogonality, a weak version of ALI (WALI). We found that under WALI the only element for which the Empirical Bayes estimator is zero, it is the zero element. Hence the WALI resolve an indeterminacy problem, because if there was more information in there, it would be impossible to recover it using the Empirical Bayes estimator. We found that it is possible to have latent variables “living” among all the possible observed scores, thus they are unobserved and not unobservable. Also, using the identification result of Székely and Rao (2000), we identify the marginal distributions of latent variables underlying the observed scores, including the error terms. By expanding our Hilbert space to vectorial random variables, we generalized all these results to multidimensional latent variables too. Then we implement the non-parametric estimator proposed by Bonhomme and Robin (2010). In order to evaluate under which conditions the estimator was able to recover the marginal distribution of the latent variables, we perform a simulation study. Finally, we used real data

and found that there are marginal distributions different from the standard normal distribution ( $\sim N(0, 1)$ ).

# Resumen

En los modelos de Teoría de Respuesta al Ítem (TRI) las variables latentes son especificadas por el Axioma de Independencia Local (ALI). De este modo las variables latentes explican las variaciones no aleatorias de los puntajes observados en una prueba. Ellas se utilizan para representar elementos más substantivos de alguna teoría. Por ejemplo, en modelos TRI son interpretadas como habilidades cognitivas de las personas en algunas ocasiones. Decisiones de alto impacto son tomadas a partir de las estimaciones de estos modelos, tanto en Chile como en el extranjero. Sin embargo, a pesar de su importancia, en el proceso de estimación de puntajes la distribución marginal de ellas es asumida, no es deducida desde los puntajes observados. Más aún, se ha demostrado que en los modelos TRI la distribución de las variables latentes no es identificada. El principal objetivo de esta tesis fue identificar la distribución de las variables latentes utilizadas en un conjunto de modelos TRI, y estimar esta distribución.

Nosotros enmarcamos todos los modelos de TRI que pueden ser formulados como modelos generalizados lineales mixtos en un espacio de Hilbert. En este enfoque, el ALI es reemplazado por el Axioma de Ortogonalidad Condicional (WALI), una versión débil de ALI. Encontramos que pueden haber variables latentes tomando valores de entre los puntajes observados para las cuales WALI es cierto. Es decir pueden haber variables latentes no-observadas y no no-observables. Además, utilizando el resultado de identificación de Székely and Rao (2000), identificamos la distribución de todas las variables latentes involucradas en los modelos TRI, incluidos los errores y variables latentes multidimensionales. Expandiendo nuestro espacio de Hilbert a variables aleatorias vectoriales, generalizamos todos nuestros resultados a variables latentes multidimensionales. Luego implementamos el estimador no parámetrico propuesto por Bonhomme and Robin (2010). Para conocer bajo qué condiciones el estimador es capaz de recuperar las distribuciones marginales de las variables latentes

realizamos un estudio de simulación. Finalmente aplicamos este procedimiento de estimación con datos reales, y encontramos que la distribución marginal de la variable latente que representa el constructo de interés sicológico-educacional no es una distribución normal estándar ( $\sim N(0, 1)$ ).

# I. Introduction

The latent variable concept has been widely used in Educational and Political decisions. For example, in value-added models, a latent variable is interpreted as the school effect (Raudenbush & Willms, 1995). In IRT models, latent variables are used to represent information about persons (Embretson & P. Reise, 2000) . Based in the estimates values of those latent variables, high impact decisions are made. Ranking of schools, selection of students, scholarships and resource distributions are just some examples (Biesta, 2014; Raudenbush, 2004; Koretz, 2008; Hargreaves & Braun, 2013). Therefore, it is important to study these kinds of models to improve the rightness and fairness of those decisions. In this work we studied the role of Latent variables in a set of IRT models. Specifically, on those which can be formulated as generalized linear mixed models (GLMM).

Latent variables have a major importance in IRT. The collected datas from a test application are represented by observed variables, typically called observed scores, whereas some aspects of a substantive theory are represented by latent variables. Latent variables are intended to explain the observed variables. Sometimes they are related with cognitives attributes of persons, depending on the items (or questions) used, they can represent math or language skills, anxiety ... etc. Using latent variables, IRT (and social science in general) can model a structure within the observed data, and not just describe the observed information through a small amount of parameters. This structure is described by the relation between the observed information and the latent variables (Pearl, 2009; Koopmans & Reiersol, 1950; Mouchart, Russo, & Wunsch, 2010). ALI specified the latent variables of interest by describing the structure of the datas. Due to this structure the joint distribution of the ob-

served information and the latent variables is decomposed as the product between a particular (marginal) probability distribution of the latent variables and the conditional distribution of the observed scores respect to the latent variables. Thus, this marginal-conditional decomposition is taking into account that latent variables explain the non random variations in the observed score, variations in latent variables precede variations in the observed information. Hence, it is (at least) possible to discuss if latent variable cause observed information (Sobel, 2011). But, in most cases, not even the importance of latent variables prevent us of assuming the marginal distribution of them.

Knowing the marginal distribution of latent variables ( $G_\theta$ ) is important. In the estimation process usually is assumed a standard normal distribution (Harwell, Baker, & Zwarts, 1988). However, in a Rasch model, if the real  $G_\theta$  ( $G_\theta$  in the population) of interest is not normal, the accuracy of the estimates values decrease when this assumption is made. This is true for either items parameters or person parameters (Woods & Lin, 2009; Finch & Edwards, 2016). On the other hand, a necessary condition for the estimated parameters in a model to be meaningful, it is the identifiability of those parameters in that model (Koopmans & Reiersol, 1950; Manski, 2007). The identification problems are different than the estimation problems. The former looks for the conclusions that can be made for a given a population and a model, when there is not restriction in the size of the sample. In other words, it solves the problem of which parameters are estimable such that, they have an interpretation which is deduced from the available datas. While the estimation problems are related with how to estimate the parameters given a sample. Usually, estimation problems are solved once a large sample is available. Thus, in any psychometric investigation (and in any investigation) identifiability is a necessary conditions for estimations, even when a Bayesian approach is used (San Martín, 2018; San Martin, Jara, Rolin, & Mouchart, 2011). However, to the best of our knowledge, all the improvements made in the estimation of  $G_\theta$  (Finch & Edwards, 2016; Karabatsos & Walker, 2009) were based

on IRT models that can be formulated as GLMM. And  $G_\theta$  is not identified in those models (San Martín, Rolin, & Castro, 2013; San Martín, González, & Tuerlinckx, 2015; San Martín et al., 2011).

The main goal of this thesis was to identify and estimate de marginal distribution of the latent variables in a set of IRT models. To this end, we framed all the IRT models that can be formulated as GLMM in a Hilbert space. Zimmerman (1975) explained how suitable Hilbert spaces are to frame psychometric models. Some common assumptions such as independence of the true score (conditional expected value) and error, are natural properties of linear operators on these spaces. Our approach to use Hilbert spaces geometry for the study of latent variables can be understood as inspired by Zimmerman's work (Zimmerman, 1975). However, is important to highlight a difference with his work (and some others). We do not restrict the latent variable to be the expected value of some observed random variable conditional on a randomly selected person (random point). This restricts latent variables to the sigma algebra generated from persons as a random variable (the random variable being the process of picking persons). We start from something more general, using only the restriction that random variables have finite variance (observed and latent). Next, we explore the meaning of the Axiom of Local Independence (ALI) defined on two observed scores respect to a latent variable. Our proposal to analyze the information from test data is very general. There are no restrictions with respect to the format of the observed information. The scores can be binary variables, sums of binary variables, etc. Any function (i.e. random variable) of the raw observed scores with finite variance qualifies. We worked with random variables from any probability distribution. This approach allows to work with unidimensional as well as with multidimensional latent variables. We have coined our approach as *Neo-Classical Test Theory*, because it can be understood as a new formulation and an extension of Classical Test Theory (CTT)(Holland & Hoskens, 2003; Dunn, 1992).

It is important to mention the term *Neo-Classical Test Theory* was already used

by Drewes (2017), however there is a fundamental difference with his work. Drewes (2017) define the true score directly as the latent variable representing some attribute of a person, but, how was already explained, we use the conditional expected value of the observed information on the latent variable of interest as the true score. In this way we analyze the relation between the observed score and the latent variables. In particular we studied this relation under WALI.

The structure of this manuscript is as follow. In the first chapter we formally define some keys concept in this work (observed scores, latent variable, subspaces...etc.). Then, by embedding IRT models in an orthogonal general orthogonal decomposition, we studied some properties of them and the circumstances under which the latent variables are identified. In the second chapter we use the identification result to estimate the marginal distribution of latent variable. We show some simulations and an application with real datas obtained from the PSU. The mathematical details of some definitions, proofs and used codes are presented in Appendixes.

## II. Neo-Classical Test Theory: A Geometrical Approach

In this chapter we present the theoretical bases of our approach. It is briefly explained what a Hilbert space is. It includes explanations of how random variables are represented in a Hilbert space. And by doing so, we introduce concepts as Observed Subspaces and Latent Subspaces. Then we made use of the geometry of Hilbert spaces to analyze IRT models.

### 2.1 Observed score, Latent variables and Hilbert space

Let  $H$  denote the Hilbert space of all real random variables with finite variance (Small, 1994). Thus  $H$  is the probabilistic space (see Appendix 1 for details on probabilistic spaces and  $L^2$ )

$$L^2(M, \mathcal{M}, P) = \{f : (M, \mathcal{M}) \rightarrow (\mathbb{R}, \mathcal{B}) \text{ is a random variable: } E(|f|^2) < \infty\}. \quad (2.1)$$

$M$  contains all the relevant information on the observed scores and the latent variables. This includes all the possible outcomes from a test and the items, along with

all the (raw) information about the latent variable. The sigma algebra generated by a random variable contains all the events (associated with this random variable) such that it is possible to assign a probability to them (see Appendix 2 for a formal definition of sigma algebra). All the linear combinations of the elements in the sigma algebra of the observed scores and the latent variable are contained in  $\mathcal{M}$ , the sigma algebra of our probabilistic space. If we denote  $\sigma(y)$  and  $\sigma(\theta)$  the sigma algebras generated by the observed scores and the latent variables respectively, and  $\sigma(y) \vee \sigma(\theta)$  the set of linear combinations among the elements in  $\sigma(y)$  and  $\sigma(\theta)$  (it is a subspace actually), then the above affirmation is written as  $\sigma(y) \vee \sigma(\theta) \subset \mathcal{M}^1$ .  $\mathbb{R}$  and  $\mathcal{B}$  are the Reals and the Borel  $\sigma$ -algebra respectively. Note  $H$  is the space of all random variables with domain  $(M, \mathcal{M})$  and having as range all the Borel sets in the real line. Thus  $H$  is the space of  $\mathcal{B}$ -measurable random variables with finite variance. Observe that any function of a random variable in  $H$  is also an element of  $H$ , even if this function is not linear. We will be back on this point when we present how different IRT models can be embedded in this approach.

Now let  $Y$  and  $\Theta$  be two subspaces of  $H$ , thus two Hilbert spaces included in  $H$  (see Appendix 3 for a formal definition of a subspace). The former is the subspace containing all the observed test information. These are all possible real value functions of the raw test score which belong to  $H$ . Thus  $Y$  contains all the random variables with finite variance defined on (i.e. with domain in)  $(M_y, \sigma(y))$ . We call these random variables observed scores, and are denoted by the letter  $y$ .  $M_y$  contains all the relevant information from the observed scores. And  $\sigma(y)$  is the sigma algebra of this subspace such that, if  $B$  is any Borel set in the real line and  $y$  is any element (random variable) in  $Y$ , then  $y^{-1}(B) \in \sigma(y)$ . Hence,  $Y$  contains all the possible random variables with finite variance defined on the observed information, such that it is possible to assign a probability to them. Due to  $Y$  is a vectorial subspace (see

---

<sup>1</sup>In this work we use the convention that a subspace can be contained in itself, which allows for  $B \subset B$ . This convention is important in the interpretation of Theorem 1 about reliability, and in the interpretation of Equation 2.33.

Appendix 4), any linear combinations of two or more random variables is also a random variable (closed under the sum). Thus,  $Y$  also contains all the linear combinations among the random variables defined over the raw test score. Now let  $j$  be an integer subindex indicating a specific test (or item, we do not make a difference between them), then  $Y_j$  is the subspace of  $Y$  containing all the observed scores  $y_j$  (with finite variance) defined on the test (or item)  $j$ .

Table 2.1 shows an example of five observed scores defined in a test denoted by the number 1. The test had six items, each one with four alternatives. The first column shows a realization of the raw test information, thus, a pattern of answers from one person. The other five columns shows the realizations of five different observed scores. From left to right, the first observed score  $y_1$  is obtained according to some criterion relating alternatives to numbers, a Likert Scale for example. The second observed is a binary variable assigning ones and zeros. The third observed score is the sum of the first observed score. This observed score is usually used as an informative variable of the learning level reached. The fourth observed score is a sum of the first three and the last three items separately. And the last observed score is the mean of the components of the fourth observed score.

Table 2.1. *Examples of Observed Scores.*

Raw test Information	$y_1$	$y'_1$	$y''_1$	$y'''_1$	$y''''_1$
b	0	0			
d	3	1		4	
a	1	0			
a	2	1	7		3.5
b	0	0		3	
b	1	0			

*Note.* The first column shows a pattern of answers from a examinee in a test of six items, each one with four alternatives. The other columns are the realizations of five different observed scores defined on the test 1.

In addition, we denote by  $\Theta$  the subspace containing all the information of the

latent variable. This is the subspace containing the information of interest to the researcher. It is the subspace containing all the random variables with finite variance defined on  $(M_\theta, \sigma(\theta))$ . Again,  $\sigma(\theta)$  is such that  $\theta^{-1}(B) \in \sigma(\theta)$  where  $\theta$  is any element in  $\Theta$  and  $B$  is any Borel set in the real line, so that a probability measure can be assigned to every element (random variable)  $\theta$ . Hence  $\Theta$  is the subspace containing the information on the attribute for which the test was constructed. We do not make any assumption on the intersection of these two subspaces ( $Y \cap \Theta$ ). Depending on what is in this intersection, we can conclude that the information in  $\Theta$  is not observable, not observed or something of both. If the intersection contains only the zero element<sup>2</sup>, we conclude latent variables are unobservable, but if  $\Theta$  is completely contained in  $Y_j$ , we conclude latent variables are unobserved but not unobservable, because they take the same values as an (any) observed score. It is worth to mention that any Hilbert subspace is a Hilbert space too, thus  $Y$  and  $\Theta$  are Hilbert spaces as well and hence, a unique probability measure is defined on each one of them.

It is important to clarify the different meanings of the term *dimension* we use in this work. There is the geometrical dimension of any element in a Hilbert Space and the dimension of the traits they represent. Regarding the geometrical dimension, in the subspaces  $Y$  and  $\Theta$  each random variable (i.e. each function) is a vector, where the components are given by the possible values that the function can take. Thus, if the cardinality of the range of a function is infinity, the dimension of the vector representing that function could be as well. Because each realization of a  $\theta \in \Theta$  (and  $y \in Y$ ) corresponds to one sampled value, the dimension of any element  $\theta$  (and  $y$ ) is given by the cardinality of his sample space. The dimension of the subspace  $\Theta$  (and  $Y$ ) equals the dimension of the element  $\theta$  in  $\Theta$  ( $y$  in  $Y$ ) with the largest dimensionality. Hence, even if only one element in  $\Theta$  (or  $Y$ ) is infinite-dimensional, the dimension of  $\Theta$  (and  $Y$ ) would be as well. Then, the sample space where any element of  $\Theta$  is defined is a subset of  $M_\theta$ . Now, regarding the dimensionality of

---

<sup>2</sup>The zero element is a function which map every element in the sample space to zero.

the attributes represented by the elements in  $\Theta$ , note we used functions with ranges in  $\mathbb{R}$  (not in  $\mathbb{R}^n$ , with  $n$  as an integer greater than one), and thus, the attributes represented by the elements in  $\Theta$  are unidimensional. In order to generalize our results to multidimensional latent traits, we need to expand our Hilbert space to random variables with ranges in  $\mathbb{R}^n$ , with  $n$  being an integer greater than one. In this work the multidimensionality of the attributes come from multiple attributes explaining the observed scores in one test. Once we incorporate some relevant results, we expand the Hilbert space to multidimensional random variables, i.e., multiple attributes.

Also note that the sample space is not restricted to the sample of examinees. It is the total space to sample from. Therefore, for example, the variance of  $E(w|x)$ , where  $w$  and  $x$  are random variables, must to be interpreted in general as the variation of  $E(w|x)$  across different values of  $x$ , and not necessarily across different persons or within a person. However, if the datas come from a group of persons, then the realizations of the variable (sampled values) will come from different persons, and then the variance must be interpreted as the variance between persons. Because this is commonly the case, and to avoid confusion, we refer only to this type of applications. Also, it is important to highlight that we use one probability measure in each subspace, which means that we assume homogeneity across persons, so that each realization of the latent variable comes from an identical independent probability distribution (*iid*). Without this assumption estimation procedures could only be applied to data which came from one person. However, it is discussable whether the density function of the latent variable is the same for all persons, but also whether it remains constant during a test application. Without the *iid* assumption, we would have subsets of datas from an unknown number of different probability distributions, so that it would be impossible to recover information from any of those density functions. Furthermore, it is in these distributions where the information on the attributes of interest resides. Hence, when we use the ALI to specify a latent variable,

we refer to a latent variable which we do not know his substantive meaning, but we know that it represents an attribute which is common across the persons, and persons with equivalent levels in the latent trait are indistinguishable with respect to their tendency to produce observed scores in a test. This does not mean that equal levels of  $\theta$  imply equal observed scores, but the distribution of the observed scores is determined by the levels of  $\theta$ <sup>3</sup>(this is the definition of ALI). Now that the assumptions are clarified we present some features of Hilbert spaces.

A Hilbert space is a complete inner product space (Halmos, 1957). This means: a) all Cauchy sequences (see Appendix 5) from elements in  $H$  converge to an element in  $H$ ; it can be said that there are no “holes” or emptiness in it; b) because Hilbert spaces are provided with inner products (see Appendix 6), concepts like length and angles can be used. An inner product is used to induce a norm (length) and the sense of angle in a Hilbert space. This last sentence will become more clear in the next paragraph.

The expected value of the product between two real-valued random variables is an inner product in  $H$  (Florens, Marimoutou, & Pegin-Féissolle, 2007; Lindquist & Picci, 2015) (see Appendix 7). If  $w$  and  $x$  are two random variables in  $H$ , we denote the inner product between them as  $\langle w, x \rangle = E(w \cdot x)$ . The norm of a random variable  $x$  induced by this inner product is given by  $E(x^2) = \langle x, x \rangle = \|x\|^2$ . Then, assume that  $E(x) = 0$  implies that  $Var(x) = \|x\|^2$ . Also, the inner product allows a definition of orthogonality. Two random variables  $w$  and  $x$  are orthogonal, denoted by  $w \perp x$ , if and only if  $\langle w, x \rangle = 0$ , thus in  $H$  if  $\theta_1 \perp \theta_2$  then  $E(\theta_1 \cdot \theta_2) = 0$ . The orthogonal complement of any subspace  $X$  of  $H$  is denoted by  $X^\perp$ . It is the subspace in  $H$  in which all the elements are orthogonal to each element in  $X$ . Thus,

$$X^\perp := \{h \in H : \langle h, x \rangle = 0, \forall x \in X\}. \quad (2.2)$$

---

<sup>3</sup>For more statistical details on this comments see Gyenis, Hofer-Szabó, and Rédei (2017).

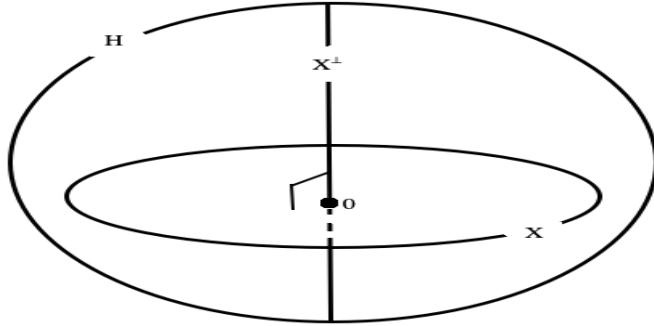
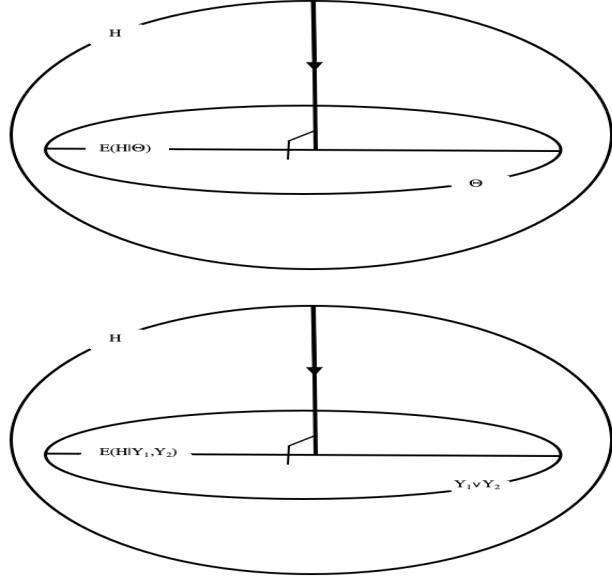


Figure 2.1. Orthogonal complement of a subspace  $X$ , the point zero indicate the zero element.

An illustration of the orthogonal complement of  $X$  is showed in Figure 2.1. According with the notation of Stone (1991), the lines in the intersection point indicate  $90^\circ$  with the entire space  $X$ . A subspace can be denoted either for a line or using an ellipse up to whether or not it is desirable to indicate other structures within the subspace. Arrows will indicate operators.

On the other hand, the conditional expected value is an orthogonal projection (Florens et al., 2007; Lindquist & Picci, 2015) (see Appendix 8). Then for any  $y$  in  $Y$ ,  $E(y|\Theta)$  is the orthogonal projection from the vector  $y$  to  $\Theta$ , it is *the* closest element of  $\Theta$  to  $y$ . There is only one element in  $\Theta$  which is the closest to  $y$ . The sense of length is given by the norm induced from the inner product. The vector from the orthogonal projection of  $y$  onto  $\Theta$  to  $y$ , belongs to  $\Theta^\perp$  (note that a vector is a random variable in our  $H$ ). Because of expected values properties, if  $y = y_1, y_2, \dots, y_n$  is a  $n$  dimensional observed score (a  $n$  dimensional random vector), then, without restoring in any argument of orthogonality,  $E(y_1, y_2, \dots, y_n|\Theta) = E(y_1|\Theta), E(y_2|\Theta), \dots, E(y_n|\Theta)$  (Rencher & Schaalje, 2008). We can also perform an orthogonal projection from an element  $\theta$  onto an observed subspace  $Y_j$ . If we use two observed scores to approximate  $\theta$ , then  $j = 1, 2$ . This correspond to an orthogonal projection of  $\theta$  onto all the linear



*Figure 2.2.* Two orthogonal projection, one onto the subspace  $\Theta$  and the other onto the span generated from  $Y_1$  and  $Y_2$ .

combinations among the elements in  $Y_1$  and  $Y_2$  (the span generated from  $Y_1$  and  $Y_2$ ,  $Y_1 \vee Y_2$ ), which we denote by  $E(\theta|Y_1, Y_2)$ . Figure 2.2 illustrate an orthogonal projection from  $H$  onto  $\Theta$  and from  $H$  onto the span generated by  $Y_1$  and  $Y_2$ .

Finally, the cosine of the angle between two subspaces (Bouldin, 1973) is given by,

$$\cos\alpha := \sup \left\{ \frac{\langle f, g \rangle}{\|f\| \|g\|}; f \in M, g \in N \right\}, \quad (2.3)$$

where  $\alpha$  is the angle between the subspaces  $M$  and  $N$  taking values from the interval  $[-\pi/2, \pi/2]$  (in radians), and  $\sup$  denotes the supremum. In what follows we use only random variables with zero mean. Because concepts like angles and length are invariants under translations, this assumption do not means a lose of generality.

Following the notation of Stone (1991) we provided some figures to illustrate the geometric interpretation of this work. With all these elements we can rescue the greek tradition and do *geo-metric* with psychometric models.

## 2.2 Neo-Classical Test Theory

From the axiomatization of Lord and Novick (1968), we can rewrite the CTT as

$$y_j = E(y_j|\Theta) + (y_{ij} - E(y_j|\Theta)) \quad (2.4)$$

where  $y_j$  is an element of  $Y_j$  representing an observed scores in a test  $j$ , it is one among all the random variables with finite variance that can be defined over the outcomes in a test  $j$ . Equation 2.4 is a typical orthogonal decomposition in  $H$  (see Appendix 5). Note that  $E(y_j|\Theta) \in \Theta$  and  $(y_j - E(y_j|\Theta)) \in \Theta^\perp$ . Because  $E(y_j|\Theta)$  is an orthogonal projection,  $E(y_j|\Theta)$  is *the* element in  $\Theta$  which minimize the size of  $y_j - E(y_j|\Theta)$  given by  $\|y_j - E(y_j|\Theta)\|$ . The error term  $(y_j - E(y_j|\Theta))$  is a consequence of  $Y_j$  being different from  $\Theta$ , in other words, this is the result of try to measure  $\Theta$  trough  $Y_j$  by the true score  $E(y_j|\Theta)$ , and not directly. As in CTT, Decomposition 2.4 make persons equivalents, but respect to latent variable realizations instead of observed score realizations. Equation 2.4 is the starting point of the Neo-Classical Test Theory (N-CTT). This is why we may refer to Equation 2.4 as N-CTT. Figure 2.3 illustrate this decomposition in the entire subspace  $\Theta$ . Note in this figure the arrow is used to indicate the orthogonal projection operator from a subspace  $Y$  onto  $\Theta$ .

Because always in  $H$   $E(y_j|\Theta) \perp (y_j - E(y_j|\Theta))$ , and taking into account that

Following the notation of (Stone, 1991), this is an illustration of the orthogonal decomposition.

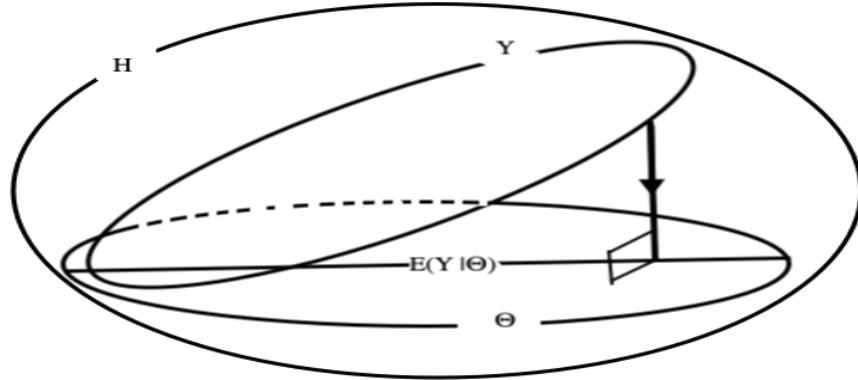


Figure 2.3. The set of all true score in our space  $H$ .

variance is the square of the length of a vector, it is possible to see that

$$\langle y_j, y_j \rangle = \langle E(y_j|\Theta) + y_j - E(y_j|\Theta), E(y_j|\Theta) + y_j - E(y_j|\Theta) \rangle \quad (2.5)$$

$$\|y_j\|^2 = \|E(y_j|\Theta)\|^2 + \|y_j - E(y_j|\Theta)\|^2. \quad (2.6)$$

Because the variance is the norm of a vector (a random variable) in this framework, more variability is equivalent to a larger  $y_j$ .

All this may seem very familiar. What is new in our formulation is that Equation 2.4 is a general decomposition, for discrete as well as for continuous observed variables. From a psychometric point of view N-CTT is a general decomposition. Some IRT models can be embedded in this framework if the true score can be formulated as generalized linear mixed models (de Boeck & Wilson, 2004). What changes from one model to another is the link function (Krzanowski, 2002) used in the *true score*  $E(y_{ij}|\Theta)$ . For example, in the Rasch model the link function is the logit function.

When a link function is used to model the true score in N-CTT, we take a decision on which subspace of  $\Theta$  we are orthogonally projecting. For instance, if we use the identity link function and an unidimensional latent variable, then we are project onto the subspace of  $\Theta$  composed of linear functions of an element  $\theta$  i.e.  $\phi(\theta) = \beta_0 + \beta_1\theta^4$ . Hence the orthogonal projection interpretation still holds for different latent variables. We believe that it is beneficial to study psychometric models at this level of generality.

### 2.2.1 Classical and Dual reliabilities

Before analyzing the reliability associated to a test from the Hilbert space framework, let's look the next variance decomposition of the observed scores,

$$Var(y_j) = Var[E(y_j|\Theta)] + E[Var(y_j|\Theta)]. \quad (2.7)$$

We can see there is a classification problem underlying the CTT. The latent variable induce groups in the realizations of an observed score, and those groups depends on the distribution of the observed score conditional to the latent variable. Each group contains a set of observed scores realizations associated with a given realization of  $\theta$ . Hence, the *total* variance can be partitioned in two parts: the true score variance  $Var[E(y_j|\Theta)]$  (*between* groups); and the variance within realizations of  $\theta$ , which is represented as the expected value across the observed score variance given a realization of  $\theta$   $E[Var(y_j|\Theta)]$  (*within* groups)<sup>5</sup>. The psychometric problem behind

---

<sup>4</sup>In a more formal way, and using the standard notation in probabilities, observe that  $\phi(\theta) = \phi(\theta(\omega)) = \phi \circ \theta(\omega) = g(\omega)$ , hence the random variable  $\phi$  (a function), proved that has finite variance, belong to our Hilbert space.

<sup>5</sup>It is important to pointed out a mistake of Holland and Hoskens (2003) in the Appendix B. They said that  $Var(y_j - E(y_j|\Theta))$  is different from  $E[Var(y_j|\Theta)]$ , but comparing Equation 2.6 and

this is: once we have a realization of the observed score, we want to know from which level in the latent trait they came.

Now, the reliability  $\eta_{y_j|\theta}$  in the test  $j$  is defined by the ratio (Holland & Hoskens, 2003)

$$\eta_{y_j|\theta} := \frac{Var[E(y_j|\Theta)]}{Var(y_j)} \left( = \frac{\|E(y_j|\Theta)\|^2}{\|y_j\|^2} \right), \quad (2.8)$$

for every  $y_j$  in  $Y_j$ . This is the proportion of the observed score variance which correspond to the true score variance (Lord & Novick, 1968; McDonald, 2011). From the decomposition in Equation 2.7 it follows that  $\eta_{y_j|\theta} \in [0, 1]$ . By specifying  $y_j$ , the definition given by Equation 2.8 is open to a dependency on the score formula used. Note that if  $Y_j \perp \Theta$  then  $E(y_j|\Theta) = 0 \forall y_j \in Y_j$ , therefore  $\eta_{y_j|\theta} = 0$ . The same holds in the opposite implication order, this is,  $\eta_{y|\theta} = 0$  implies  $Y_j \perp \Theta$ . Thus, it is possible to group the realizations of the observed scores in values of true scores if and only if  $Y_j$  is not orthogonal to  $\Theta$ . On the other hand if  $\eta_{y_j|\theta} = 1$  then  $\|E(y_j|\Theta)\| = \|y_j\|$ , which implies that  $Y_j$  is contained in  $\Theta$  ( $Y_j \subset \Theta$ ). Hence, every observed scores  $y_j$  will have the same outcomes that some  $\theta$  in  $\Theta$ . Therefore the reliability can be one only if every observed score defined in a test  $j$ , almost sure<sup>6</sup>, is equal to a latent variables in  $\Theta$ .

The main goal of a psychometric model is to capture information of some  $\theta$  in  $\Theta$  from the observed score in a test  $j$ . That is why the decomposition in Equation 2.9 is of interest:

---

Equation 2.7 they are not different.

<sup>6</sup>See Appendix 7 for an explanation of why we use an almost sure equality and not just an equality.

$$\theta = E(\theta|Y_j) + (\theta - E(\theta|Y_j)). \quad (2.9)$$

Here  $E(\theta|Y_j) \perp (\theta - E(\theta|Y_j))$ . Recall  $E(\theta|Y_j)$  is the best approximation of  $\theta$  from the subspace  $Y_j$  in a minimal error size sense. This is the EB of  $\theta$ . The variance of  $(\theta - E(\theta|Y_j))$  is the prediction error. Also note that

$$\|E(\theta|Y_j)\| \leq \|\theta\|, \quad (2.10)$$

where equality holds when the norm of the prediction error is zero (i.e.,  $\Theta \subset Y_j$ ), and how will be explained in the section 2.2.3, when ALI is true. Thus, unless one of these conditions is true, the EB always shows shrinkage compared with the latent variable. It has no sense correct for shrinkage because there is an inherent loss of information in projecting, and when no information is lost, the EB does not show shrinkage.

By decomposing the variance of  $\theta$  as,

$$Var(\theta) = Var[E(\theta|Y_j)] + E[Var(\theta|Y_j)] \quad (2.11)$$

it is showed that  $E[Var(\theta|Y_j)]$  is the quantification of the prediction error, and not

$Var(\theta|Y_j)$  as had been proposed by people working in generalized linear models (Skrondal & Rabe-Hesketh, 2009; Welham, Cullis, Gogel, Gilmour, & Thompson, 2004; McCulloch & Neuhaus, 2011). Based on Equation 2.11 it is possible to define a dual reliability in a test  $j$  as,

$$\eta_{\theta|y_j} := \frac{Var[E(\theta|Y_j)]}{Var(\theta)} \quad \left( = \frac{\|E(\theta|Y_j)\|^2}{\|\theta\|^2} \right). \quad (2.12)$$

In analogy with the classical reliability definition in Equation 2.8, from Equation 2.12 it follows that:

1.  $\eta_{\theta|y_j} \in [0, 1]$ ,
2.  $\Theta \perp Y_j \Leftrightarrow \eta_{\theta|y_j} = 0 \quad \forall y_j \in Y_j$  and
3.  $\eta_{\theta|y_j} = 1$  if every latent variable in  $\Theta$  is almost sure equal to an observed score  $y_j$  in  $Y_j$  ( $\Theta \subset Y_j$ ).

### 2.2.2 Relation between Reliabilities

The symmetry between the dual and *classical* reliabilities suggest a relation between them. The following theorems summarize this relation.

**Theorem 1** *The classical and dual reliabilities are equals.*

To prove theorem 1 we need the following lemma.

**Lemma 1** *The classical reliability for a given observed score is the square cosine of the angle between the observed score and the true score.*

In order to proof lemma 1 observe that

$$\begin{aligned}
\langle y_j, E(y_j|\Theta) \rangle &= \langle E(y_j|\Theta) + (y_j - E(y_j|\Theta)), E(y_j|\Theta) \rangle \\
&= \langle E(y_j|\Theta), E(y_j|\Theta) \rangle.
\end{aligned} \tag{2.13}$$

Because we only use random variables with zero expected value, the numerator in Definition 2.8 is the last term in the right side of Equation 2.13. Hence, the variance of the true score can be represented as the inner product between the observed score and the true score. Thus the classical reliability can be rewritten as

$$\eta_{y_j|\theta} = \frac{\langle y_j, E(y_j|\Theta) \rangle^2}{\|y_j\|^2 \|E(y_j|\Theta)\|^2}. \tag{2.14}$$

This is the square cosine of the angle between the observed score  $y_j$  and the true score  $E(y_j|\Theta)$ . When the angle between these two vectors is close to zero, the reliability tends to one. For different applications of a test in different populations of persons, the observed scores and the true scores will be different. As a result, does not make sense to force this angle to be invariant across populations (of persons). The reliability as defined in Equation 2.8 (and 2.14), does not only depends on the score formula used, but also on the population of examinees. Therefore, the reliability is not a property of a test, but of the scores formula used together with the population of examinees. When we expand  $H$  to the space of vectorial random variables (functions in  $R^n$ ), Definition 2.14 will apply to multidimensional latent variables too, and these observations will still be valid.

Now we are closer to proof theorem 1. In first place note that we can use the

same procedure of lemma 1 in Definition 2.12. Thus the dual reliability is

$$\eta_{\theta|y} = \frac{\langle \theta, E(\theta|Y_j) \rangle^2}{\|\theta\|^2 \|E(\theta|Y_j)\|^2}, \quad (2.15)$$

this is the square cosine of the angle between the latent variable  $\theta$  and the EB for each value of  $\theta$ . For notation convenient, and to simplify Equation 2.14 and Equation 2.15, let  $\hat{h}$  denote  $h/\|h\|$ , where  $h$  is any vector in  $H$  (any random variable), and  $\hat{h}$  is a vector of norm 1 in the same direction as  $h$ . Thus

$$\eta_{y_j|\theta} = \langle \hat{y}_j, \widehat{E(y_j|\Theta)} \rangle^2 \quad \text{and} \quad \eta_{\theta|y_j} = \langle \hat{\theta}, \widehat{E(\theta|Y_j)} \rangle^2. \quad (2.16)$$

On the other hand, it can be proved that any subspace  $X$  can be orthogonality decomposed by using another subspace  $W$  as

$$X = \bar{E}(W|X) \oplus X \cap W^\perp. \quad (2.17)$$

Where  $\bar{E}(W|X)$  is the smallest subspace which includes all the orthogonal projections from  $W$  to  $X$ , and  $X \cap W^\perp$  is the subspace composed from all the elements in common between  $X$  and the orthogonal complement of  $W$ , thus every element in this intersection is orthogonal to  $W$ . In Equation 2.17, and in all equations,  $\oplus$

denote direct sum (see Appendix 10 for details on this decomposition).

Now, let us denote by  $y_{j\perp}$  and  $\theta_\perp$  any element in  $\Theta \cap Y_j^\perp$  and  $Y_j \cap \Theta^\perp$  respectively. Using the decomposition in Equation 2.17 for the elements  $\hat{\theta}$  and  $\hat{y}_j$  we obtain,

$$\langle \hat{y}_j, \hat{\theta} \rangle = \langle \hat{y}_j, E(y_j|\Theta) \oplus y_{j\perp} \rangle = \langle \hat{y}_j, E(y_j|\Theta) \rangle \quad (2.18a)$$

$$\langle \hat{y}_j, \hat{\theta} \rangle = \langle E(\theta|Y_j) \oplus \theta_\perp, \hat{\theta} \rangle = \langle E(\theta|Y_j), \hat{\theta} \rangle. \quad (2.18b)$$

This procedures is valid for any  $y_j$  in  $Y_j$  and for any  $\theta$  in  $\Theta$ . Also, because in real values random variables  $\langle x, w \rangle = \langle w, x \rangle$ , and as a consequence of the angles invariance respect to changes in the norms of vectors, it is concluded that

$$\eta_{y_j|\theta} = \eta_{\theta|y_j} \quad \square. \quad (2.19)$$

Both reliabilities are the same because each of them is the angle between the observed score and the latent variable. Theorem 1 implies that in test theory  $\eta = 1$  if and only if  $\Theta = Y$ . Thus, a reliability equal to one is possible only if we have a direct “look” on what we are interesting in. Furthermore,  $\eta = 1$  means that the realizations of a  $y_j$  are in one to one correspondence with the realizations of some latent variable  $\theta$  in  $\Theta$ . Thus, there will be one true score for each realization of  $y_j$ .

Given Theorem 1, we propose the next definition of reliability in a test  $j$  applied in a given population,

$$\eta_j := \sup \left\{ \langle \hat{y}_j, \hat{\theta} \rangle \right\}. \quad (2.20)$$

Due to the definition in Equation 2.3, Definition 2.20 is the square cosine of the angle between  $Y_j$  and  $\Theta$ . Definition 2.20 is the maximum possible reliability across the observed scores that can be defined in a test  $j$  and the latent variables in a given population. The maximum occurs when the angle between the observed score and the latent variable is the closest to zero, or zero.

Now let consider two observed subspaces  $Y_1$  and  $Y_2$  intended to obtain information of the same latent variable  $\theta$ . The instruments associated with  $Y_1$  and  $Y_2$  are not necessarily parallel. A dual decomposition of this situation is

$$\theta = E(\theta|Y_1, Y_2) + (\theta - E(\theta|Y_1, Y_2)), \quad (2.21)$$

where the EB  $E(\theta_i|Y_1, Y_2)$  is the orthogonal projection onto all the possible linear combinations between the elements in  $Y_1$  and  $Y_2$ .

We can use another orthogonal decomposition by partitioning the true score as follows

$$\theta = E(\theta|Y_1) + (E(\theta|Y_1, Y_2) - E(\theta|Y_1)) + (\theta - E(\theta|Y_1, Y_2)) \quad (2.22)$$

where all the quantities in parenthesis are orthogonal to each other. Note that the first two terms at the right side of the equation correspond to the EB estimator.

Now let decompose the variance of  $\theta$  using Equation 2.21 and Equation 2.22. Using the fact that variance is the norm of a vector, we get:

$$\|\theta\|^2 = \|E(\theta|Y_1)\|^2 + \|E(\theta|Y_1, Y_2) - E(\theta|Y_1)\|^2 + \|\theta - E(\theta_i|Y_1, Y_2)\|^2 \quad (2.23)$$

$$\|\theta\|^2 = \|E(\theta|Y_1, Y_2)\|^2 + \|\theta - E(\theta|Y_1, Y_2)\|^2. \quad (2.24)$$

From Equations 2.23 and 2.24 we can obtain the reliability as the norm of the EB over the norm of the latent variable.

$$\frac{\|E(\theta_i|Y_1, Y_2)\|^2}{\|\theta\|^2} = \frac{\|E(\theta|Y_1)\|^2}{\|\theta\|^2} + \frac{\|E(\theta|Y_1, Y_2) - E(\theta|Y_1)\|^2}{\|\theta\|^2}. \quad (2.25)$$

The first term at the right hand of the equation corresponds to the reliability associated with the scores of the first test. The second term is always positive (even if  $Y_1$  and  $Y_2$  are negatively correlated) hence, for every  $y_1, y_2$  in  $Y_1, Y_2$ , Equation 2.25 implies

$$\eta_{\theta|y_1, y_2} \geq \eta_{\theta|y_1}. \quad (2.26)$$

It is clear from the result that the dual reliability improves with the number of measures even if the measures contain errors, and even if the true scores and the reliabilities are different between them. Considering theorem 1, this is a generalization of the spearman-brown equation because we do not restored on parallels tests (Holland & Hoskens, 2003; Zimmerman & Williams, 1966).The inequality in equation 2.26 has other implications too.

**Proposition 1** *The precision of the Empirical Bayes estimator improves with the numbers of measurements.*

As consequence of Equation 2.11 the dual reliability improves when the prediction error is smaller, and thus, the precision of the EB improves. Hence, Proposition 1 is the result of Equations 2.19 and 2.26.

So far we have shown a new perspective to analyze psychometric models. We showed generalizations and demonstrations of some already known or intuitable results. Now we go one step further in the Hilbert space geometry. We introduce some new insights of the role of ALI when latent variables are studied from observed information, as in tests applications.

### 2.2.3 Splitting Subspaces and Axiom of Local Independence

The ALI is of great importance in social science. In psychometric, latent variables underlying a set of observed random variables are specified by the ALI. This means that after conditioning on it, the observed scores are independent. In other words, they explain the associations (correlation for example) among the observed scores. In linear subspaces, such as the observed and latent subspaces, this condition is replaced by conditional orthogonality. This is a weaker assumption than ALI (Florens, Mouchart, & Rolin, 1993; Florens & Mouchart, 1985), we call it WALI. Is weaker in the sense that ALI implies WALI, but WALI does not implies ALI (unless the joint distribution of the orthogonal variable is normal). If WALI is true for a latent

variable  $\theta$  respect to two observed scores  $y_1$  and  $y_2$ , then  $\theta$  explains the correlation between  $y_1$  and  $y_2$ , hence after conditioned on  $\theta$  the observed scores are no longer correlated. Recall that here observed scores are observed random variables from any probability distribution, such that they have mean equal to zero and finite variance. In what follows we expand this concept to subspaces.

Suppose we have two observed scores from two tests (or items)  $y_1$  and  $y_2$  related to the same latent variable  $\theta$ . For a given  $\theta$ , different realizations of  $y_1$  and  $y_2$  are possible. If those realizations are orthogonal for any value of  $\theta$ , we say they are orthogonal conditional on  $\theta$ , and this is denoted by  $y_1 \perp y_2 | \theta$ . If this is true for all possible observed scores defined on the outcomes in the tests (or items) 1 and 2, and for all possible latent variables (with finite variance and zero mean and from any probability distribution), we will say the subspaces  $Y_1$  and  $Y_2$  are orthogonal conditional on the subspace  $\Theta$ . This is denoted by

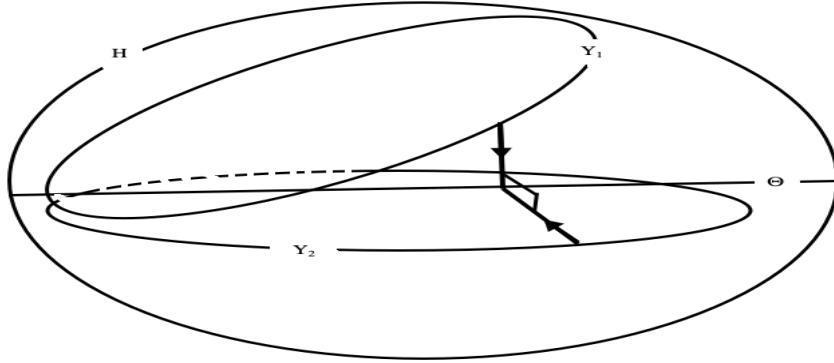
$$Y_1 \perp Y_2 | \Theta. \quad (2.27)$$

Lindquist, Picci, and Ruckebusch (1979) called splitting subspace the spaces that behave like  $\Theta$  respect to  $Y_1$  and  $Y_2$ . This means that for any  $y_1$  in  $Y_1$  and for any  $y_2$  in  $Y_2$ ,

$$\langle y_1 - E(y_1 | \Theta), y_2 - E(y_2 | \Theta) \rangle = 0. \quad (2.28)$$

Figure 2.4 illustrate the orthogonality of the errors.

Using the notation of Lindquist et al. (1979), let  $Y_0$  be the subspace formed from



*Figure 2.4.* The latent subspace  $\Theta$  is a minimal splitting subspace respect to the observed subspaces  $Y_1$  and  $Y_2$ .

the sum of  $Y_1$  and  $Y_2$  (i.e., all possible linear combinations between the elements in  $Y_1$  and  $Y_2$  denoted as  $Y_1 \vee Y_2$ ), then we say that  $\Theta$  splits  $Y_0$  into two parts  $Y_1$  and  $Y_2$  such that 2.28 hold. If  $\Theta$  is a subspace of  $Y_0$  then is called an internal splitting subspace, and if there is not a proper subspace of  $\Theta$  satisfying 2.27 (or 2.28) then  $\Theta$  is a minimal splitting subspace. Recall that the space where we work on is  $H$ , which is defined in Equation 2.1, section 2.3.1. Thus any subspace that we use will be a subspace of  $H$ .

**Lemma 2** *If  $\Theta$  is a minimal splitting subspace respect to  $Y_1$  and  $Y_2$ , the only element  $\theta$  for which  $E(\theta|Y_1) = E(\theta|Y_2) = 0$  is the zero element.*

In order to proof Lemma 2, we need to considerer that find all the  $\Theta$ 's which are minimal subspaces respect to  $Y_1$  and  $Y_2$ , is equivalent to find all the splitting subspaces satisfying (see Appendix 11)

$$\Theta \cap Y_1^\perp = 0 \quad (2.29a)$$

$$\Theta \cap Y_2^\perp = 0. \quad (2.29b)$$

On the other hand, we can apply the orthogonal decomposition in Equation 2.17 on  $\Theta$  with respect to  $Y_1$  and  $Y_2$  as follows

$$\Theta = \bar{E}(Y_1 | \Theta) \oplus \Theta \cap Y_1^\perp \quad (2.30a)$$

$$\Theta = \bar{E}(Y_2 | \Theta) \oplus \Theta \cap Y_2^\perp. \quad (2.30b)$$

Now let  $\Theta'$  be the subspace of  $\Theta$  defined as:

$$\Theta' = \{\theta \in \Theta : E(\theta | Y_j) = 0\} \quad j = 1, 2. \quad (2.31)$$

$\Theta'$  is the kernel of the EB based on  $Y_1$  and  $Y_2$ , i.e., is the subspace of  $\Theta$  formed from all the elements in  $\Theta$  which are orthogonal to  $Y_1$  and  $Y_2$ . Thus,  $\Theta' = \Theta \cap Y_j^\perp$  for  $j = 1, 2$ . Therefore, if  $\Theta$  is a minimal splitting subspace respect to  $Y_1$  and  $Y_2$ , the only element in  $\Theta'$  will be the  $\{0\}$ . In other words, the only element of  $\Theta$  that cannot be approximated from the observed information will be the zero element. Thus, when WALI holds, we maximize how much we can known about the latent variables from the observed information  $\square$ .

The WALI solves an indeterminacy problem. By measuring  $\theta$  through the EB estimates, one may fear to lose hidden information. The WALI ensures that only the zero element is in the kernel of the EB, and thus, no information is lost (no shrinkage). This point is clearer when decomposition in Equation 2.17 is applied on  $\theta$ , and replaced in  $E(\theta|Y)$  as follows:

$$E(\theta|Y) = E\left[E(y_i|\Theta)|Y\right] + E(\theta_{y^\perp}|Y) \quad (2.32)$$

where  $\theta_{y^\perp}$  denotes the part of the latent variable which is in the orthogonal complement of  $Y$ . The orthogonal projection of any  $\theta_{y^\perp}$  onto  $Y$  is always zero, no matter how much of  $\Theta$  is shared with  $Y^\perp$ . There is no way to recover information from  $\theta_{y^\perp}$  through the EB estimator. If WALI holds, only the zero element is in there, so that the indeterminacy problem is solved.

It is possible to go further using the geometry of Hilbert spaces and construct boundaries on  $\Theta$  when it is a minimal splitting. It can be proved (see Appendix 12) that if  $\Theta$  is a minimal splitting subspace with respect to  $Y_1$  and  $Y_2$ , such that  $\Theta$  contains  $Y_2$ <sup>7</sup>, then

$$Y_2 \subset \Theta \subset (Y_2 \vee Y^\square) \oplus Y_0^\perp. \quad (2.33)$$

Where  $Y^\square$  is defined as  $E(Y_1|Y_2) \vee E(Y_2|Y_1)$ , the subspace composed of all the linear combinations between  $E(Y_1|Y_2)$  and  $E(Y_2|Y_1)$ . Here  $E(Y_1|Y_2)$  represents the set of all possible orthogonal projections from  $Y_1$  onto  $Y_2$ , and similarly for the set  $E(Y_2|Y_1)$ . Hence, because  $Y_1$  and  $Y_2$  are subspaces of observed scores,  $Y^\square$  contains observable information from the tests. On the other hand,  $Y_0^\perp$  stands for the orthogonal complement of  $Y_0$  (remember that  $Y_0 := Y_1 \vee Y_2$ ). This means  $Y_0^\perp$  does not contain observed scores, and hence, represent unobservable information from the tests. Keeping these

---

<sup>7</sup>If this assumption is replaced for  $Y_1 \subset \Theta$  interchange every  $Y_2$  for  $Y_1$  in the boundaries.

last observations in minds, from Equation 2.33 we can see that if a minimal splitting latent subspace exists in:

- a)  $Y_2 \subset \Theta \subset (Y_2 \vee Y^\square) \oplus Y_0^\perp$ , means latent variables could take unobservable values from the tests. We say they may be unobservable and not just unobserved. This is a result of the term  $Y_0^\perp$  which contains unobservable information from the tests.
- b)  $Y_2 \subset \Theta \subset (Y_2 \vee Y^\square)$ , means latent variables are unobserved but not unobservable. These bounds show latent variables can *only* take values among the possible observed scores defined in the two tests, thus they are actually constituted by observable information from the tests, it's just that we may or may not see this information during an application of the tests. Thus,  $\Theta$  is an internal minimal splitting and it is possible to construct tests with perfect reliability.

Also it can be shown that if  $\Theta$  hold ALI respect to  $Y_1$  and  $Y_2$ , and  $\Theta$  is contained in  $Y_2$  (or  $Y_1$ ) then the minimal splitting subspace is  $\overline{E}(Y_1|Y_2)$  (or  $\overline{E}(Y_2|Y_1)$ , see Appendix 13). This means the orthogonal projections of the observed scores in test 1 onto the observed scores in test 2, is the minimal subspace of  $Y_2$  for which  $Y_1$  is conditional orthogonal to  $Y_2$ . Of course this also correspond to not observed latent variables.

On the other hand, in the section Classical and Dual Reliabilities, was mentioned that reliability equals to one implies that the latent variable is a function of the observed score. Hence, it is important to check if  $\Theta$  being a minimal splitting subspace implies that reliability is always equal to one. In order to clarify this, remember the reliability is related with the angle between the observed and the latent subspaces (Equations 2.18a and 2.18b). Now, if  $\Theta$  is a minimal splitting subspace respect to  $Y_1$  and  $Y_2$ , Equation 2.30a implies that all the elements in  $\Theta$  will be the true scores obtained from  $Y_1$  (and  $Y_2$ ). Thus,  $\theta$  can be replaced by the true scores  $E(y_1|\Theta)$  (or  $E(y_2|\Theta)$ ). In this way the reliability of the scores associated to  $Y_1$  is given by

$$\eta = \frac{\|E(y_j | \Theta)\|^2}{\|y_j\|^2} \leq 1.$$

This shows that it is possible to have values of reliability less than one if  $\Theta$  is a minimal splitting subspace.

Finally, from Equations 2.29a and 2.29b, and from decompositions 2.30a and 2.30b, it can be seen that under WALI the true scores of any two elements  $y_1 \in Y_1$  and  $y_2 \in Y_2$  is

$$E(y_1 | \Theta) = \theta \quad (2.34a)$$

$$E(y_2 | \Theta) = \theta, \quad (2.34b)$$

therefore, the N-CTT (Equation 2.4) decomposition of  $y_1$  and  $y_2$  can be expressed as

$$y_1 = \theta + e_1 \quad (2.35a)$$

$$y_2 = \theta + e_2 \quad (2.35b)$$

where  $e_1$  and  $e_2$  are the error terms. By induction, all the results based on WALI can be easily generalized to  $J$  items (or tests), where  $J$  can be any integer. A very

important fact about 2.35a and 2.35b is that Kotlarski (1967) proved that the joint distribution of  $y_1$  and  $y_2$ , determines the marginal distributions of  $\theta$ ,  $e_1$  and  $e_2$  up to a translation of the curve. Thus, on this framework the shape of the marginal distributions of the latent variables and the error terms are all identified. There is no need to assume the marginal distributions of either of these three random latent variables. Note that denoting by  $y_{i1}$  the realizations of the random variable  $y_1$  from the person  $i$ , and similarly for  $y_{i2}$ ,  $\theta_i$ ,  $e_{i1}$  and  $e_{i2}$ , Equation 2.35a and Equation 2.35b implies

$$y_{i1} = \theta_i + e_{i1} \quad (2.36a)$$

$$y_{i2} = \theta_i + e_{i2}. \quad (2.36b)$$

In Equations 2.36a and 2.36b we use the fact that any random variable in our Hilbert space (i.e.  $y_1, y_2, \theta, e_1$  and  $e_2$ ) is a vector whose components are each possible realization of the random variable (the sample space). And thus, considering that a vector sum is component a component, Equation 2.35a can be understood as

$$\begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \end{bmatrix} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{21} \\ \vdots \\ e_{n1} \end{bmatrix}, \quad (2.37)$$

where  $n$  is the sample space cardinality. The same representation can be used in Equation 2.35b. Thus, when the realizations are one per each person  $i$ , Equations 2.36a and 2.36b hold.

## 2.2.4 Multidimensional Latent Variables

The inner product from which we define angles and distances is  $\langle x, w \rangle = E(xw)$ , the expectation of the product between two random variables. And all the (random) latent variables used in this work are described as  $f : (M, \mathcal{M}) \rightarrow (\mathbb{R}, \mathcal{B})$ , with finite variance and zero mean. Hence, each realization of any of these random variables is associated with a single value on the Real Line ( $\mathbb{R}$ ). This means all the latent variables used represent one-dimensional traits<sup>8</sup>.

To explore multidimensional latent variables, we need to expand our Hilbert space to random variables with range in  $\mathbb{R}^n$ , where  $n$  is any positive integer bigger than one. For example if  $n = 4$ , then we are able to work up to four-dimensional random variables. We need to define an inner product for this kind of random variables, so we can make use of geometry.

For example, for any integer  $n$ , the  $n$ -dimensional observed score corresponding to a person  $i$  in a test (or item)  $j$  is denoted by  $y_{ij} = (y_{ij1}, y_{ij2}, \dots, y_{ij(n-1)}, y_{ijn})$ , and the realization of a  $n$ -dimensional latent variable in a person  $i$  is denoted by  $\theta_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{i(n-1)}, \theta_{in})$ . Because  $n$  one-dimensional observed scores could be arranged in one  $n$ -dimensional observed score, it could be more appropriate to drop the subindex  $j$ . The inner product that we propose is as follows:

$$\langle \theta, y \rangle = \langle \theta_1, y_1 \rangle + \langle \theta_2, y_2 \rangle + \dots + \langle \theta_n, y_n \rangle, \quad (2.38)$$

where each inner product at the right side of Equation 2.38 is as before. For example

---

<sup>8</sup>Remember the distinction made at the beginning of the chapter between: the geometrical dimension of a random variable in  $H$ , which is related with how many values can take (all the possible realizations); and the trait (represented by the latent variable) dimension, which is related with the dimension of each realization of the random variable.

$\langle \theta_1, y_1 \rangle = E(\theta_1 y_1)$ . It is important to highlight that each component of the vectorial random variable belongs to the original Hilbert space. Recall that two vectors are orthogonal if the inner product between them is zero. Thus when multidimensional observed scores and latent variables are used, the error term will still be orthogonal to the true score. This is easy to check by noting that each inner product at the right side of Equation 2.38 will be zero.

Now let analyze a test from which we obtain a one dimensional observed scores, this is a single observed score per person. Let assume these observed scores are explained, just for the sake of simplicity, for three latent variables and not for  $n$ . We know the reliability is the square cosine of the angle between the observed scores and latent variable (see theorem 1). Also the cosine of the angle between any two vectors is defined as the inner product between them over their norms (Equation 2.3). To avoid large expressions, we use the same notation used in Equations 2.16 ( $\hat{h} = h/\|h\|$ ). Because we consider a three-dimensional latent variable,  $\theta$  is represented by the triplet  $(\theta_1, \theta_2, \theta_3)$  and the observed scores by  $y_1 = (y_1, y_1, y_1)$  (because it is a one dimensional observed score). Thus, the reliability would be

$$\eta = \langle \hat{y}, \hat{\theta} \rangle^2 \quad (2.39a)$$

$$= (\langle y_1, \theta_1 \rangle + \langle y_1, \theta_2 \rangle + \langle y_1, \theta_3 \rangle)^2. \quad (2.39b)$$

In the right side of Equation 2.39b, each term is smaller than the square root of the reliability of the observed score respect to the latent variables  $\theta_1$ ,  $\theta_2$  and  $\theta_3$  respectively, because they are divided by the products of the norms of  $y_1$  with  $\theta$  (multidimensional). Hence, if we apply the square root on both sides, we obtain

$$\sqrt{\eta} < \sqrt{\eta_1} + \sqrt{\eta_2} + \sqrt{\eta_3}. \quad (2.40)$$

Equation 2.40 is a superior limit in the reliability of a multidimensional test. If one of the reliabilities  $\eta_1$ ,  $\eta_2$  or  $\eta_3$  is equal to one, the others must be zero. This is because the multidimensional inner product that we used assume  $\theta_1 \perp \theta_2 \perp \theta_3 \dots \perp \theta_n$ , thus if a reliability is one, the observed score must be perpendicular to the other components of the latent variable. Hence reliability is related with which latent variable are generated the observed scores, in this sense reliability cannot be distinguished from validity. The fact that the multidimensional reliability is less than one, is a consequence of the Cauchy–Schwarz inequality (see Appendix 14).

The result presented in Equations 2.34a and 2.34b is valid for multidimensional latent traits too. They are implied by Equations 2.30a, 2.30b, 2.29a and 2.29b; and the proofs of these equations (Appendices 10 and 11) are valid for any Hilbert space, thus are valid for the Hilbert space of real random vectors with the inner product proposed in Equation 2.38. To illustrate how would looks Equations 2.34a and 2.34b with a multidimensional latent variable instead, we continue with the example of a three-dimensional latent trait, but this time using two one-dimensional observed scores. Let denote  $\Theta^3$  the latent subspace of three dimensional latent variables (each component of the three-dimensional latent variable with zero mean and finite variance). Because we have two observed scores, we can put them together in a two-dimensional observed score as  $y = (y_1, y_2)$ , in this way,

$$E((y_1, y_2)|\Theta^3) = [E(y_1|\Theta^3), E(y_2|\Theta^3)] \quad \text{with} \quad (2.41a)$$

$$E(y_1|\Theta^3) = \theta_1, \theta_2, \theta_3 \quad (2.41b)$$

$$E(y_2|\Theta^3) = \theta_1, \theta_2, \theta_3. \quad (2.41c)$$

The identification result of Kotlarski (1967) is not valid in this case, but Székely and Rao (2000) generalized his result to multi-factor linear models, as the one presented in Equations 2.42a and 2.42b

$$y_1 = \theta_1 + \theta_2 + \theta_3 + e_1 \quad (2.42a)$$

$$y_2 = \theta_1 + \theta_2 + \theta_3 + e_2. \quad (2.42b)$$

In the next chapter are more details about the restrictions of these identification results.

It is important to highlight that this framework allows to understand a multidimensional error. In geometric terms, is *the* orthogonal vector from the “plane” (spam)<sup>9</sup> formed for the components of  $\theta$  to the observed score. And in more substantive terms, is all the information with none correlation with any of the components of  $\Theta^n$  ( $n$  dimensional Latent subspace), such that, explain the variations in  $Y_j$  that can no be explained by any components of  $\Theta^n$ .

---

<sup>9</sup>In other fields usually these multidimensional planes are called hyperplanes.

## 2.2.5 Final Comments

The boundaries presented in Equation 2.33 are functions of the observed subspaces and information not contained in the tests. This implies that the existence or size of  $\Theta$  could depend on the items or tests used. If the observed subspaces are multidimensional, this may no be a problem, because it is clear that more dimensions in the observed score, the more likely that we measure a multidimensional latent trait. But when the observed scores are unidimensional could seem counterintuitive, because this suggest that the bigger of the subspace  $Y_2$ , the bigger the subspace  $\Theta$  is. And, because  $\Theta$  is meant to represent some attribute of persons, one would expect that  $\Theta$  will not depend in the size of  $Y_2$ . Thus, we need to look again careful the meaning of “size” in subspaces of real random variables (not vectorial).

In  $H$ , our original Hilbert space defined in Equation 2.1, each function is a vector. And each possible outcome of the function represents a component of the vector. Thus, if the range of a function is infinite, the real numbers for example, then his dimension will be infinite as well. The “size” of the subspaces used here (subspaces of  $H$ ) is related with all the possibles realizations of a random variable, and hence, with the dimension of the function with the biggest range. Thus,  $\Theta \subset Y_j$  for one side means that latent variables can take the same values as an observed score. And it also means that the function with the largest range in  $Y_j$  have a larger or equal range than the function with the largest range in  $\Theta$ . Now let us consider the next illustration. Someone wants to measure the length of different pencils in a classroom. In this example, the length represents a latent variable and the pencils represent the examinees. Probably, their lengths will be similar, but not equal, let us say most of them will be around 20 cm. If one would measure the pencils lengths with a stick that had marks only at 5, 10, 15, etc., centimeters (each mark in the stick represents a possible observed score), due to lack of precision, one could not distinguish between the lengths of the pencils. But if a stick with millimeters marks is used, we may be able to differentiate among different lengths. Therefore, even if vastly admitted that

the length of any object does not depend on whether it is measured<sup>10</sup>, how much we can know about it depends on what is used to measure it. Thus, even if  $\Theta$  could be ontologically independent of  $Y$ , it is reasonable to expect an increase in  $\Theta$  when some  $Y_2$  does it, how Equation 2.33 shows.

Now let take a deeper look at the example above. One can say that the length of an object can be defined as how much space is covered in a given direction. This is the attribute that we measure. However, when somebody asks for the length of something, we give information in the set of positive real numbers, we give a number. The variable length is a function from this attribute into the real numbers. How was defined  $\Theta$  (as the subspace of random variable from  $(M_\theta, \sigma(\theta))$  to  $(\mathbb{R}, \mathcal{B})$ ) does not represent the quality, but the information that we can know regards this attribute,  $\Theta$  is constituted by (mathematical) functions (Appendix 7 have more comments related with this). We use WALI or ALI to ensure that the information (the numbers) in  $\Theta$  are meaningful respect the variation of the observed scores. But we do not know from where these numbers come, in statistical terms, we do not have information about the sample space of  $\Theta$ . It is important to identify with which specific person's (or societies's) aspects these values are related with. In order to come up with a concept relating the values of some  $\theta$  with something in the “universe” we need to identify from where these values came from.

---

<sup>10</sup>Leaving aside speeds close to the speed of light and the microscopical world.

# III. Marginal Distribution

## Estimation of a Latent Variable

In this chapter we use the identification result showed in Equations 2.42b and 2.42b. The identification result is based in a deconvolution. There is a long tradition of estimators based in deconvolution (Chen, Hong, & Nekipelov, 2011; Delaigle & Gijbels, 2002, 2004; Delaigle, Hall, & Meister, 2008; Li & Vuong, 1998). Here, we propose the method of Bonhomme and Robin (2010) to estimate the marginal distribution of the latent variables. The main three advantages of this method are: a) they did not assume a distribution of the errors; b) neither assume errors with equal distributions (but they must be orthogonal); and c) is valid for multidimensional latent variables. In this chapter we discuss the limitations and assumptions of the chosen method. Finally we present an application using real datas.

Some keys ideas in this chapter are: a) the characteristic function is the Fourier transform of the density function (of a random variable), hence identify-estimate the characteristic function is equivalent to identify-estimate the density function. For each density function there is only one characteristic function; b) Non vanishing characteristic functions means characteristic functions without zeros in their image. The characteristic function is a complex value function, when we refer to not vanishing characteristic functions, we are referring for both, the real and imaginary component; c) because all the integrals and derivate are numerical, it is important to properly choose a bandwidth, this is, the distance between two values in the independent variable. If this distance is too large, we would not recover the real function, and if it is too small, we would include more error.

### 3.1 Model Specifications

The identification result presented by Kotlarski (1967) is as follows. Consider Equations 2.35a and 2.35a from the chapter one showed next,

$$y_1 = \theta + e_1$$

$$y_2 = \theta + e_2.$$

In these equations,  $y_1$  and  $y_2$  are the only observed variables. Kotlarski (1967) proved that the joint distribution of  $y_1$  and  $y_2$  determines the distributions of  $\theta$ ,  $e_1$  and  $e_2$  up to a change of location (see Appendix 15). In this proof he assumes that: a) the characteristic functions of  $\theta$ ,  $e_1$  and  $e_2$  do not vanish; and b)  $\theta$ ,  $e_1$  and  $e_2$  are mutually independent. This identification result is valid if instead of two observed random variables,  $y_1$  and  $y_2$ , there are  $J$  observed random variables ( $J$  is an integer bigger than 2), and  $J$  linear combinations  $y_1 = \theta + e_1$ ,  $y_2 = \theta + e_2$ , ...,  $y_J = \theta + e_J$ . In Hilbert spaces, independence is equivalent to orthogonality, thus the second assumption do not suppose any additional restriction in our model. In order to explain this assertion, remember that  $E(y|\Theta) = \theta$  is a consequence of the orthogonality of the errors (WALI), and because  $E(y|\Theta)$  is an orthogonal projection,  $\theta$  is orthogonal to any error term. But the first assumption impose an additional restriction. Some characteristic functions have zeros in their image. For example, a Bernoulli distribution with parameter bigger or equal than 0.5. This means that it would be impossible to recover the marginal distribution of binary latent variables with parameters bigger or equal than 0.5. Also, the identification result of Kotlarski (1967) is valid only for unidimensional random variables. Regarding these limitations, is important to mention that there have been two major advances in this identification result (Evdokimov &

White, 2012; Székely & Rao, 2000). One of them relax the assumption of not vanish characteristic function, and the other is valid for multidimensional latent variables.

Evdokimov and White (2012) relaxed the assumption of non-vanishing characteristic functions used by Kotlarski (1967). Let denote by  $\varphi_\theta$ ,  $\varphi_{e_1}$  and  $\varphi_{e_2}$  the characteristic functions of  $\theta$ ,  $e_1$  and  $e_2$  respectively, and let  $\varphi'$  denote the first derivate of the characteristic function  $\varphi$ . In the work of Evdokimov and White (2012)  $\varphi_{e_1}$  can have real zero, as long as: a)  $\varphi'_{e_1}$  at those points is not zero; and b)  $\varphi_{e_2}$  have countable zeros. There are not restriction on the zeros of  $\varphi_\theta$ . In other words, is possible for  $\varphi_{e_1}$  to have real zeros, as long they are not a local minimum and  $\varphi_{e_2}$  is not zero in some finite segment of  $t$  ( $t \in \mathbb{R}$ ). These are less restrictive assumptions than those used by Kotlarski (1967). For instance, continuing with the example of binary latent variables, it would be possible to identified his marginal distribution if the parameter is different from 0.5 (and not less than 0.5).

On the other hand, Székely and Rao (2000), without relaxing the assumptions of Kotlarski (1967), generalized the identification result including multidimensional dimensional random variables. They proved that the maximum number  $n$  of latent variables that can be identified, depends on the numbers of linear combinations  $J$  and the number of known moments  $m$  as

$$\binom{J+m}{m+1}, \quad (3.1)$$

the binomial coefficient between  $l + m$  and  $m + 1$ . We are using random variables with zero mean, thus we known one moment,  $m = 1$ . If we had three observed scores, we would have three linear combinations,  $J = 3$ . Thus the maximum number  $n$  of identified latent variables would be 6. Therefore, we would identified the

marginal distributions of a three dimensional latent variable  $\theta = (\theta_1, \theta_2, \theta_3)$  and the distributions of three error terms  $e_1, e_2$  and  $e_3$ .

The estimator that we propose is based in the work of Székely and Rao (2000), thus we use the restriction of non vanish characteristic functions. However, we restrict the simulations and applications to one dimensional latent variables. There are others estimator based in the restrictions of Kotlarski (1967) and Székely and Rao (2000) which could be useful to estimate one dimensional latent variables, but those estimators are not explicitly written as a function of the observed information, or they assume errors with known or equals distributions (Dattner, Goldenshluger, & Juditsky, 2011; Delaigle et al., 2008; Li & Vuong, 1998). Observe that in our approach the error terms are unavoidable, unless we use an operational definition of the latent variable, in which case  $\Theta \subset Y$ , and then, the errors would be zero. Hence, in our approach there is not a priori reason to believe that the errors are normally distributed. Due to this, is important for us to manipulate the distribution of the error terms too.

## 3.2 Method of Estimation

How was already mentioned, we propose the estimator developed by Bonhomme and Robin (2010), which is based in the identification result of Kotlarski (1967) and Székely and Rao (2000). The main idea is to estimate the characteristic function of each latent variable ( $\theta, e_1$  and  $e_2$ ). Once the characteristic function is estimated, it only remains to perform a Fourier transform to obtain the distribution.<sup>1</sup>

The method is as follows. The characteristic function of a random variable  $x$  with distribution  $f(x)$  is given by

---

<sup>1</sup>Each marginal distribution is uniquely determined by his characteristic function by the Fourier transform.

$$\varphi_x(t) := \int \exp(itx) xf(x) dx, \quad (3.2)$$

where  $i = \sqrt{(-1)}$ . From Equations 2.35a and 2.35a, the joint characteristic function of the observables  $y_1$  and  $y_2$   $\varphi_y$ , can be written as

$$\varphi_y(t_1, t_2) = \varphi_{e_1}(t_1)\varphi_{e_2}(t_2)\varphi_\theta(t_1 + t_2). \quad (3.3)$$

Now, following the notation of Bonhomme and Robin (2010), let denote  $\kappa_x$  the cumulant characteristic function of a random variable  $x$ , defined as  $\ln(\varphi_x)$ . Due to the restriction of non zeros,  $\kappa_x$  is defined in every point in the image of a characteristic function. Observe that the joint  $\kappa_y$  is a linear combination of  $\kappa_\theta$ ,  $\kappa_{e_1}$  and  $\kappa_{e_2}$ , thus there is one equation and three unknowns functions. What they did then is differentiate two times, so they have more equations, and the matrix associated with the system of equations become invertible. The same reasoning can be applied to multidimensional latent variables. Putting aside the error terms, the maximum number of identifiable latent variable is  $J(J - 1)/2$ , with  $J$  being the number of linear combinations as before.

As an example, if we use four observed score to estimate the latent variable  $\theta$ , the estimator will be,

$$\kappa_\theta(t) = 0.167 \int_0^t \int_0^v (\partial_{12}^2 + \partial_{13}^2 + \partial_{14}^2 + \partial_{23}^2 + \partial_{24}^2 + \partial_{34}^2) \kappa_y \left( \frac{u}{4}, \frac{u}{4}, \frac{u}{4}, \frac{u}{4} \right) du dv, \quad (3.4)$$

where  $\partial_{jm}^2$  is the second order partial derivative respect to  $j$  and  $m$ . The constant 0.167 depends in the number of observed scores used (see Appendix 16), if it would have been used ten observed scores, then the constant would be 0.022. As Bonhomme and Robin (2010) explained, the speed of convergence of the estimation depends on the assigned weights to each components in the argument of  $\kappa_y$ , and this choice is arbitrary. We always chosen the one which maximize the speed of convergence.

### 3.3 Simulation Study

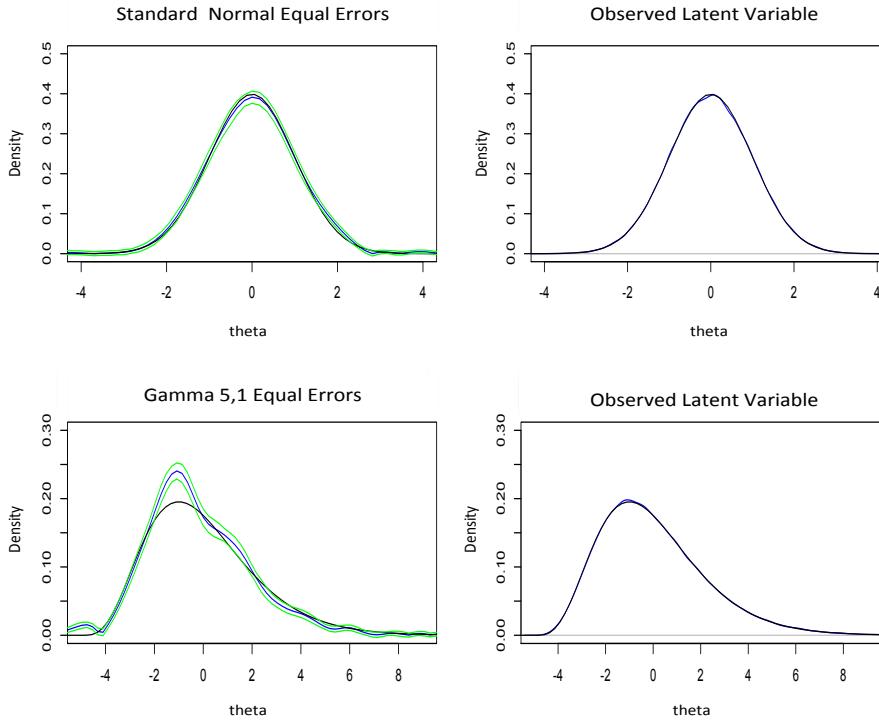
In this section we present a simulation study in order to test the estimator in different situations. We use the version 3.5 of the statistical software R (R Core Team, 2018). In Appendixes 18 to 21 are some codes of the simulations. Following the protocol of Bonhomme and Robin (2010) we use a second order kernel to smooth the densities. The optimum bandwidth in terms of the mean integrated squared error (MISE), was chosen using an adaptation of the plug-in method. This method was explained in details in Delaigle and Gijbels (2002). The adaptation was suggested by Bonhomme and Robin (2010)<sup>2</sup>. Following the simulations protocol used in Bonhomme and Robin (2010), in each simulation we used 100 samples, each one of size 1000. The numerical integration method used was the trapezoid with 201 nodes in each integral. All the random variables simulated had zero mean.

Figure 3.1 and Figure 3.2 shows the marginal distribution of  $\theta$  (one dimensional) estimated from two observed scores. Assuming the WALI, the observed scores can be modeled as in Equations 2.35a and 2.35a. In all the graphs the green line represent a 95% confidence interval, the blue lines are the estimated densities and the black lines represent some density of reference. In Figure 3.1 both error terms were sampled

---

<sup>2</sup>In Appendix 17 is described in details the application of the method.

from a standard normal distribution. From left to right, the figure 3.1 shows the estimated densities using the proposed method, and then as the latent variable  $\theta$  were observed<sup>3</sup>. From top to bottom, the figure 3.1 shows the estimated densities when  $\theta$  is sampled from a standard normal distribution, and from a gamma distribution with shape parameter 5 and scale parameter 1. In this circumstances, it can be seen that when  $\theta$  came from a standard normal distribution, the estimator works very well. But when  $\theta$  came from a gamma distribution of parameters (5, 1), the estimator recovers the general shape of the curve, but fails in the top.



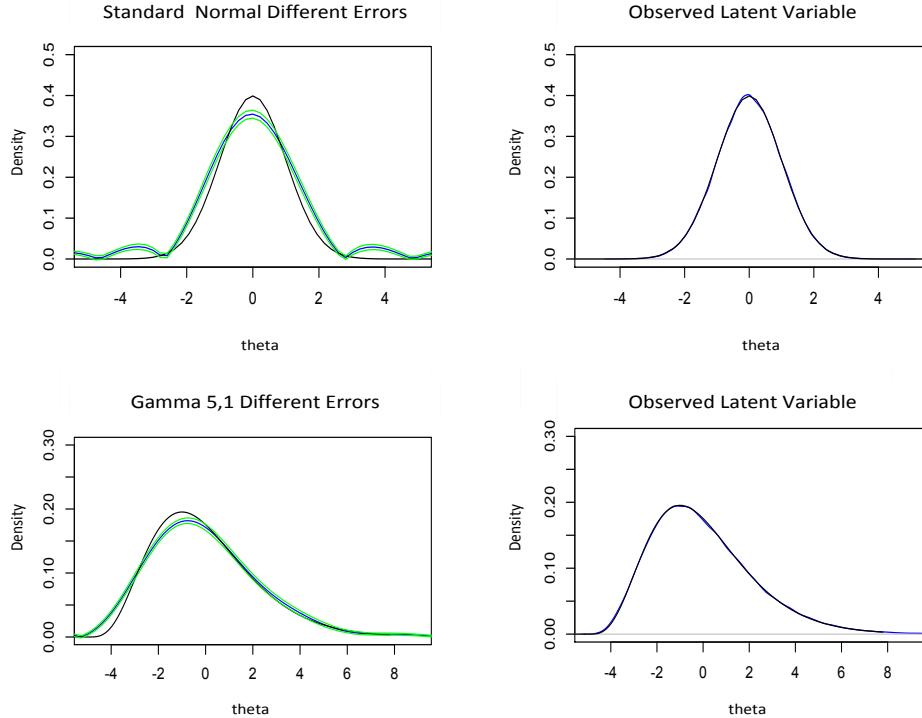
*Figure 3.1.* From left to right and top to bottom, estimated distributions when each error was sampled from a standard normal distribution, and:  $\theta \sim \text{Normal}(0, 1)$  using the proposed method;  $\theta \sim \text{Normal}(0, 1)$  if we could observe it (kernel estimator);  $\theta \sim \text{Gamma}(5, 1)$  using the proposed method; and  $\theta \sim \text{Gamma}(5, 1)$ , if we could observe it (kernel estimator)

Figure 3.2 is similar to 3.1, but  $e_1$  is a random sample from a gamma distributions of parameters (5, 1), and  $e_2$  is a random sample from a normal distribution with zero

---

<sup>3</sup>This correspond to a kernel density estimated with a gaussian smoothing.

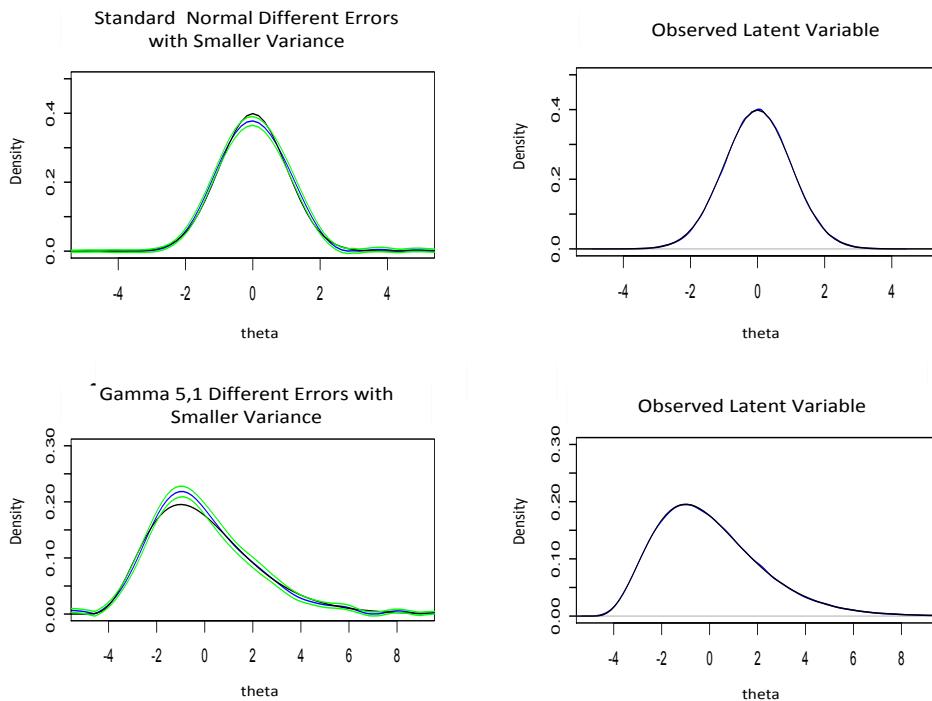
mean and variance equal to nine. In both cases the proposed estimator it seems to fail to fully recover the shape of the distribution. But still is able to distinguish between a normal and a gamma distribution.



*Figure 3.2.* From left to right and top to bottom, estimated distributions when  $e_1$  was sampled from a  $\text{Gamma}(5, 1)$  distribution and  $e_2$  from a  $\text{Normal}(0, 9)$  distribution, and:  $\theta \sim \text{Normal}(0, 1)$  using the proposed method;  $\theta \sim \text{Normal}(0, 1)$  if we could observe it (kernel estimator);  $\theta \sim \text{Gamma}(5, 1)$  using the proposed method; and  $\theta \sim \text{Gamma}(5, 1)$ , if we could observe it (kernel estimator)

However, this *bad* behavior in the estimator it could be expected. Because the bigger the variance (longer norm of the errors), the less of the reliability of the test (angle closer to  $90^\circ$  between  $\theta$  and  $y$ ). A test with low reliability would not provide good information about  $\theta$ , and thus the estimations is not as good as with bigger reliabilities. To see this point, Figure 3.3 shows the same conditions that Figure 3.2, except with less spread errors. Here,  $e_1$  was sampled from a gamma distribution with shape parameter 2 and scale parameter 1, and  $e_2$  from a normal distribution with

zero mean and variance 4. In Figure 3.3 the estimated distribution match better the real distribution than in Figure 3.2. Again, the more significant discrepancy occurs at the top of the distributions. When  $\theta$  is normally distributed, the true density slightly goes out of the confidence interval for the top, but when  $\theta$  follows a gamma distribution, goes out for the bottom. Thus the estimated mode will be significantly biased in both cases. Overestimated when the  $\theta$  is normally distributed and underestimated when  $\theta$  follows a gamma distributions.



*Figure 3.3.* From left to right and top to bottom, estimated distributions when  $e_1$  was sampled from a  $\text{Gamma}(2, 1)$  distribution and  $e_2$  from a  $\text{Normal}(0, 4)$  distribution, and:  $\theta \sim \text{Normal}(0, 1)$  using the proposed method;  $\theta \sim \text{Normal}(0, 1)$  if we could observe it (kernel estimator);  $\theta \sim \text{Gamma}(5, 1)$  using the proposed method; and  $\theta \sim \text{Gamma}(5, 1)$ , if we could observe it (kernel estimator)

Table 3.1 report the MISE of the proposed estimator of the simulations. MISE measure the distance between the estimated function and the real functions, when the distance is understood as in Chapter 1 (see Appendix 20 for details of the code).

Table 3.1. *MISE*

	$\theta \sim N(0, 1)$	$\theta \sim \Gamma(5, 1)$
$e_1 \sim N(0, 1)$ and $e_2 \sim N(0, 1)$	0.016	0.032
$e_1 \sim N(0, 9)$ and $e_2 \sim \Gamma(5, 1)$	0.041	0.035
$e_1 \sim N(0, 4)$ and $e_2 \sim \Gamma(2, 1)$	0.031	0.024

*Note.* MISE in the simulations presented in Figure 3.1, Figure 3.2 and Figure 3.3

Thus, the closer to zero the better the estimation.

Table 3.1 is coherent with the graphs. When the latent variable  $\theta$  came from a gamma distribution, the estimation was better when one of the errors had a gamma distribution too. This can be seen by comparing the first and the third rows of the second column in Table 3.1. The best approximation occurs for normal standard errors and standard normal  $\theta$ . This is good, because in case that actually the latent variable and errors follows a standard normal distribution, the estimator will perceive that. Also, even if for other distributions the estimator would not have a good fit, probably it would reproduce the real distribution better than assuming a normal distribution. Therefore, it may be advisable to use the estimator in order to assign fairer scores to students.

From Table 3.1 it can be see that MISE get bigger along with the variance of the errors. Figure 3.4 shows how the values of MISE increase with the standard deviations of the errors. Here, the latent variable was sampled from a standard normal distribution, and both errors from normal distributions with equal variances. Each point correspond to a different error variance.

Another case of interest it may be a test with binaries observed scores. There are high impact test in this format. SIMCE and PSU are just two examples from Chile among others. We simulated this kind of test and estimates the errors distributions. In this case the approach was different. Before we took a random sample of  $\theta$  from a known distribution, and did the same with the errors. Then each pair  $(\theta, e_j)$ , with

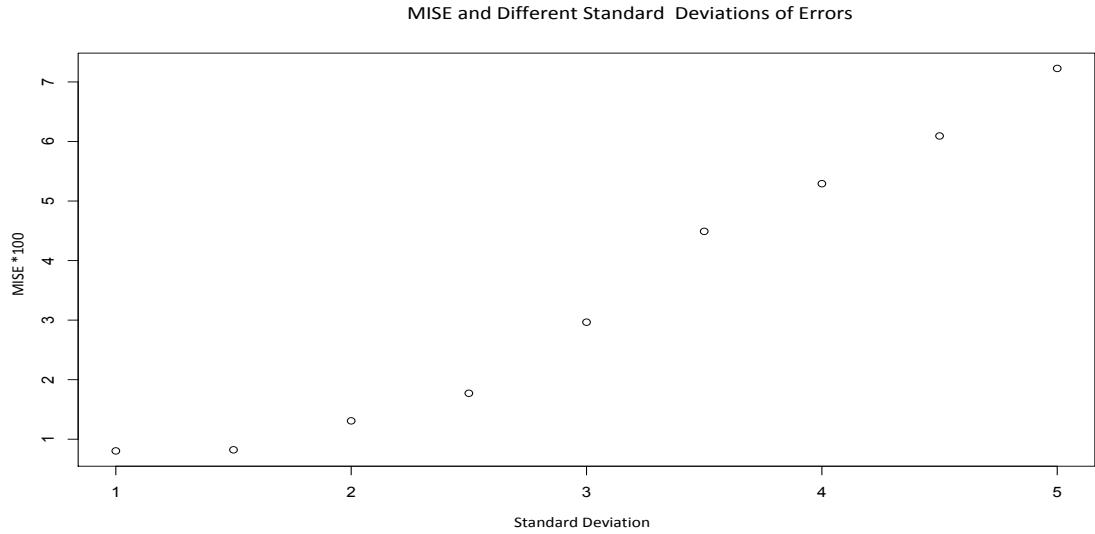


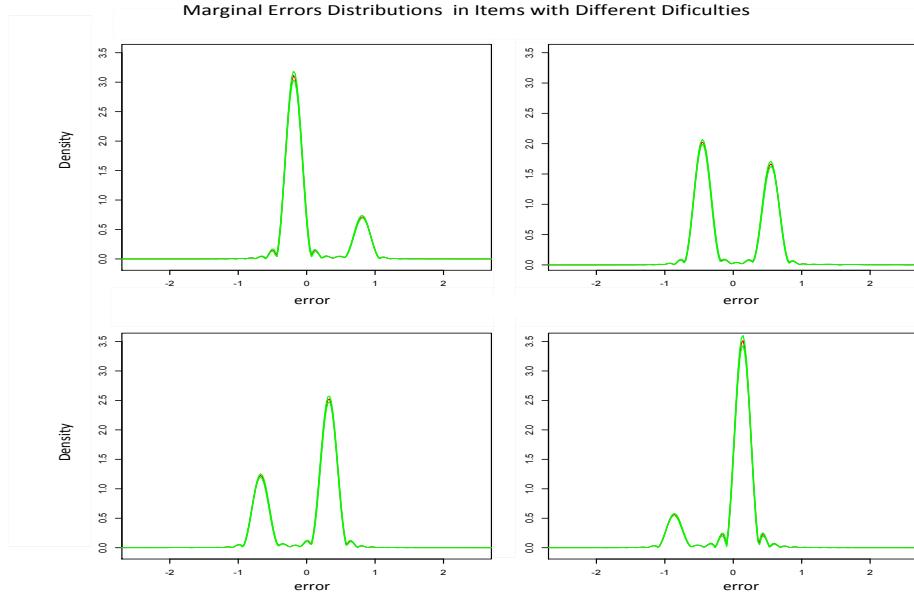
Figure 3.4. The values of MISE where multiplied by 100 to benefit the reading of the data.

$j=\{1,2\}$ , was summed to get the observed score  $y_1$  and  $y_2$ . And then we see if we could recover the known distribution of  $\theta$  from  $y_1$  and  $y_2$ . Regarding the simulation of binaries items, the binaries observed scores were directly sampled from a Bernoulli distribution of parameter  $\pi$ . We used four observed scores,  $y_1$ ,  $y_2$ ,  $y_3$  and  $y_4$ . The (closed) intervals of  $\pi$  from where the observed scores  $y_1$ ,  $y_2$ ,  $y_3$  and  $y_4$  were sampled, respectively were  $(0.02, 0.35)$ ,  $(0.25, 0.65)$ ,  $(0.5, 0.85)$  and  $(0.755, 0.97)$ . Thus,  $y_4$  was the easiest item and  $y_1$  the more difficult<sup>4</sup>.

Figure 3.5 shows the distributions. In the top row at the left, there is the distribution of the errors in the observed score  $y_1$ , in the same row at the right, the distribution of errors in  $y_2$ . At the bottom from left to right, the distributions of the errors in  $y_3$  and  $y_4$  respectively. Again, the green lines represent a 95% confidence interval. The estimation was very precise. Observe that most of the errors are grouped

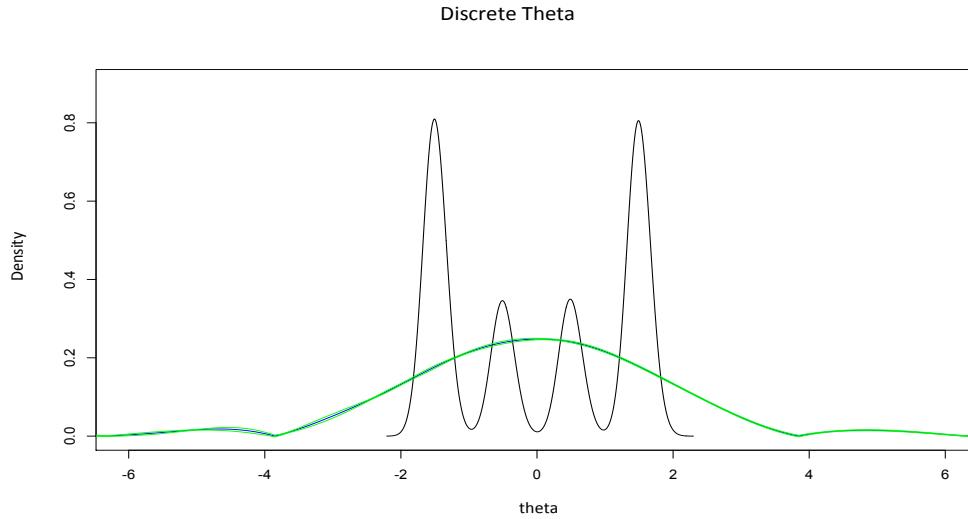
---

<sup>4</sup>The parameter  $\pi$  is represented as the fraction between the “ability” of a person ( $\theta$ ) and the difficulty of an item ( $\beta$ ) (San Martín, 2015). In this simulation we assume that the same person answered the four items, and the ability of each person to respond the four items is the same. That is why we interpret the  $\pi$  parameter as an indicator of the difficulty of the items.



*Figure 3.5.* Error distributions for binaries observed scores. From top to bottom and left to right, the items are ordered from hardest to easiest.

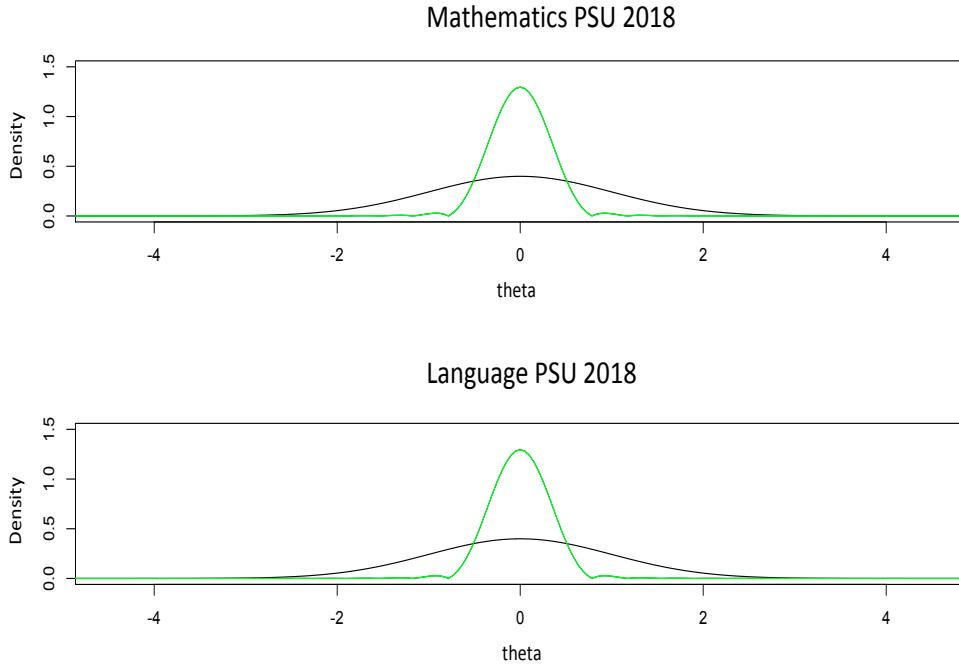
around two values. Thus, the errors are almost binaries too. They are not perfectly grouped around zero and one because in the simulation process the mean of each sample was subtracted. Thus, observed scores were binaries variables with one of the values being less than zero and bigger than minus one, and the other bigger than zero and less than one. Due to observed scores become from a binary distribution, the distance between the peaks of a distribution give none relevant information. All the estimated distributions are bimodal, but in the easiest and the hardest items most of the errors take values around zero and one respectively. This make sense because one may expect that item 1, the more difficult, is not sensible enough to differentiate among persons with lower levels of  $\theta$ , which is the people that more likely had a zero in  $y_1$ . The same reasoning can be applied to the other three items. This suggest that error distributions give information about the difficulty of binary items. Also from their distributions (and from their characteristic functions) we can have a qualitative idea about the variance of the errors, and thus have a qualitative



*Figure 3.6.* Estimated distributions of a discrete  $\theta$ . The blue line represents the proposed estimator, the green lines represents a 95% confidence interval and the black represents the estimations as  $\theta$  was observed.

idea about the reliability of a test. For instance, if the peak closer to one is wider than the other, it means that the test is less reliable in that region.

As a final comment, it is important to highlight that it is possible to distinguish between a discrete and a continuous random variables from his characteristic function. The characteristic function of a discrete random variable is a discrete Fourier transform, and hence is a periodic function. This means the characteristic function, after fixed increasing of his argument, repeat his values. However, when  $\theta$  is discrete, the proposed estimator fails to recover it, as shows Figure 3.6. Here,  $\theta$  can take only four values, 1, 2, 3 and 4, each one with probabilities 0.35, 0.15, 0.15 and 0.35 respectively. The black line indicate the estimated distribution if we could have seen  $\theta$  (using a kernel estimator with normal window). The blue line is the estimated distribution with the proposed estimator and, as before, the green lines indicate a 95% confidence interval.

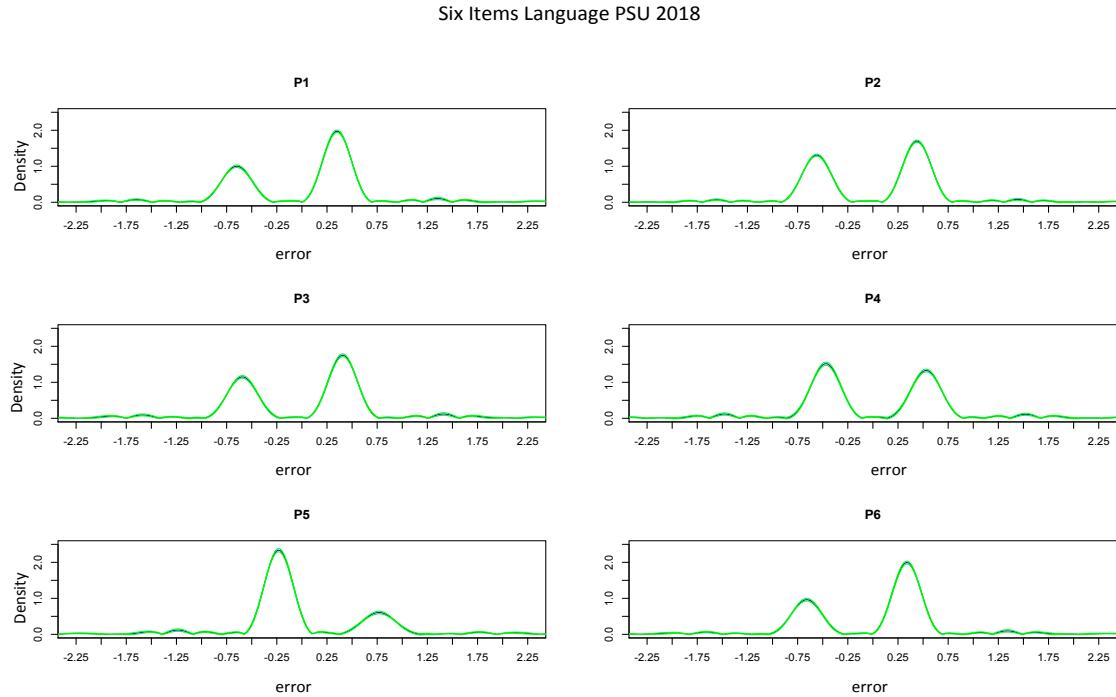


*Figure 3.7.* Estimated distributions of  $\theta$  from real datas. From top to bottom the distributions in a math and language test. The black line is a standard normal density, the blue line is the estimated distribution and the green lines are a 95% confidence interval.

### 3.4 Application with real datas

In this section we shows the estimated distributions of  $\theta$  using binaries observed scores. The datas correspond to an application of two tests in Chile at the end of 2018, a mathematic and a language test. Both tests were an admission requirement at most universities, and the imputed scores were used to select students in the most prestigious universities in Chile. The name of the test is “Prueba de Selección Universitaria”(PSU). It is applied at the end of each year. To estimate the distribution of  $\theta$  and the distributions of the error terms we used six items of each test. The six questions of language were applied to 66603 students, and the six question of mathematic to 66075 students. We took 100 samples of 1000 students.

Figure 3.7 shows the estimated marginal distributions of  $\theta$  using the items from the mathematic and language tests. Both densities are very similar between them, but quite different from a standard normal. When a standard normal density is assumed, the graph suggest that the variance of  $\theta$  is overestimated on both cases. This means, considering only the six selected questions per test, both tests are less sensible to differentiate among different levels of  $\theta$  than it is assumed when a standard normal distribution is used. In order to capture more variance of  $\theta$  it may be advisable to use polytomous items. It is important for high impact tests to capture differences of  $\theta$  in the population (great variance), otherwise decisions could be taken under unfair criteria. PSU does not use the IRT framework to assign scores, but if it would be used, then it should considerer to use polytomous items.



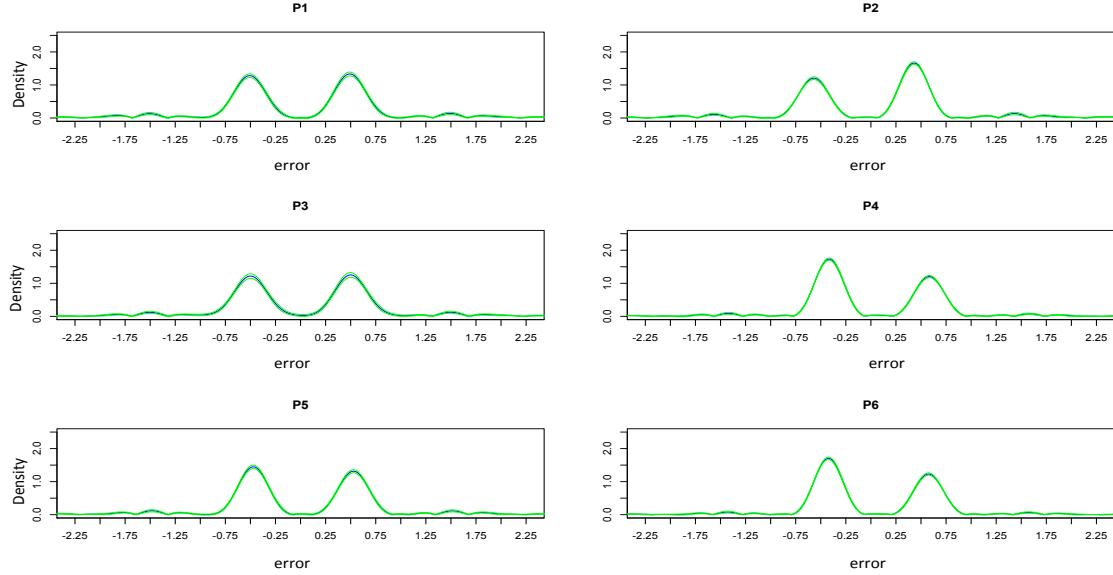
*Figure 3.8.* Estimated error distributions of the six selected from the Language test. The blue line is the estimated distribution and the green lines are a 95% confidence interval.

The Figure 3.8 and Figure 3.9 presents the error estimated distributions of the

six selected items from the Language and Mathematic test respectively. How was explained before, these distributions gave us information about the difficulty of the items (see Figure 3.5). In general, the Language selected items were easier than the Mathematic selected items. For instance, in Language, the items P1, P2 and P6 were easy for most students. Because of that, most of the errors were grouped around the lowest levels of  $\theta$ . The exception is the item P5, which was very difficult for the students. This item works good to differentiate among high levels of  $\theta$ , hence most errors are around the incorrect answer (remember these are binaries items). The items P2 and P4 had a good metric behavior in the sense that the errors were balanced between the wrong and correct answers. Thus, they were able to differentiate among the medium levels of  $\theta$  in the population. Regarding the Mathematic items, the items P4 and P6 were slightly difficult, and the item P2 was the exception, it was slightly easy. In general the distributions of the errors were more balanced than the same distributions in Language. It is reasonable to expect less variance in  $\theta$  if the errors distributions in a test are unbalanced in the same direction (either to the highest or the lowest levels of  $\theta$ ).

It is important to mention that, with real datas, the plug-in method fails to select the best bandwidth. The method return the biggest possible bandwidth. This may happened because at two equidistant points from the origin, the characteristic function jump from one value to other near to zero, is not smooth in those points. So its behavior is similar to a discrete function. It is well known that the characteristic function (Fourier transform) is a linear operator, and in infinite dimensional Hilbert spaces, the linear operators are bounded if and only if they are continuous. Hence, the method may be looking for an infinity bandwidth. We opt for truncated the limits of the integrals near to the points where the characteristic function abruptly goes near to zero. Due to properties of the Fourier transform, when we truncated the limits in the integral of the characteristic function, we only could be overestimating the variance of the latent variable, but not underestimating it. Thus, from Figure

Six Items Mathematics PSU 2018



*Figure 3.9.* Estimated error distributions of the six selected from the Mathematic test. The blue line is the estimated distribution and the green lines are a 95% confidence interval.

3.7, we can conclude that may be, the difference between the estimated density and the standard normal is even more appreciable. However, there are deconvolution's estimators not based in Fourier transform, but in those works the error densities are assumed (Lee et al., 2013).

As a final comment, the results presented in this work are also valid for longitudinal applications of tests. In this case we should use one Hilbert space (and one of the subspaces  $\Theta$  and  $Y_j$ ) per person, if the interest is in one person. A Hilbert space for this kind of procedures could have a different probability measure. This bring the possibility of empirically answer the problem regarding the interpretation of  $\theta$  described by Borsboom, Mellenbergh, and Heerden (2003), which is as follow: Imagine two persons A and B, respectively a level  $\theta_a$  and  $\theta_b$  was assigned to A and B in  $\theta$ , such that  $\theta_b > \theta_a$ . Hence, for a given item, probably B will have a better

score than A. It is valid question to ask, is this means that an increase from  $\theta_a$  to  $\theta_b$ , would result in A having the same probability of  $B$  of having a given observed score in a fixed item? or is this just means that all the persons, in the moment of examination, with the same level in  $\theta$  will have the same probability of generate a given observe score in a given item? This problem is important. Whether we are measuring something constructed among the persons, or something constructed at the person level depends on the answer to the problem. We could empirically solve this problem. For example, we could apply several times some tests to a person during the period that she (or he) is learning to sum. Let say she goes for three levels of learning to sum. Then we could apply the same tests to a sample of persons representing the same levels of learning. If the estimated marginal distributions of  $\theta$  are very similar in these two cases (with similars errors too), this will suggest that, if A increase his level to  $\theta_b$ , she (or he) will be equivalent to B respect to the probability of have a given score in a given item. This experiment will certainly give us new insights about the ontological status of  $\theta$ .

# IV. Conclusion

This thesis present a framework where different IRT models can be embedded. The geometry of Hilbert spaces allowed to explore the psychometric models from two sides, the observed and not observed information. In this framework was possible to define a Multidimensional Reliability. Also we were able to interpret the errors involved when a multidimensional  $\theta$  is used. We proved that under WALI, there is not lost information about  $\theta$  when the EB estimator is used. We found boundaries to the subspace of latent variables  $\Theta$  when WALI is assumed. Those boundaries show us that is possible for a latent variable to take the same values than an observed score. Hence, they can be not observed and not observable random variables. Finally we accomplish the main objectives which were: a) identified the marginal distribution of the latent variables involved in some IRT models, b) and implement a method to estimate it.

Because of the work of Kotlarski (1967), Equation 2.35a and Equation 2.35b implies that there is no need to make assumptions regarding the marginal distributions of the latent variables. It is important to highlight that  $E(y_{ij}|\Theta) = \theta_i$  is presented as a consequence of WALI and not as a definition of the latent variable (see Appendix 11). This result allows to propose an estimator of the marginal distributions of the latent variables involved in Equation 2.35a and in Equation 2.35b. This would add fairness to score assignment, which is really important considering how many high impact decisions are made based on the estimation of latent variables.

Regarding the multidimensional latent variables, the inner product used implies that the covariance between two components of a latent variables is zero. This is a restrictive assumption, because if latent variables are indeed modeling some per-

son's cognitive attributes, there is not apparent reason to assume that two of those attributes are not correlated. However, if covariance between the components is allowed, then the entire Hilbert space should be restricted to multidimensional random variables (observed or latent), such that the variance of the sum of their components is different from zero. For example, if a latent variable is two-dimensional,  $\theta_i = (\theta_{i1}, \theta_{i2})$ , then  $Var(\theta_1) + Var(\theta_2) \neq -2Cov(\theta_1, \theta_2)$ . Therefore, it will be not a Hilbert space, because it will be no complete, and concepts as orthogonal projections may no longer be useful as they are in this work. Also the identification of the marginal distributions of multidimensional latent variables, is valid only if they are orthogonal to each other. In this sense it could be beneficial to explore mathematical tools more suitable to the problems of psychology.

On the other hand, the proposed estimator is based in the identification result of Székely and Rao (2000), which identify multidimensional latent variable, but do not relax the assumptions of none vanishing characteristic function. It would be beneficial for the study of multidimensional latent variable to generalize the identification result of Evdokimov and White (2012) to multidimensional latent variables. And also to adapt other estimators which do not assume equal densities of the errors and neither a particular density. May be, this could lead us to an estimator which not produce "jumps" in the tails of the characteristic function. However, when this restriction in multidimensional random variable is left aside (as in the large used one dimensional latent variables IRT models), Hilbert spaces represent a great mathematical framework to increase our understanding on psychometric models, and the consequences in the properties of the random variables that we use on them.

Regarding the possible theoretical contributions, the assumption of conditional orthogonality at least opens the possibility of discussion about if latent variable are the cause of the observed information in a test. It is known that causality is have been used has criterion of existence (Bhaskar, 2015). For instance, in a classical point of view (in the physical sense), the agreement of the existence of

fields like electric and gravitational is based in a causal criterion, otherwise these concepts would be as esoteric as the luck or the ether. Because we are assuming a sort of causality to estimate the distributions of the latent variables, we can say, if *something* is creating the observed scores, that *something* is distributed in this way in the population. Therefore, if we were able to correlate the estimated distributions with some substantive attribute of the population, we will be closer to discover an attribute with great potential to create a concept.

In summary, the flexibility of Hilbert spaces in included different distributions of random variables, and in embedded many IRT models, along with the utility of the proposed estimator, help us to understand in a deeper way the role of latent variables in some psychometric context. And at the same time, these new methodological tools proved to be useful to open new topics of research in psychometrics.

# References

- Bhaskar, R. (2015). *The possibility of naturalism: A philosophical critique of the contemporary human sciences*. Routledge.
- Biesta, G. (2014). Medir lo que valoramos o valorar lo que medimos? globalización, responsabilidad y la noción de propósito de la educación. *Pensamiento Educativo: Revista de Investigación Educacional Latinoamericana*, 51, 46-57. doi: 10.7764/PEL.51.1.2014.17
- Bonhomme, S., & Robin, J.-M. (2010). Generalized non-parametric deconvolution with an application to earnings dynamics. *The Review of Economic Studies*, 77(2), 491–533. Retrieved from <http://www.jstor.org/stable/40587637>
- Borsboom, D., Mellenbergh, G. J., & Heerden, J. V. (2003). The theoretical status of latent variables. *Psychological Review*, 110, 203–219.
- Bouldin, R. (1973). The product of operators with closed range. *Tohoku Math. J.* (2), 25(3), 359–363. Retrieved from <https://doi.org/10.2748/tmj/1178241337> doi: 10.2748/tmj/1178241337
- Chen, X., Hong, H., & Nekipelov, D. (2011). Nonlinear models of measurement errors. *Journal of Economic Literature*, 49(4), 901–937. Retrieved from <http://www.jstor.org/stable/23071661>
- Dattner, I., Goldenshluger, A., & Juditsky, A. (2011). On deconvolution of distribution functions. *Ann. Statist.*, 39(5), 2477–2501. Retrieved from <https://doi.org/10.1214/11-AOS907> doi: 10.1214/11-AOS907
- de Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized*

*linear and nonlinear approach.* Springer.

Delaigle, A., & Gijbels, I. (2002). Estimation of integrated squared density derivatives from a contaminated sample. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 64(4), 869–886. Retrieved from <http://www.jstor.org/stable/3088819>

Delaigle, A., & Gijbels, I. (2004). Practical bandwidth selection in deconvolution kernel density estimation. *Computational Statistics & Data Analysis*, 45, 249–267. doi: 10.1016/S0167-9473(02)00329-8

Delaigle, A., Hall, P., & Meister, A. (2008). On deconvolution with repeated measurements. *The Annals of Statistics*, 36(2), 665–685. Retrieved from <http://www.jstor.org/stable/25464642>

Drewes, D. (2017). Neo-classical test theory: A modernization. *Acta Psychopathologica*, 03. doi: 10.4172/2469-6676.100140

Dunn, G. (1992). *Design and analysis of reliability studies: The statistical evaluation of measurement errors*. New York, US: Halsted Press and John and Wiley & Sons Inc.

Embretson, S., & P. Reise, S. (2000). *Item response theory for psychologists* (Vol. 4). US: Lawrence Erlbaum Associates Publishers.

Evdokimov, K., & White, H. (2012). Some extensions of a lemma of kotlarski. *Econometric Theory*, 28(4), 925–932. doi: 10.1017/S0266466611000831

Finch, H., & Edwards, J. M. (2016). Rasch model parameter estimation in the presence of a nonnormal latent trait using a nonparametric bayesian approach. *Educational and Psychological Measurement*, 76(4), 662–684. Retrieved from <https://doi.org/10.1177/0013164415608418> doi: 10.1177/0013164415608418

Florens, J.-P., Marimoutou, V., & Pegin-Féissolle, A. (2007). *Econometric modeling and inference* (J. Perktold & M. Carrasco, Eds.). Cambridge University Press. doi: 10.1017/CBO9780511805592

- Florens, J.-P., & Mouchart, M. (1985). A linear theory for non-causality. *Econometrica*, 53(1), 157–175. Retrieved from <http://www.jstor.org/stable/1911729>
- Florens, J. P., Mouchart, M., & Rolin, J.-M. (1990). *Elements of bayesian statistics*. New York: Marcel Dekker.
- Florens, J. P., Mouchart, M., & Rolin, J. M. (1993). Noncausality and marginalization of markov processes. *Econometric Theory*, 9(2), 241–262. Retrieved from <http://www.jstor.org/stable/3532478>
- Gyenis, Z., Hofer-Szabó, G., & Rédei, M. (2017). Conditioning using conditional expectations: the borel–kolmogorov paradox. *Synthese*, 194(7), 2595–2630. Retrieved from <https://doi.org/10.1007/s11229-016-1070-8> doi: 10.1007/s11229-016-1070-8
- Halmos, P. (1957). *Introduction to hilbert space and the theory of spectral multiplicity*. Chelsea Pub. Co. Retrieved from <https://books.google.cl/books?id=cxtRAAAAMAAJ>
- Halmos, P. (2012). *A hilbert space problem book*. Springer New York. Retrieved from <https://books.google.cl/books?id=V1TxBwAAQBAJ>
- Hargreaves, A., & Braun, H. (2013). *Data-driven improvement and accountability*. Boulder, Co: National Education Policy Center. Retrieved from <http://nepc.colorado.edu/publication/data-driven-improvement-accountability/>
- Harwell, M. R., Baker, F. B., & Zwarts, M. (1988). Item parameter estimation via marginal maximum likelihood and an em algorithm: A didactic. *Journal of Educational Statistics*, 13(3), 243–271. Retrieved from <http://www.jstor.org/stable/1164654>
- Holland, P. W., & Hoskens, M. (2003, Mar 01). Classical test theory as a first-order item response theory: Application to true-score prediction from a possibly nonparallel test. *Psychometrika*, 68(1), 123–149. Retrieved from

<https://doi.org/10.1007/BF02296657> doi: 10.1007/BF02296657

Hunter, J., & Nachtergael, B. (2001). *Applied analysis*. World Scientific. Retrieved from <https://books.google.cl/books?id=o0YQVeHmNk4C>

Ito, K. (1986). *Introduction to probability theory*. Cambridge: Cambridge University Press.

Karabatsos, G., & Walker, S. G. (2009). Coherent psychometric modelling with bayesian nonparametrics. *British Journal of Mathematical and Statistical Psychology*, 62(1), 1-20. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1348/000711007X246237> doi: 10.1348/000711007X246237

Koopmans, T. C., & Reiersol, O. (1950). The identification of structural characteristics. *The Annals of Mathematical Statistics*, 21(2), 165–181. Retrieved from <http://www.jstor.org/stable/2236899>

Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Harvard University Press.

Kotlarski, I. (1967). On characterizing the gamma and the normal distribution. *Pacific Journal of Mathematics*, 20(1), 69–76. Retrieved from <https://projecteuclid.org:443/euclid.pjm/1102992970>

Krzanowski, W. (2002). *An introduction to statistical modelling*. John and Wiley & Sons, Ltda.

Lee, M., Hall, P., Shen, H., Marron, J. S., Tolle, J., & Burch, C. (2013). Deconvolution estimation of mixture distributions with boundaries. *Electron. J. Statist.*, 7, 323–341. Retrieved from <https://doi.org/10.1214/13-EJS774> doi: 10.1214/13-EJS774

Li, T., & Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis*, 65(2), 139 - 165. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0047259X98917419>

doi: <https://doi.org/10.1006/jmva.1998.1741>

Lindquist, A., & Picci, G. (2015). *Linear stochastic systems: A geometric approach to modeling, estimation and identification*. Springer-Verlag Berlin Heidelberg.

Lindquist, A., Picci, G., & Ruckebusch, G. (1979). On minimal splitting subspaces and markovian representations. *Mathematical systems theory*, 12(1), 271–279. Retrieved from <https://doi.org/10.1007/BF01776578> doi: 10.1007/BF01776578

Lord, F., & Novick, M. (1968). *Statistical theories of mental test scores*. Addison-Wesley Pub. Co.

Manski, C. (2007). *Identification for prediction and decision*. Harvard University Press.

McCulloch, C. E., & Neuhaus, J. M. (2011). Prediction of random effects in linear and generalized linear models under model misspecification. *Biometrics*, 67(1), 270-279. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0420.2010.01435.x> doi: 10.1111/j.1541-0420.2010.01435.x

McDonald, R. P. (2011). Measuring latent quantities. *Psychometrika*, 76(4), 511–536. Retrieved from <https://doi.org/10.1007/s11336-011-9223-7> doi: 10.1007/s11336-011-9223-7

Mouchart, M., Russo, F., & Wunsch, G. (2010). Inferring causal relations by modelling structures. *Statistica*, 70(4), 411–432. Retrieved from <https://rivista-statistica.unibo.it/article/view/3596> doi: 10.6092/issn.1973-2201/3596

Pearl, J. (2009). *Causality*. Cambridge University Press. doi: 10.1017/CBO9780511803161

R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>

- 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(1), 121-129. doi: 10.3102/10769986029001121
- Raudenbush, S. W., & Willms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, 20(4), 307-335. Retrieved from <http://www.jstor.org/stable/1165304> doi: 10.2307/1165304
- Rektorys, K. (1977). The space l2. In *Variational methods in mathematics, science and engineering* (pp. 32-37). Dordrecht: Springer Netherlands. doi: 10.1007/978-94-011-6450-4
- Rencher, A. C., & Schaalje, G. B. (2008). *Linear models in statistics, second edition*. Wiley-Interscience.
- San Martin, E., Jara, A., Rolin, J.-M., & Mouchart, M. (2011). On the bayesian nonparametric generalization of IRT-type models. *Psychometrika*, 76, 385-409. doi: 10.1007/s11336-011-9213-9
- San Martin, E., Rolin, J.-M., & Castro, L. (2013). Identification of the 1PL model with guessing parameter: Parametric and semi-parametric results. *Psychometrika*, 78, 341-379. doi: 10.1007/s11336-013-9322-8
- San Martín, E. (2015). Rasch models: How (In-)coherently are they showed and used? *Actualidades en Psicología*, 29, 91 - 102. Retrieved from [http://www.scielo.sa.cr/scielo.php?script=sci\\_arttext&pid=S2215-35352015000200091&nrm=iso](http://www.scielo.sa.cr/scielo.php?script=sci_arttext&pid=S2215-35352015000200091&nrm=iso) doi: 10.15517/ap.v29i119.18911
- San Martín, E. (2018). Identifiability of structural characteristics: How relevant is it for the bayesian approach? *Brazilian Journal of Probability and Statistics*, 32, 346-373. doi: 10.1214/16-BJPS346
- San Martín, E., González, J., & Tuerlinckx, F. (2015). On the unidentifiability of the fixed-effects 3PL model. *Psychometrika*, 80(2), 450-467. Retrieved from <https://doi.org/10.1007/s11336-014-9404-2> doi: 10.1007/s11336-014-9404-2

- Shilov, G. (2012). *Linear algebra*. Dover Publications.
- Skrondal, A., & Rabe-Hesketh, S. (2009). Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(3), 659-687. Retrieved from <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-985X.2009.00587.x> doi: 10.1111/j.1467-985X.2009.00587.x
- Small, C. G. (1994). *Hilbert space methods in probability and statistical inference*. New York: Wiley.
- Sobel, M. (2011, 07). Measurement, causation and local independence in latent variable models. In (p. 11-28). doi: 10.1007/978-1-4612-1842-5\_2
- Stone, M. (1991). M. stone : Coordinate-free multivariable statistics: An illustrated geometric progression from halmos to gauss and bayes. *Journal of the Royal Statistical Society, Series A, Statistics in Society*, 154(2), 365–366. Retrieved from <http://search.proquest.com/docview/1307592282/>
- Székely, G. J., & Rao, C. R. (2000). Identifiability of distributions of independent random variables by linear combinations and moments. *Sankhy: The Indian Journal of Statistics, Series A (1961-2002)*, 62(2), 193–202. Retrieved from <http://www.jstor.org/stable/25051306>
- Welham, S., Cullis, B., Gogel, B., Gilmour, A., & Thompson, R. (2004). Prediction in linear mixed models. *Australian & New Zealand Journal of Statistics*, 46(3), 325-347. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-842X.2004.00334.x> doi: 10.1111/j.1467-842X.2004.00334.x
- Woods, C. M., & Lin, N. (2009). Item response theory with estimation of the latent density using davidian curves. *Applied Psychological Measurement*, 33(2), 102-117. Retrieved from <https://doi.org/10.1177/0146621608319512> doi: 10.1177/0146621608319512
- Zimmerman, D. W. (1975). Probability spaces, hilbert spaces, and the ax-

ioms of test theory. *Psychometrika*, 40(3), 395–412. Retrieved from <https://doi.org/10.1007/BF02291765> doi: 10.1007/BF02291765

Zimmerman, D. W., & Williams, R. H. (1966). Generalization of the spearman-brown formula for test reliability: The case of non-independence of true scores and error scores. *British Journal of Mathematical and Statistical Psychology*, 19(2), 271-274. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2044-8317.1966.tb00371.x> doi: 10.1111/j.2044-8317.1966.tb00371.x

## **Appendix 1: Definition of a $L^2$ space**

It is the set of square integrable functions with domain in a measurable space (Rektorys, 1977; Hunter & Nachtergael, 2001).

## **Appendix 2: Definition of a sigma algebra**

This definition was obtained from the Florens, Mouchart, and Rolin (1990) and Ito (1986). Let  $M$  be a set (all the possible outcomes of a random variable) and  $\mathcal{M}$  be a family of subsets of  $M$  (a collection of subset of  $M$ ).  $\mathcal{M}$  is a sigma-algebra if

1  $\emptyset \in \mathcal{M}$  (the empty set is in  $\mathcal{M}$ )

2  $A \in \mathcal{M} \Rightarrow A^c \in \mathcal{M}$  (if any event  $A$  (a subset of  $M$ ) is in  $\mathcal{M}$ , then his complement in  $M$  also is in  $\mathcal{M}$ )

3  $A_n \in \mathcal{M}$  ( $n = 1, 2, 3, \dots$ )  $\Rightarrow \cup_{n=1}^{\infty} A_i \in \mathcal{M}$

## **Appendix 3: Definition of a Hilbert subspace**

Let  $M$  be a subset of a Hilbert space  $H$  such that if  $x$  and  $y$  are any elements of  $M$ , then  $\alpha x + \beta y$  also belong to  $M$ , where  $\alpha$  and  $\beta$  are scalars (reals or complex). If also  $M$  is complete, then  $M$  is a Hilbert subspace of  $H$  (Halmos, 1957). Note  $M$  is a Hilbert space on his own, hence a Hilbert subspace is a Hilbert space which is part of a bigger Hilbert space.

## **Appendix 4: Definition of a Vectorial subspace (or Linear subspace)**

It is a vectorial space which is part of a bigger vectorial space. A set of vectors  $V$  is a vectorial space over a field  $f$  (could be the reals or complex numbers) if (Shilov, 2012):

i  $x$  and  $y$  are two elements of  $V$ , then the sum rule  $x + y$  is in  $V$  (closed under sum).

ii  $\alpha$  is a scalar in  $f$  and  $x$  is any element in  $V$ , then the multiplication by a scalar rule  $\alpha x$  lead to a unique value in  $V$  (closed under multiplication by a scalar).

And for any  $x, y, z \in V$  and  $\alpha, \beta \in f$  these rules (sum and multiplication by a scalar) must obey:

- $1x = x$ .
- $(x + y) + z = x + (y + z)$ .
- There is a zero element  $0$  such that  $x + 0 = x$ .
- $\forall x$  there is an  $y$  such that  $x + y = 0$ .
- $x + y = y + x$ .
- $\alpha(\beta x) = (\alpha\beta)x$ .
- $(\alpha + \beta)x = \alpha x + \beta x$ .
- $\alpha(x + y) = \alpha x + \alpha y$ .

## Appendix 5: Definition of a Cauchy series

Let  $\{x_n\}_{n \in \mathbb{N}}$  be a sequence (of vectors or scalars).  $\{x_n\}$  is a Cauchy sequence if for all  $\varepsilon > 0$  in the reals, there is a positive integer  $N$  such that for  $m, n > N$  in the natural numbers

$$\|x_m - x_n\| < \varepsilon \quad (0.1)$$

In this study the norm  $\|\cdot\|$  is induced for the inner product. A Cauchy sequence is a sequence where his elements become arbitrary close to each other, but not necessarily converge to the same number (Halmos, 1957).

## Appendix 6: Definition of inner product

Restricted to vectorial spaces defined over the reals, an inner product is a real-valued function acting on the cartesian product of a vectorial space  $V$  (in this research was  $H$ ) and itself:

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R} \quad (0.2)$$

having the next properties (Halmos, 1957):

- symmetry,  $\langle x, y \rangle = \langle y, x \rangle$
- linearity in the first argument,  $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$  and  $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$
- Positive definiteness,  $\langle x, x \rangle \geq 0$  and  $\langle x, x \rangle = 0 \Leftrightarrow x = \vec{0}$

for all  $x, y$  and  $z$  in  $V$ , and for any real number  $\alpha$ .  $\vec{0}$  represents the zero vector i.e., a vector with all his components equals to zero.

## Appendix 7: Proof, expected value as an inner product

To demonstrate that the expected value of the product between two scalars random variables is an inner product, we must to check the properties described in the definition of inner product (Appendix 2). The expected value of the product between two random variables is

$$E(xy) = \int xy p(x, y) dx dy. \quad (0.3)$$

Because  $xy = yx$  and the joint distribution of  $x$  and  $y$  is the same that  $y$  and  $x$ ,  $E(xy)$  is symmetric. Regarding linearity in the first argument, using the function  $g(x) = \alpha x$

$$E[g(x)z] = \int \alpha x y p(x, y) dx dy = \alpha E(xz). \quad (0.4)$$

Also,

$$E[(x + y)z] = \int (x + y)p(x + y, z) d(x + y) dz \quad (0.5)$$

$$= \int xp(x + y, z) dx dz + \int yp(x + y, z) dy dz. \quad (0.6)$$

Note in the first integral that  $y$  is constant and in the second integral  $x$  is constant, thus they only translate the curve of the probability function, but do not alter its form. Thus, arbitrarily, we can fix  $y$  and  $x$  respectively in the first and second integral to zero, in this way we get the linearity in the first argument.

The demonstration of positiveness definition is crucial to improve our understanding of the Hilbert where we have worked. The first part is painless ( $E(x^2) \geq 0$ ), but

we need to be more careful with the second part. If we choose  $g(x) = x^2$  then

$$E(x^2) = \int x^2 p(x) dx. \quad (0.7)$$

This integral is zero when almost sure  $x = 0$ , not only when  $x$  is zero. To deal with this problem in  $H$  it has been agreed that two random variables are “equals” when they have the same values almost everywhere, in this way the positiveness definition is obtained. This is why this space is called the space of equivalent functions and not the space of functions.

## Appendix 8: Proof, conditional expected value as an orthogonal projection

Previous to show that the conditional expected value is an orthogonal projection, we need the next result. If  $x$  and  $w$  are any two random variables in  $H$  and  $f(w)$  is any function of  $w$  in  $H$ , then  $f(w) \perp x - E(x|w)$ . Thus,

$$\begin{aligned} \langle f(w), x - E(x|w) \rangle &= E[f(w)(x - E(x|w))] \\ 0 &= E\{E[f(w)(x - E(x|w))|w]\} \\ 0 &= E\{f(w)E[x - E(x|w)|w]\} \\ 0 &= E\{f(w) * 0\}. \end{aligned}$$

hence  $x - E(x|w)$  is perpendicular to any function of  $w$ .

Now, to proof that  $E(x|w)$  is an orthogonal projection in  $H$  ( $L^2$ , the equivalent functions space) we need to show that  $E[x - E(x|w)]^2 \leq E[x - f(w)]^2$ , i.e., that the distance between  $x$  and any function  $f(w)$  is minimal when this function is  $E(x|w)$ .

In this way,

$$\begin{aligned} E[x - f(w)]^2 &= E[x - E(x|w) + E(x|w) - f(w)]^2 \\ &= E[x - E(x|w)]^2 - 2E\left\{[x - E(x|w)][E(x|w) - f(w)]\right\} \\ &\quad + E[E(x|w) + f(w)]^2 \end{aligned}$$

Note that the negative term is an inner product between  $x - E(x|w)$  and  $E(x|w) - f(w)$ . Also  $E(x|w) - f(w)$  is a function of  $w$ , therefore the negative term is zero. Also  $E[E(x|w) + f(w)]^2$  is always positive, thus the distance between  $x$  and any function of  $w$  is always larger than the distance between the same  $x$  and  $E(x|w)$   $\square$ .

## **Appendix 9: Proof, orthogonality between the true score and the error term**

In Appendix 4 was proved that  $x - E(x|w)$  is perpendicular to any function of  $w$ . Because the true score is a function of  $\theta$ , the error term and the true score are perpendicular.

## **Appendix 10: Proof, orthogonal decomposition**

To prove the orthogonal decomposition

$$A = \bar{E}(B|A) \oplus A \cap B^\perp \quad (0.8)$$

where  $A$  and  $B$  are two Hilbert subspaces, we follow the next procedure. First we name two subspaces  $A_1 = \bar{E}(B|A)$  and  $A_2 = A \cap B^\perp$ , and then we show that there is an orthogonal projection such that his range (the closure) is equal to  $A_1$  and his kernel is equal to  $A_2$ . To simplify the notation, during this demonstrations  $P_A b$  will denote the orthogonal projection from an element in  $B$  onto  $A$ . Hence, in this notation, the decomposition to proof can be written as

$$A = \bar{P}_A B \oplus A \cap B^\perp \quad (0.9)$$

Then, how was explained, let (the subspace)  $A_1$  be the closure of the imagine of  $B$  under the orthogonal projection  $P_A$ , and (the subspace)  $A_2$  be  $A \cap B^\perp$ .

First we prove that there is an orthogonal projection different from  $P_A$  such that the closure of his range is  $A_1$ . To this end let define a subspace  $B'$  as,

$$B^* = \{b \in B : P_A b = b\} \quad (0.10)$$

$$B' = \overline{B^*} \text{ (closure of } B^*) \quad (0.11)$$

Note that  $B' \subseteq A$ , thus for any  $b' \in B'$  and for any  $a \in A$ :

$$P_A b' = P_{B'} a = b' \quad (0.12)$$

where  $P_{B'}$  is an orthogonal projection onto  $B'$ . On the other hand, because how was defined  $A_1$ , it is contained in  $B$  and

$$\forall a_1 \in A_1$$

$$a_1 = P_A b = P_A(P_A b) = P_A(a_1), \quad (0.13)$$

hence every element of  $A_1$  is in  $B'$  too, but because  $B' \subseteq B$  then  $A_1 \supseteq B'$ . Therefore from Equation 0.12 and Equation 0.13 is concluded that  $A_1 = \overline{\text{Ran}(P_A)} = \overline{\text{Ran}(P_{B'})} = B'$ , where  $P_{B'}$  is the operator formally defined as:

$$P_{B'} : A \longrightarrow B' \quad (0.14)$$

$$a \longrightarrow b' (= a_1) \quad (\forall a \in A), \quad (0.15)$$

hence it is the operator which orthogonal project  $A$  onto  $B'$ .

Only remain to prove that  $A \cap B^\perp = \text{Ker}(P_{B'})$ . To this end let call  $A \cap B^\perp = A_2$  and  $a_2$  any element in it. We can define  $A_2$  under the condition:

$$A_2 = \{a \in A : P_B a = 0\} \quad (0.16)$$

Also we can define a subspace  $A'_2 = A \cap B'^\perp$  as:

$$A'_2 = \{a \in A : P_{B'}a = 0\} \quad (0.17)$$

Note that  $A_2$  and  $A'_2$  are the kernels of  $P_A$  and  $P_B$  respectively, hence we are interested in verifying if  $A_2 = A'_2$ . From the definition of  $B'$  we see:

$$B' \subseteq B \Rightarrow B'^\perp \supseteq B^\perp \quad (0.18)$$

$$\therefore A'_2 \supseteq A_2 \text{ or } A_2 = A'_2 \quad (0.19)$$

Also  $B'$  is the biggest subspace of  $B$  contained in  $A$ , thus all what is in  $B'^\perp$  which is not in  $B^\perp$  is contained in  $A$ . Thus  $A_2 = A'_2$  or what is the same  $\text{Ker}(P_{B'}) = A \cap B^\perp$  (which is closed).

Therefore Decomposition 2.17 is equivalent to the very well known decomposition of any subspace as the range of an orthogonal decomposition and his kernel  $\square$ .

## Appendix 11: Proof, minimal splitting and identification

This proof is a reconstruction of the paper (Lindquist et al., 1979). In order to prove Equation 2.29a and Equation 2.29b we need to prove two others equivalences first. Let us start with showing that the definition given in the Equation 2.28 is equivalent to

$$\langle y_1, y_2 \rangle = \langle E(y_1|\Theta), E(y_2|\Theta) \rangle \quad \forall y_1 \in Y_1, \quad \forall y_2 \in Y_2. \quad (0.20)$$

Note that Equation 2.28 ( $\langle y_1 - E(y_1|\Theta), y_2 - E(y_2|\Theta) \rangle = 0$ ) can be rearranged as,

$$\langle y_1, y_2 \rangle - \langle y_1, E(y_2|\Theta) \rangle - \langle E(y_1|\Theta), y_2 \rangle + \langle E(y_1|\Theta), E(y_2|\Theta) \rangle \quad (0.21)$$

But each  $y_1$  and  $y_2$  can be orthogonality decomposed as  $y_1 = E(y_1|\Theta) + (y_1 - E(y_1|\Theta))$  and  $y_2 = E(y_2|\Theta) + (y_2 - E(y_2|\Theta))$ , then,

$$\langle E(y_1|\Theta), y_2 \rangle = \langle E(y_1|\Theta), E(y_2|\Theta) \rangle + 0 \text{ and} \quad (0.22)$$

$$\langle y_1, E(y_2|\Theta) \rangle = 0 + \langle E(y_1|\Theta), E(y_2|\Theta) \rangle \quad (0.23)$$

$$\therefore \langle E(y_1|\Theta), y_2 \rangle = \langle y_1, E(y_2|\Theta) \rangle = \langle E(y_1|\Theta), E(y_2|\Theta) \rangle. \quad (0.24)$$

Therefore:

$$\langle y_1 - E(y_1|\Theta), y_2 - E(y_2|\Theta) \rangle = \langle y_1, y_2 \rangle - \langle E(y_1|\Theta), E(y_2|\Theta) \rangle = 0 \quad \square \quad (0.25)$$

Still there is one more result that we need in order to prove the identification result. We can decompose  $\Theta$  into two orthogonal subspaces as  $\Theta = \Theta_1 \oplus \Theta_2$ , ( $\oplus$  is the orthogonal direct sum), if  $\Theta$  is a splitting subspace respect to  $Y_1$  and  $Y_2$  then  $\Theta_1$  is also splitting subspace respect to  $Y_1$  and  $Y_2$  if and only if

$$E(Y_1|\Theta_2) \perp E(Y_2|\Theta_2) \quad (0.26)$$

Using Equation 0.20 we can rewrite this condition as:

$$\begin{aligned} \langle y_1, y_2 \rangle &= \langle E(y_1|\Theta), E(y_2|\Theta) \rangle \\ &= \langle E(y_1|\Theta_1), E(y_2|\Theta_1) \rangle + \langle E(y_1|\Theta_2), E(y_2|\Theta_2) \rangle \end{aligned} \quad (0.27)$$

Thus if  $\Theta_1$  is a splitting subspace necessarily:

$$\langle E(y_1|\Theta_2), E(y_2|\Theta_2) \rangle = 0 \Leftrightarrow E(Y_1|\Theta_2) \perp E(Y_2|\Theta_2) \quad \square \quad (0.28)$$

Now, with these two results, we can prove that find the splitting subspaces that obey to Equation 2.29a and Equation 2.29b, it is equivalent to find all the minimal splitting subspaces respect to  $Y_1$  and  $Y_2$ .

Let first assume that  $\Theta$  is a minimal splitting subspace. Note that  $\Theta_1 = \bar{E}(Y_1|\Theta)$

and  $\Theta_2 = \Theta \cap Y_1^\perp$  which implies  $E(Y_1|\Theta_2) = 0$ , so  $E(Y_1|\Theta_2) \perp E(Y_2|\Theta_2) = 0$ , therefore  $\Theta_1$  is a splitting subspace. But  $\Theta$  was assumed to be a minimal splitting subspace, so  $\Theta_1 = \Theta = \bar{E}(Y_1|\Theta)$ . This reasoning can easily be applied to  $\Theta$  and  $Y_2$ . Therefore  $\Theta$  being a minimal splitting subspace respect to  $Y_1$  and  $Y_2$  implies 2.29a and 2.29b.

Now let assume that  $\Theta$  obey Equation 2.29a and Equation 2.29b, and  $\Theta_1$  is any splitting subspace contained in  $\Theta$ . We need to prove that  $\Theta_2 = 0$  thus  $\Theta$  is a minimal splitting subspace. Because  $\Theta_1$  is a splitting subspace  $E(Y_1|\Theta_2) \perp E(Y_2|\Theta_2)$ . On the other hand using Decomposition 2.17,

$$\Theta = \bar{E}(Y_1|\Theta)$$

$$\Theta = \bar{E}(Y_2|\Theta).$$

Also  $\Theta_2 \subset \Theta$ , then  $\Theta_2 = \bar{E}[\Theta|\Theta_2] = \bar{E}[E(Y_1|\Theta)|\Theta_2] = \bar{E}(Y_1|\Theta_2)$ , in the same way  $\bar{E}(Y_2|\Theta_2) = \Theta_2$ . Therefore, because Equation 0.26, the only way that  $\Theta_1$  is a splitting subspace is that  $\Theta_2 = 0$ , so  $\Theta$  is a minimal splitting subspace  $\square$ . This ends the demonstration.  $\square$

## Appendix 7: Proof, boundaries of a minimal splitting subspace

This proof is also a reconstruction of the one presented by Lindquist et al. (1979). To deduce the boundaries in the Latent subspace when WALI is true, we need first two other affirmations. Let start with the next affirmation: The subspace  $\Theta \subset H$  ( $H$  is our Hilbert space defined in Equation 2.1) is a splitting subspace satisfying 2.34a ( $E(y_1|\Theta) = \theta$ ) if and only if  $\Theta = \bar{E}(Y_1|S)$  for some subspace  $S \supset Y_2$ .

To proof this we can rearrange the definition of splitting subspace in Equation

2.28 ( $\langle y_1 - E(y_1|\Theta), y_2 - E(y_2|\Theta) \rangle = 0$ ) as

$$\langle y_1 - E(y_1|\Theta), y_2 \rangle - \langle y_1 - E(y_1|\Theta), E(y_2|\Theta) \rangle = 0 \quad (0.29)$$

Recall in Equation 0.29  $E(y_1|\Theta)$  and  $E(y_2|\Theta)$  represent vectors with the true scores in the population of interest. Note that  $E(y_2|\Theta) \in \Theta$  and  $(y_1 - E(y_1|\Theta)) \perp \Theta$ . Hence:

$$\begin{aligned} & \langle y_1 - E(y_1|\Theta), y_2 \rangle = 0 \\ & = \langle y_1 - E(y_1|\Theta), y_2 + \theta_+ \rangle; \quad \forall \theta \in \Theta \text{ and } \forall y_2 \in Y_2. \end{aligned} \quad (0.30)$$

The right side of Equation 0.30 means that  $(y_1 - E(y_1|\Theta)) \perp Y_2 \vee \Theta$ . Then

$$E(y_1 - E(y_1|\Theta)|Y_2 \vee \Theta) = 0 \quad (0.31)$$

$$E(y_1|Y_2 \vee \Theta) = E(E(y_1|\Theta) | Y_2 \vee \Theta) \quad (0.32)$$

Since  $\Theta \subset Y_2 \vee \Theta$  the right side of Equation 0.32 is equal to  $E(y_1|\Theta)$ . Thus if  $Y_1 \perp Y_2|\Theta$  then,  $E(y_1|Y_2 \vee \Theta) = E(y_1|\Theta)$ . Therefore by Equations 2.29a and 2.30a ( $\theta = \bar{E}(y_1|\Theta)$  and  $\theta = E(y_1|Y_2 \vee \Theta) \forall (\theta, y_1) \in (\Theta, Y_1)$ ), and  $Y_2 \vee \Theta \supset Y_2$ , choosing  $S = Y_2 \vee \Theta$  the first part of the demonstration its ended.

For the second part we need to prove that some  $S \supset Y_2$ . If  $\Theta = \bar{E}(Y_1|S)$  if  $\Theta = \bar{E}(Y_1|\Theta)$ . From  $\Theta = \bar{E}(Y_1|S)$  it is deduced that  $\Theta \subset S$ , hence  $\bar{E}(E(y_1|S)|\Theta) = \bar{E}(y_1|\Theta)$ . Also, from the last condition,  $\bar{E}(y_1|\Theta) = \bar{E}(y_1|S)$ . On the other hand, as a result of the first part of the proof, we choose  $S \supset Y_2 \vee \Theta$  hence  $\bar{E}(y_1|Y_2 \vee \Theta) = \bar{E}(y_1|\Theta)$  which is true only if  $\Theta$  is a splitting subspace (this was proved in the first part of the demonstration). Thus,  $\Theta = \bar{E}(Y_1|S)$  and  $S \supset Y_2$  is equivalent to  $\Theta$  being a minimal splitting subspace respect to  $Y_1$  and  $Y_2$ . This end the demonstration.

Before to prove the boundaries of the latent spaces we need this second affirmation. Let  $S \supset Y_2$ , then  $\bar{E}(Y_1 \cap Y_2^\perp | S) = \bar{E}(Y_1 | S) \cap Y_2^\perp$ .

A demonstration of this result is as follows; it is direct that  $\bar{E}(Y_1 \cap Y_2^\perp | S) \subset \bar{E}(Y_1 | S)$ , but is not so clear that  $\bar{E}(Y_1 \cap Y_2^\perp | S) \subset Y_2^\perp$ . In order to prove the last inclusion let  $y$  be any element in this intersection, thus  $y \in Y_1 \cap Y_2^\perp$  so  $\bar{E}(Y_1 \cap Y_2^\perp | S) = \bar{E}(y | S)$ . From  $y \perp Y_2$  is straightforward that  $\bar{E}(y | Y_2) = 0$ , also  $Y_2 \subset S$  therefore  $\bar{E}[E(y | S) | Y_2] = 0$  so  $\bar{E}(y | S) \in Y_2^\perp \forall y \in Y_1 \cap Y_2^\perp \square$ .

Now we can prove that the boundaries of  $\Theta$  are  $Y_2 \subset \Theta \subset (Y_2 \vee Y^\square) \oplus Y_0^\perp$ , where  $Y^\square = E(Y_1 | Y_2) \vee E(Y_2 | Y_1)$ . The bottom boundary is already proved, as in all the previous results we used *for some*  $S \supset Y_2$ . Thus it only remains to deduce the upper bound. Due to Equation 2.29b implies to demonstrate that  $S \subset (Y_2 \vee Y^\square) \oplus Y_0^\perp$  if and only if  $\bar{E}(Y_1 | S) \cap Y_2^\perp = 0$ .

Because  $\bar{E}(Y_1 \cap Y_2^\perp | S) = \bar{E}(Y_1 | S) \cap Y_2^\perp$  this last condition implies  $S \subset (Y_1 \cap Y_2^\perp)^\perp$ .

In the other hand using the orthogonal decomposition 2.17 in  $Y_1$ :

$$Y_0 = Y_1 \vee Y_2 \tag{0.33}$$

$$Y_0 = Y_2 \vee E(Y_2 | Y_1) \oplus Y_1 \cap Y_2^\perp \tag{0.34}$$

$$Y_0 = Y_2 \vee E(Y_1 | Y_2) \vee E(Y_2 | Y_1) \oplus Y_1 \cap Y_2^\perp \tag{0.35}$$

$$Y_0 = Y_2 \vee Y^\square \oplus Y_1 \cap Y_2^\perp \tag{0.36}$$

Therefore  $(Y_1 \cap Y_2^\perp)^\perp = (Y_2 \vee Y^\square) \oplus Y_0^\perp$ . Where  $(Y_1 \cap Y_2^\perp)^\perp$  is the orthogonal complement in all the space  $H$ . So  $S \subset (Y_1 \cap Y_2)^\perp \Leftrightarrow S \subset (Y_2 \vee Y^\square) \oplus Y_0^\perp \square$ .

It is important to highlight that we used the result  $\overline{E}(Y_1 \cap Y_2^\perp \mid S) = \overline{E}(Y_1 \mid S) \cap Y_2^\perp$  for some  $S \supset Y_2$  in order to proof the boundaries of a minimal splitting subspace, and not  $E(Y_1 \cap Y_2^\perp \mid S) = E(Y_1 \mid S) \cap Y_2^\perp$  how was used by Lindquist et al. (1979). Florens and Mouchart (1985) pointed out this mistake. Lindquist et al. (1979), assumed that all the subspaces are closed, but an orthogonal projection, even between two closed subspace, can be open if we work on infinite dimensional subspaces. Hence, because  $\Theta$  must be closed (equivalent to complete) we need the proof of  $\overline{E}(Y_1 \cap Y_2^\perp \mid S) = \overline{E}(Y_1 \mid S) \cap Y_2^\perp$ . Recall a closed subspace contain all his limit point (it is complete), thus if  $\Theta$  where open, there will be always the risk that some true scores do not exist, or some random variables defined over some psychological attribute could partiality exist in  $\Theta$ , because some values just will be forbidden. In mathematical terms, it is a space without holes. The next is an example recovered from Halmos (2012) where is shown that an image under a projection of an infinite dimension (closed) subspace is open.

Let  $H$  be a Hilbert space (could be the one we use in this thesis, but the proof is more general). Considerer the direct sum  $H \oplus H$ . If  $T$  is an operator acting on  $H$ , such that  $0 \leq \|T\| \leq 1$  ( $\|T\|$  is the norm of the operator  $T$ ) then  $P$  is a projection acting on  $H \oplus H$ ,

$$P = \begin{bmatrix} T & \sqrt{T(1-T)} \\ \sqrt{T(1-T)} & 1-T \end{bmatrix}.$$

Because  $0 \leq \|T\| \leq 1$  the square root is positive. To prove that  $P$  is a projection

we must show that  $P$  is idempotent, thus if  $v$  is any vector in  $H \oplus H$ , then  $Pv = PPv$ .

We can decompose  $v$  as  $v = x + y$ , where  $x$  and  $y$  are vectors in the *first* and *second*  $H$  respectively in the expression  $H \oplus H$ . Hence

$$P \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} Tx + \sqrt{T(1-T)}y \\ \sqrt{T(1-T)}x + (1-T)y \end{bmatrix}$$

applying  $P$  again and after reordering terms we get

$$P \begin{bmatrix} Tx + \sqrt{T(1-T)}y \\ \sqrt{T(1-T)}x + (1-T)y \end{bmatrix} = \begin{bmatrix} Tx + \sqrt{T(1-T)}y \\ \sqrt{T(1-T)}x + (1-T)y \end{bmatrix}$$

thus  $P$  is idempotent therefore it is a projection.

Now we can choose an operator  $T$  in  $l^2$  (the Hilbert space of sequences because the easier to calculate a norm) with the desired qualities ( $0 < \|T\| < 1$ ). We choose  $T$  to be:

$$T = \text{diag} \left\{ \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots, \frac{1}{k} \right\},$$

where  $k$  can be infinite. It is possible to make a projection  $P$  out of  $T$ . To prove that the image under a projection can be no closed, we are going to apply  $P$  on all the vectors of the form  $(x, 0)^t$  (zero component in the second  $H$  of  $H \oplus H$ ). Recall that if the limit of a sequence of vectors in a given inner product space do not belong to the inner product space, then the inner product space is not closed (a Hilbert space is a complete inner space). Then, applying  $P$  to a vector of the wanted form we get,

$$P \begin{bmatrix} x \\ 0 \end{bmatrix} = \begin{bmatrix} Tx \\ \sqrt{T(1-T)}x \end{bmatrix}$$

Thus the image under  $P$  of the vectors of the form  $(x, 0)^t$  is  $(Tx, \sqrt{T(1-T)}x)^t$ . Just for simplicity, if  $k$  was four and we choose a three-dimensional vector  $x = (x_0, x_1, x_2)^t$  to applied  $T$ , we get,

$$Tx = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{4} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{x_0}{2} \\ \frac{x_1}{3} \\ \frac{x_2}{4} \end{bmatrix}$$

Hence, in this example a vector  $v = (v_0, v_1, v_2)^t$  belong to the image of  $T$ , if  $x_0 = 2v_0$ ,  $x_1 = 3v_1$  and  $x_2 = 4v_2$ . Hence (because  $T$  act on  $l^2$ ) if an infinite-dimensional vector  $v$  belong to the range  $T$  then,

$$\sum_{n=0}^{\infty} (n+2)^2 |v_n|^2 < \infty \quad (0.37)$$

in a similar way if the infinite-dimensional vector  $v$  belong to the range of  $\sqrt{T(1-T)}$  then,

$$\sum_{n=0}^{\infty} \frac{(n+2)^2}{n+1} |v_n|^2 < \infty \quad (0.38)$$

It only remains to find a sequence of vectors  $v_i$  in  $l^2$  such that the limit of the sequence composed by the image of each  $v_i$  does not belong to the range of  $P$ . One example of this is

$$v_i = \left( \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{3}}, \dots, \frac{1}{\sqrt{i+1}}, 0 \dots \right)^t \quad (0.39)$$

When  $i$  tends to infinity both series in Equation 0.37 and Equation 0.38 approach to infinity, hence it does not belong to the range of  $P$ . Therefore the range of the projection  $P$  in this example is not closed, and thus the range of a projection can be no closed in an infinite-dimensional vector space.

### **Appendix 13: Proof, minimal splitting subspace contained in $Y_2$**

This proof was recover from Florens et al. (1990). Let call  $\Theta$  a splitting subspace respect to two observed scores subspaces  $Y_1$  and  $Y_2$  such that  $\Theta \subset Y_2$ . Equation 0.32 implies that  $E(y_1|Y_2 \vee \Theta) = E(y_1|\Theta)$ . Then, because  $\Theta \subset Y_2$   $E(y_1|Y_2) = E(y_1|\Theta)$ . Therefore  $\overline{E}(Y_1|Y_2) \subset \Theta$ . Hence, for any splitting subspace  $\Theta$  contained in  $Y_2$ , it is true that:

$$\overline{E}(Y_1|Y_2) \subset \Theta \subset Y_2 \quad (0.40)$$

Thus the minimal splitting subspace contained in  $Y_2$  is  $\overline{E}(Y_1|Y_2)$ . The same procedure can be done changing  $Y_2$  for  $Y_1$   $\square$ .

## Appendix 14: Definition of Cauchy-Schwarz inequality

If  $x$  and  $y$  are any two vectors in a inner subspace ( $H$  and all his subspaces are inner product space) the Cauchy-Schwarz inequality is (Halmos, 1957):

$$\langle x, y \rangle \leq \|x\| \|y\| \quad (0.41)$$

## Appendix 15: Proof, identification result of Kotlarski

Kotlarski (1967) proved that if

$$y_1 = \theta + e_1 \quad \text{and} \quad y_2 = \theta + e_2 \quad (0.42)$$

The joint distribution of  $y_1$  and  $y_2$  completely determines the distributions of  $\theta$ ,  $e_1$  and  $e_2$ . The proof is as follow.

Let denote by  $\varphi(t_1, t_2)$  the joint characteristic function of  $(y_1, y_2)$ , and  $\varphi_\theta(t)$ ,  $\varphi_{e_1}(t)$  and  $\varphi_{e_2}(t)$  the characteristic functions of  $\theta$ ,  $e_1$  and  $e_2$  respectively, then

$$\varphi(t_1, t_2) = E \exp \left\{ i [t_1 y_1 + t_2 y_2] \right\} \quad (0.43)$$

$$\varphi(t_1, t_2) = E \exp \left\{ i [t_1 (\theta + e_1) + t_2 (\theta + e_2)] \right\} \quad (0.44)$$

$$\varphi(t_1, t_2) = \varphi_\theta(t_1) \varphi_{e_1}(t_1) \varphi_\theta(t_2) \varphi_{e_2}(t_2) \varphi_\theta(t_1 + t_2). \quad (0.45)$$

Now let denote by  $\tau$ ,  $\epsilon_1$  and  $\epsilon_2$  another three independent real random variables

with no vanishing characteristics functions  $\psi_\tau$ ,  $\psi_{\epsilon_1}$ ,  $\psi_{\epsilon_2}$  respectively. Considering  $x_1 = \tau + \epsilon_1$ ,  $x_2 = \tau + \epsilon_2$ , just as before

$$\psi(t_1, t_2) = \psi_{\epsilon_1}(t_1)\psi_{\epsilon_2}(t_2)\psi_\tau(t_1 + t_2). \quad (0.46)$$

Any distribution is completely determined by the characteristic function. Then, if  $\psi(t_1, t_2) = \varphi(t_1, t_2)$  implies that each characteristic function are equals up to a change of scale, the result will be demonstrated. Proceeding in this way,

$$\psi_{\epsilon_1}(t_1)\psi_{\epsilon_2}(t_2)\psi_\tau(t_1 + t_2) = \varphi_{e_1}(t_1)\varphi_{e_2}(t_2)\varphi_\theta(t_1 + t_2). \quad (0.47)$$

Now, in Equation 0.47 let use  $\psi_\tau(t) = \varphi_\theta(t)p_\theta(t)$ ,  $\psi_{\epsilon_1}(t) = \varphi_{e_1}(t)p_{e_1}(t)$  and  $\psi_{\epsilon_2}(t) = \varphi_{e_2}(t)p_{e_2}(t)$ , with  $p(t)$  being a complex function that satisfies  $p(0) = 1$  and  $t \in \mathbb{R}$ . Replacing in Equation 0.47 we get,

$$p_{e_1}(t_1)p_{e_2}(t_2)p_\theta(t_1 + t_2) = 1. \quad (0.48)$$

To solve equation 0.48, let evaluate it in  $t_2 = 0$  and then in  $t_1 = 0$ . By doing this we get,

$$p_{e_1}(t_1) = p_\theta^{-1}(t_1) \quad ; \quad p_{e_2}(t_2) = p_\theta^{-1}(t_2) \quad (0.49)$$

Replacing these results in 0.48

$$p_\theta(t_1 + t_2) = p_\theta(t_1)p_\theta(t_2) \quad (0.50)$$

The only complex continuous function which is a solution of 0.50 under the condition  $p_\theta(0) = 1$  is the exponential function, thus  $p_\theta(t) = e^{ct}$ . Then, for each  $\psi(t)$ ,  $\varphi(t)$  and  $p(t)$ ,  $\psi(t) = e^{ct}\varphi(t)$ . And because each  $\psi(t)$  is a characteristic function,  $\psi(-t) = \bar{\psi}(t)$ , then  $c = ib$ . Where  $i^2 = -1$  and  $b \in \mathbb{R}$ . Therefore the distribution of  $y_1$  and  $y_2$  determines the distributions of  $e_1, e_2$  and  $\theta$  up to a change of location  $\square$ .

## Appendix 17: Plug-In method

Let call  $\theta$  the random variable of interest. In order to get the optimum bandwidth to estimate the distribution of  $\theta$ , at some point is necessary to assume some distribution, usually the normal. The method is iterative, more steps ensure less influence of the assumed distribution used. But more steps increase the variance of the parameter of interest ( $R$ ). Because the increase in variance is more important with the numbers of steps, and following the recommendation of Delaigle and Gijbels (2004), we used two steps. The procedure is as follows:

1. Compute

$$R_4 = \frac{8!}{2^9 Var^9(\theta) 4! \sqrt{\pi}} \quad (0.51)$$

The variance is estimated from the characteristic function of  $\theta$ ,  $\varphi_\theta(t)$ . For this step is chosen the bigger wider bandwidth used in the method. This decision is justified by theorem 2 in Bonhomme and Robin (2010).

2. Replace  $\hat{R}_4$  in the expression

$$-\frac{\mu R_4}{T_3^2} + \frac{1}{2\pi N} \int v^6 \left| \varphi_{H2} \left( \frac{v}{T_3} \right) \right|^2 |\varphi_e(v)|^{-2} dv, \quad (0.52)$$

and then minimize the expression respect to  $T_3$  to obtain  $\hat{T}_3$ .

3. Replace  $\hat{T}_3$  in the next expression to estimate  $R_3$

$$R_3 = \frac{1}{2\pi} \int v^6 \left| \varphi_{H2} \left( \frac{v}{T_3} \right) \right|^2 \left| \frac{\varphi_y(v)}{\varphi_e(v)} \right|^2 dv. \quad (0.53)$$

4. Replace the estimated value of  $R_3$  in

$$-\frac{\mu R_3}{T_2^2} + \frac{1}{2\pi N} \int v^4 \left| \varphi_{H2} \left( \frac{v}{T_3} \right) \right|^2 |\varphi_e(v)|^{-2} dv, \quad (0.54)$$

and minimize it respect to  $T_2$  to obtain  $\hat{T}_2$ .

5. Replace  $\hat{T}_2$  in the next expression to estimate  $R_2$ ,

$$R_2 = \frac{1}{2\pi} \int v^4 \left| \varphi_{H2} \left( \frac{v}{T_2} \right) \right|^2 \left| \frac{\varphi_y(v)}{\varphi_e(v)} \right|^2 dv. \quad (0.55)$$

6. Finally, replace  $\hat{R}_2$  in

$$MISE(T_N) = -\frac{\mu R_2}{4T_N^4} + \frac{1}{2\pi N} \int \left| \varphi_{H2} \left( \frac{v}{T_3} \right) \right|^2 |\varphi_e(v)|^{-2} dv, \quad (0.56)$$

and minimize it respect to  $T_N$ . This estimated value goes in the limits of the integrals to estimate the characteristics function. Due to we use a fixed number of nodes, find the limit  $T_N$  is equivalent to find the bandwidth.

Delaigle and Gijbels (2002, 2004) treated as know the distribution of the error (s)e. One aspect of the modification made by Bonhomme and Robin (2010) is related with this distribution. They propose replace  $\varphi_e(v)$  by

$$\frac{\varphi_y(\mathbf{v})}{\varphi_\theta(v)}. \quad (0.57)$$

The bold letter were used to distingue the joint characteristic function of all the observed scores from the characteristic function of a single observed score.

In all these expressions  $N$  is sample size,  $H_2$  is the kernel of order 2, and  $\mu$ , under the condition considerers here, is equal to 6.

## Appendix 17: Bonhomme estimator

In order to estimate the marginal distributions of the errors and latent variables we use the estimator developed by Bonhomme and Robin (2010). In this paper they considered:

- $Y = (y_1, y_2, \dots, y_l, \dots, y_L)^T$ , is a vector of L observed random variables with zero mean (Observed scores).
- $X = (x_1, x_2, \dots, x_k, \dots, x_K)^T$ , is a vector of K mutually independent latent variable with zero mean and finite variance (Latent trait and errors).
- $Y = AX$  where  $A$  is a  $L \times K$  matrix of scalar parameters and any two rows are linearly independent.

If we were interested in two measurements

$$y_1 = \theta_1 + e_1 \quad \& \quad y_2 = \theta_1 + e_2,$$

$$Y = (y_1, y_2)^T, \quad X = (\theta, e_1, e_2)^T$$

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

Thus

$$Y = AX$$

Before continuing, it could be useful to give some notation,

- $\mathbf{A}_k$  is the  $k$ th column of the matrix  $\mathbf{A}$ .
- $\varphi_{x_k}(\tau)$  is the characteristics function of  $x_k$ ,  $\varphi_{x_k}(\tau) = E(e^{ix_k\tau})$ .
- $\kappa_{x_k}(\tau) = \ln(\varphi_{x_k}(\tau))$  is the cumulant generating function of  $x_k$ .
- And similarly for the multivariate case  $\kappa_Y(\mathbf{t})$  and  $\varphi_Y(\mathbf{t})$ , where  $\mathbf{t} \in \mathbb{R}^L$ .
- $\partial_l \kappa(\mathbf{t})$ ,  $\partial_{lm}^2 \kappa(\mathbf{t})$  are the first and second order partial derivates with respect to  $t_l$  and  $t_m$ , respectively.  $\nabla \kappa(\mathbf{t})$  is the gradient vector and  $\nabla \nabla^T \kappa(\mathbf{t})$  is the Hessian matrix.

Because the  $X_k$  were assumed independent, the cumulant generating function of  $\mathbf{Y}$  is given by,

$$\kappa_Y(\mathbf{t}) = \sum_{k=1}^K \kappa_{x_k}(\mathbf{t}^T \mathbf{A}_k) \quad (0.58)$$

$$\nabla \kappa_Y(\mathbf{t}) = \sum_{k=1}^K \kappa'_{x_k}(\mathbf{t}^T \mathbf{A}_k) \mathbf{A}_k \quad (0.59)$$

$$\nabla \nabla^T \kappa_Y(\mathbf{t}) = \sum_{k=1}^K \kappa''_{x_k}(\mathbf{t}^T \mathbf{A}_k) \mathbf{A}_k \mathbf{A}_k^T \quad (0.60)$$

Bonhomme and Robin (2010) also called *vech* the matrix operator that acts on symmetric matrices, by selecting only the components below or on the diagonal. For example,

$$M = \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix}$$

$$\text{vech}(M) = (a, b, c, d, e, f)$$

And define a matrix Q as

$$Q = [vech(\mathbf{A}_1 \mathbf{A}_1^T), \dots, vech(\mathbf{A}_K \mathbf{A}_K^T)]$$

with dimensions  $L(L+1)/2 \times K$ .

Using the *vech* operator and the Q matrix, Equation 0.60 can be expressed as

$$\begin{aligned} vech(\nabla \nabla^T \kappa_Y(\mathbf{t})) &= \sum_{k=1}^K \kappa''_{x_k}(\mathbf{t}^T \mathbf{A}_k) vech(\mathbf{A}_k \mathbf{A}_k^T) \\ &= Q \begin{pmatrix} \kappa''_{x_1}(\mathbf{t}^T \mathbf{A}_1) \\ \vdots \\ \kappa''_{x_K}(\mathbf{t}^T \mathbf{A}_K) \end{pmatrix}. \end{aligned}$$

Therefore the cumulant generating function of the latent variables are

$$\begin{pmatrix} \kappa''_{x_1}(\mathbf{t}^T \mathbf{A}_1) \\ \vdots \\ \kappa''_{x_K}(\mathbf{t}^T \mathbf{A}_K) \end{pmatrix} = Q^- vech(\nabla \nabla^T \kappa_Y(\mathbf{t}))$$

where  $Q^- = (Q^T Q)^{-1} Q^T$  is a pseudo inverse of Q.

Note that for any  $\theta \in \mathbb{R}^L \setminus \{0\}$  and  $\tau \in \mathbb{R}$ ,

$$t = \frac{\tau \theta}{\theta^T A_k}.$$

Thus,

$$\kappa''_{x_k}(\tau) = Q_k^- vech \left( \nabla \nabla^T \kappa_Y \left( \frac{\tau \theta}{\theta^T A_k} \right) \right).$$

Then

$$\kappa_{x_k}(\tau) = \int_0^\tau \int_0^v Q_k^- vech \left( \nabla \nabla^T \kappa_Y \left( \frac{u\theta}{\theta^T \mathbf{A}_k} \right) \right) dudv$$

When we refer to the choice of a direction of integration, we are referring to the choice of  $\theta$ , because in principle its arbitrary. But Bonhomme and Robin (2010) proved that the rate of convergence of the estimator depends on the choice of  $\theta$ , and there is a choice which maximize this rate of convergence.

What follows is an example of the estimator when two items (or tests) are used.

Starting form

$$vech(\nabla \nabla^T \kappa_Y(\mathbf{t})) = Q \begin{pmatrix} \kappa''_{x_1}(\mathbf{t}^T \mathbf{A}_1) \\ \vdots \\ \kappa''_{x_K}(\mathbf{t}^T \mathbf{A}_K) \end{pmatrix}$$

$$Q = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

$$vech(\nabla \nabla^T \kappa_Y(\mathbf{t})) = vech \begin{pmatrix} \partial_{11} \kappa_y(t_1, t_2) & \partial_{12} \kappa_y(t_1, t_2) \\ \partial_{21} \kappa_y(t_1, t_2) & \partial_{22} \kappa_y(t_1, t_2), \end{pmatrix}$$

Hence

$$\begin{pmatrix} \partial_{11} \kappa_y(t_1, t_2) \\ \partial_{12} \kappa_y(t_1, t_2) \\ \partial_{22} \kappa_y(t_1, t_2) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \kappa''_\theta(t_1, t_2) \\ \kappa''_{e_1}(t_1) \\ \kappa''_{e_2}(t_2) \end{pmatrix}$$

And after integrate,

$$\begin{aligned}
\kappa_\theta &= \int_0^\tau \int_o^u \partial_{12}^2 \kappa_Y \left( \frac{v}{2}, \frac{v}{2} \right) dv du \\
\kappa e_1 &= \kappa_{y_1}(\tau) - \int_0^\tau \partial_2 \kappa_Y(u, 0) du \\
\kappa e_2 &= \kappa_{y_2}(\tau) - \int_0^\tau \partial_1 \kappa_Y(0, u) du
\end{aligned}$$

The constant mentioned regarding Equation 3.4 is one here. It appears after perform the pseudo inverse on the matrix  $Q$ .

## Appendix 18: Code, Plug-In method

```

s_s<- 1000 #sample size
s_r<- 100 #number of samples
nod<- 201 #intervals in the integrals
n_tn<- 30 #Bandwidths tested in the plug-in method

#pa<-c(0.35,0.15,0.15,0.35)
#theta<- replicate(s_r,sample(c(1,2,3,4),s_s,replace=FALSE,prob=pa))
theta<-t(replicate(s_r, rnorm(s_s,0,1)))
u1<- t(replicate(s_r, rnorm(s_s,0,1)))
u2<- t(replicate(s_r,rnorm(s_s, 0,1)))
Y1<- theta+u1
Y2<- theta+u2 #matrix s_r x s_s
#####
ch_ntn<-function(t)
{
tt<-match(0,t[1,])

```

```

Y1_s<-array(,c(s_s,nod,s_r))
for(i in 1:s_r){Y1_s[,,i]=replicate(nod,Y1[i,])}
Y2_s<-array(,c(s_s,nod,s_r))# same of Y1.
for(i in 1:s_r){Y2_s[,,i]=replicate(nod,Y2[i,])}

En<-array(,c(s_s,nod,s_r))
for (i in 1:s_r) {En[,,i]=exp(1i*(Y1_s[,,i]+Y2_s[,,i])*(t/2))}
#derivates
ENXA<-Y1_s*En
ENXB<-Y2_s*En
ENXE<-Y1_s*Y2_s*En
ENXA<-apply(ENXA,c(2,3),mean)
ENXB<-apply(ENXB,c(2,3),mean)
ENXE<-apply(ENXE,c(2,3),mean)
ENXC<-apply(En,c(2,3),mean)
DENX<-((ENXA*ENXB)/(ENXC*ENXC))-(ENXE/ENXC)
rm(En,ENXA,ENXB,ENXE,ENXC,Y1_s,Y2_s)

#first integral

iden<-replicate(s_r,1)
Int1<-array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{Int1[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}

Int1<-ifelse(is.na(Int1),0,Int1)
for(i in 1:(nod-tt)) {Int1[,,i]=Int1[,,i]*DENX}
Int1<-2*Int1
for(i in 1:(nod-tt)) {Int1[(tt+i),,i]=0.5*Int1[(tt+i),,i]}

```

```

Int1<-apply(Int1,c(2,3),sum)
Int1<-(max(t)/(nod-1))*Int1
IntN1<-array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{IntN1[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}

IntN1<-ifelse(is.na(IntN1),0,IntN1)
for(i in 1:(nod-tt)) {IntN1[,,i]=IntN1[,,i]*DENX}
IntN1<-2*IntN1
for(i in 1:((nod-tt))) {IntN1[i,,i]=0.5*IntN1[i,,i]}

IntN1<-apply(IntN1,c(2,3),sum)
IntN1<-(max(t)/(nod-1))*IntN1
Int1<-t(cbind(IntN1,replicate(s_r,0),Int1))
rm(IntN1)

```

#Second integral

```

Int<-array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{Int[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}

Int<-ifelse(is.na(Int),0,Int)
for(i in 1:(nod-tt)) {Int[,,i]=Int[,,i]*Int1}
Int<-2*Int
for(i in 1:(nod-tt)) {Int[(tt+i),,i]=0.5*Int[(tt+i),,i]}

Int<-apply(Int,c(2,3),sum)
Int<-(max(t)/(nod-1))*Int
IntN<-array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{IntN[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}
```

```

IntN<-ifelse(is.na(IntN),0,IntN)

for(i in 1:(nod-tt)) {IntN[,,i]=IntN[,,i]*Int1}

IntN<-2*IntN

for(i in 1:(nod-tt)) {IntN[i,,i]=0.5*IntN[i,,i]}

IntN<-apply(IntN,c(2,3),sum)

IntN<-(max(t)/(nod-1))*IntN

Int<-t(cbind(IntN,replicate(s_r,0),Int))

rm(IntN,Int1)

#Characteristics function

ch<- exp(Int)

rm(Int)

ch[Mod(ch)>1.1]=0

tau<- t[1,]

Krnlp<- (1-(tau/max(tau))^2)^3

Krnlp<- replicate(s_r, Krnlp)

ch<- ch*Krnlp

ch<- apply(ch,1,mean)

return(ch)

}

t<- array(,c(s_s,nod,n_tn))

for(i in 1:n_tn)

{t[,,i]=t(replicate(s_s, seq(-1.2*(i-0.5),1.2*(i-0.5), length.out=nod)))}

for (i in 1:n_tn){

t[abs(t[,,i])==min(abs(t[1,,i]))]<-0

}

```

```

n_tn_ch<- apply(t,3,ch_ntn)
n_tn_ch[Mod(n_tn_ch)>1.1]=0

#first step

der<- numeric(length(n_tn_ch[,n_tn])-1)
delta<- t[1,3,(n_tn)]-t[1,2,(n_tn)]
for(i in 1:length(der)){
  der[i]=n_tn_ch[i+1,(n_tn)]-n_tn_ch[i,(n_tn)]
}
der<-der/delta
der2<- numeric(length(der)-1)
for(i in 1:length(der2)){
  der2[i]= der[i+1]-der[i]
}
der2<- der2/delta
tt<-numeric(n_tn)
for (i in 1:n_tn)tt[i]=match(0,t[1,,i])
variance<- Mod(der2[tt[(n_tn)]]-1)
R_4<- (factorial(8))/((2^9)*(sqrt(variance)^9)*factorial(4)*sqrt(pi))

#second step

Y1_s<- array(,c(s_s,nod,n_tn))
for(i in 1:n_tn){Y1_s[,,i]=replicate(nod,t(Y1)[,1])}
Y2_s<- array(,c(s_s,nod,n_tn))
for(i in 1:n_tn){Y2_s[,,i]=replicate(nod,t(Y2)[,1])}
chY<- array(,c(s_s,nod,n_tn))
for (i in 1:n_tn) {chY[,,i]=exp(1i*((Y1_s[,,i]+Y2_s[,,i])/2)*t[,,i])}

```

```

chY<- apply(chY,c(2,3),mean)
chY[Mod(chY)>1.1]=0
chu<- chY/n_tn_ch
chu[Mod(chu)>1.1]=1
chu<- (Mod(chu))^{(-2)}
Krnlp<- matrix(,nod,n_tn)
for (i in 1:n_tn) {Krnlp[,i]=(1-(t[1,,i]/max(t[1,,i]))^2)^3}
Krnlp<- (Mod(Krnlp))^2
v<- matrix(,nod,n_tn)
for (i in 1:n_tn){v[,i]=t(t[1,,i])}
v<- v^6
int_s2<- v*Krnlp*chu
int_s2<- (1/(2*pi*s_s))*apply(int_s2,2,sum)
delta<-numeric(n_tn)
for (i in 1:n_tn){ delta[i]=t[1,2,i]-t[1,1,i]}
int_s2<-int_s2*delta
T_3<- numeric(n_tn)
for (i in 1:n_tn) {T_3[i]=t[1,nod,i] }
a<- (6*R_4)/T_3^2
a_2 <- abs(int_s2-a)
min<-which(a_2==min(a_2))
T_3<- T_3[min]

#third step
chy1<- matrix(s_s,nod)
for (i in 1:nod) {chy1[,i]=exp(1i*(Y1_s[,i,1])*t[,i,min])}
chy1<- apply(chy1,2,mean)
chy1[Mod(chy1)>1.1]=0

```

```

chy1<- Mod(chy1)
chy1<- chy1^2
chu1<- chu[,min]
chy1<- chy1*chu1
Krn1P_t<- Krn1P[,min]
v<- numeric(nod)
for (i in 1:nod){v[i]=t[1,i,min])}
v<- v^6
delta<- t[1,2,min]-t[1,1,min]
R_3<- v*Krn1P_t*chy1
R_3<- ifelse(is.na(R_3),0,R_3)
R_3<- delta*sum(R_3)
R_3

#quarter step
v<- matrix(,nod,n_tn)
for (i in 1:n_tn){v[,i]=t[1,,i])}
v<- v^4
int_s3<- v*Krn1P*chu
int_s3<- (1/(2*pi*s_s))*apply(int_s3,2,sum)
delta<-numeric(n_tn)
for (i in 1:n_tn){ delta[i]=t[1,2,i]-t[1,1,i]}
int_s3<-int_s3*delta
T_2<- numeric(n_tn)
for (i in 1:n_tn) {T_2[i]=t[1,nod,i] }
a<- (6*R_3)/T_2^2
a_2<- abs(int_s3-a)
min<-which(a_2==min(a_2))

```

```

T_2<-T_2[min]

#Fifth step
chy1<- matrix(,s_s,nod)
for (i in 1:nod) {chy1[,i]=exp(1i*(Y1_s[,i,1])*t[,i,min])}
chy1<- apply(chy1,2,mean)
chy1[Mod(chy1)>1.1]=0
chy1<-Mod(chy1)
chy1<-chy1^2
chu1<-chu[,min]
chy1<-chy1*chu1
Krn1P_t<-Krn1P[,min]
v<- numeric(nod)
for (i in 1:nod){v[i]=t(t[1,i,min])}
v<- v^4
delta<-t[1,2,min]-t[1,1,min]
R_2<-v*Krn1P_t*chy1
R_2<-ifelse(is.na(R_2),0,R_2)
R_2<-delta*sum(R_2)

#sixth step
int_s6<- Krn1P*chu
int_s6<- (1/(2*pi*s_s))*apply(int_s6,2,sum)
delta<-numeric(n_tn)
for (i in 1:n_tn){ delta[i]=t[1,2,i]-t[1,1,i]}
int_s6<-int_s6*delta
TN<- numeric(n_tn)
for (i in 1:n_tn) {TN[i]=t[1,nod,i] }

```

```

a<- (36*R_2)/(4*(TN^2))
a_2<- abs(int_s6-a)
min<-which(a_2==min(a_2))
TN<-TN[min]#this is the limit in the integral

sub<-which(t[1,nod,]==TN)

rm(Y1_s,Y2_s,n_tn_ch,chY,chu,chy1,chu1,KrnlP,KrnlP_t,v,int_s2,
int_s3,int_s6,R_2,R_3,R_4,T_2,T_3,a,a_2,min,der,der2)

```

## Appendix 19: Code, density estimation and graphs

```

t<-t[,sub]#sub was obtained from plug-in code.
tt<-match(0,t[1,])
Y1_s<-array(c(s_s,nod,s_r))
for(i in 1:s_r){Y1_s[,i]=replicate(nod,Y1[i,])}
Y2_s<-array(c(s_s,nod,s_r))
for(i in 1:s_r){Y2_s[,i]=replicate(nod,Y2[i,])}

En<-array(c(s_s,nod,s_r))
for (i in 1:s_r) {En[,i]=exp(1i*(Y1_s[,i]+Y2_s[,i])*(t/2))}
#derivates
ENXA<-Y1_s*En
ENXB<-Y2_s*En
ENXE<-Y1_s*Y2_s*En
ENXA<-apply(ENXA,c(2,3),mean)
ENXB<-apply(ENXB,c(2,3),mean)

```

```

ENXE<-apply(ENXE,c(2,3),mean)
ENXC<-apply(En,c(2,3),mean)
DENX<-((ENXA*ENXB)/(ENXC*ENXC))-(ENXE/ENXC)
rm(En,ENXA,ENXB,ENXE,ENXC,Y1_s,Y2_s)

#first integral

iden<-replicate(s_r,1)
Int1<-array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{Int1[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}
Int1<-ifelse(is.na(Int1),0,Int1)
for(i in 1:(nod-tt)) {Int1[,,i]=Int1[,,i]*DENX}
Int1<-2*Int1
for(i in 1:(nod-tt)) {Int1[(tt+i),,i]=0.5*Int1[(tt+i),,i]}
Int1<-apply(Int1,c(2,3),sum)
Int1<-(max(t)/(nod-1))*Int1
IntN1<-array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{IntN1[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}
IntN1<-ifelse(is.na(IntN1),0,IntN1)
for(i in 1:(nod-tt)) {IntN1[,,i]=IntN1[,,i]*DENX}
IntN1<-2*IntN1
for(i in 1:((nod-tt))) {IntN1[i,,i]=0.5*IntN1[i,,i]}
IntN1<-apply(IntN1,c(2,3),sum)
IntN1<-(max(t)/(nod-1))*IntN1
Int1<-t(cbind(IntN1,replicate(s_r,0),Int1))
rm(IntN1)

```

```

#Second integral

Int<-array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{Int[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}
Int<-ifelse(is.na(Int),0,Int)
for(i in 1:(nod-tt)) {Int[,,i]=Int[,,i]*Int1}
Int<-2*Int
for(i in 1:(nod-tt)) {Int[(tt+i),,i]=0.5*Int[(tt+i),,i]}
Int<-apply(Int,c(2,3),sum)
Int<-(max(t)/(nod-1))*Int
IntN<-array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{IntN[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}
IntN<-ifelse(is.na(IntN),0,IntN)
for(i in 1:(nod-tt)) {IntN[,,i]=IntN[,,i]*Int1}
IntN<-2*IntN
for(i in 1:((nod-tt))) {IntN[i,,i]=0.5*IntN[i,,i]}
IntN<-apply(IntN,c(2,3),sum)
IntN<-(max(t)/(nod-1))*IntN
Int<-t(cbind(IntN,replicate(s_r,0),Int))
rm(IntN,Int1)

#Characteristics function

ch=exp(Int)
rm(Int)

```

```

ch[Mod(ch)>1.1]=0
tau=t[,]
Krn1P=(1-(tau/max(tau))^2)^3
Krn1P=replicate(s_r, Krn1P)
ch=ch*Krn1P

#Fourier Transform

invf=array (,c(nod,nod,s_r))
for(i in 1:s_r){invf[,i]=replicate(nod,ch[,i])}
x=t(replicate(nod, seq(min(t),max(t), length.out=nod)))
t_1=t(replicate(nod, seq(min(t),max(t), length.out=nod)))
invf2=array(,c(nod,nod,s_r))
for(i in 1:s_r){invf2[,i]=invf[,i]*exp(-1i*x*t(t_1))}
invf2=2*invf2
invf=(max(t)/(2*(nod-1)*pi))*apply(invft,c(2,3), sum)
rm(invft)

xp=seq(min(x),max(x), length.out=nod)
edensity=apply(invft,1,mean)
den<-density(theta)#if we could observe theta
sd=apply(Mod(invft),1,sd)
error=qnorm(0.975)*sd/sqrt(s_r)
Upper=Mod(edensity)+error
Lower=Mod(edensity)-error

#real density
a<-dnorm(xp,0,1)

```

```

#graphs
quartz()
par(mfrow = c(2, 1))
plot(xp, Mod(edensity), type="l", col="blue",
      main= "Standard Normal", xlim=c(-6,6), ylim=c(0,0.80), xlab="theta",
      ylab="Densities")
lines(xp-1,a, col="black")
lines(xp,Upper, col="green")
lines(xp,Lower, col="green")
plot(den, type="l", col="blue", main="Standard Normal",
      xlim=c(-6,6), ylim=c(0,0.80), xlab="theta", ylab="Densities")
lines(xp-1,a, col="black")

```

## Appendix 20: Code, MISE

```

dif<-which(a==max(a))-which(Mod(edensity)==max(Mod(edensity)))
ise=matrix(, (nod-dif), s_r)
for (i in 1:(nod-dif)){
  for (j in 1:s_r){
    ise[i,j]=(Mod(invf[i,j])-a3[i+dif])^2
  }
}
delta=xp3[2]-xp3[1]

ise2=numeric(s_r)
for (i in 1:s_r){
  ise2[i]=(ise[1,i]+ise[s_r,i])/2
}

```

```
MISE=mean(delta*apply(ise,2,sum)-ise2)
```

## Appendix 21: Code, distributions of the errors in one item with real datas

```
data<-read.csv("Libro",header=TRUE,sep=",")  
data$X=NULL  
  
s_s<- 1000 #sample size (all sample have the same size)800  
s_r<- 100 #number of samples70  
nod<- 201 #amount of intervals in the integrals  
t=t(replicate(s_s, seq(-23,23, length.out=nod)))  
  
  
  
Y1<- t(replicate(s_r, sample(data[, "Item1"] , s_s, replace=FALSE)))  
Y1<- Y1-apply(Y1,1,mean)  
Y2<- t(replicate(s_r, sample(data[, "Item2"] , s_s, replace=FALSE)))  
Y2<- Y2-apply(Y2,1,mean)  
Y3<- t(replicate(s_r, sample(data[, "Item3"] , s_s, replace=FALSE)))  
Y3<- Y3-apply(Y3,1,mean)  
Y4<- t(replicate(s_r, sample(data[, "Item4"] , s_s, replace=FALSE)))  
Y4<- Y4-apply(Y4,1,mean)  
Y5<- t(replicate(s_r, sample(data[, "Item5"] , s_s, replace=FALSE)))  
Y5<- Y5-apply(Y5,1,mean)  
Y6<- t(replicate(s_r, sample(data[, "Item6"] , s_s, replace=FALSE)))  
Y6<- Y6-apply(Y6,1,mean)  
  
tt<-match(0,t[1,])  
Y1_s<-array(,c(s_s,nod,s_r))
```

```

for(i in 1:s_r){Y1_s[,,i]=replicate(nod,Y1[i,])}
Y2_s<-array(,c(s_s,nod,s_r))
for(i in 1:s_r){Y2_s[,,i]=replicate(nod,Y2[i,])}
Y3_s<-array(,c(s_s,nod,s_r))
for(i in 1:s_r){Y3_s[,,i]=replicate(nod,Y3[i,])}
Y4_s<-array(,c(s_s,nod,s_r))
for(i in 1:s_r){Y4_s[,,i]=replicate(nod,Y4[i,])}
Y5_s<-array(,c(s_s,nod,s_r))
for(i in 1:s_r){Y5_s[,,i]=replicate(nod,Y5[i,])}
Y6_s<-array(,c(s_s,nod,s_r))
for(i in 1:s_r){Y6_s[,,i]=replicate(nod,Y6[i,])}

EnU6=array(,c(s_s,nod,s_r))
for (i in 1:s_r) {EnU6[,,i]=exp(1i*Y6_s[,,i]*t)}
EnU5=array(,c(s_s,nod,s_r))
for (i in 1:s_r) {EnU5[,,i]=exp(1i*Y5_s[,,i]*t)}
EnU4=array(,c(s_s,nod,s_r))
for (i in 1:s_r) {EnU4[,,i]=exp(1i*Y4_s[,,i]*t)}
EnU3=array(,c(s_s,nod,s_r))
for (i in 1:s_r) {EnU3[,,i]=exp(1i*Y3_s[,,i]*t)}
EnU2=array(,c(s_s,nod,s_r))
for (i in 1:s_r) {EnU2[,,i]=exp(1i*Y2_s[,,i]*t)}
EnU1=array(,c(s_s,nod,s_r))
for (i in 1:s_r) {EnU1[,,i]=exp(1i*Y1_s[,,i]*t)}
#first order derivate
ENU6=log(apply(EnU6,c(2,3),mean))
ENU6=log(apply(EnU5,c(2,3),mean))
ENU5=log(apply(EnU5,c(2,3),mean))

```

```

ENU4=log(apply(EnU4,c(2,3),mean))
ENU3=log(apply(EnU3,c(2,3),mean))
ENU2=log(apply(EnU2,c(2,3),mean))
ENU1=log(apply(EnU1,c(2,3),mean))
ENU21<- Y2_s*EnU1
ENU21<- apply(ENU21,c(2,3),mean)
ENU21<- 1i*(ENU21/exp(ENU1))
ENU23<- Y3_s*EnU1
ENU23<- apply(ENU23,c(2,3),mean)
ENU23<- 1i*(ENU23/exp(ENU1))
ENU24<- Y4_s*EnU1
ENU24<- apply(ENU24,c(2,3),mean)
ENU24<- 1i*(ENU24/exp(ENU1))
ENU25<- Y5_s*EnU1
ENU25<- apply(ENU25,c(2,3),mean)
ENU25<- 1i*(ENU25/exp(ENU1))
ENU26<- Y6_s*EnU1
ENU26<- apply(ENU26,c(2,3),mean)
ENU26<- 1i*(ENU26/exp(ENU1))
rm(EnU2,EnU3,EnU4,EnU5,EnU6)

```

```

#the second order derivates
ENXA213<- Y2_s*EnU1
ENXB213<- Y3_s*EnU1
ENXE213<- Y2_s*Y3_s*EnU1
ENXA213<- apply(ENXA213,c(2,3),mean)
ENXB213<- apply(ENXB213,c(2,3),mean)
ENXE213<- apply(ENXE213,c(2,3),mean)

```

```
ENXC<- apply(EnU1,c(2,3),mean)
DENX213<- ((ENXA213*ENXB213)/(ENXC*ENXC))-(ENXE213/ENXC)
rm(ENXA213,ENXB213,ENXE213)
```

```
ENXA214<- Y2_s*EnU1
ENXB214<- Y4_s*EnU1
ENXE214<- Y2_s*Y4_s*EnU1
ENXA214<- apply(ENXA214,c(2,3),mean)
ENXB214<- apply(ENXB214,c(2,3),mean)
ENXE214<- apply(ENXE214,c(2,3),mean)
DENX214<- ((ENXA214*ENXB214)/(ENXC*ENXC))-(ENXE214/ENXC)
rm(ENXA214,ENXB214,ENXE214)
```

```
ENXA215<- Y2_s*EnU1
ENXB215<- Y5_s*EnU1
ENXE215<- Y2_s*Y5_s*EnU1
ENXA215<- apply(ENXA215,c(2,3),mean)
ENXB215<- apply(ENXB215,c(2,3),mean)
ENXE215<- apply(ENXE215,c(2,3),mean)
DENX215<- ((ENXA215*ENXB215)/(ENXC*ENXC))-(ENXE215/ENXC)
rm(ENXA215,ENXB215,ENXE215)
```

```
ENXA216<- Y2_s*EnU1
ENXB216<- Y6_s*EnU1
ENXE216<- Y2_s*Y6_s*EnU1
ENXA216<- apply(ENXA216,c(2,3),mean)
ENXB216<- apply(ENXB216,c(2,3),mean)
ENXE216<- apply(ENXE216,c(2,3),mean)
```

```
DENX216<- ((ENXA216*ENXB216)/(ENXC*ENXC))-(ENXE216/ENXC)
rm(ENXA216,ENXB216,ENXE216)
```

```
ENXA234<- Y3_s*EnU1
ENXB234<- Y4_s*EnU1
ENXE234<- Y3_s*Y1_s*EnU1
ENXA234<- apply(ENXA234,c(2,3),mean)
ENXB234<- apply(ENXB234,c(2,3),mean)
ENXE234<- apply(ENXE234,c(2,3),mean)
DENX234<- ((ENXA234*ENXB234)/(ENXC*ENXC))-(ENXE234/ENXC)
rm(ENXA234,ENXB234,ENXE234)
```

```
ENXA235<- Y3_s*EnU1
ENXB235<- Y5_s*EnU1
ENXE235<- Y3_s*Y5_s*EnU1
ENXA235<- apply(ENXA235,c(2,3),mean)
ENXB235<- apply(ENXB235,c(2,3),mean)
ENXE235<- apply(ENXE235,c(2,3),mean)
DENX235<- ((ENXA235*ENXB235)/(ENXC*ENXC))-(ENXE235/ENXC)
rm(ENXA235,ENXB235,ENXE235)
```

```
ENXA236<- Y3_s*EnU1
ENXB236<- Y6_s*EnU1
ENXE236<- Y3_s*Y6_s*EnU1
ENXA236<- apply(ENXA236,c(2,3),mean)
ENXB236<- apply(ENXB236,c(2,3),mean)
ENXE236<- apply(ENXE236,c(2,3),mean)
DENX236<- ((ENXA236*ENXB236)/(ENXC*ENXC))-(ENXE236/ENXC)
```

```
rm(ENXA236,ENXB236,ENXE236)
```

```
ENXA245<- Y4_s*EnU1  
ENXB245<- Y5_s*EnU1  
ENXE245<- Y4_s*Y5_s*EnU1  
ENXA245<- apply(ENXA245,c(2,3),mean)  
ENXB245<- apply(ENXB245,c(2,3),mean)  
ENXE245<- apply(ENXE245,c(2,3),mean)  
DENX245<- ((ENXA245*ENXB245)/(ENXC*ENXC))-(ENXE245/ENXC)  
rm(ENXA245,ENXB245,ENXE245)
```

```
ENXA246<- Y4_s*EnU1  
ENXB246<- Y6_s*EnU1  
ENXE246<- Y4_s*Y6_s*EnU1  
ENXA246<- apply(ENXA246,c(2,3),mean)  
ENXB246<- apply(ENXB246,c(2,3),mean)  
ENXE246<- apply(ENXE246,c(2,3),mean)  
DENX246<- ((ENXA246*ENXB246)/(ENXC*ENXC))-(ENXE246/ENXC)  
rm(ENXA246,ENXB246,ENXE246)
```

```
ENXA256<- Y5_s*EnU1  
ENXB256<- Y6_s*EnU1  
ENXE256<- Y5_s*Y6_s*EnU1  
ENXA256<- apply(ENXA256,c(2,3),mean)  
ENXB256<- apply(ENXB256,c(2,3),mean)  
ENXE256<- apply(ENXE256,c(2,3),mean)  
DENX256<- ((ENXA256*ENXB256)/(ENXC*ENXC))-(ENXE256/ENXC)  
rm(ENXA256,ENXB256,ENXE256)
```

```

#integrals

iden=replicate(s_r,1)

Int1=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{Int1[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}

Int1=ifelse(is.na(Int1),0,Int1)

for(i in 1:(nod-tt)) {Int1[,,i]=Int1[,,i]*ENU21}

Int1=2*Int1

for(i in 1:(nod-tt)) {Int1[(tt+i),,i]=0.5*Int1[(tt+i),,i]}

Int3=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{Int3[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}

Int3=ifelse(is.na(Int3),0,Int3)

for(i in 1:(nod-tt)) {Int3[,,i]=Int3[,,i]*ENU23}

Int3=2*Int3

for(i in 1:(nod-tt)) {Int3[(tt+i),,i]=0.5*Int3[(tt+i),,i]}

Int4=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{Int4[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}

Int4=ifelse(is.na(Int4),0,Int4)

for(i in 1:(nod-tt)) {Int4[,,i]=Int4[,,i]*ENU24}

Int4=2*Int4

for(i in 1:(nod-tt)) {Int4[(tt+i),,i]=0.5*Int4[(tt+i),,i]}

Int5=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{Int5[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}

Int5=ifelse(is.na(Int5),0,Int5)

```

```

for(i in 1:(nod-tt)) {Int5[, , i]=Int5[, , i]*ENU25}

Int5=2*Int5

for(i in 1:(nod-tt)) {Int5[(tt+i), , i]=0.5*Int5[(tt+i), , i]}

Int6=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{Int6[((nod-tt)+2):((nod-tt)+1+i), , i]=t(replicate(i, iden))}

Int6=ifelse(is.na(Int6),0,Int6)

for(i in 1:(nod-tt)) {Int6[, , i]=Int6[, , i]*ENU26}

Int6=2*Int6

for(i in 1:(nod-tt)) {Int6[(tt+i), , i]=0.5*Int6[(tt+i), , i]}

Inte1=apply(Int1,c(2,3),sum)

Inte1=(max(t)/(nod-1))*Inte1

dim(Inte1)

Inte3=apply(Int3,c(2,3),sum)

Inte3=(max(t)/(nod-1))*Inte3

Inte4=apply(Int4,c(2,3),sum)

Inte4=(max(t)/(nod-1))*Inte4

Inte5=apply(Int5,c(2,3),sum)

Inte5=(max(t)/(nod-1))*Inte5

Inte6=apply(Int6,c(2,3),sum)

Inte6=(max(t)/(nod-1))*Inte6

IntN1=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{IntN1[(nod-tt):i, , i]=t(replicate((nod-tt+1)-i, iden))}

IntN1=ifelse(is.na(IntN1),0,IntN1)

for(i in 1:(nod-tt)) {IntN1[, , i]=IntN1[, , i]*ENU21}

```

```

IntN1=2*IntN1

for(i in 1:(nod-tt)) {IntN1[i,,i]=0.5*IntN1[i,,i]}

IntN3=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{IntN3[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}

IntN3=ifelse(is.na(IntN3),0,IntN3)

for(i in 1:(nod-tt)) {IntN3[,,i]=IntN3[,,i]*ENU23}

IntN3=2*IntN3

for(i in 1:(nod-tt)) {IntN3[i,,i]=0.5*IntN3[i,,i]}

IntN4=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{IntN4[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}

IntN4=ifelse(is.na(IntN4),0,IntN4)

for(i in 1:(nod-tt)) {IntN4[,,i]=IntN4[,,i]*ENU24}

IntN4=2*IntN4

for(i in 1:(nod-tt)) {IntN4[i,,i]=0.5*IntN4[i,,i]}

IntN5=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{IntN5[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}

IntN5=ifelse(is.na(IntN5),0,IntN5)

for(i in 1:(nod-tt)) {IntN5[,,i]=IntN5[,,i]*ENU25}

IntN5=2*IntN5

for(i in 1:(nod-tt)) {IntN5[i,,i]=0.5*IntN5[i,,i]}

IntN6=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{IntN6[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}

IntN6=ifelse(is.na(IntN6),0,IntN6)

for(i in 1:(nod-tt)) {IntN6[,,i]=IntN6[,,i]*ENU26}

```

```

IntN6=2*IntN6

for(i in 1:(nod-tt)) {IntN6[i,,i]=0.5*IntN6[i,,i]}

InteN1=apply(IntN1,c(2,3),sum)
InteN1=(max(t)/(nod-1))*InteN1
InteN3=apply(IntN3,c(2,3),sum)
InteN3=(max(t)/(nod-1))*InteN3
InteN4=apply(IntN4,c(2,3),sum)
InteN4=(max(t)/(nod-1))*InteN4
InteN5=apply(IntN5,c(2,3),sum)
InteN5=(max(t)/(nod-1))*InteN5
InteN6=apply(IntN6,c(2,3),sum)
InteN6=(max(t)/(nod-1))*InteN6

Int1=t(cbind(InteN1,replicate(s_r,0),Inte1))
Int3=t(cbind(InteN3,replicate(s_r,0),Inte3))
Int4=t(cbind(InteN4,replicate(s_r,0),Inte4))
Int5=t(cbind(InteN5,replicate(s_r,0),Inte5))
Int6=t(cbind(InteN6,replicate(s_r,0),Inte6))

rm(InteN1,InteN3,InteN4,InteN5,InteN6,IntN1,IntN3,IntN4,IntN5,
IntN6,Inte1,Inte3,Inte4, Inte5,Inte6)

#double integrals
Int113=array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{Int113[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}
Int113=ifelse(is.na(Int113),0,Int113)

```

```

for(i in 1:(nod-tt)) {Int113[, , i]=Int113[, , i]*DENX213}

Int113=2*Int113

for(i in 1:(nod-tt)) {Int113[(tt+i), , i]=0.5*Int113[(tt+i), , i]}

Inte113=apply(Int113,c(2,3),sum)

Inte113=(max(t)/(nod-1))*Inte113

Int114=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{Int114[((nod-tt)+2):((nod-tt)+1+i), , i]=t(replicate(i, iden))}

Int114=ifelse(is.na(Int114),0,Int114)

for(i in 1:(nod-tt)) {Int114[, , i]=Int114[, , i]*DENX214}

Int114=2*Int114

for(i in 1:(nod-tt)) {Int114[(tt+i), , i]=0.5*Int114[(tt+i), , i]}

Inte114=apply(Int114,c(2,3),sum)

Inte114=(max(t)/(nod-1))*Inte114

Int115=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{Int115[((nod-tt)+2):((nod-tt)+1+i), , i]=t(replicate(i, iden))}

Int115=ifelse(is.na(Int115),0,Int115)

for(i in 1:(nod-tt)) {Int115[, , i]=Int115[, , i]*DENX215}

Int115=2*Int115

for(i in 1:(nod-tt)) {Int115[(tt+i), , i]=0.5*Int115[(tt+i), , i]}

Inte115=apply(Int115,c(2,3),sum)

Inte115=(max(t)/(nod-1))*Inte115

Int116=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{Int116[((nod-tt)+2):((nod-tt)+1+i), , i]=t(replicate(i, iden))}

Int116=ifelse(is.na(Int116),0,Int116)

for(i in 1:(nod-tt)) {Int116[, , i]=Int116[, , i]*DENX216}

```

```

Int116=2*Int116

for(i in 1:(nod-tt)) {Int116[(tt+i),,i]=0.5*Int116[(tt+i),,i]}

Inte116=apply(Int116,c(2,3),sum)
Inte116=(max(t)/(nod-1))*Inte116

Int134=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{Int134[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}

Int134=ifelse(is.na(Int134),0,Int134)

for(i in 1:(nod-tt)) {Int134[,i]=Int134[,i]*DENX234}

Int134=2*Int134

for(i in 1:(nod-tt)) {Int134[(tt+i),,i]=0.5*Int134[(tt+i),,i]}

Inte134=apply(Int134,c(2,3),sum)
Inte134=(max(t)/(nod-1))*Inte134

Int135=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{Int135[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}

Int135=ifelse(is.na(Int135),0,Int135)

for(i in 1:(nod-tt)) {Int135[,i]=Int135[,i]*DENX235}

Int135=2*Int135

for(i in 1:(nod-tt)) {Int135[(tt+i),,i]=0.5*Int135[(tt+i),,i]}

Inte135=apply(Int135,c(2,3),sum)
Inte135=(max(t)/(nod-1))*Inte135

Int136=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{Int136[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}

Int136=ifelse(is.na(Int136),0,Int136)

for(i in 1:(nod-tt)) {Int136[,i]=Int136[,i]*DENX236}

Int136=2*Int136

```

```

for(i in 1:(nod-tt)) {Int136[(tt+i),,i]=0.5*Int136[(tt+i),,i]}

Inte136=apply(Int136,c(2,3),sum)

Inte136=(max(t)/(nod-1))*Inte136

Int145=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{Int145[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}

Int145=ifelse(is.na(Int145),0,Int145)

for(i in 1:(nod-tt)) {Int145[,,i]=Int145[,,i]*DENX245}

Int145=2*Int145

for(i in 1:(nod-tt)) {Int145[(tt+i),,i]=0.5*Int145[(tt+i),,i]}

Inte145=apply(Int145,c(2,3),sum)

Inte145=(max(t)/(nod-1))*Inte145

Int146=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{Int146[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}

Int146=ifelse(is.na(Int146),0,Int146)

for(i in 1:(nod-tt)) {Int146[,,i]=Int146[,,i]*DENX246}

Int146=2*Int146

for(i in 1:(nod-tt)) {Int146[(tt+i),,i]=0.5*Int146[(tt+i),,i]}

Inte146=apply(Int146,c(2,3),sum)

Inte146=(max(t)/(nod-1))*Inte146

Int156=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{Int156[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}

Int156=ifelse(is.na(Int156),0,Int156)

for(i in 1:(nod-tt)) {Int156[,,i]=Int156[,,i]*DENX256}

Int156=2*Int156

for(i in 1:(nod-tt)) {Int156[(tt+i),,i]=0.5*Int156[(tt+i),,i]}

```

```

Inte156=apply(Int156,c(2,3),sum)
Inte156=(max(t)/(nod-1))*Inte156

IntN113=array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{IntN113[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}
IntN113=ifelse(is.na(IntN113),0,IntN113)
for(i in 1:(nod-tt)) {IntN113[,,i]=IntN113[,,i]*DENX213}
IntN113=2*IntN113

for(i in 1:((nod-tt))) {IntN113[i,,i]=0.5*IntN113[i,,i]}
InteN113=apply(IntN113,c(2,3),sum)
InteN113=(max(t)/(nod-1))*InteN113

IntN114=array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{IntN114[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}
IntN114=ifelse(is.na(IntN114),0,IntN114)
for(i in 1:(nod-tt)) {IntN114[,,i]=IntN114[,,i]*DENX214}
IntN114=2*IntN114

for(i in 1:((nod-tt))) {IntN114[i,,i]=0.5*IntN114[i,,i]}
InteN114=apply(IntN114,c(2,3),sum)
InteN114=(max(t)/(nod-1))*InteN114

IntN115=array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{IntN115[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}
IntN115=ifelse(is.na(IntN115),0,IntN115)
for(i in 1:(nod-tt)) {IntN115[,,i]=IntN115[,,i]*DENX215}
IntN115=2*IntN115

for(i in 1:((nod-tt))) {IntN115[i,,i]=0.5*IntN115[i,,i]}

```

```

InteN115=apply(IntN115,c(2,3),sum)
InteN115=(max(t)/(nod-1))*InteN115
IntN116=array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{IntN116[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}
IntN116=ifelse(is.na(IntN116),0,IntN116)
for(i in 1:(nod-tt)) {IntN116[,,i]=IntN116[,,i]*DENX216}
IntN116=2*IntN116
for(i in 1:((nod-tt))) {IntN116[i,,i]=0.5*IntN116[i,,i]}
InteN116=apply(IntN116,c(2,3),sum)
InteN116=(max(t)/(nod-1))*InteN116
IntN134=array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{IntN134[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}
IntN134=ifelse(is.na(IntN134),0,IntN134)
for(i in 1:(nod-tt)) {IntN134[,,i]=IntN134[,,i]*DENX234}
IntN134=2*IntN134
for(i in 1:((nod-tt))) {IntN134[i,,i]=0.5*IntN134[i,,i]}#to
InteN134=apply(IntN134,c(2,3),sum)
InteN134=(max(t)/(nod-1))*InteN134
IntN135=array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{IntN135[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}
IntN135=ifelse(is.na(IntN135),0,IntN135)
for(i in 1:(nod-tt)) {IntN135[,,i]=IntN135[,,i]*DENX235}
IntN135=2*IntN135
for(i in 1:((nod-tt))) {IntN135[i,,i]=0.5*IntN135[i,,i]}#to
InteN135=apply(IntN135,c(2,3),sum)

```

```

InteN135=(max(t)/(nod-1))*InteN135
IntN136=array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{IntN136[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}
IntN136=ifelse(is.na(IntN136),0,IntN136)
for(i in 1:(nod-tt)) {IntN136[,,i]=IntN136[,,i]*DENX236}
IntN136=2*IntN136
for(i in 1:((nod-tt))) {IntN136[i,,i]=0.5*IntN136[i,,i]}#to
InteN136=apply(IntN136,c(2,3),sum)
InteN136=(max(t)/(nod-1))*InteN136
IntN145=array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{IntN145[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}
IntN145=ifelse(is.na(IntN145),0,IntN145)
for(i in 1:(nod-tt)) {IntN145[,,i]=IntN145[,,i]*DENX245}
IntN145=2*IntN145
for(i in 1:((nod-tt))) {IntN145[i,,i]=0.5*IntN145[i,,i]}#to
InteN145=apply(IntN145,c(2,3),sum)
InteN145=(max(t)/(nod-1))*InteN145
IntN146=array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{IntN146[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}
IntN146=ifelse(is.na(IntN146),0,IntN146)
for(i in 1:(nod-tt)) {IntN146[,,i]=IntN146[,,i]*DENX246}
IntN146=2*IntN146
for(i in 1:((nod-tt))) {IntN146[i,,i]=0.5*IntN146[i,,i]}#to
InteN146=apply(IntN146,c(2,3),sum)
InteN146=(max(t)/(nod-1))*InteN146

```

```

IntN156=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{IntN156[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}

IntN156=ifelse(is.na(IntN156),0,IntN156)

for(i in 1:(nod-tt)) {IntN156[,,i]=IntN156[,,i]*DENX256}

IntN156=2*IntN156

for(i in 1:((nod-tt))) {IntN156[i,,i]=0.5*IntN156[i,,i]}#to

InteN156=apply(IntN156,c(2,3),sum)

InteN156=(max(t)/(nod-1))*InteN156

Int113=t(cbind(InteN113,replicate(s_r,0),Inte113))
Int114=t(cbind(InteN114,replicate(s_r,0),Inte114))
Int115=t(cbind(InteN115,replicate(s_r,0),Inte115))
Int116=t(cbind(InteN116,replicate(s_r,0),Inte116))
Int134=t(cbind(InteN134,replicate(s_r,0),Inte134))
Int135=t(cbind(InteN135,replicate(s_r,0),Inte135))
Int136=t(cbind(InteN136,replicate(s_r,0),Inte136))
Int145=t(cbind(InteN145,replicate(s_r,0),Inte145))
Int146=t(cbind(InteN146,replicate(s_r,0),Inte146))
Int156=t(cbind(InteN156,replicate(s_r,0),Inte156))

rm(InteN113,InteN114,InteN115,InteN116,InteN134,InteN135,InteN136,
InteN145,InteN146,InteN156,IntN113,IntN114,IntN115,IntN116,IntN134,
IntN135,IntN136,IntN145,IntN146,IntN156,Inte113,Inte114,Inte115,Inte116,
Inte134, Inte135,Inte136,Inte145,Inte146,Inte156)

#Second integral

```

```

Int13=array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{Int13[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}

Int13=ifelse(is.na(Int13),0,Int13)
for(i in 1:(nod-tt)) {Int13[,,i]=Int13[,,i]*Int113}

Int13=2*Int13
for(i in 1:(nod-tt)) {Int13[(tt+i),,i]=0.5*Int13[(tt+i),,i]}

Inte13=apply(Int13,c(2,3),sum)
Inte13=(max(t)/(nod-1))*Inte13

Int14=array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{Int14[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}

Int14=ifelse(is.na(Int14),0,Int14)
for(i in 1:(nod-tt)) {Int14[,,i]=Int14[,,i]*Int114}

Int14=2*Int14
for(i in 1:(nod-tt)) {Int14[(tt+i),,i]=0.5*Int14[(tt+i),,i]}

Inte14=apply(Int14,c(2,3),sum)
Inte14=(max(t)/(nod-1))*Inte14

Int15=array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{Int15[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}

Int15=ifelse(is.na(Int15),0,Int15)
for(i in 1:(nod-tt)) {Int15[,,i]=Int15[,,i]*Int115}

Int15=2*Int15
for(i in 1:(nod-tt)) {Int15[(tt+i),,i]=0.5*Int15[(tt+i),,i]}

Inte15=apply(Int15,c(2,3),sum)
Inte15=(max(t)/(nod-1))*Inte15

Int16=array(,c(nod,s_r,(nod-tt)))

```

```

for (i in 1:(nod-tt))

{Int16[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}

Int16=ifelse(is.na(Int16),0,Int16)

for(i in 1:(nod-tt)) {Int16[,,i]=Int16[,,i]*Int116}

Int16=2*Int16

for(i in 1:(nod-tt)) {Int16[(tt+i),,i]=0.5*Int16[(tt+i),,i]}

Inte16=apply(Int16,c(2,3),sum)

Inte16=(max(t)/(nod-1))*Inte16

Int34=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{Int34[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}

Int34=ifelse(is.na(Int34),0,Int34)

for(i in 1:(nod-tt)) {Int34[,,i]=Int34[,,i]*Int134}

Int34=2*Int34

for(i in 1:(nod-tt)) {Int34[(tt+i),,i]=0.5*Int34[(tt+i),,i]}

Inte34=apply(Int34,c(2,3),sum)

Inte34=(max(t)/(nod-1))*Inte34

Int35=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{Int35[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}

Int35=ifelse(is.na(Int35),0,Int35)

for(i in 1:(nod-tt)) {Int35[,,i]=Int35[,,i]*Int135}

Int35=2*Int35

for(i in 1:(nod-tt)) {Int35[(tt+i),,i]=0.5*Int35[(tt+i),,i]}

Inte35=apply(Int35,c(2,3),sum)

Inte35=(max(t)/(nod-1))*Inte35

Int36=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

```

```

{Int36[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}

Int36=ifelse(is.na(Int36),0,Int36)

for(i in 1:(nod-tt)) {Int36[,,i]=Int36[,,i]*Int136}

Int36=2*Int36

for(i in 1:(nod-tt)) {Int36[(tt+i),,i]=0.5*Int36[(tt+i),,i]}

Inte36=apply(Int36,c(2,3),sum)

Inte36=(max(t)/(nod-1))*Inte36

Int45=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{Int45[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}

Int45=ifelse(is.na(Int45),0,Int45)

for(i in 1:(nod-tt)) {Int45[,,i]=Int45[,,i]*Int145}

Int45=2*Int45

for(i in 1:(nod-tt)) {Int45[(tt+i),,i]=0.5*Int45[(tt+i),,i]}

Inte45=apply(Int45,c(2,3),sum)

Inte45=(max(t)/(nod-1))*Inte45

Int46=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{Int46[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}

Int46=ifelse(is.na(Int46),0,Int46)

for(i in 1:(nod-tt)) {Int46[,,i]=Int46[,,i]*Int146}

Int46=2*Int46

for(i in 1:(nod-tt)) {Int46[(tt+i),,i]=0.5*Int46[(tt+i),,i]}

Inte46=apply(Int46,c(2,3),sum)

Inte46=(max(t)/(nod-1))*Inte46

Int56=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{Int56[((nod-tt)+2):((nod-tt)+1+i),,i]=t(replicate(i, iden))}
```

```

Int56=ifelse(is.na(Int56),0,Int56)
for(i in 1:(nod-tt)) {Int56[,,i]=Int56[,,i]*Int156}
Int56=2*Int56
for(i in 1:(nod-tt)) {Int56[(tt+i),,i]=0.5*Int56[(tt+i),,i]}
Inte56=apply(Int56,c(2,3),sum)
Inte56=(max(t)/(nod-1))*Inte56
IntN13=array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{IntN13[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}
IntN13=ifelse(is.na(IntN13),0,IntN13)
for(i in 1:(nod-tt)) {IntN13[,i]=IntN13[,,i]*Int113}
IntN13=2*IntN13
for(i in 1:((nod-tt))) {IntN13[i,,i]=0.5*IntN13[i,,i]}
InteN13=apply(IntN13,c(2,3),sum)
InteN13=(max(t)/(nod-1))*InteN13
IntN14=array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{IntN14[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}
IntN14=ifelse(is.na(IntN14),0,IntN14)
for(i in 1:(nod-tt)) {IntN14[,i]=IntN14[,,i]*Int114}
IntN14=2*IntN14
for(i in 1:((nod-tt))) {IntN14[i,,i]=0.5*IntN14[i,,i]}
InteN14=apply(IntN14,c(2,3),sum)
InteN14=(max(t)/(nod-1))*InteN14
IntN15=array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{IntN15[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}
IntN15=ifelse(is.na(IntN15),0,IntN15)

```

```

for(i in 1:(nod-tt)) {IntN15[, , i]=IntN15[, , i]*Int115}

IntN15=2*IntN15

for(i in 1:((nod-tt))) {IntN15[i, , i]=0.5*IntN15[i, , i]}

InteN15=apply(IntN15,c(2,3),sum)

InteN15=(max(t)/(nod-1))*InteN15

IntN16=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{IntN16[(nod-tt):i, , i]=t(replicate((nod-tt+1)-i, iden))}

IntN16=ifelse(is.na(IntN16),0,IntN16)

for(i in 1:(nod-tt)) {IntN16[, , i]=IntN16[, , i]*Int116}

IntN16=2*IntN16

for(i in 1:((nod-tt))) {IntN16[i, , i]=0.5*IntN16[i, , i]}

InteN16=apply(IntN16,c(2,3),sum)

InteN16=(max(t)/(nod-1))*InteN16

IntN34=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{IntN34[(nod-tt):i, , i]=t(replicate((nod-tt+1)-i, iden))}

IntN34=ifelse(is.na(IntN34),0,IntN34)

for(i in 1:(nod-tt)) {IntN34[, , i]=IntN34[, , i]*Int134}

IntN34=2*IntN34

for(i in 1:((nod-tt))) {IntN34[i, , i]=0.5*IntN34[i, , i]}

InteN34=apply(IntN34,c(2,3),sum)

InteN34=(max(t)/(nod-1))*InteN34

IntN35=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{IntN35[(nod-tt):i, , i]=t(replicate((nod-tt+1)-i, iden))}

IntN35=ifelse(is.na(IntN35),0,IntN35)

for(i in 1:(nod-tt)) {IntN35[, , i]=IntN35[, , i]*Int135}

```

```

IntN35=2*IntN35

for(i in 1:(nod-tt)) {IntN35[i,,i]=0.5*IntN35[i,,i]}

InteN35=apply(IntN35,c(2,3),sum)

InteN35=(max(t)/(nod-1))*InteN35

IntN36=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{IntN36[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}

IntN36=ifelse(is.na(IntN36),0,IntN36)

for(i in 1:(nod-tt)) {IntN36[,,i]=IntN36[,,i]*Int136}

IntN36=2*IntN36

for(i in 1:(nod-tt)) {IntN36[i,,i]=0.5*IntN36[i,,i]}

InteN36=apply(IntN36,c(2,3),sum)

InteN36=(max(t)/(nod-1))*InteN36

IntN45=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{IntN45[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}

IntN45=ifelse(is.na(IntN45),0,IntN45)

for(i in 1:(nod-tt)) {IntN45[,,i]=IntN45[,,i]*Int145}

IntN45=2*IntN45

for(i in 1:(nod-tt)) {IntN45[i,,i]=0.5*IntN45[i,,i]}

InteN45=apply(IntN45,c(2,3),sum)

InteN45=(max(t)/(nod-1))*InteN45

IntN46=array(,c(nod,s_r,(nod-tt)))

for (i in 1:(nod-tt))

{IntN46[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}

IntN46=ifelse(is.na(IntN46),0,IntN46)

for(i in 1:(nod-tt)) {IntN46[,,i]=IntN46[,,i]*Int146}

IntN46=2*IntN46

```

```

for(i in 1:(nod-tt)) {IntN46[i,,i]=0.5*IntN46[i,,i]}

InteN46=apply(IntN46,c(2,3),sum)
InteN46=(max(t)/(nod-1))*InteN46
IntN56=array(,c(nod,s_r,(nod-tt)))
for (i in 1:(nod-tt))
{IntN56[(nod-tt):i,,i]=t(replicate((nod-tt+1)-i, iden))}
IntN56=ifelse(is.na(IntN56),0,IntN56)
for(i in 1:(nod-tt)) {IntN56[,,i]=IntN56[,,i]*Int156}
IntN56=2*IntN56
for(i in 1:(nod-tt)) {IntN56[i,,i]=0.5*IntN56[i,,i]}

InteN56=apply(IntN56,c(2,3),sum)
InteN56=(max(t)/(nod-1))*InteN56

Int13<- t(cbind(InteN13,replicate(s_r,0),Inte13))
Int14<- t(cbind(InteN14,replicate(s_r,0),Inte14))
Int15<- t(cbind(InteN15,replicate(s_r,0),Inte15))
Int16<- t(cbind(InteN16,replicate(s_r,0),Inte16))
Int34<- t(cbind(InteN34,replicate(s_r,0),Inte34))
Int35<- t(cbind(InteN35,replicate(s_r,0),Inte35))
Int36<- t(cbind(InteN36,replicate(s_r,0),Inte36))
Int45<- t(cbind(InteN45,replicate(s_r,0),Inte45))
Int46<- t(cbind(InteN46,replicate(s_r,0),Inte46))
Int56<- t(cbind(InteN56,replicate(s_r,0),Inte56))

rm(InteN13,InteN14,InteN15,InteN16,InteN34,InteN35,InteN36,InteN45,
InteN46,InteN56,InteN13,InteN14,InteN15,InteN16,InteN34,InteN35,InteN36,
InteN45,InteN46,InteN56,Inte13,Inte14,Inte15,Inte16,Inte34,Inte35,Inte36,
Inte45,Inte46,Inte56)

```

```

ch<- exp(ENU1-0.1666667*(Int1+Int3+Int4+Int5+Int6+
Int13+Int14+Int15+Int16+Int34+Int35+Int36+Int45+Int46+Int56))
rm(Int1,Int3,Int4,Int5,Int6,Int13,Int14,Int15,Int16,Int34,Int35,
Int36,Int45,Int46,Int56)
ch[Mod(ch)>1.1]=0
tau<- t[1,]
Krn1P<- (1-(tau/max(tau))^2)^3
Krn1P<- replicate(s_r, Krn1P)
ch=ch*Krn1P
chmean<- apply(ch,1,mean)

invf=array (,c(nod,nod,s_r))
for(i in 1:s_r){invf[,i]=replicate(nod,ch[,i])}
x=t(replicate(nod, seq(min(t)+20,max(t)-20, length.out=nod)))
t_1=t(replicate(nod, seq(min(t),max(t), length.out=nod)))
invf2=array(,c(nod,nod,s_r))
for(i in 1:s_r){invf2[,i]=invf[,i]*exp(-1i*x*t(t_1))}

invf2=2*invf2
invf=(max(t)/(2*(nod-1)*pi))*apply(invft,c(2,3), sum)
rm(invft)

xp=seq(min(x),max(x), length.out=nod)
edensity=apply(invft,1,mean)
sd=apply(Mod(invft),1,sd)
error=qnorm(0.975)*sd/sqrt(s_r)
Upper=Mod(edensity)+error
Lower=Mod(edensity)-error

```

```

area<- sum(edensity3)*(xp[2]-xp[1])
#area under the curve of the estimated distribution using
the trapezoidal rule.

#graph
quartz()
par(mfrow = c(1, 1))
plot(xp, Mod(edensity), type="l", col="blue",
main="Distribution of the Errors", xaxt="n", xlim=c(-2.5,2.5),
ylim=c(0,2.5), xlab="Errors", ylab="Density")
axis(1, at = seq(-2, 2, by = 0.25))
lines(xp,Upper, col="green")
lines(xp,Lower, col="green")
legend(0.7,2.4,c("Estimated", "Confidence Interval"),lty=c(2,1),
col=c("blue", "green"))

```