Alma Mater Studiorum - Università di Bologna

Scuola di Dottorato in Scienze Economiche e Statistiche
Dottorato di ricerca in

Metodologia Statistica per la Ricerca Scientifica
XXIV ciclo

A Multilevel Model with Time Series Components for the Analysis of Tribal Art Prices

Lucia Modugno

Dipartimento di Scienze Statistiche "Paolo Fortunati"

Gennaio 2012

Alma Mater Studiorum - Università di Bologna

Scuola di Dottorato in Scienze Economiche e Statistiche
Dottorato di ricerca in

Metodologia Statistica per la Ricerca Scientifica
XXIV ciclo

A Multilevel Model with Time Series Components for the Analysis of Tribal Art Prices

Lucia Modugno

Coordinatore:
Prof.ssa Daniela Cocchi

Tutor:
Prof. Rodolfo Rosa

Co-Tutor:
Dott. Simone Giannerini
Dott.ssa Silvia Cagnone

Settore Disciplinare: SECS-S/01
Settore Concorsuale: 13/D1

Dipartimento di Scienze Statistiche "Paolo Fortunati"
Gennaio 2012

Acknowledgments

I am really grateful to Simone Giannerini and Silvia Cagnone that have made possible this work with crucial insights and discussions, but, most of all, have supported me and believed in me. A special thank also to my mum and to Matteo for the great patience and for always encouraging me in my moments of discouragement.

Abstract

In the present work we perform an econometric analysis of the Tribal art market. To this aim, we use a unique and original database that includes information on Tribal art market auctions worldwide from 1998 to 2011. In literature, art prices are modelled through the hedonic regression model, a classic fixed-effect model. The main drawback of the hedonic approach is the large number of parameters, since, in general, art data include many categorical variables. In this work, we propose a multilevel model for the analysis of Tribal art prices that takes into account the influence of time on artwork prices. In fact, it is natural to assume that time exerts an influence over the price dynamics in various ways. Nevertheless, since the set of objects change at every auction date, we do not have repeated measurements of the same items over time. Hence, the dataset does not constitute a proper panel; rather, it has a two-level structure in that items, level-1 units, are grouped in time points, level-2 units. The main theoretical contribution is the extension of classical multilevel models to cope with the case described above. In particular, we introduce a model with time dependent random effects at the second level. We propose a novel specification of the model, derive the maximum likelihood estimators and implement them through the E-M algorithm. We test the finite sample properties of the estimators and the validity of the own-written R-code by means of a simulation study. Finally, we show that the new model improves considerably the fit of the Tribal art data with respect to both the hedonic regression model and the classic multilevel model.

Contents

List of Tables vi

List of Figures vii

Abstract 1

1 Introduction 1

1.1 Review on price indexes in the art market 3

1.2 The first database of Tribal artworks 5

1.2.1 Descriptive analysis 7

2 The multilevel model 17

2.1 Introduction of multilevel analysis 17

2.2 Conventional approaches for multilevel data 19

2.2.1 "Rough" strategies 19

2.2.2 Fixed-effects models 20

2.2.3 Interactive models 20

2.2.4 Random-effects models 21

2.3 Model specification 23

2.4 Model estimation 26

2.4.1 Maximum likelihood estimation 26

2.4.2 Iterative GLS 30

2.5 Predicting the random effects 31

2.5.1 BLUP 32

2.5.2 Prediction of expected responses 34

2.5.3 Empirical Bayes prediction 35
2.5.4 Shrunken estimates 35
2.5.5 Standard errors of the predictors 38
2.6 Developments and applications of the multilevel model 39

3 Hedonic regression model and multilevel model for Tribal art prices 43
3.1 Models with no covariates 44
3.2 Models with covariates 49
3.2.1 Robust standard errors 62
3.2.2 Considerations 64

4 A multilevel model with time series components 67
4.1 Model specification 68
4.2 Full maximum likelihood estimation through the EM algorithm 70
4.3 Simulation study 78
4.4 The new model and Tribal art prices 83

5 Conclusions 95

References 97

List of Tables

1.1 Tribal art data: physical variables 6
1.2 Tribal art data: historical variables 7
1.3 Tribal art data: market variables 8
1.4 Tribal artworks for continent 9
1.5 Tribal artworks for type of object 10
1.6 Tribal artworks for material 11
1.7 Tribal artworks for patina 12
1.8 Tribal artworks for auction house and venue 12
1.9 Tribal artworks for type of last owner 14

3.1 Results of the null model, FE-intercept and RE-intercept models 48
3.2 Results of the FE-hedonic and RE-hedonic models 52
3.3 Shapiro-Wilk normality test for the residuals of the FE-hedonic and the RE-hedonic models 60

4.1 Scheme of the Monte Carlo study. 78
4.2 Scenarios 1 and 2 of the MC study 79
4.3 Scenarios 3 and 4 of the MC study 79
4.4 Scenarios 5 and 6 of the MC study 80
4.5 Simulated forecast 82
4.6 Results of the AR-RE-hedonic model 84
4.7 Shapiro-Wilk normality tests for the residuals of the AR-RE-hedonic model. 91
4.8 Forecasting 94

List of Figures

1.1 Density histogram of hammer prices 9
1.2 Time series of prices 10
1.3 Yearly turnover and percentage of sold items 11
1.4 Prices for type of illustration 13
1.5 Prices for type of description on the catalogue 14
1.6 Prices for quotations 15
1.7 Prices for exhibition 15

3.1 Plot of the time-specific intercepts for the FE-intercept and RE-intercept model 49
3.2 Boxplots of residuals by semester of the null model, FE-intercept and RE-intercept models 50
3.3 Plot of the time-specific intercepts for the FE-hedonic and RE-hedonic model 56
3.4 Residuals versus fitted values of the FE-hedonic and RE-hedonic models 59
3.5 Boxplots of residuals by semester for the FE-hedonic and the RE-hedonic models. 59
3.6 Standard deviations of the level-1 residuals by semester of the RE-hedonic model 60
3.7 Normal probability plot of residuals of the FE-hedonic and the RE-hedonic models. 61
3.8 Plots of autocorrelation functions of the residuals of the RE-hedonic model 62

4.1 Residuals versus fitted values of the AR-RE-hedonic model. . . 90

4.2 Normal probability plots of residuals of the AR-RE-hedonic model. 90
4.3 Residuals by semester for the AR-RE-hedonic and RE-hedonic model (3.5). 91
4.4 Plots of autocorrelation functions of the residuals of the AR-RE-hedonic model. 92
4.5 Standard deviations of the level-1 residuals by semester of the AR-RE-hedonic. 93
4.6 Plot of the forecasting capabilities. 94

Chapter 1

Introduction

Investing in artworks allows to obtain high returns and greater fiscal advantages with respect to investing in financial or housing markets. Also, the art market can be considered less volatile than other assets and, for this reason, artworks can be considered *alternative investment* items. Hence, evaluating the convenience of investing in the art market requires the knowledge of some pieces of information: the general market trend and the specific segment trend, the investment return in comparison with other assets, but also the investment risk and its correlation with other financial instruments.

Since artwork items are considered investment goods in the same way as stocks, bonds and real estates, in the past, the analysis of this new market was performed by resorting to tools for the analysis of financial markets. However, these tools miss some essential aspects, mainly because, with respect to the stocks that are exchanged a high number of times in each instant of time, artworks are one-off pieces in their kind, hardly comparable with each other, and they pass through the market only a handful of times (usually only one). Therefore, the art market trend is more difficult to evaluate than the stock market trend (Figini, 2007).

Anderson (1974) and Stein (1977) were the first to study the investment in the art market. Later, Brumal (1986)'s results on the very low art gain in the long term (only 0.55% between 1600 and 1950) paved the way to numerous studies on different art markets, the Impressionist art among all.

The same interest has not been shown in studying the Tribal art market, until recent years (Figini, 2007).

In the present thesis we perform an econometric analysis of the Tribal art market. The relevant data come from the first world database of Tribal art prices that contains more than 20000 records of items sold from 1998 to 2011 by the most important auction houses. At each auction date, the ensemble of objects to be sold are put together in a catalogue. The selling price and the most important characteristics of the whole catalogue are recorded, as presented in section 1.2. The database has been built by a team of researchers of the University of Bologna, Faculty of Economics – Rimini and it is a unique source of information.

Among the existing methods to build indexes for prices of artworks, that will be briefly go over in the next section, the most suitable for fitting our data is the hedonic regression model. It is a fixed-effect model that regresses the price on object features and includes fixed time-effects.

This thesis has two innovative contributions. First, we propose a multilevel model for the analysis of Tribal art prices. To our knowledge, this modelling framework has not been applied yet to this kind of problems. In chapter 2, a literature review of the multilevel model is presented. As we will show in chapter 3, such approach will give a substantial advantage over fixed-effect models, especially in terms of degrees of freedom, parsimony and interpretability.

Now, it is natural to assume that time exerts an influence over the price dynamics in various ways. In our case, this assumption has been verified through the diagnostic analysis performed on the classical multilevel model that postulates independent effects over time. Nevertheless, since the set of objects change at every auction date we do not have repeated measurements of the same items over time. Hence, the dataset does not constitute a proper panel.

The main theoretical contribution of the thesis is the extension of classical multilevel models to cope with the case under investigation. In particular, we introduce a model with time dependent random effects at the second level. In chapter 4, we propose a novel specification of the model, derive the maximum likelihood estimators for it and implement the whole framework and the E-M algorithm in R without resorting to third-party software.

We have tested the finite sample properties of the estimators and the validity of the software implementation through a simulation study.

Finally, the new model has been fit successfully to the Tribal art database and the main conclusions are drawn.

1.1 Review on price indexes in the art market

The literature on the art economy has proposed some methods to build indexes for prices of artworks, especially for paintings. In the following, we mention the most important proposals.

Sotheby's Art Index (and similar others) is constructed by taking the average price of a group of artworks considered representative of the market in that moment. Obviously, it is not objective since it involves selecting the representative sample by some experts.

The **average painting methodology** (Stein, 1977) constructs the index on paintings with certain "average" characteristics. Also this index is not objective since it requires choosing the characteristics considered average, although the degree of subjectivity is smaller than that of the previous index.

The **representative painting method** (Candela and Scorcu, 1997) constructs the index on a sample of paintings selected according to their price pattern, rather than to certain characteristics. This method is less subjective, since it is based on empirical and statistical arguments.

The **repeated sales regression** (Goetzmann, 1993) considers the prices of the same object exchanged twice. However, rarely artworks are resold, therefore, the resold objects are not exactly representative of the whole market. One of the most famous indexes constructed in this way is that by Mei and Moses (2002).

The **hedonic regression**, called also the *grey painting method*, was born to pricing houses (Rosen, 1974). Some applications to the art market are for example in Chanel (1995) and Locatelli-Biey and Zanola (2005).
This approach assumes that the price of an artwork depends both on the market trend and on the effect of certain object characteristics. The prices of the relevant features are regressed against past information. These prices can be used to forecast the price of another object by summing the prices of its characteristics.

More formally, the hedonic regression is a multiple linear regression:

$$y_{it} = \beta_{0t} + \beta_1 X_{1it} + \ldots + \beta_k X_{kit} + \epsilon_{it}$$

where y_{it} is the price of the object i sold in the period t, the X_{it}'s are the k object features (the artist's name, the type of object, the material, the technique, etc.), and ϵ_{it} is the error process with zero mean and constant variance. The estimated coefficients $\hat{\beta}_j$, for $j = 1, \ldots, k$, are interpreted as the price of each characteristic, the so-called *shadow prices*, assumed to be constant in time. The estimated coefficient $\hat{\beta}_{0t}$ expresses, instead, the value of the *grey painting* in the period t, that is, the value of an artwork created by a standard artist, through standard techniques, with standard dimensions, etc. (Candela and Scorcu, 2004). The market price index is built from the prices of this *grey painting* in different periods, $\hat{\beta}_{01}, \ldots, \hat{\beta}_{0t}, \ldots, \hat{\beta}_{0T}$.

This method has been widely used for the painting market since, contrarily to other proposals, it solves the problem of artwork heterogeneity by explaining prices through object features (Figini, 2007). Moreover, it allows to construct a market price index by neutralizing the effect of quality.

Nevertheless, the hedonic regression has some drawbacks. First of all, it is difficult to account for all the relevant characteristics for determining the price of an object, so that this method can explain only a part of the price. Moreover, most of the object features are categoric, such as, for example, the artist's name that in Western market affects strongly the price of artworks. Therefore, the regression equation will contain many dummy variables: if, for example, the database includes observations from 100 artists, then there will be 99 indicator variables only for this characteristic, and this will be repeated for each dummy inserted in the model. This means that there will be a high number of parameters to be estimated, so that the resulting models are not parsimonious.

1.2 The first database of Tribal artworks

Constructing price indexes requires data on sales of artworks. In the art market, the only available information is that coming from auction exchanges. Nowadays, there are companies[1] that publish, through the web, information about auctions, price indexes, and provide art valuations and other services. However, most of these companies deals with the Western art. In this scenario, for a long time, there has not been a database on the Ethnic art. However, in recent years, the Tribal art market underwent increases in turnover, so that it attracted the interest of investors and economists.

In order to fill the lack of information in this neglected segment of art, given the confidence on its commercial potential, the first database on Ethnic artworks has been created from the agreement of four institutions: the Department of Economics of the University of the Italian Switzerland, the Museum of the Extraeuropean cultures in Lugano, the *Museo degli Sguardi* in Rimini, and the Faculty of Economics of the University of Bologna, campus of Rimini[2]. The database includes more than 20000 observations auctioned by the major auction houses during 1998-2011. The information has been collected from the paper catalogues released from the auction houses before the auctions.

For each object, 37 features have been gathered; these include physical, historical and market characteristics (see Figini (2007)), listed respectively in

[1] Artenet.com, Artinfo.com, Arsvalue.com, Artprice.com are some of them.

[2] The project managers are the economists Guido Candela and Antonello E.Scorcu, and the statistician Simone Giannerini.

Table 1.1, 1.2 and 1.3.

Table 1.1: Artworks *physical* variables (labels in parentheses) with corresponding levels.

Variable	**Levels**
Type of object (OGG)	Furniture, Sticks, Masks, Religious objects, Ornaments, Sculptures, Musical instruments, Tools, Clothing, Textiles, Weapons, Jewels
Condition (CDCA)	Passable, Good, Very good
Gaps and repairs (CLIO)	Yes, No
Material (MATP)	Ivory, Vegetable fibre, Wood, Metal, Gold, Stone, Precious stone, Terracotta, ceramic, Silver, Textile and hides Seashell, Bone, horn, Not indicated
Patina (CPAT)	Not indicated, Pejorative, Present, Appreciative
Height, width, diameter, length (MISA, MISL, MISD, MISN)	(quantitative)

Among the existing proposals for modelling art prices, the hedonic regression method is the one that better fits Tribal art data. Rather, this approach seems more suitable on the Ethnic art than on the Western art data. An ethnic object, in fact, is characterized by its ethnic provenance, rather than by the artist's name that is unknown (The Tribal art is considered an *anonymous art*). Since the number of ethnic groups is generally smaller than the number of artists' names, the hedonic model for the Tribal art has less dummy variables than the same model applied to another segment of art. Moreover, also the amount of iconographic subjects and materials is more limited. Therefore, the main drawback of the hedonic regression method is less severe in the Tribal segment (Figini, 2007).

Table 1.2: Artworks *historical* variables (labels in parentheses) with corresponding levels.

Variable	**Levels**
Dating (DATA)	
Continent, Region, Stylistic area, Ethnic group (CONT, REG, ACSTI, ETNIA)	4 Cont., 8 Reg., 89 Styl. Areas 158 Ethnic groups
Illustration on the catalogue (CAIL)	Absent, black/white ill., col. illustr.
Illustration width (CAAI)	Absent, Miscellaneous ill., Quarter page, Half page, Full page, More than one ill., Cover
Description (CATD)	Absent, Short visual descr., Visual descr., Broad visual descr., Critical descr., Broad critical descr.
Specialized bibliography (CABS) Comparative bibliography (CABC) Exhibition (CAES)	Yes, No
Historicization (CAST)	Absent, Museum certification, Relevant museum certification, Simple certification
Last owner (CAUA)	Unknown, Museum, Private individual, Other, Art gallery, Company

1.2.1 Descriptive analysis

In this subsection, some descriptive analysis of the Tribal art dataset are presented. Hammer prices have been deflated through the HICP (Harmonized Index of Consumer Prices) and transformed in euro.
Figure 1.1 shows the comparison between the density curve of the logarithm (base 10) of hammer prices and a Gaussian curve with the same mean and variance as the distribution of prices: it looks a little bit leptokurtic with respect to the normal distribution and quite symmetric.
Figure 1.2 displays the series of prices aggregated by year. It reveals that the most unsatisfying year for the Ethnic art market has been 2003, but it has

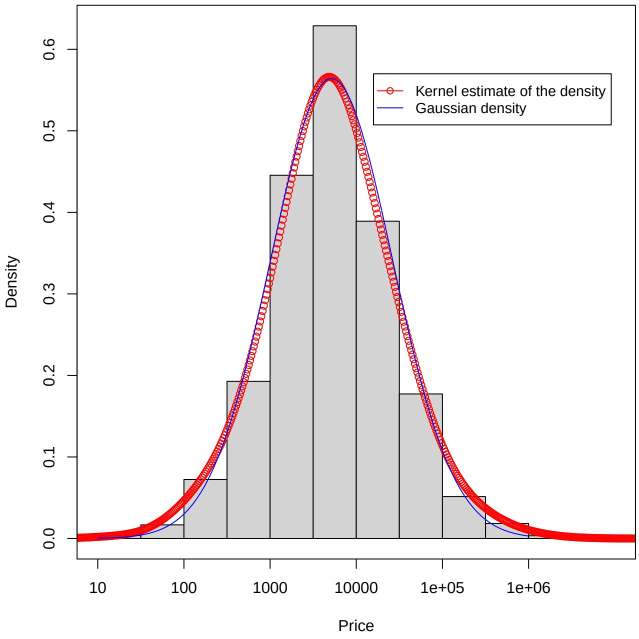
Table 1.3: Artworks *market* variables (labels in parentheses) with corresponding levels.

Variable	**Levels**
Auction date (ASDA)	
Venue (ASLU)	New York, Paris, Zurich, Amsterdam
Auction house (ASNC)	Sotheby's, Christie's, Encheres Rive Gauche, Piasa, Koller, Bonhams
Importance of the auction house (ASRC)	National, International
State of business (ASSA)	Ceased, Existent
Auction type (ASTP)	Heterogeneous, Single collection, Homogeneous
Auction title (ASTT)	
# of items on the catalogue (CANC)	(quantitative)
Number of lot (CANO)	(quantitative)
Currency (CAVS)	Euro, Dollar, Franc
Minimum and Maximum estimation (VESM, VESX)	(quantitative)
Hammer price (PRICE)	(quantitative)
Buy-in and Return of the object (VENB, VEOR)	Yes, No

been also the year with the highest number of sold artworks, both in absolute and percentage terms, as shown by the plot in Figure 1.3. After this period, the market has recorded a gradually increase in prices and overall turnover (plot in Figure 1.3). This positive trend gives an idea of the great potential of the Tribal art market.

The Ethnic artworks made from African ethnic groups are the most sold in auction, and, as can be read in Table 1.4, are the items with the highest average hammer price, with respect to America, Eurasian and Oceanic artworks. By looking at the coefficient of variation, it seems that the African artworks are also the most risky investment items. However, aggregating the objects by continent can only give an idea of the importance of each macro-group, since Tribal artworks are characterized by the Ethnic groups, as the Modern

Figure 1.1: Density histogram of hammer prices on logarithmic scale (base 10), overlaid by a normal curve of the same mean and variance and a kernel estimate of the density.



artworks by the artist's name and the artistic style.

About the type of auctioned object, Table 1.5 reveals that sculptures, tools

Table 1.4: Tribal artworks for continent which they come from. CV is the coefficient variation defined as the ratio between standard deviation and arithmetic mean, $\frac{\text{sd}}{|\bar{x}|}$.

	# obj. auctioned	% of sales	Average price	Median	CV
Eurasia	426	75	7028	1532	2.58
Africa	10604	67	24474	4000	5.07
Oceania	3091	77	21471	4461	3.31
America	5706	75	12918	4699	3.24

and masks are the most sold pieces. Sculptures and masks, together with religious objects, are also the most priced with respect to other type of objects.

As evident from Table 1.6, most of the Ethnic artworks are made principally with wood. However, the stone objects seem precious and also less risky than

Figure 1.2: Time series of prices in logarithmic scale (base 10). The amount of sold items are shown within the boxes.

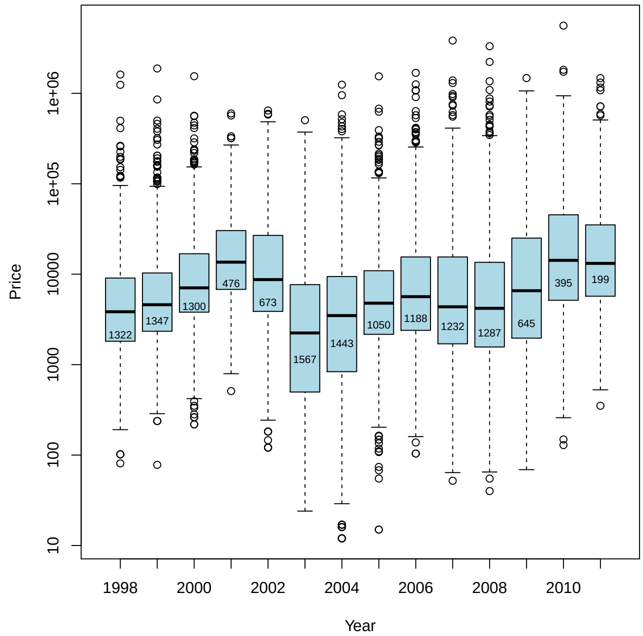
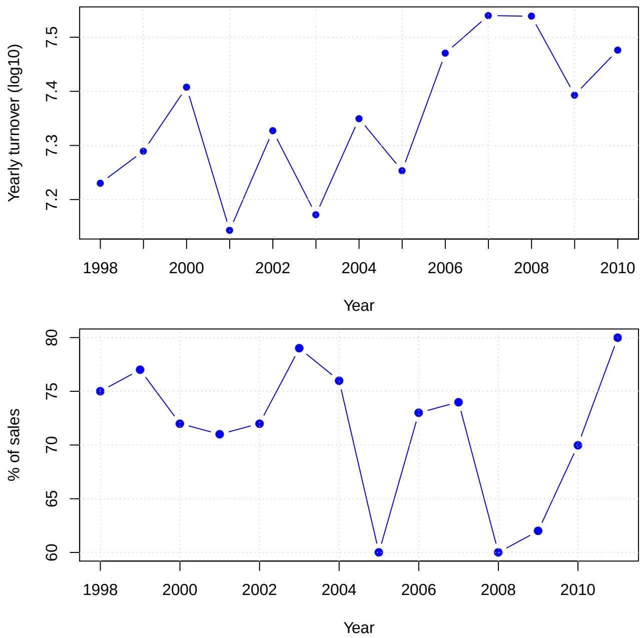


Table 1.5: Tribal artworks for type of object. CV is the coefficient variation defined as the ratio between standard deviation and arithmetic mean, $\frac{\text{sd}}{|\bar{x}|}$.

	# obj. auctioned	% of sales	Average price	Median	CV
Jewels	442	71	9406	2569	3.29
Weapons	750	72	8608	2840	2.78
Sticks	763	74	13047	3318	3.63
Musical instruments	399	68	17222	3824	4.00
Ornaments	1561	74	15737	3886	2.93
Tools	4134	77	11055	4224	3.86
Clothing	387	72	17073	4224	3.04
Furniture	663	72	35427	5207	7.61
Textiles	338	80	14262	5977	2.70
Masks	3083	64	23772	6211	2.94
Sculptures	5927	69	34687	6769	4.06
Religious objects objects	1380	72	33367	6910	3.29

others. Also the artworks made with precious stone, wood and ceramic are sold at high prices. Moreover, the most common and priced materials are the main materials of the most auctioned and the most payed types of object,

Figure 1.3: Yearly turnover (euro) in logarithmic scale (base 10) and yearly percentage of sold items with respect to the total amount of objects auctioned. The year 2011 has not been included in the first plot since the database contains information on it only for one semester.



namely wooden masks, stone or ceramic tools and sculptures. Many Ethnic

Table 1.6: Tribal artworks for material. CV is the coefficient variation defined as the ratio between standard deviation and arithmetic mean, $\frac{\text{sd}}{|\bar{x}|}$.

	# obj. auct.	% of sales	Av. price	Median	CV
Metal	829	72	25372	1758	6.88
Bone, horn	263	80	7513	2028	2.94
Seashell	155	74	11204	2421	3.05
Silver	137	67	5686	2591	1.88
Ivory	649	74	14753	3549	4.83
Not indicated	136	82	9061	3861	2.15
Vegetable fibre, paper, plumage	565	77	11469	4023	3.09
Gold	612	73	10571	4180	1.89
Textile and hides	849	75	14319	4830	3.02
Wood	11733	69	28520	5387	4.29
Terracotta, ceramic	2625	76	13004	5617	3.06
Precious stone	521	69	30355	7336	3.41
Stone	753	72	13836	7505	1.51

artworks have a patina that can assume a different interpretation depending on the type of object and its original function. When it is interpreted as a sign of consumption or genuineness, the patina adds value to the object, as indicated in Table 1.7 for the "appreciative patina".

Table 1.7: Tribal artworks for patina. CV is the coefficient variation defined as the ratio between standard deviation and arithmetic mean, $\frac{\text{sd}}{|\bar{x}|}$.

	# obj. auctioned	% of sales	Average price	Median	CV
Absent	11111	72	18359	4451	5.21
Present	4390	69	19127	5135	3.26
Pejorative	141	81	16034	5239	2.23
Appreciative	4185	72	38151	7210	3.89

The market features of the objects concern the organization and the general functioning of the Tribal art market. As shown in Table 1.8, the most important venue for this segment of art are Paris and New York. Christie's and Sotheby's are the dominant auction houses, as in the Modern and Contemporaneous arts. They, in fact, entered in the market respectively in 1970 and 1967, long before the other auction houses working in this sector nowadays. A noteworthy observation is that the organization of auctions is oriented to exploit economies of scale, that is to concentrate auctions in time and space in order to reduce unit costs.

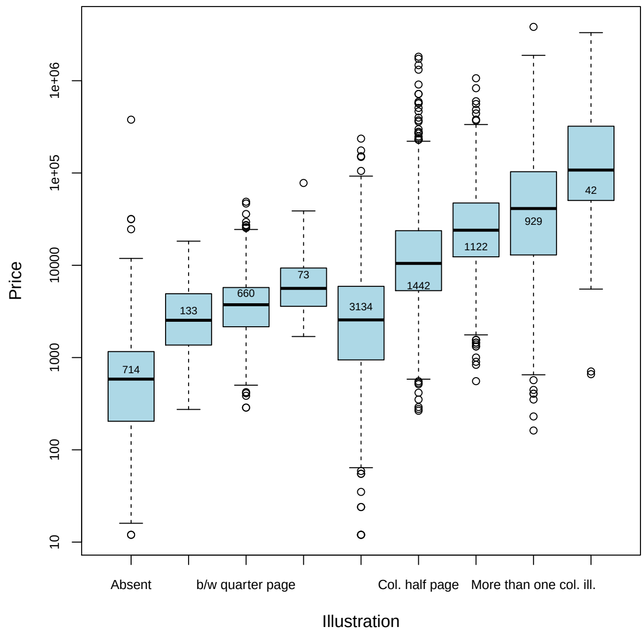
The marketing actions, made by auction houses through catalogues, seem

Table 1.8: Tribal artworks for auction house and venue. CV is the coefficient variation defined as the ratio between standard deviation and arithmetic mean, $\frac{\text{sd}}{|\bar{x}|}$.

	# obj. auct.	% of sales	Average price	Median	CV
Koller-Zurich	1396	47	3569	1116	4.89
Christie's-Amsterdam	654	100	6536	2048	3.80
Encheres Rive Gauche -Paris	64	52	4549	2218	1.55
Christie's-Paris	4751	75	10354	2363	4.27
Bonhams-New York	282	38	6239	3208	1.49
Christie's-New York	539	81	19579	4350	3.77
Piasa-Paris	69	61	16732	5498	3.13
Sotheby's-Paris	3159	68	43000	7874	3.92
Sotheby's-New York	8913	73	26997	7902	4.08

important in fetching good prices. In fact, the boxplots in Figure 1.4 highlight that prices tend to increase as the importance given to the object on the catalogue through illustrations increases. In particular, artworks which have been dedicated a coloured wide illustration are priced on average more than those without or with black and white figures on the catalogue.
A similar effect on prices is due to the type of description dedicated on the

Figure 1.4: Boxplots of prices (in logarithmic scale) for type of illustration on the catalogue. The amount of sold items are shown within the boxes.



catalogue to each object (boxplots in Figure 1.5). Moreover, a critical description is more valuated than a visual description.

The pedigree of an artwork can be also constituted by quotations on bibliography that can be object-specific or just comparative. The boxplots in Figure 1.6 reveal that investors tend to pay more for objects boasting citations, and, as expected, more for the specific rather than comparative bibliography. Few artworks have been previously exhibited and this fact is positively valuated by art collectors that tend to offer more for having those objects. In fact, it seems that an object exhibition by a museum, for example, certificates its value.

Figure 1.5: Boxplots of prices (in logarithmic scale) for type of description on the catalogue. The amount of sold items are shown within the boxes.

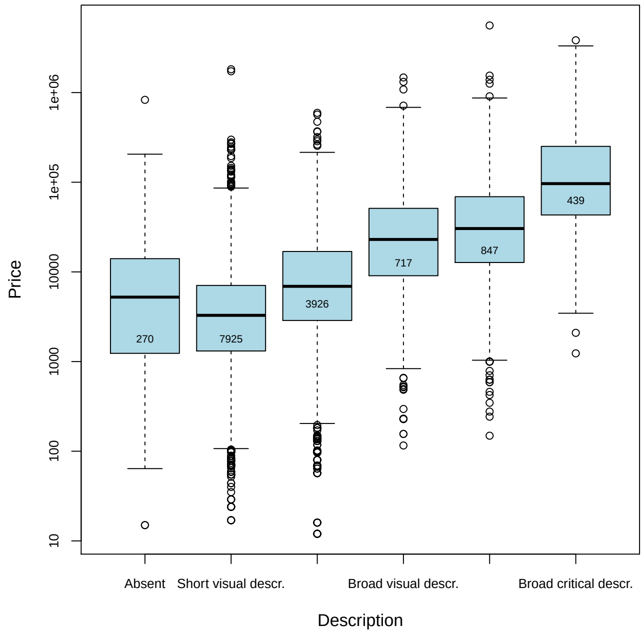


Table 1.9: Tribal artworks for type of last owner. CV is the coefficient variation defined as the ratio between standard deviation and arithmetic mean, $\frac{\text{sd}}{|\bar{x}|}$.

	# obj. auctioned	% of sales	Average price	Median	CV
Unknown	5641	72	8635	3054	4.19
Art gallery	307	63	44342	4052	6.64
Other	173	71	19167	5546	2.71
Private individual	13505	71	27881	6306	4.12
Museum	77	79	62683	6336	2.39
Company	124	83	34939	17304	1.32

Finally, Table 1.9 shows some statistics about the prices for type of last owner of the object. The companies are the most paid seller. However, in general, the table discloses that knowing who has been the last owner is important for the buyer.

Figure 1.6: Boxplots of prices (in logarithmic scale) for quotations. The amount of sold items are shown within the boxes.

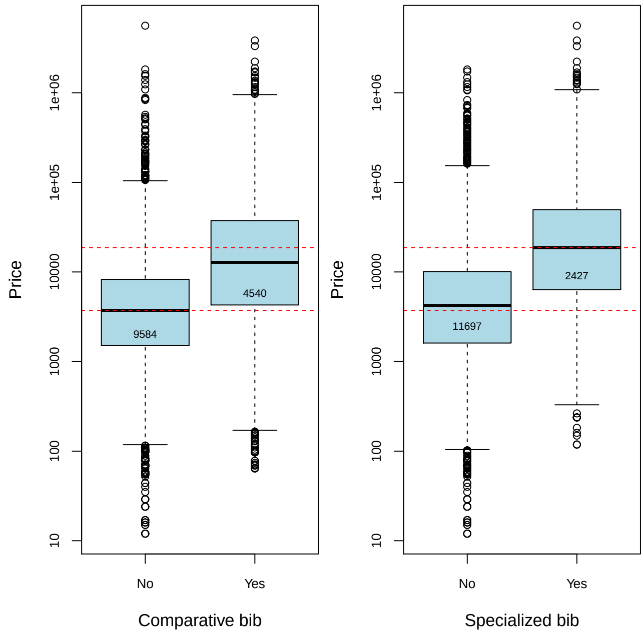
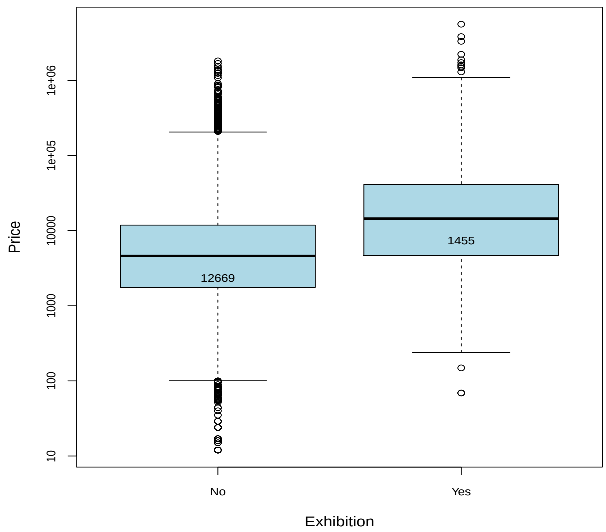


Figure 1.7: Boxplots of prices (in logarithmic scale) for exhibition. The amount of sold items are shown within the boxes.



Chapter 2

The multilevel model

2.1 Introduction of multilevel analysis

Multilevel data consist of units of analysis of different type, one hierarchically clustered within the other. In a strictly nested data structure, the term *levels* represents the different types of unit of analysis, i.e. the various type of groupings; in particular, the most detailed level is called the first (or the lowest) level. The sense of the hierarchy is as follows: there are individuals described by some variables (level-1 observations), and they are also grouped into larger units (higher level observations), which in turn could be described by other variables.

The leading example of multilevel data comes from studies on educational achievement, in which pupils, teachers, classrooms, schools, district, and so on, are clustered one within the other, and they might all be units of analysis, each described by own variables. Another well-known example is about organizational studies, where, generally, data are represented by employees grouped in departments and firms. Moreover, hierarchical data often occur in social sciences: economists and political scientists frequently work with data measured at multiple levels in which individuals are nested in geographic divisions, institutions or groups, and so forth (Jones et al., 1992). Further, other particular structures of data can be thought as multilevel: the repeated measurements over time on an individual, the respondents to the same interviewer and also subjects within a particular study among those of a meta-analysis can be considered groups of observations, and, consequently, be treated as multilevel data.

The idea behind modelling multilevel data, coming from sociological studies (DiPrete and Forristal, 1994), is that living environments ("macro level" in the sociological field) affect (and can be affected by) individual behaviours ("micro level"), and, contextual effects are due to social interactions within an environment. In general, individuals both can influence and be influenced by various type of contexts, mentioned above: spatial, temporal, organizational and socio-economic-cultural.
As Kreft and Leeuw (1998) put it, "the more individuals share common experiences due to closeness in space and/or in time, the more they are similar, or, to a certain extent, duplications of each other"; in other words, performances of pupils in the same classroom or those of employees in the same department tend to be more similar than those from different groups because of sharing contexts.

The specificity of multilevel data cannot be ignored, first of all because of an important statistical motivation: the observations within one group are not independent of each other, as traditional models require. This means that each individual from the same group may provide less additional information than a new individual in a new group. If standard statistical analysis, which generally assume independent observations, is performed on multilevel data (the so-called *naive pooling* strategy), results may be misleading. In fact, a positive correlation among observations within a group, referred to as *intra-class correlation* (ICC), usually causes the underestimate of standard errors because the analysis assume that there is more information in the data than there really is. The case of negative intra-class correlation is less frequent: it could occur only when the individuals within a context are in competition, and this may make them less similar to each other. Therefore, generally, a non-null intra-class correlation biases traditional statistical inference.

Indeed, multilevel structure has not to be treated only as a statistical nuisance that just needs to be considered for obtaining correct statistical estimations, but a key concept that yields important information by itself. In

fact, in addition to statistical motivations, there also are important substantive reasons for considering information from all levels of analysis.
First of all, multilevel model allows to combine multiple levels of analysis in a single model by specifying predictors at different levels. This can be useful, for example, to determine whether variables measured at one level affect relations occurring at another level.[1].
Second, as we will better see, the multilevel analysis allows to decompose the overall variance in within-group and between-group variances, and, therefore, to know how much the groups are responsible for the variability of the outcome, the so-called Variance Partition Coefficient (VPC).

2.2 Conventional approaches for multilevel data

2.2.1 "Rough" strategies

When looking for statistical techniques capable of taking into account the correlation structure of multilevel data, one could think, at first, of two simple procedures: either to disaggregate the higher level variables (e.g. school-level variables) to the individual level or, conversely, to convey the analysis on the higher level after the aggregation of the individual observations to the higher level through a single *summary statistic*. Both these strategies are obviously unsatisfactory: while the first one does not take into account the dependence of the observations within a group, the second approach, which may be referred to as *data resolution*, is inefficient because, even though it avoids the over-inflation of the apparent size of the dataset, not only it wastes huge pieces of information, but also, it may produce misleading results (Burton et al., 1998; Raudenbush and Bryk, 2002). In fact, this is best known as *ecological fallacy* or *aggregation fallacy* (Goldstein, 2009).

[1]The multilevel analysis has been applied, for example, by some demographers, to examine how differences in national economic development, an information gathered at the national level, interact with adult educational achievement to influence fertility rates, which, conversely, are households information (Raudenbush and Bryk, 2002).

2.2.2 Fixed-effects models

A simple way to represent the dependence and the variations in outcome, that may be due to differences between groups and/or to individual differences within a group, is by including group-specific terms in the model (dummy variables). This approach has been borrowed from longitudinal data analysis. Since they show a dependence among the elements, longitudinal data can be thought of as two-level data with occasion i at level 1 and units j at level 2 (Skrondal and Rabe-Hesketh, 2008). While in the latter context the fixed-effect model is often referred to as *least squares dummy variable* (LSDV) model, in the field of experimental design, it is called, instead, *analysis of covariance* model (ANCOVA).
On the one hand, these models perfectly capture the clustered structure of multilevel data, since the dummy variables account for differences among groups. On the other hand, however, they do not allow to include level-2 (or higher) covariate to explain the differences among the groups. Moreover, since the model without a constant term includes as many dummy variables as the groups are (in the model with a constant term, instead, for reasons of collinearity, one group will be designated baseline category, thus it will not have the corresponding dummy variable), a further argument against the use of the dummy variable model, from a statistical point of view, is the high number of parameters to be estimated.

2.2.3 Interactive models

In political science, there has been another noteworthy attempt to modelling multilevel data structures through models sometimes referred to as *interactive models* (Steenbergen and Jones, 2002). These models include in the regression model both higher level predictors and interaction terms. The first capture the groups differences in the intercepts; the latter, instead, consisting of group-specific terms interacting with lower level predictors, represent the differences in the partial slopes of these predictors. This is the strength of the interactive models compared to dummy variable models. However, they are based on the strong assumption that differences among groups are completely captured by group predictors without error terms. Since this assumption is often proved to be false, also the interactive models do not fully address the requirements of multilevel data modelling.

2.2.4 Random-effects models

Whereas the older statistical models were fixed-effects regression models, the specification of the regression coefficients as random effects has become a common practice since the 80's.
In order to understand the benefits of random-effect models compared to fixed-effect models, particularly in the context of multilevel analysis, consider a simple one-level model with one regressor:

$$y_i = \beta_0 + \beta_1 x_{1i} + \epsilon_i$$

with $i = 1, \ldots, n$, under the usual assumption of independent Gaussian errors with zero mean and variance σ^2. To deal with the groupings, we let intercept and slope coefficients vary among groups

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + \epsilon_{ij} \tag{2.1}$$

where y_{ij} is the response of the level-1 unit i $(= 1, \ldots, n_j)$ nested in the level-2 unit j $(= 1, \ldots, J)$, and x_{1ij} is the level-1 covariate.
In a fixed-effects regression model, the group varying coefficients, β_{0j} and β_{1j}, are treated as fixed but unknown parameters to be estimated. Indeed, in a strictly dummy variable model, the intercept is group varying and the slope is the same for all groups, $\beta_{1j} = \beta_1$ for all j.

Random-effects models, in contrast, contain error terms at higher levels, and this implies a more complex error framework capable of modelling the heteroscedastic structure. The coefficients in Eq. (2.1) can be re-expressed in the following way:

$$\beta_{0j} = \beta_0 + u_{0j} \quad \beta_{1j} = \beta_1 + u_{1j} \tag{2.2}$$

where β_0 and β_1 are *fixed*, i.e. they do not vary among groups, and the u's are random variables with

$$\mathrm{E}(u_{0j}) = 0, \quad , \mathrm{Var}(u_{0j}) = \sigma^2_{u0}, \quad \mathrm{Var}(u_{1j}) = \sigma^2_{u1}, \quad \mathrm{Cov}(u_{0j}, u_{1j}) = \sigma_{u01}.$$

For this specification, these models are also called *mixed-effects models*.

The use of random coefficients for modelling grouped data presents several advantages. First of all, usually, we deal only with a sample from a larger population, with groups at any level being sub-samples from the whole populations of such type of groups. Thus, it makes more sense to treat parameters as drawn randomly from a larger "population" of parameters.
Second, specifying a model with random coefficients for grouped data is important also for predictions. Consider, for example, a model of test scores for students within schools. Since a fixed-coefficients model contains a parameter for each school, we cannot do a prediction for a new student in a new school, because there is not an indicator for this school in the model (Goldstein, 2010).
According to these two first points, in general, effects should be fixed if they are interesting in themselves, or random if there is interest in the underlying population (Searle et al., 1992).

Another common argument against using only fixed effects is the high number of parameters to be estimated, which, in addition, increases with the number of groups. This results in a loss of a substantial number of degrees of freedom. Moreover, rarely it is possible to give a meaningful interpretation to these parameters. In the random-coefficients models, instead, only the variance components need to be estimated and interpreted.
Moreover, the complex error structure of a model specification with random coefficients, allows to decompose the variance of the response in different components, and this provides important insights. In a two-level model, for example, the total variance has three components: the first variance term allows groups to differ in their mean values (intercepts) on the dependent variable; the second variance term allows slopes between independent and

dependent variables to differ across groups (single-level regression models, instead, generally assume that the relationship between the independent and dependent variable is constant across groups); a third variance term reflects the within-group variation, that is the degree to which an individual variable differs from its predicted value within a specific group.
Further, the random-coefficients model allows to specify a different linear regression model for any level, so that the model will have many nested linear models; in particular, Eq. (2.1) is the level-1 regression model, and Eq. (2.2) is the level-2 regression model.
Finally, a further justification to using random group-effects is the fact that they represent the ignorance at a certain level, as the residuals represent the general ignorance.

In summary, important assumptions made in usual regression analysis are, among others, independence and homoscedasticity of individual responses. Since these assumptions are violated by data with a multilevel structure, the results of the classical regression analysis on these data would be biased. The multilevel regression analysis deals with the dependence of the outcome variable among individuals within groups through random effects.

2.3 Model specification

Multilevel models are referred to in numerous ways: contextual-effects models (Blalock, 1984) or multilevel linear models (Goldstein, 2010) mainly in sociological researches, random-coefficients models in econometric literature, hierarchical mixed linear models or random-effect models in biometric applications (Laird and Ware, 1982), hierarchical linear models in Bayesian contexts.
The multilevel model is an extension of the random-coefficients model and the interactive model presented in sections 2.2.4 and 2.2.3, since it takes into account not only the dependence among elements and the hierarchical structure, but it also allows to incorporate variables from all levels.

A multilevel model can be specified in two stages (Skrondal and Rabe-

Hesketh, 2004; Raudenbush and Bryk, 2002) or in reduced form (Goldstein, 2010). We chose the first type for a better specification and interpretation of each single model.

The level-1 model with just one covariate of a two-level linear model expresses the response y_{ij} of the level-1 unit i in the group (level-2 unit) j, for $i = 1, \dots, n_j$ and for $j = 1, \dots, J$, as

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \epsilon_{ij}, \qquad \epsilon_{ij}|x_{ij} \sim \text{NID}(0, \sigma^2) \tag{2.3}$$

where β_{0j} is the group-specific intercept, β_{1j} is the group-specific slope, and ϵ_{ij} are level-1 error terms. In the second level model, β_{0j} and β_{1j} are modeled as

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}z_j + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}z_j + u_{1j} \\ \mathbf{u}_j|\mathbf{x}_j, z_j &= \left[\begin{array}{c} u_{0j} \\ u_{1j} \end{array}\right]\Bigg|\mathbf{x}_j, z_j \sim \text{NID}\Big(\mathbf{0}, \mathbf{\Sigma}\Big), \quad \mathbf{\Sigma} = \left[\begin{array}{cc} \sigma^2_{u0} & \sigma_{u01} \\ \sigma_{u01} & \sigma^2_{u1} \end{array}\right], \end{aligned} \tag{2.4}$$

where the γ's are fixed-coefficients, z_j represents the level-2 covariates, and u_{0j} and u_{1j} are the group-level error terms. Moreover, the random effects for the group j, $\mathbf{u}_j$, are assumed independent of the within-group errors, ϵ_{ij}. From now on, the conditioning on the covariates is omitted but implicit.

By substituting the level-2 model (2.4) in the level-1 model (2.3), the reduced form of the model is obtained:

$$\begin{aligned} y_{ij} &= \gamma_{00} + \gamma_{01}z_j + u_{0j} + (\gamma_{10} + \gamma_{11}z_j + u_{1j})x_{ij} + \epsilon_{ij} \\ &= \gamma_{00} + \gamma_{01}z_j + \gamma_{10}x_{ij} + \gamma_{11}z_jx_{ij} + u_{0j} + u_{1j}x_{ij} + \epsilon_{ij}. \end{aligned} \tag{2.5}$$

When, $\beta_{1j} = \gamma_{10}$, the model becomes a two-level random-intercept model and the variance of the responses is composed by the sum of two variance components: the within-group variance, σ^2, and the between-group variance, σ^2_{u0}. The responses of two units in the same group are correlated since they share the same random intercept. The correlation within a group takes the following form:

$$\text{ICC} = \text{Corr}(y_{ij}, y_{i'j}) = \frac{\text{Cov}(y_{ij}, y_{i'j})}{\text{Var}(y_{ij})} = \frac{\sigma^2_{u0}}{\sigma^2_{u0} + \sigma^2}.$$

Therefore, the ICC can also be interpreted as the proportion of total variability in the response due to the between-group variance.

The presence of the random slope makes the variance of the responses dependent on the covariates having random coefficients, that is

$$\text{Var}(y_{ij}) = \sigma^2_{u0} + 2\sigma_{01}x_{ij} + \sigma^2_{u1}x^2_{ij} + \sigma^2.$$

Then, in the more general case, the intra-class correlation is not equal to the proportion of variability explained by the second-level variance, that, in order to avoid confusion, is called by someone (for example Goldstein (2010)) Variance Partition Coefficient (VPC).
By adopting the Skrondal and Rabe-Hesketh (2004)'s notation, the two-level model in the reduced form (2.5) is simplified in the following way:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + u_{0j} + u_{1j} z_{1ij} + \epsilon_{ij}$$

where $\beta_0 = \gamma_{00}$, $\beta_1 = \gamma_{01}$, $\beta_2 = \gamma_{11}$, $x_{1ij} = z_j$ and $x_{ij} = z_{1ij}$.
Writing the transformed reduced model in matrix notation, the whole (n $\times$ 1) response vector takes the more general form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}, \tag{2.6}$$

where, $\mathbf{u} = (\mathbf{u}_1, \ldots, \mathbf{u}_J)^{\mathrm{T}}$, $\mathbf{Z}$ is a $(n \times p * J)$ block-diagonal matrix with each block equal to

$$\begin{bmatrix} 1 & z_{11j} & \ldots & z_{p1j} \\ 1 & z_{12j} & \ldots & z_{p2j} \\ 1 & \vdots & \vdots & \vdots \\ 1 & z_{1n_jj} & \ldots & z_{pn_jj} \end{bmatrix},$$

and p is the number of random effects.
Therefore, the response vector is distributed as

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Omega}) \tag{2.7}$$

where $\boldsymbol{\Omega} = \mathbf{Z}\boldsymbol{\Gamma}\mathbf{Z}^{\mathrm{T}} + \sigma^2\mathbf{I}_n$ and

$$\boldsymbol{\Gamma} = \mathrm{Var}(\mathbf{u}) = \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Sigma} \end{bmatrix}$$

.

2.4 Model estimation

Inference for the linear multilevel model is conducted on the effects, both fixed and random, and the variance components. It can be based either on least squares approach (Goldstein, 2010), maximum likelihood methods (Searle et al., 1992; Laird and Ware, 1982; Pinheiro and Bates, 2000; Raudenbush and Bryk, 1986), or on Bayesian methodology (Seltzer et al., 1986).

2.4.1 Maximum likelihood estimation

Maximum likelihood is the most used estimation method in multilevel modelling. It produces, in fact, estimators that are asymptotically efficient, consistent and, for large sample size, robust against violations of the non-normality assumption for the errors.
Two different likelihood functions can be optimized, each corresponding to a specific approach: the Full Maximum Likelihood (FML) and the Restricted Maximum Likelihood (REML). The substantial difference between them is that, in the latter method, the likelihood function does not include the re-

gression coefficients.

Full maximum likelihood

Call $\boldsymbol{\theta}$ the vector of variance components. Under the assumptions of normality and independence for $\mathbf{u}_j$ and $\boldsymbol{\epsilon}_j$, the group-response vectors, $\mathbf{y}_j$, are independent and normally distributed with mean $\mathbf{X}_j\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{\Omega}_j = \mathbf{Z}_j\boldsymbol{\Sigma}\mathbf{Z}_j^{\mathrm{T}} + \sigma^2\mathbf{I}_{n_j}$. Therefore, the full likelihood function associated with the response vector $\mathbf{y}$ of the model (2.6) is:

$$\begin{aligned}
\mathrm{L}(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{y}) &= \prod_{j=1}^{J} f(\mathbf{y}_j; \boldsymbol{\beta}, \theta) \\
&= \prod_{j=1}^{J} \frac{|\boldsymbol{\Omega}_j|^{-1/2} \exp\{-\frac{1}{2}(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\Omega}_j^{-1}(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta})\}}{(2\pi)^{\frac{n_j}{2}}}.
\end{aligned} \tag{2.8}$$

Maximum likelihood estimates for the fixed effects and the variance components are the values maximizing the likelihood function (2.8), or equivalently, the logarithm of the likelihood:

$$\begin{aligned}
\ell(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{y}) &= \ln \mathrm{L}(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{y}) \\
&= -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{j=1}^{J} \ln|\boldsymbol{\Omega}_j| - \frac{1}{2}\sum_{j=1}^{J}\left[(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\Omega}_j^{-1}(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta})\right]
\end{aligned} \tag{2.9}$$

The derivatives of the log-likelihood with respect to the parameters are as follows:

$$\frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{y})}{\partial \boldsymbol{\beta}} = \sum_{j=1}^{J}\left[\mathbf{X}_j^{\mathrm{T}}\boldsymbol{\Omega}_j^{-1}(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta})\right] \tag{2.10}$$

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{y})}{\partial \boldsymbol{\theta}} &= -\frac{1}{2}\sum_{j=1}^{J} \mathrm{tr}\left(\boldsymbol{\Omega}_j^{-1}\frac{\partial \boldsymbol{\Omega}_j}{\partial \boldsymbol{\theta}}\right) \\
&\quad + \frac{1}{2}\sum_{j=1}^{J}\left[(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\Omega}_j^{-1}\frac{\partial \boldsymbol{\Omega}_j}{\partial \boldsymbol{\theta}}\boldsymbol{\Omega}_j^{-1}(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta})\right],
\end{aligned} \tag{2.11}$$

obtained by using the following properties of matrices derivatives (Magnus and Neudecker, 2007):

(i) $\frac{\partial \ln |\mathbf{A}|}{\partial \mathbf{A}} = \mathrm{tr}(\mathbf{A}^{-1})$

(ii) $\frac{\partial (\mathbf{x}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{x})}{\partial \mathbf{A}} = -\mathbf{A}^{-1}(\mathbf{x}\mathbf{x}^{\mathrm{T}})\mathbf{A}^{-1}$,

where A is a square matrix.
Maximizing the log-likelihood function (2.9) consists in equating the partial derivatives (2.10) and (2.11) to zero and solving the resulting equations for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. However, the equation for $\boldsymbol{\theta}$ is nonlinear in the variance vector, so that it is not possible to achieve closed form expressions for $\boldsymbol{\theta}$. Therefore, solutions have to be obtained through iterative algorithms.

One approach to obtain iteratively maximum likelihood estimates consists in computing a GLS estimate for the fixed-effect vector $\boldsymbol{\beta}$, that is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{y}. \tag{2.12}$$

This is the maximum likelihood estimator of the regression coefficient vector, that is the same obtained by equating to zero the score function with respect to $\boldsymbol{\beta}$ (2.10). Then, the variance components are obtained by optimizing the profiled likelihood, that is the likelihood function achieved by substituting the estimate of $\boldsymbol{\beta}$ from the previous step. This optimization usually is accomplished through the Expectation-Maximization algorithm (Dempster et al., 1981; Laird et al., 1987). This is an iterative procedure to compute maximum likelihood estimates in the case of incomplete-data problem. Instead of maximizing the marginal likelihood associate to the data vector $\mathbf{y}$, Eq. 2.8, the EM algorithm maximizes the joint likelihood of the *complete dataset*, $(\mathbf{y}, \mathbf{u})$, that is that composed by the observed data and the unobserved (or missing) data. Starting with some initial values for the parameters, the EM algorithm consists of iterating two steps: the *Expectation* step and the *Maximization* step.

1. In the E-STEP, the current vector of estimated parameters, $\hat{\boldsymbol{\theta}}^{(h)}$, is used to evaluate the distribution of the unobserved data conditional

to the observed data, $\mathbf{u}|\mathbf{y}$. Hence, the expected value of the joint log-likelihood, conditional to the data vector $\mathbf{y}$, is obtained.

2. The M-STEP consists in computing a new value of $\hat{\boldsymbol{\theta}}^{(h+1)}$ by maximizing the expected value of the likelihood function evaluated in the E-step.

E-STEP and M-STEP are iterated until convergence. A more formal explanation of the EM algorithm is provided in section 4.2.

Restricted maximum likelihood

The problem with the maximum likelihood estimates of the variance components is that the procedure does not take into account the degree of freedom lost in estimating $\boldsymbol{\beta}$. That is, in the variance estimators, the vector of true parameters $\boldsymbol{\beta}$ is replaced by its estimator $\hat{\boldsymbol{\beta}}$. However, in this way, the uncertainty about the regression coefficients is not taken into account. As a result, the variance components are estimated with bias in a downward direction (Harville, 1977; Patterson and Thompson, 1971).[2]

In order to produce unbiased variance estimates, Harville (1977) and Patterson and Thompson (1971) proposed a variant of the FML method, the Restricted maximum likelihood.

The REML method obtains the estimate of the variance components by optimizing the likelihood of $\boldsymbol{\theta}$ not corresponding to $\mathbf{y}$ but to a complete set of error contrasts for the responses, that is on a set of $n-k$ linearly independent error contrasts, $\mathbf{a}^{\mathrm{T}}\mathbf{y}$, where k is the number of fixed-effects, such that the function does not contain any fixed effects.[3] More specifically, the vector $\mathbf{a}$ is chosen such that $\mathbf{a}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta} = \mathbf{0}$, and, therefore, $\mathbf{a}^{\mathrm{T}}\mathbf{y} = \mathbf{a}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta} + \mathbf{a}^{\mathrm{T}}\mathbf{Z}\mathbf{u}$ does not contain any terms in $\boldsymbol{\beta}$.

[2] Raudenbush and Bryk (2002) found out that the maximum likelihood estimator of the between-group variance in a random-intercept model has approximately a bias factor of $(J-k)/J$, where J is the number of groups and k is the number of fixed effects.

[3] It is easy to show that the log-likelihoods corresponding to different sets of error contrasts contain the same information and, as a consequence, the same maxima in that they differ only by a constant (Longford, 1993).

In practice, the difference between FML and REML estimates becomes negligible as the number of groups increases. Moreover, the Restricted likelihood function cannot be used to construct a chi-square test to compare two nested models, and it is computationally less easy than the FML.
The REML theory can rely also on Bayesian arguments.

2.4.2 Iterative GLS

The most common method among those based on least squares is the Iterative Generalized Least Squares (Goldstein, 2010, 1986). The GLS estimate for $\boldsymbol{\beta}$ is given by:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{y}$$

that is the best linear unbiased estimator (BLUE). However, since the variance-covariance matrix, $\boldsymbol{\Omega}$, is typically unknown, this formula is not computable. For this reason, the GLS procedure is conducted iteratively. Starting from some initial values for the fixed parameters (generally the OLS estimates), the "raw" residuals are constructed as:

$$\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}.$$

By noting that

$$\mathrm{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathrm{T}}) = \mathrm{Var}(\mathbf{y}) = \boldsymbol{\Omega},$$

the procedure applies the GLS method on the following model:

$$\mathrm{vec}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathrm{T}}) = f(\theta) + \mathbf{r}$$

where $\mathrm{vec}(\mathbf{e}\mathbf{e}^{\mathrm{T}})$ constitutes the dependent variable, $\mathbf{r}$ is the residual vector, and $f(\theta)$ is a vector of functions containing coefficients to be estimated, that are the variance components, and, as covariates, vectors of 0 and 1 depending on the structure of $\boldsymbol{\Omega}$. Therefore, the variance components in θ, estimated through GLS, are then used to obtain new estimates of the fixed effects. The procedure goes on until convergence.
Under the assumption of normality, the IGLS procedure produces maximum

likelihood estimates. Moreover, since the IGLS estimates of variance components are biased, as the ML estimates, a Restricted version exists, the Restricted Iterative Generalized Least Squares (RIGLS).

2.5 Predicting the random effects

The random effects, $\mathbf{u}_j$, are not parameters for the statistical model. Nevertheless, numerical values are assigned to them based on the available information, namely the data vector $\mathbf{y}$. Since they are random variables, instead of parameters, the assigned values will not be properly *estimates* but *predictions* of their unobservable values (Henderson, 1953; Searle et al., 1992). These values can be useful, for example, for conducting inference on particular groups, or model diagnostic, such as for finding outlying groups or checking assumptions about random effects.

Skrondal and Rabe-Hesketh (2009) provide an overview on the large existing literature about prediction of random effects and responses in multilevel linear model. When the model parameters are (or treated as) known, they list four different philosophical approaches, two Bayesian and two frequentist.

- Bayesian approach (Lindley and Smith, 1972; Fearn, 1975; Smith, 1973): inference regarding random effects is based on their posterior distribution given the observed data and the prior distribution representing uncertainty about $\mathbf{u}$.

- Empirical Bayesian approach (Strenio et al., 1983; Morris, 1983): inference regarding the random effects is conducted by jointly sampling $\mathbf{u}$ and $\mathbf{y}$. In this case, the distribution of random effects represents their variation in the population.

- Frequentist prediction (Searle et al., 1992): inference regarding the random effects is viewed as prediction of the unobserved realizations of random variables. This approach allows also an empirical Bayesian justification.

- Frequentist estimation (Henderson, 1953; Searle et al., 1992): the random effects are treated as fixed parameters, thus inference on them typically consists of maximum likelihood estimation.

In the following, the most commonly used predictions of the random effects are presented.

2.5.1 BLUP

The problem is to assign a value, *prediction*, to the unobserved realization of the random vector $\mathbf{u}$. In order to derive the *best predictor*, the criterion of minimum variance used for estimating parameters, since they are fixed values, is replaced by the minimum mean square criterion for the realized value of a random variable. In other words, the best predictor for the random vector $\mathbf{u}$ is the value that minimizes

$$\mathrm{E}\big[(\hat{\mathbf{u}}-\mathbf{u})^{\mathrm{T}}(\hat{\mathbf{u}}-\mathbf{u})\big].$$

The best predictor corresponds to the conditional mean of the random effects vector given the data, $\hat{\mathbf{u}} = \mathrm{E}(\mathbf{u}|\mathbf{y})$. The minimum mean square can be rewritten as

$$\begin{aligned}\mathrm{E}\big[(\hat{\mathbf{u}}-\mathbf{u})^{\mathrm{T}}(\hat{\mathbf{u}}-\mathbf{u})\big] &= \mathrm{E}\big[(\hat{\mathbf{u}}-\mathrm{E}(\mathbf{u}|\mathbf{y})+\mathrm{E}(\mathbf{u}|\mathbf{y})-\mathbf{u})^{\mathrm{T}}(\hat{\mathbf{u}}-\mathrm{E}(\mathbf{u}|\mathbf{y})+\mathrm{E}(\mathbf{u}|\mathbf{y})-\mathbf{u})\big]\\ &= \mathrm{E}\Big[\big(\hat{\mathbf{u}}-\mathrm{E}(\mathbf{u}|\mathbf{y})\big)^{\mathrm{T}}\big(\hat{\mathbf{u}}-\mathrm{E}(\mathbf{u}|\mathbf{y})\big)\Big]\\ &+ \mathrm{E}\Big[\big(\mathrm{E}(\mathbf{u}|\mathbf{y})-\mathbf{u}\big)^{\mathrm{T}}\big(\mathrm{E}(\mathbf{u}|\mathbf{y})-\mathbf{u}\big)\Big]\\ &+ 2\mathrm{E}\Big[\big(\hat{\mathbf{u}}-\mathrm{E}(\mathbf{u}|\mathbf{y})\big)^{\mathrm{T}}\big(\mathrm{E}(\mathbf{u}|\mathbf{y})-\mathbf{u}\big)\Big].\end{aligned} \tag{2.13}$$

By expressing the latter term as

$$\mathrm{E}_{\mathbf{y}}\Big\{\mathrm{E}\Big[\big(\hat{\mathbf{u}}-\mathrm{E}(\mathbf{u}|\mathbf{y})\big)^{\mathrm{T}}\big(\mathrm{E}(\mathbf{u}|\mathbf{y})-\mathbf{u}\big)|\mathbf{y}\Big]\Big\},$$

given the result $\mathrm{E}(\mathbf{u}) = \mathrm{E}_{\mathbf{y}}\big[\mathrm{E}(\mathbf{u}|\mathbf{y})\big]$, it is trivial to note that it is equal to zero. By noting also that the second term of (2.13) does not depend on $\hat{\mathbf{u}}$, minimizing the mean square with respect to $\hat{\mathbf{u}}$ is equivalent to minimize

$$\mathrm{E}\Big[\big(\hat{\mathbf{u}} - \mathrm{E}(\mathbf{u}|\mathbf{y})\big)^{\mathrm{T}}\big(\hat{\mathbf{u}} - \mathrm{E}(\mathbf{u}|\mathbf{y})\big)\Big].$$

Therefore, the best predictor for $\mathbf{u}$ is

$$\hat{\mathbf{u}} = \mathrm{E}(\mathbf{u}|\mathbf{y}). \tag{2.14}$$

Note that the best predictor is unbiased, not in the classical sense, but in the sense that its expected value is equal to that of the random variable that is predicting:

$$\mathrm{E}_{\mathbf{y}}(\hat{\mathbf{u}}) = \mathrm{E}_{\mathbf{y}}\big[\mathrm{E}(\mathbf{u}|\mathbf{y})\big] = \mathrm{E}(\mathbf{u}).$$

Since $\mathbf{u} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Gamma})$ and $\mathbf{y} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Omega})$,[4] their jointly distribution is

$$\mathbf{u}, \mathbf{y} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{X}\boldsymbol{\beta} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Gamma} & \boldsymbol{\Gamma}\mathbf{Z}^{\mathrm{T}} \\ \mathbf{Z}\boldsymbol{\Gamma} & \boldsymbol{\Omega} \end{bmatrix}\right),$$

where

$$\boldsymbol{\Gamma}\mathbf{Z}^{\mathrm{T}} = \mathrm{Cov}(\mathbf{u}, \mathbf{y}^{\mathrm{T}}) = \mathrm{Cov}(\mathbf{u}, \mathbf{u}^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}) = \mathrm{Var}(\mathbf{u})\mathbf{Z}^{\mathrm{T}}.$$

Then, the best linear predictor for $\mathbf{u}$, $\mathrm{BLP}(\mathbf{u})$, is

$$\hat{\mathbf{u}} = \mathrm{E}(\mathbf{u}|\mathbf{y}) = \boldsymbol{\Gamma}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \tag{2.15}$$

for the property of the joint normal distribution.[5]

[4] It can be showed that the same expression for the best linear predictor is valid without assumption of normality (Searle et al., 1992).

[5] If

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix}\right),$$

then the conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2$ is

$$\mathbf{X}_1|\mathbf{X}_2 \sim N\Big(\boldsymbol{\mu}_1 + \mathbf{V}_{12}\mathbf{V}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2), \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\Big). \tag{2.16}$$

2.5.2 Prediction of expected responses

In a mixed-effects model, the main interest is estimating or, better, predicting the responses y_{ij} for some values of the covariates, that is the prediction the linear combination of the fixed effects and the realized unobservable value of the random effects, $\mathbf{X}_0^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{Z}_0^{\mathrm{T}}\mathbf{u}$, for some known covariate matrices $\mathbf{X}_0$ and $\mathbf{Z}_0$ (Harville, 1977). We will wish finding the Best Linear Unbiased Predictor (BLUP) of the linear combination, that is, a predictor minimizing the mean squared prediction error, having a linear form in $\mathbf{y}$, $\mathbf{a} + \mathbf{B}\mathbf{y}$, and being unbiased in the sense that

$$\mathrm{E}(\mathbf{X}_0^{\mathrm{T}}\widehat{\boldsymbol{\beta} + \mathbf{Z}}_0^{\mathrm{T}}\mathbf{u}) = \mathrm{E}(\mathbf{X}_0^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{Z}_0^{\mathrm{T}}\mathbf{u}) = \mathrm{E}(\mathbf{X}_0^{\mathrm{T}}\boldsymbol{\beta}).$$

From the unbiasedness, it follows that

$$\mathrm{E}(\mathbf{a} + \mathbf{F}\mathbf{y}) = \mathbf{a} + \mathbf{F}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}_0^{\mathrm{T}}\boldsymbol{\beta},$$

so $\mathbf{a} = 0$ and $\mathbf{X}_0^{\mathrm{T}} = \mathbf{F}\mathbf{X}$. Consequently, the predictor results

$$\mathbf{X}_0^{\mathrm{T}}\widehat{\boldsymbol{\beta} + \mathbf{Z}}_0^{\mathrm{T}}\mathbf{u} = \mathbf{F}\mathbf{y}.$$

Therefore, the problem is choosing $\mathbf{F}$ that minimizes the mean squared prediction error subjected to $\mathbf{F}\mathbf{X} = \mathbf{X}_0^{\mathrm{T}}$ (Searle et al., 1992). The solution to this problem is

$$\mathrm{BLUP}(\mathbf{X}_0^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{Z}_0^{\mathrm{T}}\mathbf{u}) = \mathrm{BLUE}(\mathbf{X}_0^{\mathrm{T}}\boldsymbol{\beta}) + \mathrm{BLUP}(\mathbf{Z}_0^{\mathrm{T}}\mathbf{u}) \tag{2.17}$$

where $\mathrm{BLUE}(\mathbf{X}_0^{\mathrm{T}}\boldsymbol{\beta}) = \mathbf{X}_0^{\mathrm{T}}\hat{\boldsymbol{\beta}} = \mathbf{X}_0^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{y}$ is the Best Linear Unbiased Estimate of $\mathbf{X}_0^{\mathrm{T}}\boldsymbol{\beta}$. As a generalized version of the Gauss-Markov theorem states, $\mathbf{X}_0^{\mathrm{T}}\hat{\boldsymbol{\beta}}$ is BLUE of $\mathbf{X}_0^{\mathrm{T}}\boldsymbol{\beta}$ in the sense that if $\mathbf{c} + \mathbf{G}\mathbf{y}$ is any other linear unbiased estimator of $\mathbf{X}_0^{\mathrm{T}}\boldsymbol{\beta}$, $\mathrm{Var}(\mathbf{X}_0^{\mathrm{T}}\hat{\boldsymbol{\beta}}) \leq \mathrm{Var}(\mathbf{c} + \mathbf{G}\mathbf{y})$ (Harville, 1977). Moreover,

$$\text{BLUP}(\mathbf{u}) = \mathbf{\Gamma}\mathbf{Z}^{\mathrm{T}}\mathbf{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

is the Best Linear Unbiased Predictor of $\mathbf{u}$ with $\mathbf{X}\boldsymbol{\beta}$ replaced by its BLUE.

2.5.3 Empirical Bayes prediction

Given the maximum likelihood estimates for the parameters, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$, the Empirical Bayes predictors of the random effects $\mathbf{u}$ are the means of their posterior distributions. The posterior distribution is named *empirical* since the parameter estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ are plugged in.
By using the Bayes theorem, the empirical posterior distribution is obtained as

$$\pi(\mathbf{u}|\mathbf{y};\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}}) = \frac{f(\mathbf{y}|\mathbf{u};\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}})f(\mathbf{u};\hat{\mathbf{\Sigma}})}{f(\mathbf{y};\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}})}$$

where the prior distribution of the random effects, $f(\mathbf{u};\hat{\mathbf{\Sigma}})$, is combined with the data $\mathbf{y}$. The difference between the Bayesian approach and the Empirical Bayesian approach is that, in the former, prior distributions for the parameters are specified, instead of plugging in their estimates; then the posterior distribution of the random effects is obtained marginally with respect to these parameters. Thus, the Empirical Bayes predictor for the random effects is:

$$\hat{\mathbf{u}}^{EB} = \mathrm{E}(\mathbf{u}|\mathbf{y};\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}}) = \int \mathbf{u}\pi(\mathbf{u}|\mathbf{y};\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}})\mathrm{d}\mathbf{u}.$$

that is equivalent to the expression of the BLUP (2.15).

2.5.4 Shrunken estimates

Consider the simplest 2-level random-intercept model

$$\begin{gathered} y_{ij} = \beta_0 + u_j + \epsilon_{ij} \\ u_j \sim N(0,\sigma_u^2) \qquad \epsilon_{ij} \sim N(0,\sigma^2), \end{gathered} \tag{2.18}$$

which in matrix form is

$$\mathbf{y}_j = \mathbf{1}_{n_j}\beta_0 + \mathbf{1}_{n_j}u_j + \boldsymbol{\epsilon}_j$$

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_J \end{bmatrix} \sim N(\mathbf{0}, \sigma_u^2\mathbf{I}_J) \qquad \boldsymbol{\epsilon}_j \sim N(\mathbf{0}, \sigma^2\mathbf{I}_{n_j}).$$

This is also called *variance components model* (Goldstein, 2009) because the variance of the response is the sum of the within-group variance and the between-group variance:

$$\mathrm{Var}(y_{ij}) = \mathrm{Var}(u_j + \epsilon_{ij}) = \sigma_u^2 + \sigma^2$$

The random effects u_j in this model represent the deviation of the j-th group average from the overall mean β_0. Therefore, their prediction is provided by the data through the mean for the jth group, $\bar{y}_j = \sum_i^{n_j} y_{ij}/n_j$. In fact, intuitively, if $\bar{y}_j$ is higher than the overall average $\bar{y}$, u_j should be positive (Searle et al., 1992). So, a reasonable predictor of u_j is given by

$$\hat{u}_j = E(u_j|\bar{y}_j).$$

Under the normality assumptions in (2.18), the joint distribution is:

$$\begin{bmatrix} u_j \\ \bar{y}_j \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ \beta_0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 + \frac{\sigma^2}{n_j} \end{bmatrix} \right)$$

where $\mathrm{Cov}(u_j, \bar{y}_j) = \mathrm{E}(u_j \sum_{i=1}^{n_j} y_{ij})/n_j = \mathrm{E}(u_j y_{ij}) = \mathrm{E}(u_j^2) = \sigma_u^2$. Applying the property (2.16) results in

$$E(u_j|\bar{y}_j) = \frac{n_j\sigma_u^2}{\sigma^2 + n_j\sigma_u^2}(\bar{y}_j - \beta_0).$$

The term $\bar{y}_j - \hat{\beta}_0$ represents the mean "raw" or total residual for the j-th group.

As shown in the previous section, in order to obtain the best linear unbiased predictor (BLUP) of the random effects, the parameters β_0 has to be replaced by its GLS or maximum likelihood estimate. So that the best linear unbiased estimate is:

$$\text{BLUP}(u_j) = \frac{n_j\sigma_u^2}{\sigma^2 + n_j\sigma_u^2}(\bar{y}_j - \hat{\beta}_0).$$

Moreover, it may be interesting to predict also the linear combination $\mu_j = \beta_0 + u_j$ representing the *overall* average for the j-th group. The best linear unbiased prediction for it is given by

$$\text{BLUP}(\mu_j) = \hat{\mu}_j = \text{BLUE}(\beta_0) + \text{BLUP}(u_j) = \hat{\beta}_0 + \frac{n_j\sigma_u^2}{\sigma^2 + n_j\sigma_u^2}(\bar{y}_j - \hat{\beta}_0).$$

The factor $\frac{n_j\sigma_u^2}{\sigma^2+n_j\sigma_u^2}$ is often called "shrinkage factor" (Goldstein, 2009) and BLUP($\mathbf{u}_j$) *shrinkage estimate.* The shrinkage factor can be interpreted as the estimated reliability of the mean raw residual as a predictor of u_j. The name *shrinkage* is due to the fact that, since the factor takes values between 0 and 1 in absolute value, it shrinks the group mean, $\hat{\mu}_j$, toward the overall mean, β_0, by an amount depending on n_j and the variance components: as n_j increases and σ^2 decreases with respect to σ_u^2, the factor tends to one, and the group mean tend to dominate in magnitude on the population mean; as, instead, the group size decreases and σ^2 increases compared with σ_u^2, the reliability decreases since the "shrinkage factor" becomes closer to zero, and the group mean tends to the population mean. In Bayesian terms, the shrinkage factor pulls the Empirical Bayes predictor towards the mean of the prior distribution of the random effects, that is 0. Hence, when n_j decreases, the conditional distribution of the responses $f(\mathbf{y}_j; \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$ (that is the likelihood associated with the j-th group) becomes flat and uninformative compared with the prior distribution $f(\mathbf{u}_j; \hat{\mathbf{\Sigma}})$.
The *shrinkage estimates* are also often referred as *James-Stein's type estimates* (James and Stein, 1961) or *empirical Bayes estimates.*

Both the BLUP and the empirical Bayes predictor are biased conditionally to the random effect. In particular, for the variance component model (2.18), the conditional expectation of the predictor given the random intercept is

$$\mathrm{E}(\hat{\mathbf{u}}_j|\mathbf{u}_j) = \frac{n_j\sigma_u^2}{\sigma^2 + n_j\sigma_u^2}\mathbf{u}_j$$

In general, as shown in section 2.5.1, the best linear prediction and the empirical Bayes predictor are, instead, unconditionally unbiased.

Finally, the best linear predictor of the random effects requires point estimation of the fixed effects and the variance components to be plugged in. Though for special cases some corrections have been derived to adjust for the bias deriving by substituting estimates for parameters (see for references Skrondal and Rabe-Hesketh (2009)), when the estimates are consistent, this bias is expected to be small as the sample size becomes large.

2.5.5 Standard errors of the predictors

From the expression (2.15) for the BLUP, by using the identities

$$\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = [\mathbf{I} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}]\mathbf{y} = \mathbf{M}\mathbf{y}$$

and $\boldsymbol{\Omega}\mathbf{M}\boldsymbol{\Omega}^{-1} = \mathbf{M}$, where $\mathbf{M}$ is a symmetric and idempotent matrix, the standard error for the BLUP of $\mathbf{u}$ is derived as

$$\begin{aligned}\mathrm{Var}(\hat{\mathbf{u}}) &= \boldsymbol{\Gamma}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathrm{Var}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})\boldsymbol{\Omega}^{-1}\mathbf{Z}\boldsymbol{\Gamma} \\ &= \boldsymbol{\Gamma}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{M}\boldsymbol{\Omega}\mathbf{M}\boldsymbol{\Omega}^{-1}\mathbf{Z}\boldsymbol{\Gamma} = \boldsymbol{\Gamma}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{M}\mathbf{Z}\boldsymbol{\Gamma} \\ &= \boldsymbol{\Gamma}\mathbf{Z}^{\mathrm{T}}[\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}]\mathbf{Z}\boldsymbol{\Gamma}.\end{aligned} \tag{2.19}$$

This is the unconditional variance-covariance matrix of the random-effects prediction vector. Note that, through the second term, the expression in (2.19) takes into account the sampling variation of the estimates of the fixed coefficients, which is negligible in large samples (Goldstein, 2009).
However, Laird and Ware (1982) pointed out that this expression ignores the

variation of the random effects, and, for this reason, it would understate the variation in $\hat{\mathbf{u}}_j - \mathbf{u}_j$. Then, to assess the error of prediction of the random effects, they suggested to use instead

$$\begin{aligned} \mathrm{Var}(\hat{\mathbf{u}} - \mathbf{u}) &= \mathrm{Var}(\hat{\mathbf{u}}) + \boldsymbol{\Gamma} - 2\mathrm{Cov}(\hat{\mathbf{u}}, \mathbf{u}^{\mathrm{T}}) = \boldsymbol{\Gamma} - \mathrm{Var}(\hat{\mathbf{u}}) \\ &= \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{Z}\boldsymbol{\Gamma} + \boldsymbol{\Gamma}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{Z}\boldsymbol{\Gamma} \end{aligned} \tag{2.20}$$

since, from (2.19),

$$\mathrm{Cov}(\hat{\mathbf{u}}, \mathbf{u}^{\mathrm{T}}) = \mathrm{E}(\hat{\mathbf{u}}\mathbf{u}^{\mathrm{T}}) = \boldsymbol{\Gamma}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{M}\mathrm{E}(\mathbf{y}\mathbf{u}^{\mathrm{T}}) = \boldsymbol{\Gamma}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{M}\mathbf{Z}\boldsymbol{\Gamma} = \mathrm{Var}(\hat{\mathbf{u}}).$$

The form (2.20) represents the conditional or "comparative" variance-covariance matrix of $\hat{\mathbf{u}}$, as Goldstein (2009) called it because of its main use in making comparisons among groups. In fact, the comparative standard error is preferred to the unconditional one (2.19) to make inferences regarding the realized values of $\mathbf{u}$. On the other hand, for diagnostic purposes, such as standardizing the estimated residuals and detecting outlying clusters, $\mathrm{Var}(\hat{\mathbf{u}})$ should be used.

Both the standard errors take into account the sampling variability of the regression coefficient estimates, but they ignored the sampling variability associated to the variance parameter estimates, when they are plugged in. Only for simple models, there are some approximations that take into account the uncertainty about them (see Skrondal and Rabe-Hesketh (2009) for references). However, if consistent estimators for the variance components are used, the correction terms become small as the number of groups become large. Moreover, resampling procedure may be used to estimate standard errors of the prediction.

2.6 Developments and applications of the multilevel model

The general specification of multilevel models allows a large variety of applications. In the following, we briefly discuss some extensions of the models that may be particularly relevant to the case under scrutiny.

First of all, multilevel models are extended to binary, count, ordered categorical and multi-categorical responses (Stiratelli et al., 1984; Wong and Mason, 1985; Skrondal and Rabe-Hesketh, 2004). The multilevel model is also generalized to cross-classified data and multiple membership structures. These are cases where observations belong to multiple contexts simultaneously and these can be either simply nested or crossed. Multiple membership occurs when a level-1 unit belongs to more than one level-2 unit. Instead, the data show a crossed structure if, for example, students are grouped in schools and also in neighbourhoods which they come from; in this case neither "school" nor "neighbourhood" is above the other one in a hierarchical sense. Cross-classified random-effect models address this situation and allow the inclusion of predictors of more than one "classification" variable (Raudenbush, 1993; Goldstein, 1995).

Another interesting application of the multilevel model regards the multivariate case in which multivariate response variables belongs to the same model. A class of multivariate outcome models is that dealing with missing data problems, when, in particular, predictors are missing at random. These are latent variable models, where latent variables represent the unobserved data "completing" the observed data. Thus, the incomplete data is used to draw valid inferences about parameters that generate the complete data. It is possible to manage this missing data problem through hierarchical linear models, by considering the variables for each individual as "occasions of measurements", the level-1 units and individuals as level-2 units (Raudenbush and Bryk, 2002; Skrondal and Rabe-Hesketh, 2004). Moreover, when data are missing by design rather than at random, the multivariate multilevel model takes automatically into account this feature and allows to avoid special procedures for handling missing data. Also repeated measures data can be view as a specific case of the multilevel multivariate data. These, also called longitudinal or panel data, can be thought, in fact, as two-level data with occasions i at level-1 and units j at level 2. The main difference between the panel data and simple clustered data is that the level-1 units are ordered chronologically. In this case, often the usual assumption of independence among level-1 residuals does not hold, especially when measurements on the same unit are taken close together in time. This case can be handled by including correlation structures at level-1 (Goldstein, 2010). Moreover, it is possible also to allow heteroscedastic within-group errors through variance functions (Davidian and Giltinan, 1995).

Thus, in literature, the applications of multilevel models to longitudinal data consider occasions, that is the points in time, as the lowest level units and individuals as higher units. Therefore, any time dependence structure is assessed at the first level. How can we manage the case where different individuals are observed in different occasions? Our proposal is to treat individuals as level-1 units and time-points as level-2 units, that is to reverse the structure of (pseudo) panel data. However, this is not costless in that the assumption of independence among random effects is likely to not hold because they represent time-effects. For this reason, the classic multilevel framework needs to be modified to cope with this case. Chapter 4 tries to extend the multilevel modelling framework to deal with time dependence at the second level, and this represents the main theoretical contribution of the present thesis.

Actually, Browne and Goldstein (2010) remove the independence assumption among level-2 disturbances and model the correlation between pairs of clusters through an explicit function of distance. Although this approach is especially suitable in presence of spatial correlations, in theory, it could be used to model our data. However, it has been developed in a Bayesian framework that requires the specification of prior distributions for the parameters and estimation through MCMC methods; these are computer intensive and can be unfeasible when the number of level-2 units is large. We chose, instead, to adopt a frequentist approach by deriving and implementing maximum likelihood estimators with known desirable properties.

Chapter 3

Hedonic regression model and multilevel model for Tribal art prices

In literature, art prices are modelled through the hedonic regression model, a classic fixed-effect model, and, as remarked in section 1.2, also Tribal art prices can be modelled in this way.
Our idea, instead, is to consider the influence of time effects on prices through a different approach. Since we observe different artworks sold at every auction, Tribal art data do not constitute either a panel or a time series. Rather, they have a two-level structure in that items, level-1 units, are grouped in time points, level-2 units. Hence, we propose to exploit the multilevel model to explain heterogeneity of prices among time points. We have chosen to take the semesters rather than the auction dates as time points, and this choice is due mainly to three reasons:

1. the auctions are organized in two sessions, one during the winter and one during the summer; each of them is constituted by from two to four auctions organized quite close in time;

2. in general, the stakeholders look at the performance of the previous semester;

3. the auction dates are not equally spaced in time and, as we will see in chapter 4, this feature will result important to handle time dependence.

In this chapter, we will apply and compare extensively the traditional hedonic regression model and the multilevel model on the Tribal art dataset. We will see that the two-level model and the hedonic regression model produce similar results in terms of estimates and residuals.

3.1 Models with no covariates

Let us initially ignore the grouping structure of the data and assume the simple model

$$\log_{10}(y_{it}) = \beta_0 + \epsilon_{it}, \quad i = 1, \ldots, n_t, \quad t = 1, \ldots, T, \quad \epsilon_{it} \sim N(0, \sigma^2) \quad (3.1)$$

where the dependent variable is the logarithm of the observed hammer price for the observation i in the semester t, namely $y_{it} = \text{PRICE}_{it}$, and β_0 is the overall mean price. The number of semesters included in the database is $T = 27$; n_t is the number of items sold in the semester t and varies between 119 and 915; $n = \sum_{t=1}^{T} n_t = 14124$ is the total sample size.[1]
As said before, Tribal art data do not constitute a proper panel where y_{it} and $y_{i(t+1)}$ represent the price of item i observed at successive time points. Rather, since different objects are sold at every auction, y_{it} and $y_{i(t+1)}$ indicate the prices of two different objects at time points t and $t + 1$. In particular, some objects are observed at the same time t, so that, for example, y_{11} indicates the price of the first object observed at time 1, y_{21} the price of the second object observed at time 1, $y_{n_1 1}$, the price of the last object observed at time 1, y_{12} the price of the first object observed at time 2, and so on.

The model (3.1), that we call "null model", is fitted through the maximum likelihood method in order to compare the results with those obtained from fitting the multilevel models. The results are shown in the first column of Table 3.1. The other two columns of Table 3.1 contain the results of the

[1] The database includes more than 20000 observations, but only the sold items are used in the analysis.

models (3.2) and (3.3). The standard errors of the estimates of all models in this work, reported in parentheses, are computed through a wild bootstrap procedure, explained in subsection 3.2.1.

The problem with ignoring the grouped structure of data is clear from the boxplots of the residuals grouped by semester displayed in Figure 3.2a: the "group effects" are incorporated into the residuals, so that the within-group variability, σ^2, is overestimated.

The semester-effects may be accounted by allowing the mean price of each group to be represented by a separate parameter. This leads to the following *fixed-effect* model:

$$\log_{10}(y_{it}) = \beta_t + \epsilon_{it} \tag{3.2}$$

where β_t is the mean price of the semester t and $y_{it} = \text{PRICE}_{it}$. We will call it "FE-intercept".
As shown in the second column of Table 3.1, the estimated residual variability obtained for the FE-intercept model (3.2) is smaller than that in the null model (0.499 versus 0.443). Moreover, the boxplots in Figure 3.2b show that the residuals are centered around zero and smaller than those in Figure 3.2a. Hence, the fixed-effect model has somehow accounted for the semester-effects. Nevertheless, as also explained in section 2.2.2, the main drawback of the fixed-effect model, is that it includes many parameters. On the contrary, in a random-effect model that treats the group-effects as random variations around a population mean, the number of parameters does not increase with the number of groups. Also, an estimate for the between-group variability is provided.

A simple random-effect model for our data is:

$$\begin{aligned} \log_{10}(y_{it}) &= \beta_{0t} + \epsilon_{it} \\ \beta_{0t} &= \beta_0 + u_t \end{aligned} \tag{3.3}$$

where, $y_{it} = \text{PRICE}_{it}$, u_t is a random variable representing the deviation from the population mean of the mean price for the t-th semester. This is a multilevel model, where items, labeled by i, are the first-level observations, and the semesters, labeled by t, are the second-level observations.

The usual assumptions for this model are:

$$\begin{aligned} u_t &\sim \text{NID}(0, \sigma_u^2), \\ \epsilon_{it} &\sim \text{NID}(0, \sigma^2), \\ u_t &\perp \epsilon_{it}, \end{aligned}$$

for all $i = 1, \ldots, n_t$ and for all $t = 1, \ldots, T$, where the sign "$\perp$" stands for independence and "NID" for white noise process.
By fitting the two-level random intercept model (3.3), that we call "RE-intercept", through the maximum likelihood method, we obtain the parameter estimates in the third column of Table 3.1.
Let us compare the FE-intercept and the RE-intercept models. First of all, the within-group variability, σ^2, has the same estimate for both models, but the multilevel model picks out also the between-group variance, σ_u^2. We carried out a likelihood ratio test between the null model and the RE-intercept model to assess the significance of the between-group variance. The likelihood ratio test is asymptotically χ^2-distributed with one degree od freedom. However, since the null hypothesis of zero variance is on the boundary of the feasible parameter space, the p-value to be used is half the one obtained from the tables of the chi-squared distribution (Self and Liang, 1987). In this case, the test gives a value of 1572.582 which is highly significant, and this assures that the between-group variance is significantly different from zero.
The additional variance component in the RE-intercept model allows to calculate the proportion of the total variability of prices explained by the variability among semesters. In this case, it results $100 * \sigma_u^2/(\sigma_u^2 + \sigma^2) = 13.5\%$, that, in a two-level random-intercept model, corresponds to the *Intra-class correlation* (ICC), the correlation between two observations in the same semester. As already mentioned, the existence of a non-zero intra-class correlation reveals the inadequacy of traditional modelling framework (Goldstein, 2010).

The plot in Figure 3.1 highlights the closeness of the semester-effects of the two models. These values are estimated coefficients for the fixed-effect model, $\hat{\beta}_t$, and BLUP (subsection 2.5.1) for the linear combination of the random effects and the overall intercept for the random-intercept model, $\hat{\beta}_0 + \hat{u}_t$. The fact that they overlap can be justified analytically. In the FE-intercept model, the estimates of the fixed effects β_t represent the means of each cluster. In fact, by considering the cluster means of the model (3.2),

$$\bar{y}_t = \beta_t + \bar{\epsilon}_t,$$

the estimates of the time-specific intercepts correspond to

$$\hat{\beta}_t = \bar{y}_t.$$

On the other hand, the group mean for the RE-intercept model is obtained as

$$\hat{\beta}_{0t} = \hat{\beta}_0 + \hat{u}_t = \hat{\beta}_0 + \hat{\lambda}_t(\bar{y}_t - \hat{\beta}_0),$$

where

$$\hat{\lambda}_t = \frac{n_t \hat{\sigma}_u^2}{\hat{\sigma}^2 + n_t \hat{\sigma}_u^2}$$

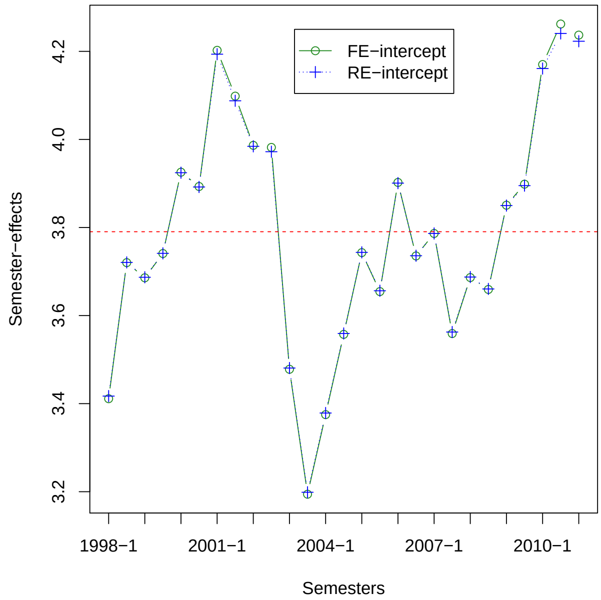
is the *shrinkage factor*, as derived in subsection 2.5.4, that pulls the group mean towards the overall mean by an amount depending on n_t and the variance components. In practice, the plot of the intercepts of the random-effect model should appear *smoother* because of the shrinkage. Instead, since, in our case, the sample sizes of the groups are big compared to the variance components, the shrinkage factor tends to one for each t. In particular, $\hat{\lambda}_t = 0.95$ for the smallest group that includes 119 units and $\hat{\lambda}_t = 0.99$ for the biggest group with 915 units. Therefore, each group-specific mean dominates in magnitude on the population mean so that $\hat{\beta}_t \simeq \hat{\beta}_{0t}$ and the two plots are almost superimposed.

Finally, the three models have been compared through the information criteria, the *Akaike Information Criterion*, AIC $= -2\log L(\theta) + 2p$, and the *Bayesian Information Criterion*, BIC $= -2\log L(\theta) + p\log n$, where $L(\theta)$ is the likelihood corresponding to the parameter set θ of dimension p, and n is the total sample size. In general, the model with the smallest AIC/BIC is chosen as the one which fits best. Both the FE-intercept model and the

Table 3.1: Parameter estimates for the null model (3.1), the FE-intercept model (3.2) and the RE-intercept model (3.3). Bootstrap standard errors are indicated in parentheses.

	Null model Estimate (s.e.)	**FE-intercept** Estimate (s.e.)	**RE-intercept** Estimate (s.e.)
$\hat{\sigma}^2$	0.499 (0.000)	0.443 (0.000)	0.443 (0.002)
$\hat{\sigma}_u^2$	-	-	0.069 (0.006)
ICC	-	-	0.135
AIC	30289	28625	28718
BIC	30304	28837	28741
# param.	2	28	3
$\hat{\beta}_0$	3.713 (0.005)	-	3.79 (0.052)
1998-1	-	3.411 (0.027)	-0.373 (0.024)
1998-2	-	3.720 (0.019)	-0.070 (0.019)
1999-1	-	3.686 (0.021)	-0.103 (0.020)
1999-2	-	3.741 (0.020)	-0.049 (0.019)
2000-1	-	3.926 (0.021)	0.134 (0.021)
2000-2	-	3.893 (0.019)	0.102 (0.020)
2001-1	-	4.202 (0.028)	0.403 (0.026)
2001-2	-	4.098 (0.039)	0.297 (0.034)
2002-1	-	3.987 (0.031)	0.194 (0.029)
2002-2	-	3.982 (0.052)	0.182 (0.042)
2003-1	-	3.478 (0.028)	-0.310 (0.029)
2003-2	-	3.194 (0.024)	-0.592 (0.024)
2004-1	-	3.375 (0.028)	-0.412 (0.032)
2004-2	-	3.557 (0.026)	-0.231 (0.027)
2005-1	-	3.743 (0.024)	-0.047 (0.025)
2005-2	-	3.654 (0.031)	-0.134 (0.029)
2006-1	-	3.902 (0.033)	0.110 (0.030)
2006-2	-	3.735 (0.023)	-0.055 (0.023)
2007-1	-	3.786 (0.025)	-0.004 (0.025)
2007-2	-	3.559 (0.036)	-0.228 (0.035)
2008-1	-	3.686 (0.035)	-0.103 (0.032)
2008-2	-	3.659 (0.028)	-0.130 (0.026)
2009-1	-	3.851 (0.038)	0.059 (0.034)
2009-2	-	3.898 (0.058)	0.105 (0.052)
2010-1	-	4.170 (0.039)	0.371 (0.040)
2010-2	-	4.262 (0.071)	0.450 (0.059)
2011-1	-	4.237 (0.045)	0.433 (0.043)

Figure 3.1: Plot of the time-specific intercepts: $\hat{\beta}_t$ for the FE-intercept and $\hat{\beta}_{0t}$ for the RE-intercept model. The red horizontal line represents $\hat{\beta}_0$.



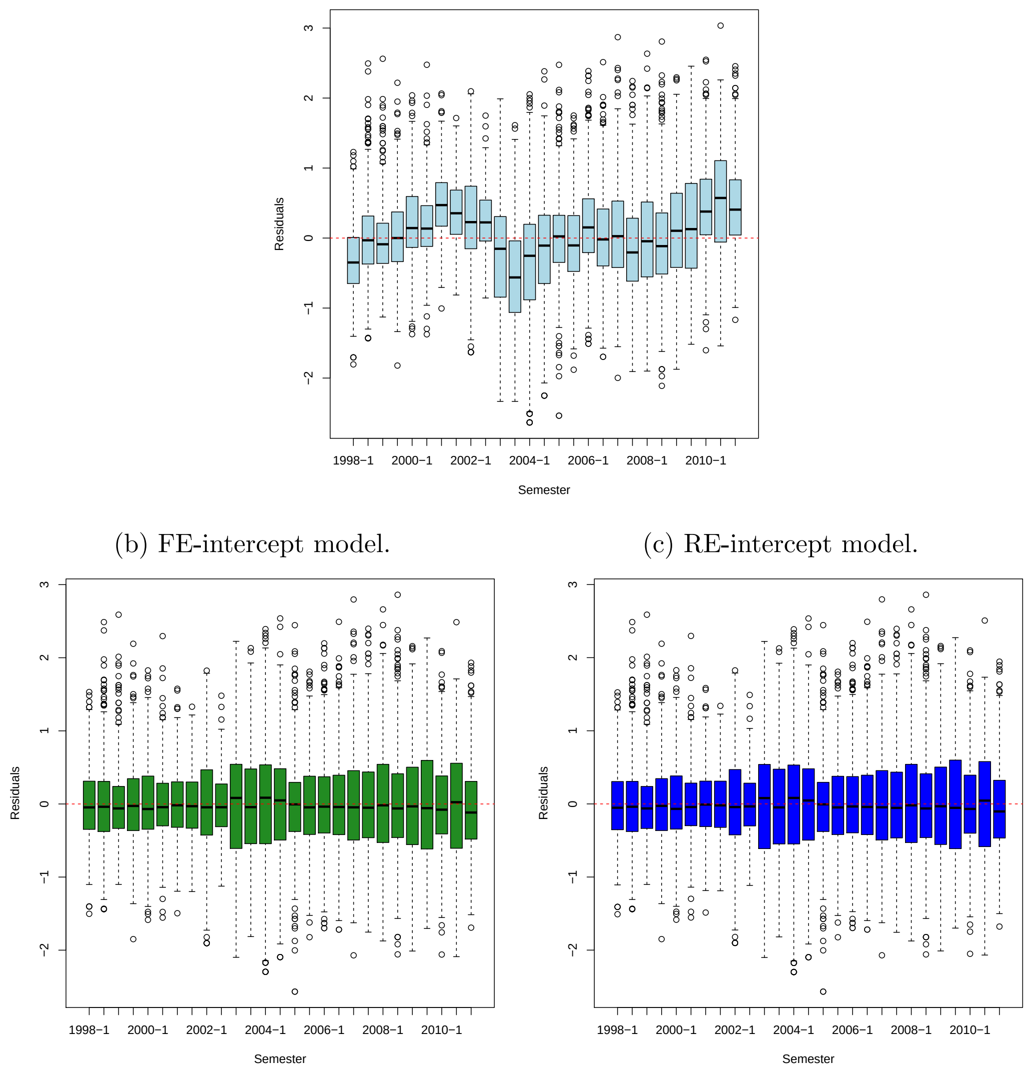
RE-intercept model have the criteria values smaller than the null model (Table 3.1). However, the AIC and the BIC for the FE-intercept model and the RE-intercept model are contrasting; the BIC tends to favour the RE-intercept model which is more parsimonious than the FE-intercept one.

3.2 Models with covariates

As a further modelling step, we have included the covariates that are assumed important into the models. We started with an extended set of variables that has been reduced through a backward elimination procedure by means of significance tests (*t-test* and *F-test*) and information criteria.

The final hedonic regression model for modelling the price of artworks, that

Figure 3.2: Boxplots of residuals by semester of the null model, FE-intercept and RE-intercept models.



(a) Null model.

(b) FE-intercept model.

(c) RE-intercept model.

we call “FE-hedonic”, is:

$$\begin{aligned} \log_{10}(y_{it}) = \beta_t &+ \beta_1 \mathrm{OGG}_{it} + \beta_2 \mathrm{REG}_{it} + \beta_3 \mathrm{MATP}_{it} \\ &+ \beta_4 \mathrm{CPAT}_{it} + \beta_5 \mathrm{CATD}_{it} + \beta_6 \mathrm{CABS}_{it} + \beta_7 \mathrm{CABC}_{it} \\ &+ \beta_8 \mathrm{CAES}_{it} + \beta_9 \mathrm{CAST}_{it} + \beta_{10} \mathrm{CAIL{:}CAAI}_{it} \\ &+ \beta_{11} \mathrm{ASNC{:}ASLU}_{it} + \epsilon_{it} \end{aligned} \tag{3.4}$$

where the meaning of the labels is reported in Tables 1.1, 1.2 and 1.3, and the sign ":" indicates the interactions between two variables. Note that all the variables in the model are categorical, therefore, each of them results in as many dummy variables as the number of levels of each covariate minus one. For example, the variable *Type of object* includes 12 levels (Table 1.1), hence OGG specifies a set of 11 dummy variables with the first level selected as the baseline, in this case the level "Furniture".

The two-level model with the same set of covariates as (3.4), that we call "RE-hedonic", has the form:

$$\begin{aligned}\log_{10}(y_{it}) &= \beta_{0t} + \beta_1 \mathrm{OGG}_{it} + \beta_2 \mathrm{REG}_{it} + \beta_3 \mathrm{MATP}_{it} \\ &+ \beta_4 \mathrm{CPAT}_{it} + \beta_5 \mathrm{CATD}_{it} + \beta_6 \mathrm{CABS}_{it} + \beta_7 \mathrm{CABC}_{it} \\ &+ \beta_8 \mathrm{CAES}_{it} + \beta_9 \mathrm{CAST}_{it} + \beta_{10} \mathrm{CAIL{:}CAAI}_{it} \\ &+ \beta_{11} \mathrm{ASNC{:}ASLU}_{it} + \epsilon_{it} \\ \beta_{0t} &= \beta_0 + u_t\end{aligned} \tag{3.5}$$

where u_t is the random intercept for the semester t. The model assumes that:

$$\begin{aligned}&u_t|\mathbf{X}_t \sim \mathrm{NID}(0, \sigma_u^2), \\ &\epsilon_{it}|X_t \sim \mathrm{NID}(0, \sigma^2), \\ &u_t \perp \epsilon_{it}\end{aligned}$$

where $\mathbf{X}_t = \begin{bmatrix} \mathrm{OGG}_t & \mathrm{REG}_t & \dots \end{bmatrix}$.

The FE-hedonic model and the RE-hedonic model (Table 3.2) are very similar in terms of estimates. First of all, the Figure 3.3 shows that the time-effects are very close for the two models, as we expected because of the very high shrinkage factors observed in the previous section.
The likelihood ratio test between the multilevel model and its unrestricted model (the hedonic model without β_{0t} plus the intercept β_0) produces a value of 1509.526 confirming that the between-semester variance is significantly different from zero.

Table 3.2: Parameter estimates for the FE-hedonic model (3.4) and the RE-hedonic model (3.5). Bootstrap standard errors are indicated in parentheses.

	FE-hedonic Estimate (s.e.)	**RE-hedonic** Estimate (s.e.)
$\hat{\sigma}^2$	0.171 (0.000)	0.171 (0.001)
$\hat{\sigma}_u^2$	-	0.026 (0.003)
ICC	-	0.133
AIC	15347	15439
BIC	16087	15991
# param.	98	73
$\hat{\beta}_0$	-	2.256 (0.121)
Semester		
1998-1	1.993 (0.074)	-0.257 (0.021)
1998-2	2.117 (0.069)	-0.137 (0.016)
1999-1	2.177 (0.072)	-0.077 (0.019)
1999-2	2.380 (0.070)	0.123 (0.017)
2000-1	2.488 (0.071)	0.230 (0.016)
2000-2	2.449 (0.070)	0.192 (0.016)
2001-1	2.425 (0.063)	0.167 (0.019)
2001-2	2.279 (0.075)	0.023 (0.024)
2002-1	2.370 (0.068)	0.114 (0.016)
2002-2	2.141 (0.080)	-0.108 (0.030)
2003-1	2.072 (0.068)	-0.180 (0.017)
2003-2	1.985 (0.067)	-0.267 (0.016)
2004-1	1.960 (0.065)	-0.291 (0.018)
2004-2	2.079 (0.069)	-0.174 (0.017)
2005-1	2.255 (0.068)	0.001 (0.018)
2005-2	2.227 (0.068)	-0.027 (0.017)
2006-1	2.244 (0.070)	-0.010 (0.018)
2006-2	2.141 (0.066)	-0.113 (0.016)

Table 3.2: continued in the next page

Table 3.2: continued from the previous page

	FE-hedonic Estimate (s.e.)	**RE-hedonic** Estimate (s.e.)
2007-1	2.182 (0.067)	-0.072 (0.016)
2007-2	2.305 (0.068)	0.049 (0.021)
2008-1	2.250 (0.064)	-0.004 (0.022)
2008-2	2.174 (0.069)	-0.080 (0.017)
2009-1	2.269 (0.067)	0.013 (0.019)
2009-2	2.491 (0.070)	0.229 (0.031)
2010-1	2.535 (0.070)	0.273 (0.033)
2010-2	2.499 (0.082)	0.233 (0.043)
2011-1	2.412 (0.075)	0.151 (0.035)
Type of object: baseline Furniture		
Sticks	-0.088 (0.026)	-0.088 (0.035)
Masks	0.109 (0.020)	0.109 (0.023)
Religious objects	-0.002 (0.025)	-0.003 (0.028)
Ornaments	-0.099 (0.027)	-0.099 (0.038)
Sculptures	0.050 (0.020)	0.050 (0.022)
Musical instruments	-0.114 (0.034)	-0.114 (0.042)
Tools	-0.082 (0.021)	-0.082 (0.023)
Clothing	-0.069 (0.038)	-0.069 (0.052)
Textiles	-0.038 (0.038)	-0.038 (0.058)
Weapons	-0.089 (0.026)	-0.089 (0.034)
Jewels	-0.050 (0.034)	-0.050 (0.049)
Region: baseline Central America		
Southern Africa	-0.158 (0.033)	-0.159 (0.039)
Western Africa	-0.105 (0.011)	-0.105 (0.018)
Eastern Africa	-0.153 (0.024)	-0.153 (0.031)
Australia	0.061 (0.053)	0.061 (0.073)
Indonesia	-0.109 (0.024)	-0.110 (0.043)
Melanesia	0.006 (0.013)	0.006 (0.029)

Table 3.2: continued in the next page

Table 3.2: continued from the previous page

	FE-hedonic Estimate (s.e.)	**RE-hedonic** Estimate (s.e.)
Polynesia	0.177 (0.015)	0.177 (0.032)
Northern America	0.229 (0.018)	0.229 (0.053)
Northern Africa	-0.371 (0.120)	-0.371 (0.140)
Southern America	0.016 (0.024)	0.015 (0.052)
Mesoamerica	0.117 (0.021)	0.116 (0.052)
Far Eastern	-0.088 (0.145)	-0.088 (0.270)
Micronesia	0.096 (0.072)	0.096 (0.070)
Indian Region	0.301 (0.105)	0.297 (0.105)
Asian Southeast	-0.070 (0.112)	-0.072 (0.142)
Middle East	-0.559 (0.088)	-0.558 (0.137)
Type of material: baseline Ivory		
Vegetable fibre, paper, plumage	-0.045 (0.028)	-0.045 (0.031)
Wood	0.073 (0.019)	0.073 (0.026)
Metal	-0.032 (0.029)	-0.032 (0.049)
Gold	0.131 (0.040)	0.130 (0.060)
Stone	0.039 (0.028)	0.039 (0.035)
Precious stone	0.053 (0.034)	0.053 (0.046)
Terracotta, ceramic	0.003 (0.026)	0.003 (0.048)
Silver	-0.083 (0.041)	-0.084 (0.073)
Textile and hides	-0.020 (0.031)	-0.020 (0.054)
Seashell	0.061 (0.049)	0.061 (0.107)
Bone, horn	-0.133 (0.033)	-0.133 (0.074)
Not indicated	0.047 (0.041)	0.046 (0.051)
Patina: baseline Not indicated		
Pejorative	0.234 (0.037)	0.233 (0.043)
Present	0.033 (0.012)	0.033 (0.025)
Appreciative	0.114 (0.013)	0.113 (0.028)
Description on the catalogue: baseline Absent		

Table 3.2: continued in the next page

Table 3.2: continued from the previous page

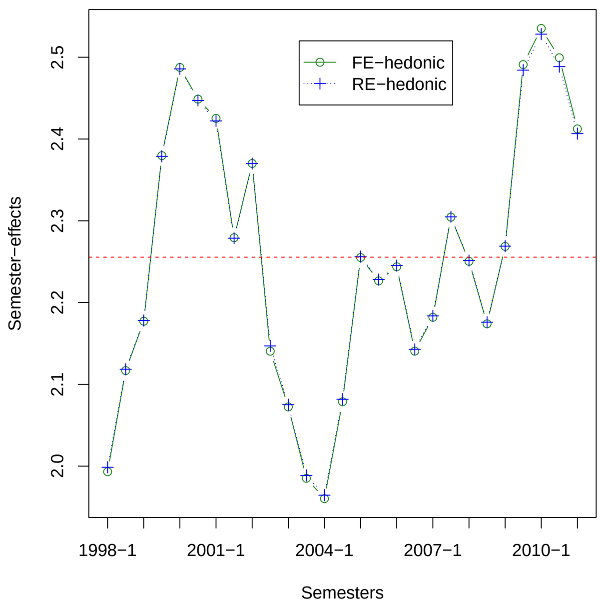
	FE-hedonic Estimate (s.e.)	**RE-hedonic** Estimate (s.e.)
Short visual descr.	-0.172 (0.033)	-0.173 (0.079)
Visual descr.	0.005 (0.035)	0.003 (0.081)
Broad visual descr.	0.238 (0.04)	0.237 (0.088)
Critical descr.	0.226 (0.037)	0.225 (0.090)
Broad critical descr.	0.598 (0.047)	0.597 (0.105)
Yes vs No		
Specialized bibliography (dummy)	0.138 (0.011)	0.138 (0.022)
Comparative bibliography (dummy)	0.120 (0.009)	0.120 (0.019)
Exhibition (dummy)	0.066 (0.013)	0.066 (0.028)
Historicization: baseline Absent		
Museum certification	0.019 (0.015)	0.019 (0.040)
Relevant museum certification	0.032 (0.014)	0.032 (0.042)
Simple certification	0.033 (0.009)	0.034 (0.026)
Illustration: baseline Absent		
Miscellaneous col. ill.	0.410 (0.021)	0.411 (0.045)
Col. cover	1.413 (0.099)	1.412 (0.181)
Col. half page	0.852 (0.024)	0.854 (0.068)
Col. full page	1.005 (0.025)	1.005 (0.070)
More than one col. ill.	1.221 (0.027)	1.221 (0.076)
Col. quarter page	0.668 (0.021)	0.669 (0.059)
Miscellaneous b/w ill.	0.404 (0.031)	0.403 (0.054)
b/w half page	0.545 (0.044)	0.546 (0.071)
b/w quarter page	0.301 (0.027)	0.303 (0.063)
Auction house and venue: baseline Bonhams-New York		
Christie's-Amsterdam	0.782 (0.052)	0.783 (0.064)
Christie's-New York	0.709 (0.051)	0.709 (0.062)
Christie's-Paris	0.600 (0.048)	0.600 (0.059)
Encheres Rive Gauche-Paris	0.531 (0.083)	0.529 (0.052)

Table 3.2: continued in the next page

Table 3.2: continued from the previous page

	FE-hedonic Estimate (s.e.)	RE-hedonic Estimate (s.e.)
Koller-Zurich	-0.005 (0.051)	-0.005 (0.089)
Piasa-Paris	0.738 (0.071)	0.736 (0.064)
Sotheby's-New York	0.881 (0.048)	0.881 (0.054)
Sotheby's-Paris	0.744 (0.047)	0.744 (0.056)

Figure 3.3: Plot of the time-specific intercepts: $\hat{\beta}_t$ for the FE-hedonic (3.4) and $\hat{\beta}_{0t}$ for the RE-hedonic model (3.5). The red horizontal line represents $\hat{\beta}_0$.



Also the estimates of the regression coefficients are similar for the two models. This means that the within-group effects are almost coincident with the total effects, and the between-group effects are negligible. This can be explained analytically. Since the model for the cluster means of the FE-hedonic model is

$$\bar{y}_t = \beta_t + \bar{\mathbf{x}}_t^{\mathrm{T}} \boldsymbol{\beta} + \bar{\epsilon}_t,$$

the vector of coefficients $\boldsymbol{\beta}$ is interpretable as the vector of the within-group effects. In fact, by following Skrondal and Rabe-Hesketh (2004), it can be estimated equivalently either from the fixed-effect model (3.4) or from the within-group regression model:

$$y_{it} - \bar{y}_t = (\mathbf{x}_{it} - \bar{\mathbf{x}}_t)^{\mathrm{T}}\boldsymbol{\beta}_W + \epsilon_{it} - \bar{\epsilon}_t.$$

The within-group estimator of the fixed regression coefficients (obtained through OLS estimation but equivalent to that obtained through maximum likelihood estimation in the case of normality) is, therefore,

$$\hat{\boldsymbol{\beta}}_W = \mathbf{W}_{xx}^{-1}\mathbf{W}_{xy}$$

where $\mathbf{W}_{xx} = (\mathbf{x}_{it} - \bar{\mathbf{x}}_t)(\mathbf{x}_{it} - \bar{\mathbf{x}}_t)^{\mathrm{T}}$ and $\mathbf{W}_{xy} = y_{it} - \bar{y}_t$. The between-group regression model, instead, has the following form

$$\bar{y}_t - \bar{y} = (\bar{x}_t - \bar{x})^{\mathrm{T}}\boldsymbol{\beta}_B + \bar{\epsilon}_t - \bar{\epsilon},$$

and the between-group estimator is

$$\hat{\boldsymbol{\beta}}_B = \mathbf{B}_{xx}^{-1}\mathbf{B}_{xy}$$

where $\mathbf{B}_{xx} = (\bar{x}_t - \bar{x})(\bar{x}_t - \bar{x})^{\mathrm{T}}$ and $\mathbf{B}_{xy} = (\bar{x}_t - \bar{x})(\bar{y}_t - \bar{y})$. On the one hand, the between-group estimator considers only the variation among groups; on the other hand, the fixed-effect model, through the within-group estimator, ignores this source of information and uses only the information within groups. The advantage of the random-effect model, instead, is that it combines the two pieces of information in one model. In fact, the GLS estimator (asymptotically equivalent to the maximum likelihood estimator but easier to handle with) of the regression coefficients of the random-effect model is an average of the within-estimator and the between-estimator weighted with the respective precisions:

$$\hat{\boldsymbol{\beta}}_{GLS} = \mathbf{V}_W^{-1}\hat{\boldsymbol{\beta}}_W + \mathbf{V}_B^{-1}\hat{\boldsymbol{\beta}}_B$$

where the matrices $\mathbf{V}$ are the respective variance-covariance matrices (Maddala, 1971). Another way to write the GLS estimator is in the following way

$$\hat{\boldsymbol{\beta}}_{GLS} = \big(\mathbf{W}_{xx} + (1-\lambda)\mathbf{B}_{xx}\big)^{-1}\big(\mathbf{W}_{xy} + (1-\lambda)\mathbf{B}_{xy}\big)$$

where $\lambda = n\sigma_u^2/(\sigma^2 + n\sigma_u^2)$ is the shrinkage factor. In our case, we saw that $\lambda \simeq 1$, so the GLS estimator is almost identical to the within-estimator and the between-group variation is ignored. Moreover, from an interpretative perspective, the fact that the between-group effects are null entails that the *contextual effects* coincide with the within-group effects. The *contextual effects* are the additional effects of group means on the responses that are not accounted for by individual levels. To identify them, we write the fixed-effect model as (Raudenbush and Bryk, 2002)

$$y_{it} = \beta_0 + (\mathbf{x}_{it} - \bar{\mathbf{x}}_t + \bar{\mathbf{x}}_t)^{\mathrm{T}}\boldsymbol{\beta} + \epsilon_{it} = \beta_0 + (\mathbf{x}_{it} - \bar{\mathbf{x}}_t)^{\mathrm{T}}\boldsymbol{\beta}_W + \bar{\mathbf{x}}_t^{\mathrm{T}}\boldsymbol{\beta}_B.$$

The contextual effects are represented by $\hat{\boldsymbol{\beta}}_B - \hat{\boldsymbol{\beta}}_W$. This means that, for example, the proportion of sculptures sold in the semester t ($\bar{x}_t$) does not affect the price of the object i in the semester t beyond the fact that the object is a sculpture, and the proportion of objects sold in the semester t by Christie's in New York does not affect the price of the object i in the semester t additionally to the fact that the object is sold by Christie's in New York.

In the following, we will have a look to the residuals of both models. The boxplots of residuals for each semester in Figures 3.5a and 3.5b look almost identical. As expected, they are better than those of the models without covariates (Figure 3.2), hence, the covariates have explained a part of the variability of artwork prices. In fact, the residual variance is 0.171 for both models, against 0.443 for the models without covariates.
The plots of residuals versus fitted values, in Figure 3.4, show that the (first level) errors are centered at zero, the variability seems to be constant, and that the points do not reveal any particular patterns. However, by plotting the standard deviation of the (level-1) residuals for each group, as shown in Figure 3.6, it is evident a time dependent heteroscedasticity.

The assumption of normality for the errors, instead, is assessed by looking at the normal plot of the residuals in Figure 3.7. The plot of level-1 residuals

Figure 3.4: Residuals versus fitted values of the FE-hedonic model (3.4) and the RE-hedonic model (3.5).

(a) FE-hedonic model. (b) RE-hedonic model.

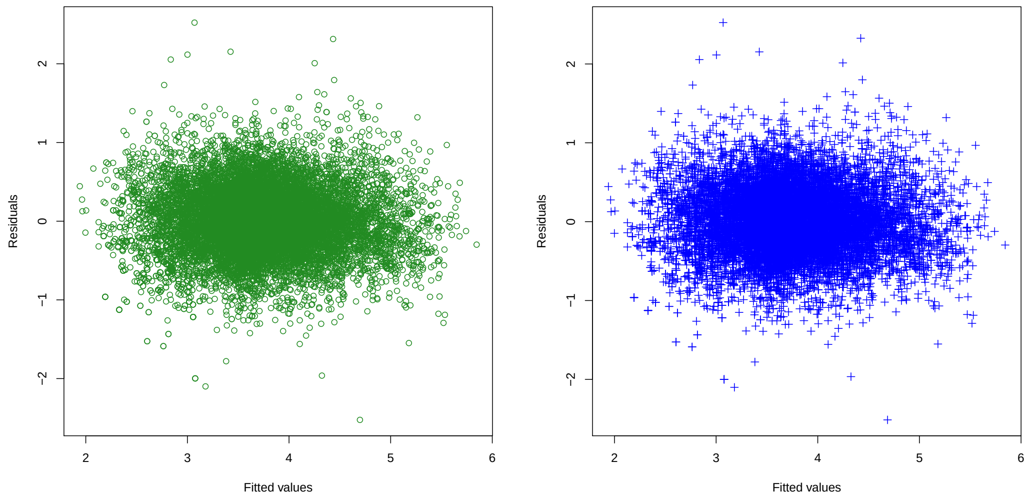
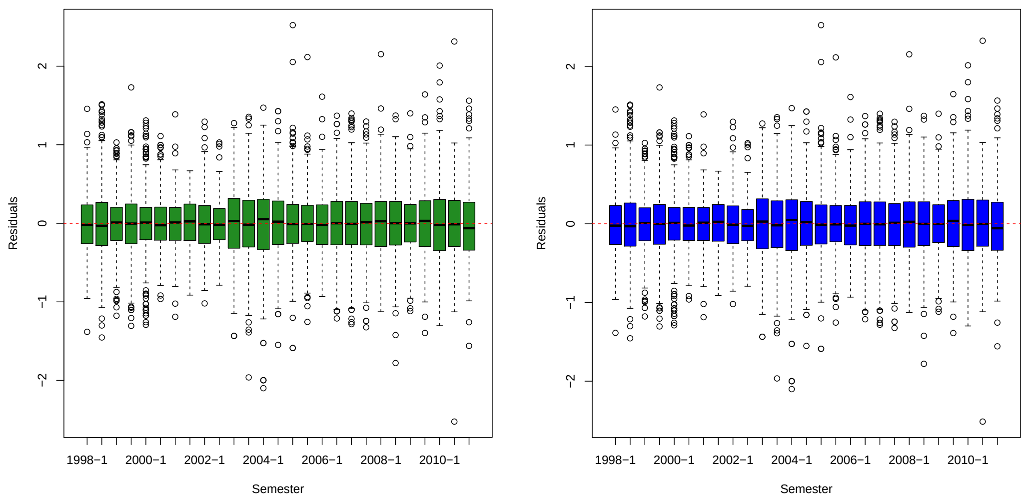


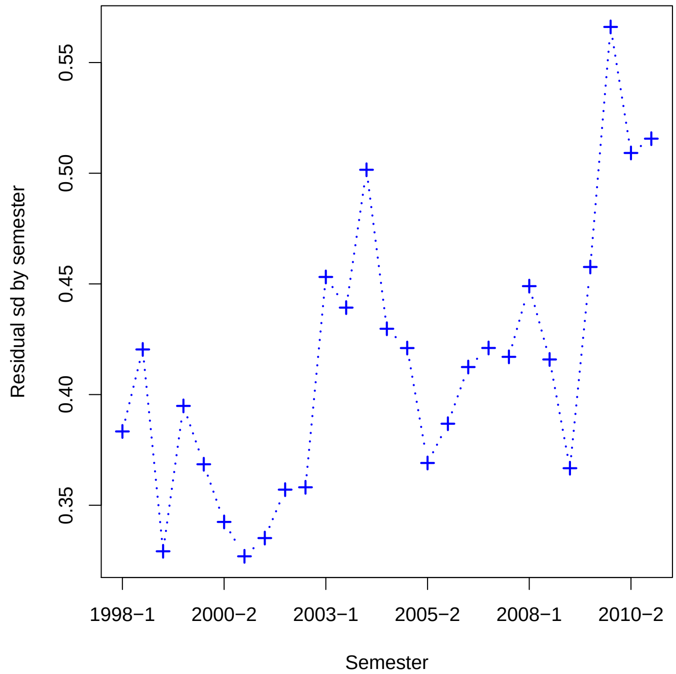
Figure 3.5: Boxplots of residuals by semester for the FE-hedonic model (3.4) and the RE-hedonic model (3.5).

(a) FE-hedonic model. (b) RE-hedonic model.



show the presence of deviation from normality in the tails. The Shapiro-Wilk test (see Table 3.3) does not reject the hypothesis of normality only for the

Figure 3.6: Standard deviations of the level-1 residuals by semester of the RE-hedonic model (3.5) (the FE-hedonic model shows an identical pattern).



second level residuals of the RE-hedonic model.
Because the assumption of homoscedasticity and also the assumption of normality for the first level errors do not hold, we calculated the standard errors of the estimates through the *Wild Bootstrap*, since it is robust to heteroscedastic errors and, being nonparametric, also to not Gaussian errors. The wild bootstrap procedure adapted to the multilevel case is briefly presented in subsection 3.2.1.

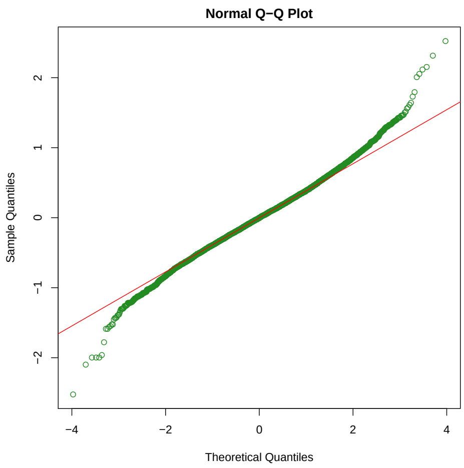
In order to test the assumption of the error processes for the RE-hedonic

Table 3.3: Shapiro-Wilk normality test for the residuals of the FE-hedonic model (3.4) and the RE-hedonic model (3.5).

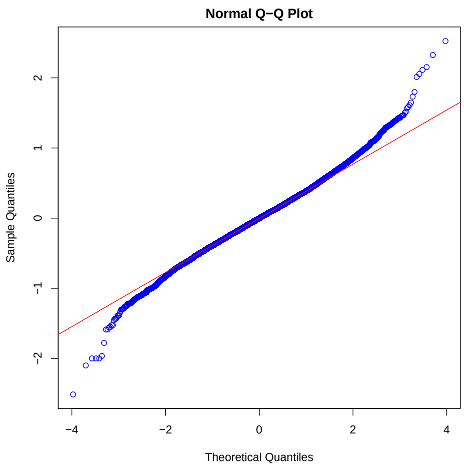
	FE-hedonic	RE-hedonic	
		lev-1	lev-2
W	0.994	0.994	0.962
p-value	0.000	0.000	0.416

Figure 3.7: Normal probability plot of residuals of the FE-hedonic model (3.4) and the RE-hedonic model (3.5).

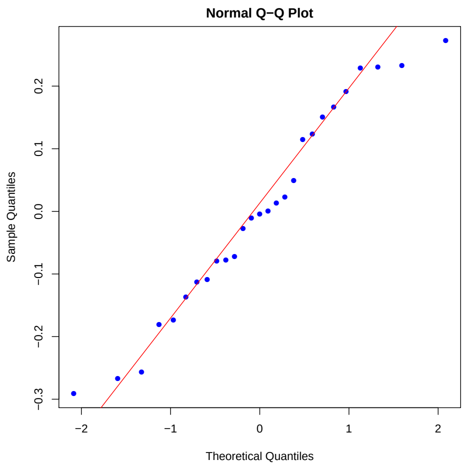
(a) FE-hedonic residuals.



(b) RE-hedonic lev-1 residuals.



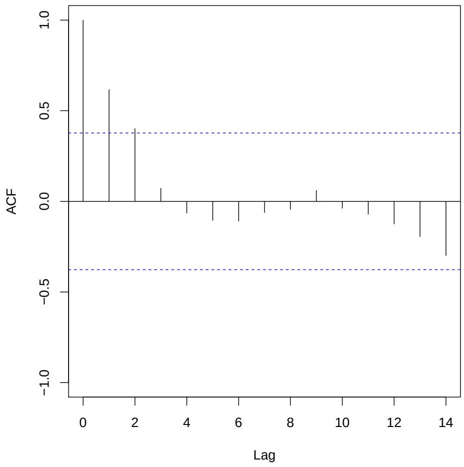
(c) RE-hedonic lev-2 residuals.



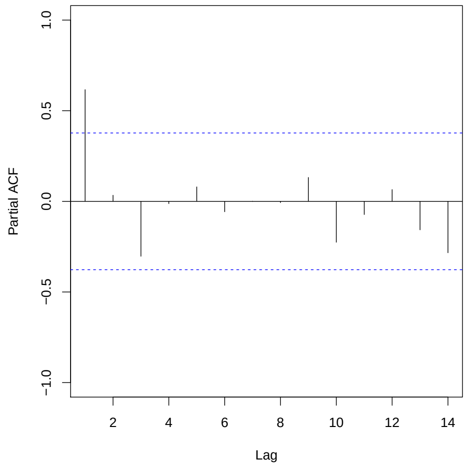
model, we have computed the autocorrelation functions (global and partial) of the means by semester of level-1 residuals (Figures 3.8a and 3.8b) and of level-2 residuals (Figures 3.8c and 3.8c). The correlograms point at an autoregressive-like structure, similar to that of an AR(1) process. This time dependence is incorporated also in the first level residual means by semester.

Figure 3.8: Plots of autocorrelation functions of the residuals of the RE-hedonic model.

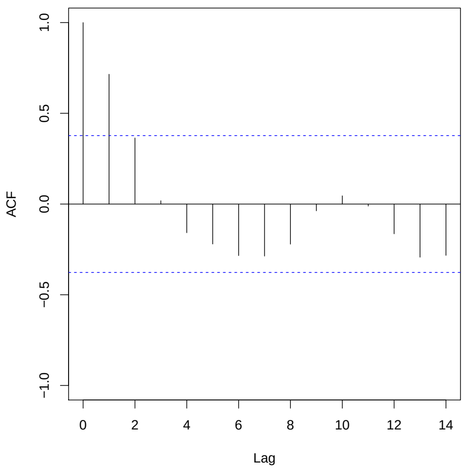
(a) Autocorrelation function of the level-1 residual means aggregated by semesters.



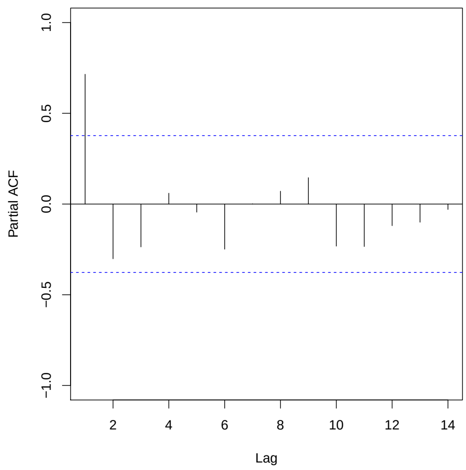
(b) Partial autocorrelation function of the level-1 residual means aggregated by semesters.



(c) Autocorrelation function of the level-2 residuals.



(d) Partial autocorrelation function of the level-2 residuals.



3.2.1 Robust standard errors

In general, asymptotic standard errors of the ML estimates are given by the square root of the diagonal elements of the inverse of the information matrix, evaluated at the ML solution. Because of the complexity of the information matrix for this model, we have recurred to the bootstrap to estimate bias and variance of estimators. In particular, we chose to perform the *wild bootstrap*, developed by Liu (1988) first, since the errors are not homoscedastic as the model assumes. We have implemented the usual wild bootstrap procedure, as explained by Davidson and Flachaire (2008), to obtain the standard errors for the estimates of the fixed-effect models. Instead, since for a two-level model a resampling scheme has to reflect the hierarchical data structure, we have adapted the procedure to this case.

Consider the classic multilevel model for the $(n_t \times 1)$ response of the generic group t:

$$\mathbf{y}_t = \beta_0 + \mathbf{X}_t\boldsymbol{\beta} + \mathbf{r}_t,$$

where

$$\mathbf{r}_t = \mathbf{1}_{n_t} u_t + \boldsymbol{\epsilon}_t,$$

for all $t = 1, \ldots, T$. The disturbances are assumed to be mutually independent and to have zero expectation, but they are allowed to be heteroscedastic. Moreover, the covariates are assumed to be strictly exogenous.
In the homoscedastic case, the variance of the residual vector $\hat{\mathbf{r}}_t$ is proportional to $\mathbf{I}_{n_t} - \mathbf{H}_t$, where $\mathbf{H}_t = \mathbf{X}_t(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}_t^T$ is the orthogonal projection matrix corresponding to design matrix $\mathbf{X}_t$. This suggests to replace $\hat{\mathbf{r}}_t$ by the vector

$$\tilde{\mathbf{r}}_t = \operatorname{diag}\big(\mathbf{I}_{n_t} - \mathbf{H}_t\big)^{-1/2} \cdot \hat{\mathbf{r}}_t,$$

where the sign "$\cdot$" indicates the element by element product of the two vectors. Then, the bootstrap procedure is as follows:

1. draw independently T values, w_t, for $t = 1, \ldots, T$, from the following two-point auxiliary distribution (Liu, 1988; Belsley and Kontoghiorghes, 2009):
$$\begin{cases} 1 & \text{with probability } 0.5 \\ -1 & \text{with probability } 0.5 \end{cases} \tag{3.6}$$
with zero mean and unitary variance;

2. generate the bootstrap samples as

$$\mathbf{y}_t^* = \hat{\beta}_0 + \mathbf{X}_t\hat{\boldsymbol{\beta}} + \mathbf{r}_t^*,$$

where the bootstrap disturbances are obtained as

$$\mathbf{r}_t^* = w_t\tilde{\mathbf{r}}_t;$$

3. compute estimates on the bootstrap sample $\mathbf{y}^*$;

4. repeat steps 1-3 B times and compute bootstrap standard errors as

$$\sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\boldsymbol{\theta}_b^* - \hat{\boldsymbol{\theta}})^2}$$

where $\hat{\boldsymbol{\theta}}$ is the vector of the ML estimates.

3.2.2 Considerations

In summary, the results of the FE-hedonic (3.4) and RE-hedonic (3.5) models are very similar in terms of estimates and residuals. The assumptions of homoscedasticity and normality for the (first level) errors of both models are not valid, but this is not a big deal, for now, since we used robust standard errors. On the other hand, the predicted random effects are normally distributed with zero mean, but they are not independent for different groups; this causes also the violation of the assumption of independence between first and second level errors.

Improving the classical multilevel model, for this case, requires relaxing the assumption of independence among random effects. Since the random effects, in the application of Tribal art data, represent time effects, the inclusion of such correlation implies treating them as a time series.
The main theoretical contribution of this thesis is the derivation of a multilevel model with time series components at a second level. As mentioned

above, the correlograms of the residuals suggest the specification of an AR(1) model. The next chapter is devoted to the theoretical derivation of such model. In particular, we will find a meaningful econometric specification for our problem. Then, we will derive and implement a maximum likelihood estimators and will perform a Monte Carlo study for assessing their finite sample behaviour. Finally, we will fit the model to the Tribal art dataset and compare the results with those presented in this chapter.

Chapter 4

A multilevel model with time series components

The previous chapter has shown that applying a classic multilevel model to the Tribal art data results in the violation of the assumption of independence among random effects and between first and second level errors. In this chapter, we propose a new extension of the classic multilevel model that consists in relaxing the assumption of independence among random effects and treating them as a time series at the second level. In particular, first we specify a multilevel model with an AR(1) process at the second level to capture the time dependence among groups. Section 4.2 contains the derivation of the maximum likelihood estimators through the E-M algorithm. The estimation algorithm has been implemented in R with an own-written code. Through a Monte Carlo study, we will find that the ML estimators have a good finite sample behaviour and that our R-code is valid. In the last section, we fit the new model to the Tribal art dataset and compare the results with those obtained by the classic multilevel model with independent random effects, presented in the previous chapter. We will see that the AR(1) process well captures the time dependence structure among group-effects. Hence, the multilevel model with time series components fits better the data.

4.1 Model specification

Consider a random intercept model with k level-1 covariates:

$$y_{it} = \beta_1 x_{1it} + \beta_2 x_{2it} + \ldots + \beta_k x_{kit} + \beta_{0t} + \epsilon_{it} \tag{4.1}$$

where $t = 1, \ldots, T$ indexes the level-2 units, and $i = 1_t \ldots, n_t$ indexes the level-1 units in the t-th level-2 unit. The intercepts β_{0t} are group-specific and random; the slopes $\beta_1, \ldots, \beta_k$, instead, are fixed. The ϵ_{it} are the level-1 residuals assumed independent for different groups given the covariates and independent of each other for the same group; also, they are normally distributed with zero mean and constant variance:

$$\epsilon_{it}|\mathbf{x}_{it} \sim \text{NID}(0, \sigma^2), \quad \forall i \quad \text{and} \quad \forall t,$$

where $\mathbf{x}_{it} = \begin{bmatrix} x_{1it} & x_{2it} & \ldots & x_{kit} \end{bmatrix}$. All the random variables in the models are to be conditioned to the design matrices, but, from now on, the conditioning will be omitted to simplify the notation.
The random intercept β_{0t} is modeled as

$$\beta_{0t} = \beta_0 + u_t,$$

where β_0 represents the mean across the population, and u_t is the deviation of the group-specific intercept β_{0t} from the overall mean.
Here, the usual assumption of independence for the random effects ($u_t \perp u_s$, for $t \neq s$) is relaxed by assuming an autoregressive process of order 1 for the level-2 errors:

$$u_t = \rho u_{t-1} + \eta_t, \qquad \eta_t \sim \text{NID}(0, \sigma^2_\eta),$$

with $|\rho| < 1$, that guarantees stationarity. Moreover, it is assumed that $\eta_t \perp u_s$ for all $s < t$ and $\eta_t \perp \epsilon_{it}$ for all $t \neq s = 1, \ldots, T$ and for all $i = 1, \ldots, n_t$. Under these assumptions the dependent variable has the following distribution:

$$y_{it} \sim N\Big(\beta_0 + \mathbf{x}_{it}\boldsymbol{\beta}, \sigma^2 + \phi_0\Big)$$

where $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 & \beta_2 & \dots & \beta_k \end{bmatrix}^{\mathrm{T}}$ is the vector of fixed slopes, and

$$\phi_0 = \mathrm{Var}(u_t) = \frac{\sigma_\eta^2}{1-\rho^2}.$$

In matrix form, the composite model for the whole response vector is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon},$$

where

$$\mathbf{Z} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \dots & \mathbf{1}_{n_T} \end{bmatrix},$$

and

$$\mathbf{b} = \begin{bmatrix} \beta_{01} & \beta_{02} & \dots & \beta_{0T} \end{bmatrix}^{\mathrm{T}}$$

is the vector of random intercepts. Since the random intercepts are not independent, the variance-covariance matrix of $\mathbf{b}$ is not diagonal, as in the classical multilevel framework. Instead, its variance-covariance matrix is that of an AR(1) process:

$$\boldsymbol{\Gamma} = \phi_0 \begin{bmatrix} 1 & \rho & \dots & \rho^{T-1} \\ \rho & 1 & \dots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \dots & 1 \end{bmatrix}.$$

Therefore, the vector $\mathbf{b}$ is normally distributed with expected value $\mathbf{1}_T\beta_0$ and variance- covariance matrix $\boldsymbol{\Gamma}$. Consequently, the probability distribution of the response vector is

$$\mathbf{y} \sim \mathrm{N}\big(\beta_0 + \mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\boldsymbol{\Gamma}\mathbf{Z}^{\mathrm{T}} + \sigma^2\mathbf{I}_n\big). \tag{4.2}$$

We can now proceed to the estimation of the parameters. We have chosen the full maximum likelihood estimation method implemented through the

E-M algorithm.

4.2 Full maximum likelihood estimation through the EM algorithm

The set of parameters of the multilevel model with AR(1) random effects to be estimated is $\boldsymbol{\theta} = \{\beta_0, \boldsymbol{\beta}, \sigma^2, \rho, \sigma_\eta^2\}$. In the following, we show the derivation of the full maximum likelihood estimators of $\boldsymbol{\theta}$ and their implementation by mean of the E-M algorithm.

The full likelihood function associated with the response vector $\mathbf{y}$, with probability distribution (4.2), is

$$\mathrm{L}(\theta; \mathbf{y}) = f(\mathbf{y}; \theta) = \frac{|\boldsymbol{\Omega}|^{-1/2}}{(2\pi)^{n/2}} \exp\left(- \frac{(\mathbf{y} - \dot{\mathbf{X}}\dot{\boldsymbol{\beta}})^{\mathrm{T}} \boldsymbol{\Omega}^{-1} (\mathbf{y} - \dot{\mathbf{X}}\dot{\boldsymbol{\beta}})}{2} \right) \tag{4.3}$$

where

$$\boldsymbol{\Omega} = \mathbf{Z}\boldsymbol{\Gamma}\mathbf{Z}^{\mathrm{T}} + \sigma^2 \mathbf{I}_n$$

is the variance-covariance matrix of $\mathbf{y}$, and

$$\dot{\mathbf{X}} = \begin{bmatrix} \mathbf{1}_n & \mathbf{X} \end{bmatrix} \qquad \text{and} \qquad \dot{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 & \boldsymbol{\beta} \end{bmatrix}^{\mathrm{T}}$$

are, respectively, the matrix design and the coefficients vector including the intercept. Maximizing $\mathrm{L}(\boldsymbol{\theta}; \mathbf{y})$ can be achieved by optimizing its logarithm, the log-likelihood

$$\ell(\boldsymbol{\theta}; \mathbf{y}) = \ln \mathrm{L}(\boldsymbol{\theta}; \mathbf{y}) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Omega}| - \frac{1}{2} (\mathbf{y} - \dot{\mathbf{X}}\dot{\boldsymbol{\beta}})^{\mathrm{T}} \boldsymbol{\Omega}^{-1} (\mathbf{y} - \dot{\mathbf{X}}\dot{\boldsymbol{\beta}}). \tag{4.4}$$

As seen for the classic multilevel model in section 2.4.1, the direct derivation of the log-likelihood (4.4) yields

$$\frac{\partial \ell}{\partial \dot{\boldsymbol{\beta}}} = \dot{\mathbf{X}}^{\mathrm{T}} \boldsymbol{\Omega}^{-1} \mathbf{y} - \dot{\mathbf{X}}^{\mathrm{T}} \boldsymbol{\Omega}^{-1} \dot{\mathbf{X}} \dot{\boldsymbol{\beta}} \tag{4.5}$$

and

$$\frac{\partial \ell}{\partial \omega_j} = -\frac{1}{2}\mathrm{tr}\Big(\mathbf{\Omega}^{-1}\frac{\partial \mathbf{\Omega}}{\partial \omega_j}\Big) + \frac{1}{2}(\mathbf{y} - \dot{\mathbf{X}}\dot{\boldsymbol{\beta}})^{\mathrm{T}}\mathbf{\Omega}^{-1}\frac{\partial \mathbf{\Omega}}{\partial \omega_j}\mathbf{\Omega}^{-1}(\mathbf{y} - \dot{\mathbf{X}}\dot{\boldsymbol{\beta}}) \qquad (4.6)$$

for $j = 1, 2, 3$, where $\omega_1 = \sigma^2$, $\omega_2 = \rho$ and $\omega_3 = \sigma^2_\eta$.
Equating the partial derivatives (4.5) and (4.6) to zero and solving the resulting equations for $\dot{\boldsymbol{\beta}}$ and $\omega_j's$ gives:

$$\dot{\mathbf{X}}^{\mathrm{T}}\mathbf{\Omega}^{-1}\dot{\mathbf{X}}\dot{\boldsymbol{\beta}} = \dot{\mathbf{X}}^{\mathrm{T}}\mathbf{\Omega}^{-1}\mathbf{y}$$

and

$$\mathrm{tr}\Big(\mathbf{\Omega}^{-1}\frac{\partial \mathbf{\Omega}}{\partial \omega_j}\Big) = (\mathbf{y} - \dot{\mathbf{X}}\dot{\boldsymbol{\beta}})^{\mathrm{T}}\mathbf{\Omega}^{-1}\frac{\partial \mathbf{\Omega}}{\partial \omega_j}\mathbf{\Omega}^{-1}(\mathbf{y} - \dot{\mathbf{X}}\dot{\boldsymbol{\beta}}) \qquad (4.7)$$

for $j = 1, 2, 3$. However, the equations (4.7) are nonlinear in the variance) components ω_j, so that, it is not possible to achieve closed form expressions for the solutions of (4.7). Therefore, we obtain solutions to these equations through the EM algorithm.

The Expectation-Maximization algorithm (Dempster et al., 1977; Laird et al., 1987) is an iterative procedure to compute maximum likelihood estimates in the case of incomplete-data problem. In the mixed-effects model, if the random effects were known, it would be possible to write a closed form of the maximum likelihood estimates of the variance parameters ω_j. This suggests to treat $\mathbf{b}$ as missing data in a EM algorithm context, so that $(\mathbf{y}, \mathbf{b})$ forms the *complete* dataset, that is the dataset composed by the vector of the observed uncomplete data, $\mathbf{y}$, and the vector of the unobserved data, $\mathbf{b}$. To simplify the notation, we separate the set of parameters of the multilevel model with AR(1) random effects in two subsets: $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$, where the subset $\boldsymbol{\theta}_1 = \{\boldsymbol{\beta}, \sigma^2\}$ includes the level-1 parameters, and $\boldsymbol{\theta}_2 = \{\beta_0, \rho, \sigma^2_\eta\}$ the level-2 parameters.

The joint or *complete* likelihood associate with the complete dataset is:

$$\begin{aligned}
\mathrm{L}(\boldsymbol{\theta};\mathbf{y},\mathbf{b}) &= f(\mathbf{y},\mathbf{b};\boldsymbol{\theta}) = f(\mathbf{y}|\mathbf{b};\boldsymbol{\theta}_1)f(\mathbf{b};\boldsymbol{\theta}_2) \\
&= (2\pi\sigma^2)^{-n/2}\exp\left[-\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{b})^{\mathrm{T}}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{b})}{2\sigma^2}\right] \\
&\times \frac{|\boldsymbol{\Gamma}|^{-1/2}}{(2\pi)^{T/2}}\exp\left(-\frac{(\mathbf{b}-\beta_0\mathbf{1}_T)^{\mathrm{T}}\boldsymbol{\Gamma}^{-1}(\mathbf{b}-\beta_0\mathbf{1}_T)}{2}\right),
\end{aligned}$$

that can be re-written as

$$\begin{aligned}
\mathrm{L}(\boldsymbol{\theta};\mathbf{y},\mathbf{b}) &= (2\pi\sigma^2)^{-n/2}\exp\left[-\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{b})^{\mathrm{T}}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{b})}{2\sigma^2}\right] \\
&\times \frac{|\mathbf{V}|^{-1/2}}{(2\pi\sigma_\eta^2)^{T/2}}\exp\left(-\frac{(\mathbf{b}-\beta_0\mathbf{1}_T)^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{b}-\beta_0\mathbf{1}_T)}{2\sigma_\eta^2}\right)
\end{aligned} \tag{4.8}$$

since the covariance matrix $\boldsymbol{\Gamma}$ is equal to $\sigma_\eta^2\mathbf{V}$, where

$$\mathbf{V} = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & \rho & \dots & \rho^{T-1} \\ \rho & 1 & \dots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \dots & 1 \end{bmatrix}.$$

Hence, the joint log-likelihood is the sum of two separate components:

$$\ell(\boldsymbol{\theta};\mathbf{y},\mathbf{b}) = \ln\mathrm{L}(\boldsymbol{\theta};\mathbf{y},\mathbf{b}) = \ell_1(\boldsymbol{\theta}_1) + \ell_2(\boldsymbol{\theta}_2) \tag{4.9}$$

where

$$\ell_1(\boldsymbol{\theta}_1) = \ln f(\mathbf{y}|\mathbf{b}) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{b})^{\mathrm{T}}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{b})}{2\sigma^2}$$

and

$$\ell_2(\theta_2) = \ln f(\mathbf{b}) = -\frac{T}{2}\ln(2\pi\sigma_\eta^2) + \frac{1}{2}\ln(1-\rho^2) - \frac{(\mathbf{b}-\beta_0\mathbf{1}_T)^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{b}-\beta_0\mathbf{1}_T)}{2\sigma_\eta^2}. \tag{4.10}$$

Note that $(1-\rho^2)$ in (4.10) comes from $|\mathbf{V}| = |\boldsymbol{\Delta}^{\mathrm{T}}\boldsymbol{\Delta}|^{-1} = (1-\rho^2)^{-1}$, where

$$\boldsymbol{\Delta} = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & \dots & 0 \\ 0 & -\rho & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -\rho & 1 \end{bmatrix},$$

and it is straightforward to show that (see for example Hamilton (1994))

$$\mathbf{V}^{-1} = \boldsymbol{\Delta}^{\mathrm{T}}\boldsymbol{\Delta} = \begin{bmatrix} 1 & -\rho & 0 & \dots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \dots & 0 & 0 \\ \vdots & & & & \vdots & \\ 0 & 0 & 0 & \dots & 1+\rho^2 & -\rho \\ 0 & \dots & & & -\rho & 1 \end{bmatrix}.$$

The estimation of $\boldsymbol{\theta}$ is obtained through an iterative procedure involving the complete log-likelihood. Starting with some initial values for the parameters, each iteration consists in two steps.

E-STEP

In the Expectation step, since the complete log-likelihood (4.9) is unobservable, the current value of $\boldsymbol{\theta}$, denoted as $\boldsymbol{\theta}^{(h)}$, is used to evaluate its expected value conditional on the observed data:

$$\mathrm{E}\big[\ell(\boldsymbol{\theta}; \mathbf{y}, \mathbf{b})|\mathbf{y}\big]. \tag{4.11}$$

The probability distribution of the unobserved data conditional to the observed data is

$$\mathbf{b}|\mathbf{y} \sim N\Big(\mathbf{1}_T\beta_0 + \boldsymbol{\Gamma}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{1}_n\beta_0 - \mathbf{X}\boldsymbol{\beta}), \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{Z}\boldsymbol{\Gamma}^{\mathrm{T}}\Big) \tag{4.12}$$

since

$$\begin{bmatrix} \mathbf{b} \\ \mathbf{y} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{1}_T\beta_0 \\ \beta_0 + \mathbf{X}\boldsymbol{\beta} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Gamma} & \boldsymbol{\Gamma}\mathbf{Z}^{\mathrm{T}} \\ \mathbf{Z}\boldsymbol{\Gamma}^{\mathrm{T}} & \boldsymbol{\Omega} \end{bmatrix} \right).$$

By exploiting some results on Schur complements (see for example Searle et al. (1992)), the variance-covariance matrix of the conditional distribution can be simplified as follows:

$$\boldsymbol{\Gamma} - \boldsymbol{\Gamma}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{Z}\boldsymbol{\Gamma}^{\mathrm{T}} = \left(\boldsymbol{\Gamma}^{-1} + \frac{\mathbf{Z}^{\mathrm{T}}\mathbf{Z}}{\sigma^2}\right)^{-1}.$$

Since, in the next step, the modified log-likelihood (4.11) will be maximized, it is more convenient to compute the conditional expectation of the score functions from the joint log-likelihood (4.9), that is

$$\int \frac{\partial \ell(\boldsymbol{\theta}; \mathbf{y}, \mathbf{b})}{\partial \boldsymbol{\theta}} f(\mathbf{b}|\mathbf{y}; \boldsymbol{\theta}^{(h)}) \mathrm{d}\mathbf{b}.$$

The complete log-likelihood gives the following score functions

- w.r.t. $\boldsymbol{\beta}$:

$$\frac{\partial \ell_1(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\beta}} = \frac{-\mathbf{X}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{X}^{\mathrm{T}}\mathbf{y} - \mathbf{X}^{\mathrm{T}}\mathbf{Z}\mathbf{b}}{\sigma^2}, \tag{4.13}$$

- w.r.t. σ^2:

$$\frac{\partial \ell_1(\boldsymbol{\theta}_1)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})}{2\sigma^4},$$

- w.r.t. β_0:

$$\begin{aligned} \frac{\partial \ell_2(\boldsymbol{\theta}_2)}{\partial \beta_0} &= \frac{\mathbf{1}_T^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{b} - \mathbf{1}_T^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{1}_T\beta_0}{\sigma_\eta^2} \\ &= \frac{(1-\rho)(b_1 + b_T) + (1-\rho)^2\sum_{t=2}^{T-1} b_t - (1-\rho)(T - (T-2)\rho)\beta_0}{\sigma_\eta^2}, \end{aligned} \tag{4.14}$$

- w.r.t. σ_η^2:

$$\frac{\partial \ell_2(\boldsymbol{\theta}_2)}{\partial \sigma_\eta^2} = -\frac{T}{2\sigma_\eta^2} + \frac{(\mathbf{b}-\beta_0\mathbf{1}_T)^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{b}-\beta_0\mathbf{1}_T)}{2\sigma_\eta^4}; \tag{4.15}$$

- w.r.t. ρ:

$$\frac{\partial \ell_2(\boldsymbol{\theta}_2)}{\partial \rho} = -\frac{\rho}{1-\rho^2} - \frac{1}{2\sigma_\eta^2}\mathbf{u}^{\mathrm{T}}\frac{\partial \mathbf{V}^{-1}}{\partial \rho}\mathbf{u}, \tag{4.16}$$

where $\mathbf{u} = \mathbf{b} - \beta_0\mathbf{1}_T$. Note that

$$\frac{\partial \mathbf{V}^{-1}}{\partial \rho} = \begin{bmatrix} 0 & -1 & 0 & \dots & 0 & 0 \\ -1 & 2\rho & -1 & \dots & 0 & 0 \\ 0 & -1 & 2\rho & \dots & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \dots & 2\rho & -1 \\ 0 & \dots & & & -1 & 0 \end{bmatrix}$$

so that its product with $\mathbf{u}^{\mathrm{T}}$ and $\mathbf{u}$ yields a simple expression:

$$\mathbf{u}^{\mathrm{T}}\frac{\partial \mathbf{V}^{-1}}{\partial \rho}\mathbf{u} = -2\sum_{t=1}^{T-1} u_t u_{t+1} + 2\rho\sum_{t=2}^{T-1} u_t^2$$

Hence, the score function (4.16) takes the following form:

$$\begin{aligned} \frac{\partial \ell_2(\boldsymbol{\theta}_2)}{\partial \rho} &= \frac{1}{\sigma_\eta^2(1-\rho^2)}\left[\rho^3\sum_{t=2}^{T-1} u_t^2 - \rho^2\sum_{t=1}^{T-1} u_t u_{t+1} - \rho\left(\sigma_\eta^2 + \sum_{t=2}^{T-1} u_t^2\right) + \sum_{t=1}^{T-1} u_t u_{t+1}\right] \\ &= \frac{1}{\sigma_\eta^2}\left[\sum_{t=1}^{T-1} u_t u_{t+1} - \rho\sum_{t=2}^{T-1} u_t^2\right] - \frac{\rho}{1-\rho^2}. \end{aligned}$$

Now, instead of calculating the conditional expectations of the score functions, it is enough to take the conditional expectation for the sufficient statistics given the observed data (McLachlan and Krishnan, 1997), evaluate them with the current estimates of the parameters $\boldsymbol{\theta}^{(h)}$ and substitute them in the score functions. Therefore, by substituting the sufficient statistics in the score functions with their corresponding conditional expected values, the following conditional expectations of the score functions are obtained:

$$\mathrm{E}\left[\frac{\partial \ell_1(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\beta}}|\mathbf{y};\boldsymbol{\theta}^{(h)}\right] = \frac{-\mathbf{X}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{X}^{\mathrm{T}}\mathbf{y} - \mathbf{X}^{\mathrm{T}}\mathbf{Z}\hat{\mathbf{b}}}{\sigma^2}$$

$$\mathrm{E}\left[\frac{\partial \ell_1(\boldsymbol{\theta}_1)}{\partial \sigma^2}|\mathbf{y};\boldsymbol{\theta}^{(h)}\right] = -\frac{n}{2\sigma^2} + \frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\hat{\mathbf{b}})^{\mathrm{T}}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\hat{\mathbf{b}}) + \mathrm{tr}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\mathbf{B})}{2\sigma^4}$$

$$\mathrm{E}\left[\frac{\partial \ell_2(\boldsymbol{\theta}_2)}{\partial \beta_0}|\mathbf{y};\boldsymbol{\theta}^{(h)}\right] = \frac{(1-\rho)(\hat{b}_1+\hat{b}_T) + (1-\rho)^2\sum_{t=2}^{T-1}\hat{b}_t - (1-\rho)(T-(T-2)\rho)\beta_0}{\sigma_\eta^2};$$

$$\mathrm{E}\left[\frac{\partial \ell_2(\boldsymbol{\theta}_2)}{\partial \sigma_\eta^2}|\mathbf{y};\boldsymbol{\theta}^{(h)}\right] = -\frac{T}{2\sigma_\eta^2} + \frac{\mathrm{tr}(\mathbf{V}^{-1}\mathbf{B}) + \hat{\mathbf{u}}^{\mathrm{T}}\mathbf{V}^{-1}\hat{\mathbf{u}}}{2\sigma_\eta^4}$$

$$\mathrm{E}\left[\frac{\partial \ell_2(\boldsymbol{\theta}_2)}{\partial \rho}|\mathbf{y};\boldsymbol{\theta}^{(h)}\right] = \frac{1}{\sigma_\eta^2}\left[\sum_{t=1}^{T-1}(\mathbf{B}_{t,t+1} + \hat{u}_t\hat{u}_{t+1}) - \rho\sum_{t=2}^{T-1}(\mathbf{B}_{t,t} + \hat{u}_t^2)\right] - \frac{\rho}{1-\rho^2} \tag{4.17}$$

where

$$\hat{\mathbf{b}} = \hat{\beta}_0 + \hat{\mathbf{u}} \quad \text{and} \quad \mathbf{B} = \mathrm{Var}(\mathbf{b}|\mathbf{y};\boldsymbol{\theta}^{(h)})$$

are respectively the conditional expected value and variance-covariance matrix of the random vector $\mathbf{b}$, whose expressions are in (4.12).

M-STEP

The Maximization step consists in maximizing the conditional expected value of the log-likelihood (4.11) computed in the E-step, to get new values for the vector of parameters, $\boldsymbol{\theta}^{(h+1)}$.In practice, since the expected values of the score functions have been computed, it is enough to set them equal to zero and solve for the parameters. Therefore, the current value of the parameters are updated as follows:

$$\hat{\boldsymbol{\beta}}^{(h+1)} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}(\mathbf{y}-\mathbf{Z}\hat{\mathbf{b}})$$

$$(\hat{\sigma}^2)^{(h+1)} = \frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\hat{\mathbf{b}})^{\mathrm{T}}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\hat{\mathbf{b}}) + \mathrm{tr}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\mathbf{B})}{n}$$

$$\hat{\beta}_0^{(h+1)} = \frac{b_1 + b_T + (1-\rho)\sum_{t=2}^{T-1} b_t}{T - (T-2)\rho}$$

$$(\hat{\sigma}_\eta^2)^{(h+1)} = \frac{\mathrm{tr}(\mathbf{V}^{-1}\mathbf{B}) + \hat{\mathbf{u}}^{\mathrm{T}}\mathbf{V}^{-1}\hat{\mathbf{u}}}{T}$$

Since the expected score function for ρ (4.17) is a cubic function, the estimate for ρ cannot be expressed in an explicit form. Therefore, only for this case, it is necessary to use methods of numerical optimization. Because the first derivative for ρ (4.2) has a simple form, it is easy to implement the Newton-Raphson algorithm. It updates the current value of the estimate, $\rho^{(h)}$ through the formula

$$\hat{\rho}^{(h+1)} = \hat{\rho}^{h} - \mathrm{E}\Big[\frac{\partial \ell_2(\boldsymbol{\theta}_2)}{\partial \rho}|\mathbf{y};\boldsymbol{\theta}^{(h)}\Big] / \frac{\partial \mathrm{E}\Big[\frac{\partial \ell_2(\boldsymbol{\theta}_2)}{\partial \rho}|\mathbf{y};\boldsymbol{\theta}^{(h)}\Big]}{\partial \rho}$$

where

$$\frac{\partial \mathrm{E}\Big[\frac{\partial \ell_2(\boldsymbol{\theta}_2)}{\partial \rho}|\mathbf{y};\boldsymbol{\theta}^{(h)}\Big]}{\partial \rho} = -\frac{\rho}{\sigma_\eta^2}\sum_{t=2}^{T-1}(\mathbf{B}_{t,t} + \hat{u}_t^2) - \frac{1+\rho^2}{(1-\rho^2)^2}$$

is the second derivatives of the joint log-likelihood with respect to ρ, that, within the E-M algorithm, is the first derivative of the expected score function.

The E-step and the M-step are iterated until convergence, that is until

$$\ell(\boldsymbol{\theta}^{(h+1)};\mathbf{y},\mathbf{b}) - \ell(\boldsymbol{\theta}^{(h)};\mathbf{y},\mathbf{b})$$

is arbitrarily small.
In summary, the E-M algorithm consists in the following steps:

1. choosing some initial value for the parameters;
2. E-step as above;
3. M-step as above;

4. repeating steps 2 and 3 until convergence.

Although the achievement of the global maximum is not guaranteed, the monotonicity of the EM procedure, i.e.

$$\ell(\boldsymbol{\theta}^{(h+1)}; \mathbf{y}, \mathbf{b}) \geq \ell(\boldsymbol{\theta}^{(h)}; \mathbf{y}, \mathbf{b}),$$

can be demonstrated (Dempster et al., 1977).

The EM algorithm produces directly the Empirical Bayes prediction for the random effects (see section 2.5), $\mathbf{b}$, that is the mean of their conditional distribution to the observed data $\mathbf{y}$ as in (4.12), namely

$$\hat{\mathbf{b}} = \mathrm{E}(\mathbf{b}|\mathbf{y}) = \hat{\beta}_0 + \hat{\mathbf{u}},$$

where

$$\hat{\mathbf{u}} = \hat{\boldsymbol{\Gamma}}\mathbf{Z}^{\mathrm{T}}\hat{\boldsymbol{\Omega}}^{-1}(\mathbf{y} - \mathbf{1}_n\hat{\beta}_0 - \mathbf{X}\hat{\boldsymbol{\beta}}).$$

4.3 Simulation study

In order to assess the finite sample properties of the estimators and to validate the practical implementation of the algorithm, we have performed a Monte Carlo study. In the scheme, we have varied the group size n_t, and the time span T, as reported in Table 4.1.

Table 4.1: Scheme of the Monte Carlo study.

Scenario	n_t	T	n	Table
1	30	10	300	4.2
2	100	10	1000	4.2
3	100	30	3000	4.3
4	30	100	3000	4.3
5	30	50	1500	4.4
6	100	50	5000	4.4

Table 4.2: Results for scenarios 1 and 2 of the MC study. The table is separated in two row-blocks: one for the level-1 parameters and the other one for the level-2 parameters.

		Scenario 1 $n = 300, \quad n_t = 30, \quad T = 10$				Scenario 2 $n = 1000, \quad n_t = 100, \quad T = 10$			
	True	Mean	Sd	Bias	MSE	Mean	Sd	Bias	MSE
$\hat{\beta}_1$	2.000	2.007	0.197	-0.007	0.039	2.000	0.101	0.000	0.010
$\hat{\beta}_2$	4.000	3.992	0.128	0.008	0.016	4.001	0.069	-0.001	0.005
$\hat{\sigma}^2$	1.000	0.995	0.082	0.005	0.007	1.001	0.044	-0.001	0.002
$\hat{\beta}_0$	3.000	3.023	0.693	-0.024	0.479	3.011	0.683	-0.011	0.465
$\hat{\sigma}^2_\eta$	0.600	0.475	0.253	0.125	0.080	0.475	0.235	0.125	0.071
$\hat{\rho}$	0.700	0.374	0.345	0.326	0.225	0.390	0.315	0.310	0.195

Table 4.3: Results for scenario 3 and 4 of the MC study. The table is separated in two row-blocks: one for the level-1 parameters and the other one for the level-2 parameters.

		Scenario 3 $n = 3000, \quad n_t = 100, \quad T = 30$				Scenario 4 $n = 3000, \quad n_t = 30, \quad T = 100$			
	True	Mean	Sd	Bias	MSE	Mean	Sd	Bias	MSE
$\hat{\beta}_1$	2.000	1.996	0.062	0.004	0.004	2.005	0.064	-0.005	0.004
$\hat{\beta}_2$	4.000	4.000	0.040	0.000	0.002	4.003	0.039	-0.002	0.002
$\hat{\sigma}^2$	1.000	0.998	0.027	0.002	0.001	0.998	0.027	0.002	0.001
$\hat{\beta}_0$	3.000	2.975	0.421	0.025	0.178	2.998	0.257	0.002	0.066
$\hat{\sigma}^2_\eta$	0.600	0.561	0.152	0.039	0.024	0.586	0.090	0.014	0.008
$\hat{\rho}$	0.700	0.586	0.147	0.114	0.035	0.659	0.082	0.041	0.008

Table 4.4: Results for scenarios 5 and 6 of the MC study. The table is separated in two row-blocks: one for the level-1 parameters and the other one for the level-2 parameters.

		Scenario 5 $n = 1500, \quad n_t = 30, \quad T = 50$				Scenario 6 $n = 5000, \quad n_t = 100, \quad T = 50$			
	True	Mean	Sd	Bias	MSE	Mean	Sd	Bias	MSE
$\hat{\beta}_1$	2.000	2.005	0.090	-0.005	0.008	2.001	0.047	-0.001	0.002
$\hat{\beta}_2$	4.000	3.999	0.054	0.001	0.003	4.001	0.029	-0.001	0.001
$\hat{\sigma}^2$	1.000	0.999	0.037	0.002	0.001	1.000	0.018	0.000	0.000
$\hat{\beta}_0$	3.000	3.012	0.341	-0.012	0.116	2.989	0.352	0.011	0.124
$\hat{\sigma}^2_\eta$	0.600	0.581	0.129	0.019	0.017	0.568	0.109	0.032	0.013
$\hat{\rho}$	0.700	0.630	0.120	0.070	0.019	0.632	0.108	0.068	0.016

We have generated n $(= n_t * T)$ level-1 and T level-2 errors (ϵ_{it} and η_t) from a normal distribution with zero mean and variance equal to, respectively, the true values of σ^2 and σ^2_η. The vector $\mathbf{u}$ has been generated as an autoregressive process of order 1 with errors η_t. In the balanced case, the matrix $\mathbf{Z}$ takes the following form:

$$\mathbf{Z} = \mathbf{I}_T \otimes \mathbf{1}_{n_t},$$

where the sign $\otimes$ represents the Kronecker product.
As for design matrix $\mathbf{X}$, two exogenous dichotomic covariates have been considered: the first one with 90% of 0 and 10% of 1 and the second one with 68% of 0 and 32% of 1. Then, the $(n \times 1)$ response vector is computed in the following way:

$$\mathbf{y} = \beta_0 + \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \epsilon.$$

The estimation algorithm, presented in section 4.2, is applied on these data to give maximum likelihood estimates for the parameters. We have repeated this procedure for 500 times, and then, for each parameter, we have calculated the following statistics over the 500 replicates:

- mean,
- standard deviation,

- bias, as the difference between the true value of the parameter and the Monte Carlo mean,

- Mean Square Error (MSE), that is $\sum_{h=1}^{500}(\hat{\boldsymbol{\theta}}_h - \boldsymbol{\theta})^2$.

The whole estimation and simulation procedures has been implemented in R with a completely own-written code.
Table 4.2, 4.3 and 4.4 show the results for the 6 scenarios. The statistics for level-1 estimates, β_1, β_2 and σ^2, depend on the global sample size, n. In fact, in scenario 3 and 4, where n is kept constant, means, standard deviations, bias and MSE are similar. As n increases, bias, standard deviations and MSE decreases as expected.
On the other hand, level-2 estimates, β_0, σ_η^2 and ρ, depend on the number of groups, the time points T. In fact, in general, the values of the statistics decrease as T increases.

The results show that the estimators have a good finite sample performance. Therefore, we expect an even better behaviour of the estimators when applied to the Tribal art dataset, which contains much more observations than the simulated data. Also, the validity of the R implementation is confirmed.

Finally, we have extended the simulation study to assess the forecasting capability of the model. To this aim, we have simulated $n + n_t$ level-1 and $T + 1$ level-2 units through the same procedure as before. Then, we have estimated the model by using only the first n units and the first T groups. The $(n_t \times 1)$ response vector $\mathbf{y}_{T+1}$ can be predicted as:

$$\hat{\mathbf{y}}_{T+1} = \hat{\beta}_0 + \mathbf{X}_{T+1}\hat{\boldsymbol{\beta}} + \mathbf{1}_{n_t}\hat{u}_{T+1}$$

where

$$\hat{u}_{T+1} = \hat{\rho}\hat{u}_T.$$

This procedure has been repeated for 500 times, and for each replication s, we have calculated the Mean Absolute Percentage Error:

$$\text{MAPE}_s = \frac{1}{n_t}\sum_{i=1}^{n_t}\left|\frac{y_{i,T+1} - \hat{y}_{i,T+1,s}}{y_{i,T+1}}\right|$$

where $y_{i,T+1}$ is the simulated response for the i-th unit in the group $T+1$.

In order to have an element of comparison, the response vector $\mathbf{y}_{T+1}$ for each replication has been predicted also through a multilevel model with independent random effects in the following way:

$$\hat{\mathbf{y}}_{T+1} = \hat{\beta}_0 + \mathbf{X}_{T+1}\hat{\boldsymbol{\beta}}.$$

Table 4.5 reports mean and standard deviation of the MAPE over the 500 replications both for the model with autoregressive random effects (in the table “AR model”) and for the model with independent random effects (“IND model”) in three different scenarios differing for level-1 and level-2 sample sizes. First of all, it shows that, unsurprisingly, the forecasting capability of both models improves in terms of mean of MAPE as the number of level-1, n, and the number of level-2 units, T, increase. Moreover, the distribution of the MAPE reveals that the model with autoregressive random effects provides a better forecast in all the scenarios for the responses of a (out-of-sample) 1-lagged time point than the classic multilevel model.

Table 4.5: Simulated forecast for the multilevel model with autoregressive random effects (AR model) and for the model with independent random effects (IND model). For both models, the mean and the standard deviation of the MAPE are reported.

MAPE	Mean		Sd	
	AR model	IND model	AR model	IND model
n=300, $n_t = 30$, T=10	0.412	0.439	0.701	0.720
n=3000, $n_t = 100$, T=30	0.365	0.393	0.605	0.714
n=5000, $n_t = 100$, T=50	0.352	0.382	0.982	1.250

4.4 The new model and Tribal art prices

In this section, we show the results of the fit of the new model upon the Tribal art dataset. The set of covariates is the same as that of the classical multilevel model with independent random effects (3.5). The model, that we will call "AR-RE-hedonic", has the following specification:

$$\begin{aligned}\log_{10}(y_{it}) =& \beta_{0t} + \beta_1 \text{OGG}_{it} + \beta_2 \text{REG}_{it} + \beta_3 \text{MATP}_{it} \\ &+ \beta_4 \text{CPAT}_{it} + \beta_5 \text{CATD}_{it} + \beta_6 \text{CABS}_{it} + \beta_7 \text{CABC}_{it} \\ &+ \beta_8 \text{CAES}_{it} + \beta_9 \text{CAST}_{it} + \beta_{10} \text{CAIL:CAAI}_{it} \\ &+ \beta_{11} \text{ASNC:ASLU}_{it} + \epsilon_{it} \\ &\epsilon_{it} \sim \text{NID}(0, \sigma^2) \quad \forall i, \forall t\end{aligned} \tag{4.18}$$

for $t = 1, \dots, T$, where T is the number of semesters, and for $i = 1_t, \dots, n_t$, where n_t is the total number of items sold in the semester t.
At the second level we have:

$$\begin{aligned}\beta_{0t} &= \beta_0 + u_t \\ u_t &= \rho u_{t-1} + \eta_t \\ \eta_t &\sim \text{NID}(0, \sigma^2_\eta), \quad |\rho| < 1\end{aligned} \tag{4.19}$$

Moreover, it is assumed the independence between η_t and u_s for all $s < t$ and between η_t and ϵ_{it} for all $t \neq s = 1, \dots, T$ and for all $i = 1, \dots, n_t$.
The results of the fit, obtained through the EM algorithm, are in Table 4.6. The estimates and the predicted random effects are quite close to those from the RE-hedonic model reported in the second column of Table 3.2. In addition to the within-group variability and the between-group variability, a further variance component appears in the autocorrelated-random-effect model, namely the level-2 residual variability, σ^2_η. In fact, while in the classic two-level model, the level-2 residual variance coincides with the between-group variance, in the model with AR(1) random effects, the between-group vari-

ance takes the following form:

$$\text{Var}(u_t) = \frac{\sigma_\eta^2}{1-\rho^2},$$

that in this case has an estimate equal to 0.031. It is slightly bigger than that of the RE-hedonic model ($\hat{\sigma}_u^2 = 0.026$). Therefore, the proportion of variability explained by the between-semesters variance (ICC) is bigger, 15.3% against 13.3%. This confirms that, at least in part, the structure at the second level has been taken into account. Further, the estimate for the 1-lag autocorrelation is quite high, $\hat{\rho} = 0.705$. It is interesting to note that its magnitude agrees with the expectation formed by observing the plots of the autocorrelation functions relating to the residuals of the RE-hedonic model (3.8).

The likelihood ratio test between the RE-hedonic model and the AR-RE-hedonic model tests the null hypothesis that $\rho = 0$, under the alternative that $\rho \neq 0$. In fact, the two models are nested: when the 1-lag parameter is zero, the random effects are independent, and the multilevel model takes the usual characteristics. In the present case, this test confirms that ρ is significantly different from zero, and, as a result, the unrestricted model is preferable. Also according to the Information Criteria, the AR-RE-hedonic model fits our data better than the RE-hedonic model.

Table 4.6: Parameter estimates of the AR-RE-hedonic model (4.18). Bootstrap standard errors are indicated in parentheses.

	Estimate (s.e.)
$\hat{\sigma}^2$	0.171 (0.000)
$\hat{\sigma}_\eta^2$	0.016 (0.009)
$\hat{\rho}$	0.705 (0.191)
ICC	0.153
AIC	15381
BIC	15940

Table 4.6: continued in the next page

Table 4.6: continued from the previous page

	Estimate (s.e.)
# param.	74
β_0	2.238 (0.367)
Semester	
1998-1	-0.252 (0.021)
1998-2	-0.134 (0.020)
1999-1	-0.074 (0.023)
1999-2	0.124 (0.022)
2000-1	0.231 (0.020)
2000-2	0.196 (0.021)
2001-1	0.170 (0.024)
2001-2	0.036 (0.028)
2002-1	0.116 (0.021)
2002-2	-0.106 (0.034)
2003-1	-0.180 (0.022)
2003-2	-0.263 (0.022)
2004-1	-0.287 (0.023)
2004-2	-0.171 (0.022)
2005-1	0.001 (0.023)
2005-2	-0.021 (0.023)
2006-1	-0.006 (0.023)
2006-2	-0.106 (0.022)
2007-1	-0.066 (0.021)
2007-2	0.054 (0.026)
2008-1	0.002 (0.026)
2008-2	-0.076 (0.024)
2009-1	0.017 (0.024)
2009-2	0.228 (0.036)
2010-1	0.172 (0.033)
2010-2	0.447 (0.048)

Table 4.6: continued in the next page

Table 4.6: continued from the previous page

	Estimate (s.e.)
2011-1	0.163 (0.032)
Type of object: baseline Furniture	
Sticks	-0.087 (0.029)
Masks	0.109 (0.023)
Religious objects	-0.004 (0.025)
Ornaments	-0.099 (0.028)
Sculptures	0.050 (0.023)
Musical instruments	-0.116 (0.031)
Tools	-0.084 (0.023)
Clothing	-0.069 (0.035)
Textiles	-0.037 (0.037)
Weapons	-0.091 (0.027)
Jewels	-0.051 (0.039)
Region: baseline Central America	
Southern Africa	-0.161 (0.032)
Western Africa	-0.105 (0.011)
Eastern Africa	-0.151 (0.023)
Australia	0.064 (0.056)
Indonesia	-0.107 (0.025)
Melanesia	0.006 (0.017)
Polynesia	0.176 (0.019)
Northern America	0.230 (0.017)
Northern Africa	-0.361 (0.119)
Southern America	0.016 (0.024)
Mesoamerica	0.116 (0.019)
Far Eastern	-0.083 (0.119)
Micronesia	0.095 (0.070)
Indian Region	0.293 (0.102)
Asian Southeast	-0.069 (0.118)

Table 4.6: continued in the next page

Table 4.6: continued from the previous page

	Estimate (s.e.)
Middle East	-0.549 (0.091)
Type of material: baseline Ivory	
Vegetable fibre, paper, plumage	-0.046 (0.027)
Wood	0.072 (0.019)
Metal	-0.030 (0.026)
Gold	0.129 (0.032)
Stone	0.040 (0.029)
Precious stone	0.050 (0.034)
Terracotta, ceramic	0.001 (0.024)
Silver	-0.086 (0.042)
Textile and hides	-0.024 (0.033)
Seashell	0.066 (0.045)
Bone, horn	-0.130 (0.034)
Not indicated	0.042 (0.041)
Patina: baseline Not indicated	
Pejorative	0.231 (0.037)
Present	0.032 (0.010)
Appreciative	0.115 (0.011)
Description on the catalogue: baseline Absent	
Short visual descr.	-0.168 (0.028)
Visual descr.	0.004 (0.029)
Broad visual descr.	0.236 (0.034)
Critical descr.	0.217 (0.034)
Broad critical descr.	0.584 (0.039)
Yes vs No	
Specialized bibliography (dummy)	0.135 (0.013)
Comparative bibliography (dummy)	0.118 (0.009)
Exhibition (dummy)	0.067 (0.014)

Table 4.6: continued in the next page

Table 4.6: continued from the previous page

	Estimate (s.e.)
Historicization: baseline Absent	
Museum certification	0.022 (0.017)
Relevant museum certification	0.037 (0.014)
Simple certification	0.037 (0.009)
Illustration: baseline Absent	
Miscellaneous col. ill.	0.411 (0.023)
Col. cover	1.429 (0.095)
Col. half page	0.867 (0.024)
Col. full page	1.013 (0.025)
More than one col. ill.	1.233 (0.029)
Col. quarter page	0.666 (0.022)
Miscellaneous b/w ill.	0.400 (0.034)
b/w half page	0.545 (0.048)
b/w quarter page	0.301 (0.025)
Auction house and venue: baseline Bonhams-New York	
Christie's-Amsterdam	0.795 (0.049)
Christie's-New York	0.721 (0.047)
Christie's-Paris	0.611 (0.046)
Encheres Rive Gauche-Paris	0.536 (0.075)
Koller-Zurich	0.006 (0.051)
Piasa-Paris	0.740 (0.055)
Sotheby's-New York	0.894 (0.047)
Sotheby's-Paris	0.742 (0.047)

Now, we focus on the residuals. The level-1 residuals versus fitted values in Figure 4.1 are centered around zero and show a quite constant variability at the first level. However, unsurprisingly, the plot of the standard deviations of the level-1 residuals for each group, shown in Figure 4.5, highlights the same heteroscedastic time-pattern as for the RE-hedonic model (Figure 3.6).

As said before, to cope with this problem, we have computed robust standard errors through the wild bootstrap procedure (subsection 3.2.1). Nevertheless, the presence of AR(1) random effects requires an extension of the wild bootstrap for time dependent data (Gonçalves and Kilian, 2004; Shao, 2010) but adapted to the multilevel structure. In particular, we have substitute the bootstrap disturbances in step 2 of subsection 3.2.1 by the following expression:

$$\mathbf{r}_t^* = u_t^* + \boldsymbol{\epsilon}_t^*,$$

$$\text{with} \quad \boldsymbol{\epsilon}_t^* = \hat{\epsilon}_t/(1-h_t)^{1/2} \cdot \mathbf{w}_t,$$

where u_t^* is an autoregressive process with disturbances equal to $w_t\hat{\eta}_t$, $\mathbf{w}_t$ is a $(n_t \times 1)$ vector of values independently drawn by the auxiliar distribution (3.6), and h_t is t-th diagonal element of the orthogonal projection matrix of $\mathbf{X}$. In this way, we have taken into account both the time dependence structure at the second level and the heteroscedasticity at the first level.

The Q-Q plot in Figure 4.2a reveals that the distribution of the residuals, though it looks symmetric, has still quite heavier tails. In fact, also in this case, the Shapiro-Wilk test rejects the null hypothesis of normality for the level-1 residuals. On the contrary, the assumption of normality for the random effects is verified by the Q-Q plot in Figure 4.2b and by the Shapiro-Wilk test in Table 4.7.
The boxplots of the residuals by each semester for the AR-RE-hedonic model are in Figure 4.3. Beside it, the boxplots for the case of independent random effects has been repeated in order to make comparisons. The figures look very similar and show that both models capture quite well the grouped structure of data. As done in the previous chapter, in order to test the assumption of error processes for the AR-RE model, we have computed the autocorrelation functions (global and partial) of the means by semester of level-1 residuals. The correlograms in Figures 4.4 do not reveal any residual dependence structure. Hence, the autoregressive behaviour of the level-2 residuals and the level-1 residuals by semester observed for the RE-hedonic model (Figure 3.8) has been completely absorbed by an AR(1) process for

the level-2 errors.

Figure 4.1: Residuals versus fitted values of the AR-RE-hedonic model.

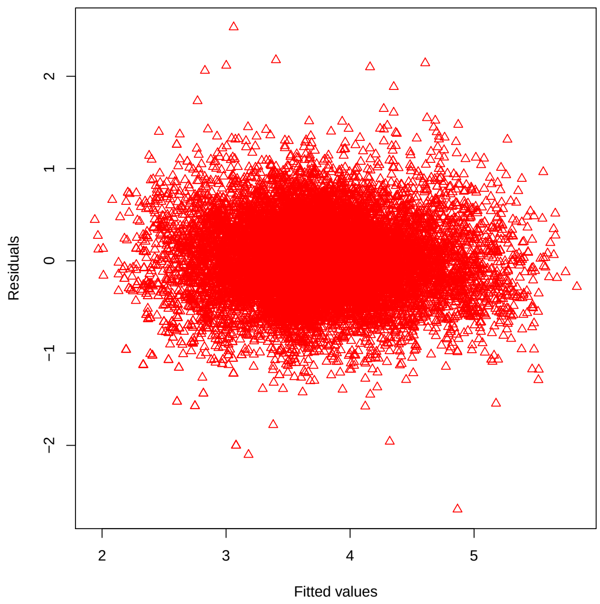
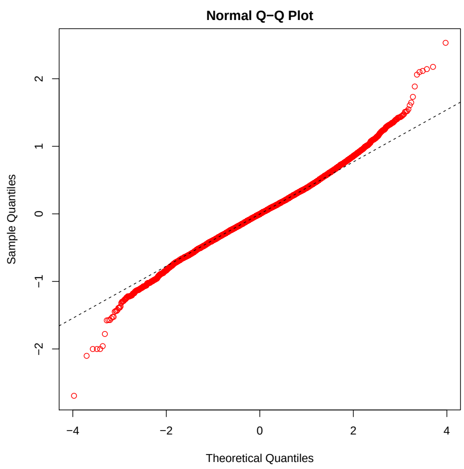


Figure 4.2: Normal probability plots of residuals of the AR-RE-hedonic model.

(a) Level-1 residuals.



(b) Level-2 residuals.

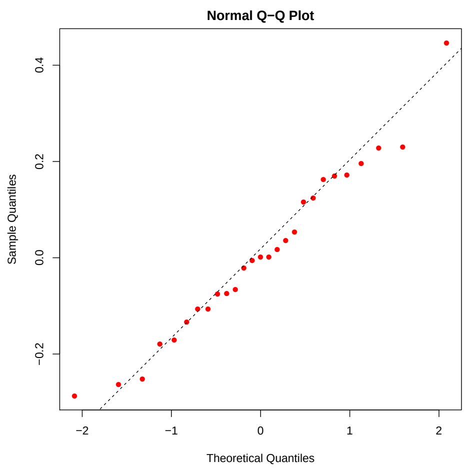


Figure 4.3: Residuals by semester for the AR-RE-hedonic and RE-hedonic model (3.5).

(a) RE-hedonic. (b) AR(1) random effects.

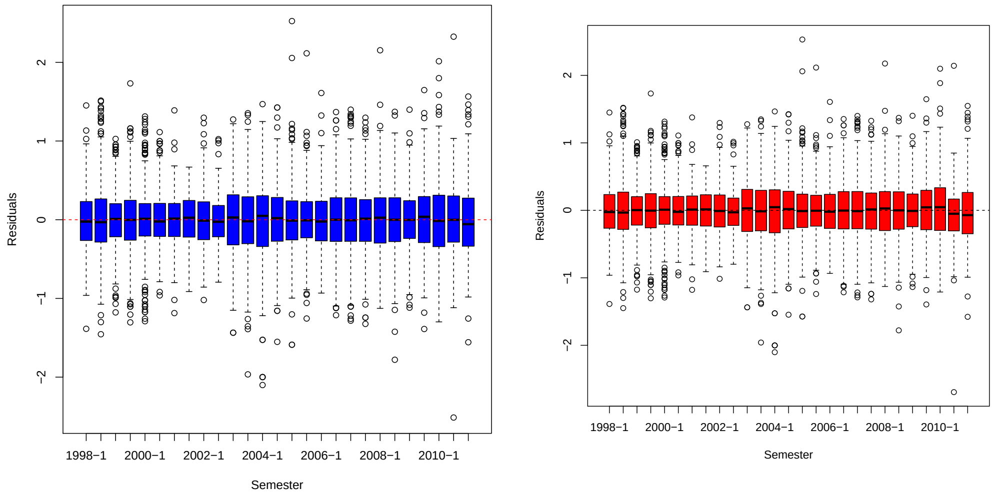


Table 4.7: Shapiro-Wilk normality tests for the residuals of the AR-RE-hedonic model.

	lev-1 residuals	lev-2 residuals
W	0.992	0.974
p-value	0.000	0.721

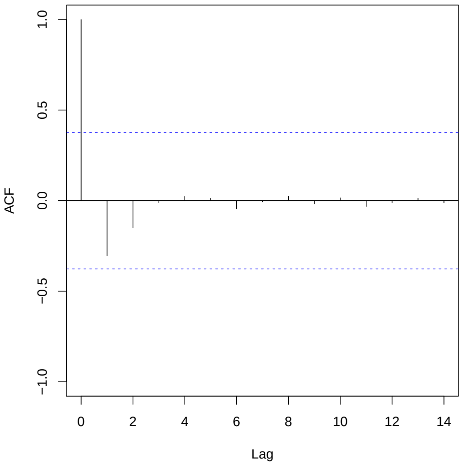
In theory, the multilevel model allows to predict both level-1 unit responses and level-2 unit responses; in other words, both the responses of units in existing groups and the responses of units in not existing groups. Table 4.8 summarizes and compares the prediction capability of the AR-RE-hedonic and RE-hedonic models for the first and second level responses through the following aggregate measures of error:

$$\text{RMSE} = \sqrt{\frac{1}{n_{T+1}} \sum_{i=1}^{n_{T+1}} \left(y_{i,T+1} - \hat{y}_{i,T+1} \right)}$$

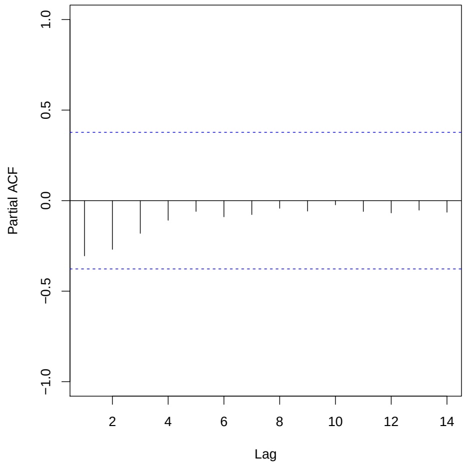
$$\text{MAPE} = \frac{1}{n_{T+1}} \sum_{i=1}^{n_t} \left| \frac{y_{i,T+1} - \hat{y}_{i,T+1}}{y_{i,T+1}} \right| .$$

Figure 4.4: Plots of autocorrelation functions of the residuals of the AR-RE-hedonic model.

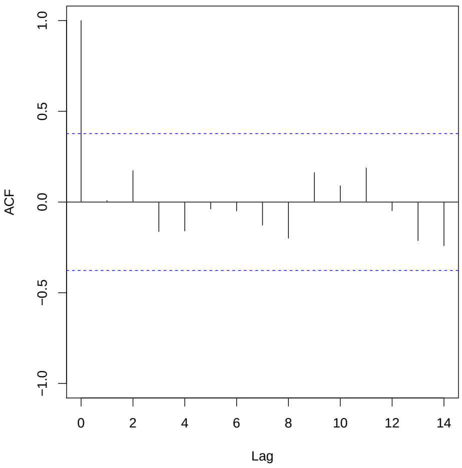
(a) Autocorrelation function of the level-1 residual mean by semesters.



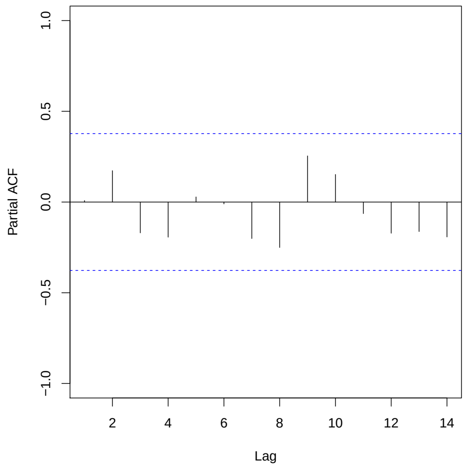
(b) Partial autocorrelation function of the level-1 residual mean by semesters.



(c) Autocorrelation function of the level-2 residuals.

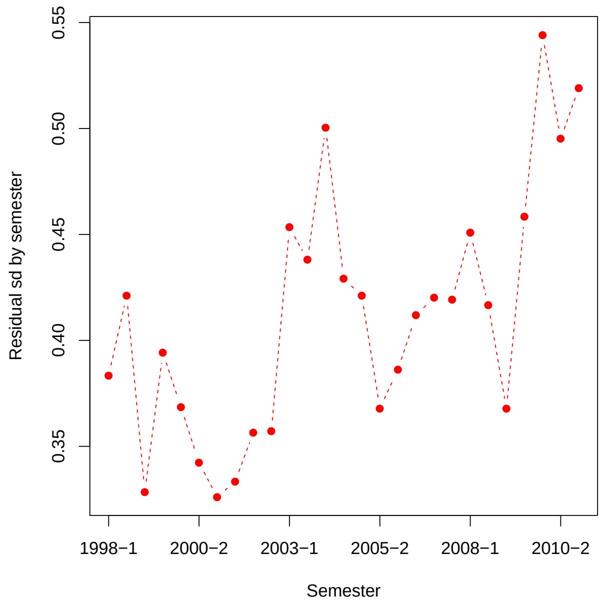


(d) Partial autocorrelation function of the level-2 residuals.



On the one hand, the models predict responses of out-of-sample level-1 units belonging to existing groups with similar performances. On the other hand, the AR-RE-hedonic model allows to forecast better the effect of a 1-lagged out-of-sample semester through the AR(1) process, and, therefore, the prices of objects sold in that semester. More formally, leaving out all the observa-

Figure 4.5: Standard deviations of the level-1 residuals by semester of the AR-RE-hedonic.



tions that belong to the last group, what we want to forecast is the price of these objects, the vector $\mathbf{y}_{T+1}$. The RE-hedonic model obtains a forecast for this vector simply as:

$$\hat{\mathbf{y}}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 \mathrm{OGG}_{it} + \hat{\beta}_2 \mathrm{REG}_{it} + \ldots + \hat{\beta}_{11} \mathrm{ASNC{:}ASLU}_{it},$$

since it does not provide a prediction for the random effect u_{T+1}. For this reason, and also because we saw that it produces the same estimates as the FE-hedonic model, the classic multilevel model and the fixed-effect model have equal forecasting capability. Instead, the AR-RE-hedonic model adds to the value obtained by the RE-hedonic model a further piece of information due to the historical memory; that is, it forecasts the response vector $\mathbf{y}_{T+1}$ as:

$$\hat{\mathbf{y}}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 \mathrm{OGG}_{it} + \hat{\beta}_2 \mathrm{REG}_{it} + \ldots + \hat{\beta}_{11} \mathrm{ASNC{:}ASLU}_{it} + \mathbf{1}_{n_{T+1}} \hat{\rho} \hat{u}_T.$$

The plot in Figure 4.6a compares the elements of the vector $\hat{\mathbf{y}}_{T+1}$ obtained by the AR-RE-hedonic model with those obtained by the RE-hedonic model, and Figure 4.6b reports the regression lines passing through those points. The two lines are parallel because the AR-RE-hedonic model shifts every point obtained by the RE-hedonic by the same quantity, $\hat{\rho}\hat{u}_T$. It is evident that the red dashed line is closer to the bisector line than the blue line, and this confirms that the AR-RE-hedonic model allows a better forecasting than the competing models.

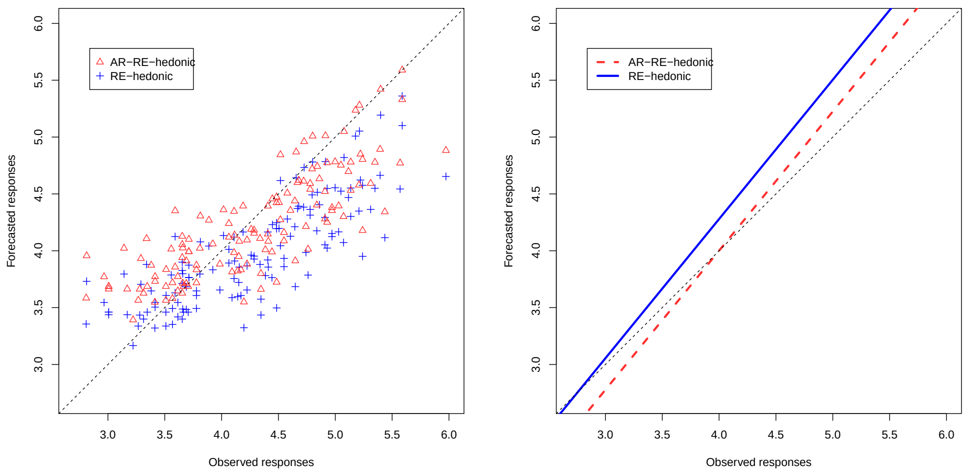
Table 4.8: Prediction of the responses of 100 out-of-sample level-1 units and forecast of the responses of units in one out-of-sample semester, $\mathbf{y}_{T+1}$.

	RE-hedonic	**AR-RE-hedonic**
100 level-1 units in existing groups		
MAPE	0.072	0.072
RMSE	0.342	0.341
units in the semester $T+1$		
MAPE	0.102	0.092
RMSE	0.393	0.367

Figure 4.6: Forecast of n_{T+1} responses corresponding to units in the out-of-sample semester $T+1$. The black dashed line represents the bisector.

(a) Plot of the forecasted vs observed responses obtained by the AR-RE-hedonic and RE-hedonic models.

(b) Regression lines passing through the points of the plot to the left.



Chapter 5

Conclusions

Nowadays, artwork items are considered investment goods in the same way as stocks, bonds and real estates. For this reason, the convenience of investing in the art market requires to be evaluated by coupling both the aesthetic and economic value. With respect to the stocks that are exchanged a high number of times in each instant of time, artworks are one-off pieces in their kind, hardly comparable with each other, and they rarely pass through the market. Therefore, the items of arts need peculiar tools of analysis. Some methods to build indexes for artworks prices have been proposed, especially for paintings. Numerous studies have been conducted on different segments of Western art, Impressionist art among all.
In the present work, we have performed an economic and econometric analysis on the Tribal art market, a segment of art that has not received the same interest as other segments by analysts and researchers, until recent years. To this aim, we used a unique and original hand-collected database that includes information on worldwide auctions of Tribal art objects in the time span 1998-2011.

Among the existent approaches for building a price index for the art market, the hedonic regression is the one that better fits Tribal art data. It is a multiple linear regression with fixed effects that takes into account the heterogeneity of artworks by explaining prices through object features and allows to construct a market price index by neutralizing the effect of quality.

However, since generally art data include many qualitative variables, the main drawback of the hedonic approach is the large number of parameters to be estimated, so that the resulting models are not parsimonious.

The first contribution of this thesis is to consider the influence of time effects on artwork prices through a different approach. Since we observe different artworks sold at every auction, Tribal art data do not constitute either a panel or a time series. Rather, they have a two-level structure in that items, level-1 units, are grouped in time points, level-2 units. Hence, we have proposed to exploit the multilevel model to explain heterogeneity of prices among time points. We have applied and compared extensively the classic hedonic regression model and the multilevel model on the Tribal art dataset. The two models provides similar results in terms of estimates and residuals. Since the assumptions of homoscedasticity and normality of the first level errors do not hold, we have computed the standard errors of the estimates through the Wild Bootstrap procedure; it is, in fact, robust to heteroscedastic errors and, being nonparametric, also to not Gaussian errors. Moreover, the time-effects do not result independent for different groups, as the classic multilevel model assumes. In general, in fact, to our knowledge, the applications of multilevel models to longitudinal data consider occasions, that is the points in time, as the lowest level units and individuals as higher units. Therefore, any time dependence structure is assessed at the first level.

The main theoretical contribution of this thesis consists in a new extension of the classic multilevel model that consists in relaxing the assumption of independence among random effects and treating them as a time series at the second level. In particular, first we have specified a multilevel model with an AR(1) process at the second level to capture the time dependence among groups. Then, we have derived and implemented in R, with an own-written code, maximum likelihood estimators through the E-M algorithm. We have conducted a Monte Carlo study that has confirmed the good finite sample behaviours of the ML estimators and the validity of our R-code. Finally, we have fitted the new model to the Tribal art dataset. We found that the AR(1) process completely captures the time dependence structure among group-effects. Moreover, with respect to the competing hedonic regression model,

the proposed model presents similar estimates and, consequently, similar interpretation of the results: the estimated regression coefficients can be still interpreted as *shadow prices* for each feature, and an index price for the art market is easily provided through the predictions of the time-effects. On the other hand, the new model has less parameters to be estimated and provides a decomposition of the total variability of the response (as the classic multilevel model). Moreover, the multilevel model with autoregressive random effects allows a better forecasting of the responses of units in a 1-lag-ahead group, that are the prices of objects that will be sold one semester later.
Therefore, the new model improves considerably the fit of the Tribal art data with respect to both the hedonic regression model and the classic multilevel model.

Many applications and further extensions of this model are possible. In fact, by treating the random effects as a time series at the second level, it allows to exploit tools of the time series analysis.

References

D. A. Belsley and E. J. Kontoghiorghes. *Handbook of Computational Econometrics*. J. Wiley & Sons, Chichester, 2009.

H. Blalock. Contextual-effects models: theoretical and methodological issues. *Annual Review of Sociology*, 10:353–72, 1984.

W. Browne and H. Goldstein. Mcmc sampling for a multilevel model with nonindependent residuals within and between cluster units. *Journal of Educational and Behavioral Statistics*, 35:453–73, 2010.

Paul Burton, Lyle Gurrin, and Peter Sly. Extending the simple linear regression model to account for correlated responses: an introduction to generalized estimating equations and multi-level mixed modelling. *Statistics in Medicine*, 17:1261–1291, 1998.

G. Candela and A. E. Scorcu. A price index for art market auctions. an application to the italian market for modern and contemporary oil paintings. *Journal of Cultural Economics*, 21:175–196, 1997.

G. Candela and A.E. Scorcu. *Economia delle arti*. Zanichelli, Bologna, 2004.

O. Chanel. Is the art market predictable? *European Economic Review*, 39: 519–527, 1995.

M. Davidian and D. M. Giltinan. *Nonlinear models for repeated measurement data*. Chapman and Hall, London, 1995.

R. Davidson and E. Flachaire. The wild bootstrap, tamed at last. *Journal of Econometrics*, 146:162–169, 2008.

A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via em algorithm. *Journal of the Royal Statistical Society, Series B*, 39:1–38, 1977.

A.P. Dempster, D.B. Rubin, and R.K. Tsutakawa. Estimation in covariance components models. *Journal of the American Statistical Association*, 76: 341–53, 1981.

Thomas A. DiPrete and Jerry D. Forristal. Multilevel models: Methods and substance. *Annual Review of Sociology*, 20:331–57, 1994.

T. Fearn. A bayesian approach to growth curves. *Biometrika*, 62:89–100, 1975.

Paolo Figini. La valutazione dell'investimento in arte: il caso dell'arte etnica. In G. Candela and M. Biordi, editors, *Arte etnica tra cultura e mercato*, pages 241–242. Skira, Milano, 2007.

W. N. Goetzmann. Accounting for taste: Art and financial markets over three centuries. *The American Economic Review*, 83:1370–1376, 1993.

Harvey Goldstein. Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, 73:43–56, 1986.

Harvey Goldstein. *Multilevel Statistical Models*. John Wiley, New York, 2nd edition, 1995.

Harvey Goldstein. *Multilevel Statistical Models*. J. Wiley & Sons, Chichester, 3rd edition, 2009.

Harvey Goldstein. *Multilevel Statistical Models*. J. Wiley & Sons, Chichester, 4th edition, 2010.

S. Gonçalves and L. Kilian. Bootstrapping autoregressions with conditional heteroscedasticity of unknown form. *Journal of Econometrics*, 123:89–120, 2004.

J. D. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton, 1994.

David A. Harville. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72:320–340, 1977.

C. R. Henderson. Estimation of variance and covariance components. *International Biometric Society*, 9:226–252, 1953.

W. James and C. Stein. Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1: 361–379, 1961.

Kelvin Jones, Ronald John Johnston, and Charles J. Pattie. People, places and regions: exploring the use of multi-level modelling in the analysis of electoral data. *British Journal of Political Science*, 22:343–380, 1992.

Ita G.G. Kreft and Jan De Leeuw. *Introducing Multilevel Modeling*. Sage, London, 1998.

Nan M. Laird and James H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38:963–974, 1982.

Nan M. Laird, Nicholas Lange, and Daniel Stram. Maximum likelihood computations with repeated measures: Application of the em algorithm. *American Statistical Association*, 82:97–105, 1987.

D. V. Lindley and A. F. M. Smith. Bayes estimates for the linear model (with discussion). *Journal of the Royal Statistical Society B*, 34:1–41, 1972.

R. Y. Liu. Bootstrap procedures under some non-i.i.d. models. *Annals of Statistics*, 16:1696–1708, 1988.

M. Locatelli-Biey and R. Zanola. The market for picasso prints: An hybrid approach. *Journal of Cultural Economics*, 29:127–136, 2005.

N. Longford. *Random coefficient models*. Clarendon, Oxford, 1993.

G.S. Maddala. The use of variance components models in pooling cross section and time series data. *Econometrica*, 39:341–58, 1971.

J.R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics - Third Edition*. Wiley, Chichester, 2007.

G.J. McLachlan and J. Krishnan. *The EM Algorithm and Extensions*. Wiley, New York, 1997.

J. Mei and M. Moses. Art as an investment and the underperformance of masterpieces. *American Economic Review*, 92:1656–1668, 2002.

C. Morris. Parametric empirical bayes inference: theory and applications. *Journal of the American Statistical Association*, 78:47–65, 1983.

H. D. Patterson and R. Thompson. Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58:545–54, 1971.

Jose C. Pinheiro and Douglas M. Bates. *Mixed-Effects Models in S and S-Plus*. Springer, 2000.

Stephen Raudenbush. A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *Journal of Educational Statistics*, 18:321–349, 1993.

Stephen W. Raudenbush and A. S. Bryk. A hierarchical model for studying school effects. *Sociology of Education*, 59:1–17, 1986.

Stephen W. Raudenbush and Anthony S. Bryk. *Hierarchical linear models: applications and data analysis methods*. Sage pubblications, Thousand Oaks, 2nd edition, 2002.

S. Rosen. Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy*, 82:34–55, 1974.

S. R. Searle, G. Casella, and C. E. McCulloch. *Variance components*. Wiley, New York, 1992.

S. G. Self and K. Liang. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. *Journal of the American Statistical Association*, 82:605–10, 1987.

M. Seltzer, W. Wong, and A. Bryk. Bayesian analysis in applications of hierarchical models: Issues and methods. *Journal of Educational and Behavioral Statistics*, 21:131–167, 1986.

X. Shao. The dependent wild bootstrap. *Journal of the American Statistical Association*, 105:218–235, 2010.

Anders Skrondal and Sophia Rabe-Hesketh. *Generalized latent variable modeling: multilevel, longitudinal, and structural equation models.* Chapman & Hall/CRC, New York, 2004.

Anders Skrondal and Sophia Rabe-Hesketh. Multilevel and related models for longitudinal data. In J. de Leeuw and E. Meijer, editors, *Handbook of Multilevel Analysis*, chapter 7. Springer, New York, 2008.

Anders Skrondal and Sophia Rabe-Hesketh. Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society A*, 172:659–687, 2009.

A. F. M. Smith. A general bayesian linear model. *Journal of the Royal Statistical Society B*, 35:67–75, 1973.

Marco R. Steenbergen and Bradford S. Jones. Modeling multilevel data structures. *American Journal of Political Science*, 46:218–237, 2002.

J. P. Stein. The monetary appreciation of paintings. *Journal of Political Economy*, 85:1021–1035, 1977.

R. Stiratelli, N. Laird, and J. Ware. Random effects models for serial observations with binary response. *Biometrics*, 40:961–971, 1984.

J. L. F. Strenio, H. I. Weisberg, and A. S. Bryk. Empirical bayes estimation of individual growth curve parameters and their relations to covariates. *Biometrics*, 39:71–86, 1983.

G. Wong and W. Mason. The hierarchical logistic regression model for multilevel analyis. *Journal of the American Statistical Association*, 80:513–524, 1985.