DECLARATION FOR THE PhD THESIS

The undersigned

SURNAME	VITALI
FIRST NAME	AGNESE
PhD Registration Number	1287705

Thesis title:

Multilevel and Spatial Approaches for the Study of Demographic Behaviours

PhD in	STATISTICS
Cycle	XXIII
Candidate's tutor	FRANCESCO C. BILLARI
Year of discussion	2012

Declares, under her responsibility, that:

1) according to the President's decree of 28.12.2000, No. 445, mendacious declarations, falsifying records and the use of false records are punishable under the penal code and special laws, should any of these hypotheses prove true, all benefits included in this declaration are automatically forfeited from the beginning;

2) the University has the obligation, according to art. 6, par. 11, Ministerial Decree of 30th April 1999 protocol no. 224/1999, to keep copy of the thesis on deposit at the Biblioteche Nazionali Centrali di Roma e Firenze, where consultation is permitted;

3) the Servizio Biblioteca Bocconi will file the thesis in its 'Archivio istituzionale ad accesso aperto' and will permit on-line consultation of the complete text;

4) in order keep the thesis on file at Biblioteca Bocconi, the University requires that the thesis be delivered by the candidate to Società NORMADEC (acting on behalf of the University) by online procedure the contents of which must be unalterable and that NORMADEC will indicate in each footnote the following information:

   - thesis: Multilevel and Spatial Approaches for the Study of Demographic Behaviours;
   - by: VITALI AGNESE;
   - discussed at Università Commerciale Luigi Bocconi – Milano in: 2012;
   - the thesis is protected by the regulations governing copyright (law of 22 April 1941, no. 633 and successive modifications). The exception is the right of Università Commerciale Luigi Bocconi to reproduce the same for research and teaching purposes, quoting the source;

5) the copy of the thesis deposited with NORMADEC by online procedure is identical to those handed in/sent to the Examiners and to any other copy deposited in the University offices on paper or electronic copy and, as a consequence, the University is absolved from any responsibility regarding errors, inaccuracy or omissions in the contents of the thesis;

6) the contents and organization of the thesis is an original work carried out by the undersigned and does not in any way compromise the rights of third parties (law of 22 April 1941, no. 633 and successive integrations and modifications), including those regarding security of personal details; therefore the University is in any case absolved from any responsibility whatsoever, civil, administrative or penal and shall be exempt from any requests or claims from third parties;

7) the PhD thesis is not the result of work included in the regulations governing industrial property, it was not produced as part of projects financed by public or private bodies with restrictions on the diffusion of the results; it is not subject to patent or protection registrations, and therefore not subject to an embargo;

Date January, 31 2012
Signed Agnese Vitali

Contents

1. Introduction 5

1.1 References 13

2. Regional differences in young Spaniards' living arrangement decisions: A multilevel approach 16

2.1 Introduction 16
2.2 The transition to adulthood in Spain 19
2.3 Hypotheses and variable selection 21
2.4 Data and Methods 28
2.5 Results 38
2.6 Concluding remarks 51
2.7 References 54
2.8 Appendix 60

3. Living arrangements of second generation immigrants in Spain: A cross-classified multilevel analysis 62

3.1 Introduction 62
3.2 Effect of the immigrants' country of origin 66
3.3 Effect of the province of residence 68
3.4 Interaction between country of origin and province of residence .. 73
3.5 Data and descriptive analyses 76
3.6 Methods 82
3.7 Results 88
3.7.1 Fixed effect estimates 88
3.7.2 Random effect estimates 92
3.7.3 Mapping provincial and sending area effects 100
3.8 Concluding Remarks 106
3.9 References 110
3.10 Appendix 119

4. A Spatial Panel Analysis of Italian Regional Fertility......................... 124

4.1 Introduction 124
4.2 How many Italies are there? A look at sub-national fertility differentials 128
4.3 Correlates of Fertility over time.. 135
4.4 Data 139
4.5 Incorporating space into regression analyses............................ 144
4.5.1 A Spatial Cross-Sectional Perspective 151
4.5.2 A Spatial Panel Perspective.. 156
4.6 Results 161
4.6.1 Cross-section: Geographically Weighted Regressions 161
4.6.2 Spatial Panel Models 168
4.7 Conclusion 172
4.8 References 175
4.9 Appendix 186

5. Concluding Remarks 190

5.1 References 196

List of Figures

Figure 3.1: Empirical Bayes predictions of province effects, immigrants' sample 102

Figure 3.2: Empirical Bayes predictions of province effects, natives' sample 102

Figure 3.3: Empirical Bayes predictions of sending area effects 104

Figure 4.1: TFR in the 20 Italian regions in year 1952 and 2009 130

Figure 4.2: Total Fertility Rate in four selected Italian regions over the period 1952–2009: Sardinia, Basilicata, Calabria (South), Lombardy, Liguria and Valle d'Aosta (North-West) 133

Figure 4.3: TFR in 99 Italian provinces in year 1999 and 2008 134

Figure 4.4: Correlation between TFR and four indicators, Italian provinces, 1999-2008.. 139

Figure 4.5: Observed values of indicators by quantile ranges, year 2007 142

Figure 4.6: Dynamic change of global Moran's I of TFR, 1999-2008 149

Figure 4.7: Results from GWRs on TFR: local parameter estimates by quintile ranges and t-values, year 2007 166

Figure 4.A: Results from GWRs on Adjusted TFR: local parameter estimates by quintile ranges and t-values, year 2007 186

Figure 4.B: TFR and Adjusted TFR in 99 Italian provinces in year 2000 and 2007... 189

List of Tables

Table 2.1: Distribution of young adults by age group and choice of living arrangement for women (W) and men (M) 30

Table 2.2: Descriptive Statistics for explanatory variables 34

Table 2.3: Three-level random-intercept multinomial logistic model, Women 43

Table 2.4: Three-level random-intercept multinomial logistic model, Men 49

Table 3.1: Proportion of second generation immigrants outside the parental home (descending order) by sending areas 79

Table 3.2: Distribution of Spanish provinces according to the proportion of immigrant and native young adults outside the parental home 81

Table 3.3: Descriptive statistics for the individual, provincial and sending area-level covariates 87

Table 3.4: Regression coefficient estimates from the cross-classified multilevel logistic model, immigrants' sample, and 2-level logistic model, natives' sample 90

Table 3.5: Random effects estimates from the multilevel logistic model...97

Table 3.6: Regression coefficient estimates from the multilevel logistic models with community effect 99

Table 3.A: List of sending countries and grouping of countries into sending areas...119

Table 3.B: Regression coefficient estimates from the 2-level logistic models...121

Table 3.C: Rank of Empirical Bayes predictions of province effects.......122

Table 3.D: Rank of Empirical Bayes predictions of sending area effects.123

Table 4.1: Estimates of the regression of fertility (TFR) on selected indicators, panel and spatial panel models with provincial fixed effects. 169

Table 4.2: Estimates of the regression of fertility (TFR) on selected indicators, spatial panel models with provincial random effects............171

Table 4.A: Estimates of the regression of fertility (Adjusted TFR) on selected indicators, panel and spatial panel models with provincial fixed effects...188

Acknowledgments

This thesis is the outcome of a fruitful, energetic and stimulating –though not always happy– Ph.D. period. The thesis benefited by brainstorming with a number of colleagues and friends plus discussions with professors and researchers with whom I had the opportunity and the chance to interact over these years. Saying that Francesco Billari has been my advisor won't make justice to his role. He has been a guide, an example, a problem-solver, an encouragement for achieving always higher goals. Learning from him has been the biggest asset of my years in Bocconi. I acknowledge stimulating discussions with Bruno Arpino, Patrick Festy, Frank Furstenberg and Ross Macmillan. The thesis also benefited from the feedback of participants to various conferences and seminars where I presented part of the work based on the thesis.

There are people whose "counselling" role has been of paramount importance. First, my parents, Loredana and Paolo, who might not have understood why I wanted to pursue a Ph.D., but who nonetheless always supported me in all decisions I took (or did not take). Special thanks go to Bruno, for his invaluable help, for transmitting me his knowledge of multilevel modelling, for supporting me in those frequent "I-won't-make-it” moments, and for his advanced “counselling” skills. Same special thanks go to Lara, office mate, roommate and BFF. Your constant encouragement cheered me up all along. Thanks to the Dondena girls, Ilaria, Maria, Alessia, Betta, Chiara, Sissi, Noemi, Elena, Léa, Alessandra, Alice, Marta, Giulia + Nicola. There are some friends who really meant a lot to me during these years. Those real friends who kept listening the same complains over and over again, for four years and a half. Just THANKS Gabriella, Alice, Cristina, Valentina, Matilde and Lucio. Thanks to Daniele, for understanding from the very beginning. Thanks to Rodolfo, for his patience and help when studying probability theory and real analysis, and for supporting me in the worst moments. Thanks to Silvia, Maria Anna, and Massimo. Thanks to Pietro Muliere, Raffaella Piccarreta and Rebecca Graziani, and to Angela Baldassarre and Silvia Acquati. Thanks to Giulia, Anna and Enrica, Joanna, Masha and Lesia, from the EDSD period. Thanks to all those who understood and cared.

1. Introduction

In recent decades the life course of individuals in developed societies has undergone unprecedented changes partly due to structural changes and partly due to a value shift and ideational change. In all advanced societies the transition toward the adult status was characterized by a rather normative sequence of five markers: finishing formal education, entering the labour market, leaving the parental home, forming a stable partnership and becoming parent. Further, the timing of events, albeit slightly different in different societies, has been fairly stable over time. This traditional pattern has now been replaced by an increasing diversification of individual trajectories. Quantum, timing and sequencing of the key markers have been altered and a great heterogeneity between and within countries has emerged.

Recent developments in the demography of the life course have led scholars to develop new substantive theoretical frameworks. Advances in demographic and sociological theories concerning the individual life course in turn raised the need for new challenging statistical methodologies able to handle the emerging complexity of the matter being studied (Gauthier, 2007; Liefbroer et al., 2010).

In this thesis I shall use different statistical methodologies to answer a particular methodological challenge posed by emerging demographic behaviours. Structural-economic, institutional circumstances and cultural values have a contextual influence on individual choices and decisions, which guide the occurrence of life course events and their timing (Lesthaeghe et al., 1988; Lesthaeghe, 1998; Teachman et al., 2002; Billari, 2004; Holdsworth et al., 2005; Gauthier, 2007). For instance, the European Fertility Project (Coale et al., 1986; Bongaarts et al., 1996) underlined the importance of cultural factors such as religion, language, ethnicity and region for explaining the fertility transition of European provinces from high to low levels during the 19^{th} and early 20^{th} century, after socio-economic variables had been controlled for (Anderson, 1986; Knodel et al., 1979; Bocquet-Appel et al., 1996). The Second Demographic Transition approach identifies key macro-level factors that contributed to the emergence of new family models, low fertility and the diversification of the life course in general.

Often, contextual influences are located at the country level. However, effects could also operate at a lower geographical level (regions, municipalities, census tracks, neighbourhoods). Influences can also operate through institutions (e.g. schools, churches, firms, welfare regimes, labour and housing markets), social contexts (households, families, ethnic group), personal networks (peers, co-workers), economic contexts (social class, occupations), normative contexts or time (Rindfuss et al., 2004).

If the context is believed to have an influence on individual attitudes or behaviours, it means that individuals belonging to the same context are more likely to show similar behaviours with respect to individuals belonging to different contexts. This characteristic implies that individuals are not independent from each other but nested within contexts.

The most common contexts considered in demographic studies are the ones induced by space. A number of contributions called the attention to the existence of spatial patterns in demographic behaviours and the need to take these into account (Boyle, 2003; Goodchild et al., 2004; Weeks, 2004; Castro, 2007; Voss, 2007; Chi et al., 2008; Lesthaeghe, 2010). Although geographically referenced data have become increasing available, it is still uncommon for demographers to explicitly account for *spatial dependence*[1] in the study of demographic behaviours. Surprisingly, spatial statistics and econometric methodologies have largely been neglected in demographic research. The availability of geographically

[1] In this thesis, the terms "spatial dependence" and "spatial autocorrelation" are used interchangeably.

referenced data, both at aggregate levels and at the individual level, together with major developments of statistical and econometric techniques able to handle estimation of spatial data, recently offered the opportunity to unravel the social mechanisms operating behind observed spatial patterns. Castro (2007) offers a review of spatial demographic applications which explicitly consider spatial effects in fertility, mortality, migration, and population models.

Demographers recognize what is called *spatial heterogeneity* i.e. the importance of different contexts, and take it into account in cross-national or sub-national studies which acknowledge between-country and within-country differences, respectively. Traditional regression models, however, assume independence between observations and therefore, when applied to clustered data, produce biased standard errors. The most widely used techniques to incorporate the existence of different contexts into statistical analyses in demographic studies are to model heterogeneity through separate analyses by country or the inclusion of dummy variables identifying groups of countries/regions/contexts (i.e. fixed effects). Multilevel analysis (Snijders et al., 1999; Goldstein, 2003) is explicitly designed for modelling hierarchical data. The hypothesis of independence among observations is relaxed, as observations are assumed to be independent across different clusters, and this corrects for biases in parameter estimates and standard errors. Apart from producing consistent standard errors, multilevel modelling allows understanding the effects of both individual and contextual macro- and/or meso-level characteristics on individual life course events as well as distinguishing the two. Multilevel modelling enables to distinguish between the effect of individual characteristics and the effect of the contextual factors to which the individual is exposed. In this way it is possible to combine macro- and micro-level influences on individual outcomes. Further, multilevel modelling allows decomposing the total variance of the outcome into components, each associated to one level of the hierarchical data structure. The relative importance of the levels on the outcome of interest can be evaluated confronting the proportion of total variance attributable to each higher level unit (between-context) and the proportion related to the variation between individuals in the same cluster (within-context).

While demographers recognize the importance of *spatial heterogeneity*, the concept of *spatial dependence* is less commonly considered. This is somehow surprising since demographic and sociological theories agree that spatial contiguity generally induces dependence in demographic behaviours via social interaction processes. In social network analysis, in fact, social influence and social learning are modelled using the concept of spatial dependence, where the space is defined on the basis of the individual network.

The rest of this thesis is organized into three chapters and a conclusion. Each chapter is an independent piece of research. What binds the chapters together is their common emphasis on the opportunities offered by available statistical methodologies which, albeit being only rarely applied, are able to tackle the new challenges brought about by developments in social demographic theories.

Individuals can belong to multiple contexts. Contexts are generally hierarchical but there are cases in which contexts overlap (Di Prete et al., 1994). The second chapter focuses on a hierarchy of three levels in which individuals are nested into municipalities, in turn nested into provinces, while the third chapter deals with a non-nested data structure.

In the second chapter I address the question of within-country heterogeneity in one aspect of the transition to adulthood. I focus on young adults' living arrangements decision in Spain, a country representative of the so called ''latest-late'' transition to adulthood (Billari et al., 2002). Micro-census data are used to investigate the relative weight that structural-contextual factors measured at the municipal level and cultural factors measured at the provincial level might have in explaining regional differences in the living arrangements decision for young women and men. I use multilevel multinomial logistic modelling on three choices of living arrangements, namely, co-residing with parents, living outside the parental home and in a partnership, and living outside the parental home but not in a partnership.

In the third chapter I present an application of cross-classified multilevel models to living arrangements of second generation immigrants in the Spanish provinces. Building on the second chapter, I analyse living arrangements of second generation immigrants in Spain simultaneously taking into account two sources of contextual heterogeneity: the immigrants' country of origin and the province of residence in Spain. Immigrants, and in particular second generation immigrants, are an example of individuals belonging to two different non-nested contexts. Immigrants "navigate" between two contexts and during their life course are exposed to two influences: the culture of the country in which they are currently residing, and that of their country of origin. In this setting, the receiving and sending contexts are non-nested, they are two intersecting levels. In fact, immigrants can be either grouped according to their country of origin, or they can be grouped according to the receiving country. The cross-classified multilevel modelling approach allows partitioning the relative importance of the two sources of heterogeneity, while testing the role of macro-level variables measured both at the country of origin and receiving country level. This type of models has received some interest in the study of immigrants' behaviours (e.g., Van Tubergen et al., 2004; Levels and Dronkers, 2008; Levels et al. 2008; Kalmijn et al., 2010), but they should deserve more consideration in this field of research because of their ability to simultaneously take into consideration the different contextual influences to which immigrants are exposed.

In the fourth chapter, I adopt a spatial perspective and explicitly account for spatial dependence. The application is performed on Italian provincial data for the period 1999–2008. it is well known that Italy is a case study in lowest-low fertility, but what is more relevant in the context of this thesis is the fact that, when a sub-national view is considered, fertility levels have varied widely across space and time. While historically fertility was higher in the South of the country, in recent years differentials have reversed. I adopt a macro-level perspective to estimate a series of spatial cross-sectional and panel regression models using fertility indicators as dependent variables, modelling spatial dependence in fertility among Italian provinces. The relative effects on fertility of a selection of indicators are evaluated.

Chapter five concludes with a summary of the main results obtained throughout the thesis, and discusses ideas for future research.

1.1 References

Anderson, B. A. (1986). Regional and cultural factors in the decline of marital fertility in Europe. In Coale, A. J. and Watkins, S. C. (eds.), *The decline of fertility in Europe*, Princeton: Princeton University Press: 293–313.

Billari, F. C., Castiglioni, M., Casto Martìn, T., Michielin, F. and Ongaro, F. (2002). Dynamics of Fertility and Partnership in Europe. In: *Insights and lessons from comparative research*. Klijzing, E. and Corijn, M. (Eds.) Volume II, United Nations, Geneva/New York, 17–41.

Billari, F. C. (2004). Becoming an Adult in Europe: A Macro(/Micro)-Demographic Perspective. Demographic Research, Special Collection 3, Article 2: 15–44.

Bongaarts, J. and Watkins, S. C. (1996). Social interactions and Contemporary Fertility Transitions. *Population and Development Review*, 22(4): 639–682.

Di Prete, T. A. and Forristal, J. D. (1994). Multilevel Models: Methods and Substance. *Annual Review of Sociology*, 20: 331–357.

Gauthier, A. H. (2007). Becoming a Young Adult: An International Perspective on the Transitions to Adulthood. *European Journal of Population*, 23: 217–223.

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Edward Arnold.

Holdsworth, C. and Morgan, D. (2005). *Transitions in context: leaving home, independence and adulthood*. Berkshire: Oxford University Press.

Knodel, J. and Van de Walle, E. (1979). Lessons from the past: policy implications of historical fertility studies. *Population and Development Review*, 5(2): 217–245.

Lesthaeghe, R. and Surkyn, J. (1988). Cultural dynamics and economic theories of fertility change. *Population and Development Review*, 14(1):1–46.

Lesthaeghe, R. (1998). On theory development: Applications to the study of family formation. *Population and Development Review*, 24(1): 1–14.

Levels M. and Dronkers J. (2008). Educational performance of native and immigrant children from various countries of origin, *Ethnic and Racial Studies*, 31(8): 1404–1425.

Levels M., Dronkers J. and Kraaykamp G. (2008). Immigrant children's educational achievement in Western countries: Origin, destination, and community effects on mathematical performance, *American Sociological Review*, 73(5): 835–853.

Liefbroer A. C. and Toulemon, L. (2010). Demographic perspectives on the transition to adulthood: An introduction. *Advances in Life Course Research*, 15(2-3): 53–58.

Rindfuss RR, Guzzo KB, Morgan SP (2004) The Changing Institutional Context of Low Fertility. *Population Research and Policy Review*, 22(5-6): 411–438.

Snijders, T. and Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modelling*. London: Sage.

Teachman, J. and Crowder, K. (2002). Multilevel models in family research: Some conceptual and methodology issues. *Journal of Marriage and the Family*, 64(2): 280-94.

Watkins, S.C. (1987). The fertility transition: Europe and the third world compared. *Sociological Forum*, 2(4): 645–75.

2. Regional differences in young Spaniards' living arrangement decisions: A multilevel approach[2]

2.1 Introduction

During the past few decades, developed countries have witnessed common structural changes in the economic development, policy system, labour market and educational system. Beside these contextual changes, recent decades have also brought about a major revolution in partnership formation patterns, as well as the emergence of new family models and new household types, which can be explained by the Second Demographic Transition theory (see, e.g., Lesthaeghe, 1983; Van de Kaa, 1987; 2001).

The coexistence and interrelationship of ideational and structural factors also acted to de-standardize the transition to adulthood process (Corijin et al., 2001). The traditional pattern, characterized by a universal

---

[2] This chapter has been published in Advances in Life Course Research, Special Issue on "Demographic perspectives on the transition to adulthood" (Vitali, 2010).

sequence of events experienced according to a normative timing, has been replaced by an increasing diversification of individual trajectories towards adulthood. The key events marking the achievement of adult status (i.e., finishing formal education, entering the labour market, leaving the parental home, forming a stable partnership and, eventually, becoming parent) have been altered in all developed societies, albeit at different speed. Indeed, heterogeneity between countries persists, partly due to institutional factors –such as the rules regulating the entrance into the labour and housing markets, the tax system, and the degree of welfare provision and social protection– and partly due to the prevailing societal norms and the strength of ties with the family of origin (Billari et al., 2001; Fernández Cordón, 1997; Reher, 1998). Recent contributions have emphasized the increasing heterogeneity in young adults' behavioural patterns throughout Europe, not only between counties, but also within countries: the quantum of the key events, their timing and sequencing all show an increasingly strong variation (see, e.g., Elzinga et al., 2007; Fussell et al., 2007).

This chapter addresses the issue of observed within-country heterogeneity in young adults' living arrangements, focusing on a single transition to adulthood, namely, the transition outside the parental home, and on a single country, Spain. The aim is to investigate whether within-country variation might be associated with macro-level characteristics observed at the local community level.

While being fully representative of the “Southern European” pattern of leaving home (Cavalli et al., 1995; Reher, 1998), Spain is a particularly interesting case study, as it is characterized by a strong regional variability in family formation processes shaping both the timing and quantum of home leaving, with historical roots which can be traced back to the 18th century (Reher, 1997).

Regional variability in the Spanish case has been widely investigated at the Autonomous Community level (Holdsworth, 1998; Reher, 1991) and at the provincial level (see, e.g., Holdsworth et al., 2002), but less so at a lower level of territorial aggregation. However, certain local structural characteristics like the labour and housing market circumstances can be expected to become important predictors of the home leaving process only when smaller, more homogeneous area aggregations are considered (cf., for a similar argument, Holdsworth et al., 2002: 996). The multilevel framework proposed by Holdsworth et al. (2002) to evaluate existing regional differences in young Spaniards’ living arrangement decisions as influenced by local economic and cultural factors, is here extended in two ways. Firstly, a three-level structure distinguishing the individual, municipal and provincial levels is considered. Secondly, a multinomial approach is implemented, allowing to separate the effects of individual and contextual characteristics on home-leaving, according to the choice of the living arrangements between living with parents, living outside the parental home and with a partner, and living outside the parental home but without a partner (cf. Holdsworth, 1998). The Spanish 2001 extract from the Integrated Public Use Microdata Series International database is used. The decomposition at both the provincial and municipal levels is exploited, merging individual characteristics with macro-level information from other national sources.

2.2 The transition to adulthood in Spain

The cluster of Mediterranean countries (Italy, Greece Spain and Portugal), is characterized by the so-called “latest-late” patterns of transition to adulthood (Billari et al., 2002). The nest leaving process in Southern Europe is strongly associated with marriage (see, e.g., Iacovou, 2002), with the gain of economic independence and the establishment of home-ownership (Holdsworth and Irazoqui Solda, 2002; Mulder, 2006). Moreover, Southern European countries, classified as belonging to the familialistic welfare regime (Esping-Andersen, 1999; Ferrera, 1996), where the family is seen as the main welfare support provider, share many other peculiarities. In fact, beside showing the “latest-late” exit from the parental home and the “lowest-low” fertility in Europe (Kohler et al., 2002), these settings also witness a high youth unemployment rate and the highest proportions of home-owners in Europe, combined with a rather difficult access to the housing market –the latter being characterized by a combination of shortage of available rented accommodations, relatively high housing prices and low mortgage access. In addition, the postponement of stable union formation universally occurring in Europe in the last decades has translated, within the Southern European setting, into a longer stay in the parental home, while elsewhere in Europe young adults tend to live alone or in other types of accommodation before eventually getting married (see, e.g., Fernández Cordón, 1997; Rossi, 1997). At the same time, the general increase in educational attainment and the prolongation of studies in the absence of policies expressively targeted for students in higher education and in combination with a low provision of student accommodations (Baizán et al., 2002), furthermore acted to delay the exit from the parental home.

The Spanish context proves particularly interesting as a case study, given the large and historically documented regional diversity in the family systems, leading to major variation with respect to the timing and the quantum of home leaving (Holdsworth, 1998; Reher, 1997; 1991). More precisely, Autonomous Communities in the North are characterized by later home-leaving if compared to Southern ones. For instance, the highest proportions of young adults aged 20-34 residing with at least one parent are found in the Autonomous Communities of Castilla y León – provinces of Zamora (69%), Palencia (68%) and Valladolid (66 %)– and Galicia–provinces of Lugo (68%) and Coruña (66%). An exception is the province of Girona (48%), in the Autonomous Community of Cataluña. On the other hand, the lowest proportions are found in the Southern Autonomous Communities of the Balearic Islands (45%), Andalucía – province of Almeria (47%)– and Comunidad Valenciana –province of Castellón/Castelló (48%) (source: own elaboration on 2001 Population and Housing Census, INE; Autonomous cities of Ceuta and Melilla excluded).

2.3 Hypotheses and variable selection

In this chapter, the status of residential dependence on parents is compared to residential independence, distinguishing between living independently with a partner and living independently in other situations. There is consensus that a series of individual background characteristics as well as characteristics of the parental home interfere with the choice of the living arrangement. In this chapter, however, the need for a conspicuous number of observations to be disaggregated geographically prevails over the accessibility to detailed personal information. Census data provide only basic demographic information on age, educational level achieved, school enrolment and employment status at the census date.

Many authors acknowledge that the transition to independent living differs between men and women in terms of influence of family structure, family background characteristics, parental resources and atmosphere in the parental home (Blaauboer et al., 2009; Canãda-Vicinay, 2005). Most importantly, young women and men differ in terms of age at leaving the parental home (Buck et al., 1993): women enter a partnership with older men and tend to leave home earlier. Hence, in terms of the choice of living arrangements, women are more likely to live independently from parents and in a partnership than men of the same age. In studying young people's living arrangements, therefore, it is appropriate to perform separate analyses for men and women. In addition, the relationship between employment and the choice of living arrangements is expected to differ for women and men. Being employed is expected to be a main predictor of independent living for men, especially if they co-reside with a partner. For women, being employed is expected to be an important predictor of independent living when they do not co-reside with a partner, because working women might less likely live with a spouse or partner (and, eventually, children). On the other hand, for women it is important to distinguish between part-time and full-time employment. Such a distinction can be inferred since the census contains information on the number of hours worked per week. Only very few men work part-time (see Table 2.2), thus no particular difference between part-time and full-time employment is expected for them, in terms of living arrangements.

Spatial variability in living arrangement decisions is addressed by focussing on two sets of indicators: factors pertaining to the structural-economical sphere and cultural factors. Both sets of factors might interfere with the choice of living arrangements by impeding residential independence in certain areas while facilitating it in certain others. As commonly done in the literature (see, e.g., Martínez-Granado et al., 2002), structural factors are considered along two dimensions: the conditions of the local labour and housing markets. The existence of structural difficulties in entering the labour market is regarded as an important constraint acting delaying the exit from the parental home. This is especially true in the Southern European context, where the low provision of state support increases economical dependence on parents. Moreover, in Mediterranean countries the independent living is strongly associated with economic independence through the achievement of a stable position on the job market. The most powerful indicator of the barriers of entry in the labour market for young people is the youth unemployment rate, which in Spain was 17.85% in 2001. However, a high unemployment rate might also encourage internal migration in order to look for an occupation elsewhere. Holdsworth et al., (2002) do not find any significant effect of unemployment rate on the probability of exiting the parental home in Spain; however, as they also argue, they measure this variable at the provincial level, whereas it would be better to consider the unemployment rate at a lower level of territorial aggregation. Additionally, since in Spain the youth unemployment rate for women (22.2%) was much higher that for men (14.4%) in 2001, gender-specific unemployment rates are considered to be better predictors of difficulties to enter the labour market for young women and men. As concerns the local housing market, information about housing prices would be informative for the purpose of modelling the difficulties that young people encounter in finding available and affordable vacant dwellings to rent or buy. However, such information is not (reliably) accessible at the municipal level, and considering average provincial prices does not seem an alternative, as large within-province variability in local housing markets exist. The proportion of owner-occupied households is therefore used as an alternative indicator for the difficulties in entering the housing market.

It is expected that in municipalities characterized by a high proportion of owner-occupied households, young adults will be less likely to live independently (cf. Mulder, 2006), as the frequency in housing turnover is lower and fewer dwellings are available for rent. The proportion of owner-occupied households needs to be further discussed in relation to the young adult choice of living arrangements. The indicator under consideration might in fact be interpreted also as a proxy for socioeconomic conditions and prosperity in the municipality of residence. If home-ownership is widespread in the municipality of residence, it is likely that the respondent's parents are themselves home-owners, thus more prone to provide monetary help to their children for the residential relocation, also in line with the existing parental economic support in the Southern European settings. Incidentally, Holdsworth and Irazoqui Solda (2002) found an association between young adults choice for rented versus owned accommodation and parental tenure. Moreover, home-ownership involves a lower cost for a couple than for young people who are not in partnership, and (prospective) couples more often opt for home-ownership rather than for rented accommodation (cf. Mulder, 2006). On this basis, the association between the proportion of owner-occupied households and the probability of living in a partnership rather than not living in a partnership versus living with parents is expected to be different: positive in the former and negative in the latter.

Structural factors alone are not thought of being able to completely explain the regional variation in young adults' living arrangements, and even more so in a complex context like the Southern European one. The explanatory power of cultural factors is modelled through an indicator measuring the "modernity" of the community context and a measure of the strength of ties with the family of origin. A modern environment is defined in opposition to places characterized by a traditional societal norm context, where living independently from parents might more frequently be perceived as an act associated with marriage, entrance into a stable occupation and home ownership. Therefore it is expected that young adults living in a relatively modern cultural context are more likely to live outside the parental home, compared to peers who have been socialized in a more tradition-oriented environment. In order to test the assumption on modernity, information on the proportion of cohabiting unions, which is here treated as synonymous with the social acceptability of modern family models, is used. Finally, strong kinship ties, a peculiar characteristic of the Southern European family system (Reher, 1998), might also be contextualized in the broader concept of the social norms environment. On the one hand, strong family ties are more frequently associated with parental support during the process of leaving home and after the residential shift has taken place (see, e.g., Holdsworth, 2004). This might be thought of as a force facilitating the process of home leaving for what concerns the economics of entering independent living, the latter inducing a possible causal effect on poverty or lack of well-being in the absence of adequate individual or household financial resources (Aassve et al., 2007). On the other hand, in a context in which family ties are strong, the nest leaving transition might be regarded in a more traditional way, with acceptability of intergenerational obligations and a tendency, for instance, to prefer marriage to cohabitation or to adapt to societal age norms (see, e.g., Billari et al., 2007). From this perspective, the strongest the ties with the family of origin, the more likely it is that young people co-reside with parents. The strength of ties with the family of origin is mirrored by the degree of intergenerational support toward the elderly. In this respect, the proportion of elderly people (aged 80 or more) co-residing with their adult children instead of living in elderly homes can be referred to as an indicator of strength of family ties.

2.4 Data and Methods

The empirical analysis relies on data from the Spanish 2001 Population and Housing Census, accessed via two different sources: individual information is gathered from the Integrated Public Use Microdata Series International –from now on IPUMS-I– (Minnesota Population Center, 2009), while the macro-level information, at both provincial and municipal level, is provided by the Spanish National Statistical Institute (Instituto Nacional de Estadistica, INE).

Despite the unavailability of retrospective/prospective information and the absence of information on values and preferences, microcensus data provide an opportunity to disaggregate large samples according to the place of residence, so that the question of within-country variability in the choice of living arrangements can be answered by multilevel modelling, while the association with macro-level variables can also be investigated. IPUMS-I collects comparable and harmonized samples of individual-level data from population censuses, which are made available for public use. The individual information refer to a 5% sample of dwellings from the Spanish 2001 Population and Housing census. The IPUMS-I sample is drawn by selecting every 5th household in the census file with a random starting point. The sample is clustered by household and stratified with strata based on key socio-demographic characteristics.

Though the IPUMS-I sample was taken from all of Spain, the smallest identifiable geographic units are municipalities with more than 20,000 inhabitants. As a consequence, smaller municipalities are grouped together into a single category within each province and it is impossible to establish the geographical location of respondents in these areas. Since the association between young adults' living arrangement and economic factors measured at the municipal level is crucial, respondents residing in small municipalities are excluded from the analyses, thus results do not take into account rural areas. The exclusion of these cases does not strongly affect the representative nature of the sample as municipalities with a population larger than 20,000 represent only 4.7% of all Spanish municipalities, but at the same time 65% of the total Spanish population (Source: own elaboration on INE).

In the analyses, only individuals in the age range 17-35 years are considered. The grouping of individuals into households enables the identification of the choice of living arrangements. The sample size allows analysing women and men separately. The distribution of young women and men according to their choice of living arrangements is presented in Table 2.1. The majority of women (51%) and men (59%) in the age range 17-35 still co-reside with their parents. Among young adults who live outside the parental home, the largest proportion co-resides with their spouse (28% of women and 20% of men), while the proportion living with an unmarried partner is much lower (5% for both women and men). The remaining category encompasses other choices of living arrangements such as living alone, or some forms of apartment-sharing or room renting.

Table 2.1: Distribution of young adults by age group and choice of living arrangement for women (W) and men (M)

Age group:	**17-23**		**24-29**		**30-35**		**All ages**	
	W	**M**	**W**	**M**	**W**	**M**	**W**	**M**
Inside the parental home	84	87	52	62	18	28	51	59
Outside the parental home, with partner	3	1	7	6	7	7	5	5
Outside the parental home, with spouse	3	1	25	15	56	47	28	20
Outside the parental home, other situation	10	10	16	17	19	19	15	16
Tot. N.	66,182	68,694	69,692	72,245	66,377	67,012	202,251	207,951

Source: Own calculations on the IPUMS-I 2001 Spanish sample

The variable of interest in this chapter is the choice of living arrangements, consisting of three categories: co-residing with parents (i.e., respondent lives in a household whose head is one of his/her parents), living outside the parental home and in partnership (i.e., respondent lives in a household whose head is not one of his/her parents and co-resides with spouse/partner) and living outside the parental home in other situations (i.e., respondent lives in a household whose head is not one of his/her parents and does not co-reside with spouse/partner).

I acknowledge that the census is not an ideal source for describing young people's living arrangements, because it only gives a snapshot of the household in a given point in time without possibility to reconstruct retrospectively the history of the household members. For this reason, there are cases which cannot distinguished in the present analysis. For example, young adults who re-entered the parental home, as well as young adults who are "living apart and together" (LAT) with parents (Billari et al., 2008) or with a partner are considered as if they never left, due to difficulties in identifying such cases in census data. Moreover, cases of residential mobility across Spanish provinces or municipalities cannot be identified. For these cases the province and municipality of residence at the census date differ from those of the parental home, hence economic and cultural indicators fail to represent the context where the decision to leave home was made. The choice of individual-level variables is limited by the information available from census databases since, for those who left home, there is no information on characteristics of the parental home. The following variables are selected: age, with categories 17-23, 24-29 (reference category) and 30-35 years old, not employed (reference category), part-time (less than 30 hours weekly worked)[3] and full-time employment status (30 or more hours weekly worked), educational attainment (primary or lower level, secondary level –ref.– and higher education completed) and school enrolment (equals one if respondent is attending school and zero otherwise).

Municipal- and provincial-level indicators were constructed from the whole 2001 census by the Spanish National Statistical Institute. Consistent with the choice of indicators formulated in the previous section, but with some adjustments because of data restrictions, the municipal-level variables considered are the gender-specific unemployment rates, computed on individuals aged 20 to 34 years old –which will be referred to

---

[3] In order to classify part-time work in Spain according to the national definition, surveys use a combination of a self-assessment question and the number of hours normally worked per week, such that persons who say they work part-time but report to work more than 35 hours are reclassified as full-time workers; conversely, persons who say they work full-time and work less than 30 hours are reclassified as part-time. Since the census only provides information on the number of hours worked per week, a threshold had to be selected to discriminate between part-time and full time employment. I follow Lamaître et al. (1997) who show that the proportion working part-time in Spain calculated according to the national definition is similar to that calculated according to the number of hours weekly worked less than 30.

as "youth unemployment rate"– and the proportion of owner-occupied households (i.e., a household member is the owner of the household and the household is owned by purchase, also with payments pending, or by inheritance or donation) versus other tenancy regimes. The provincial-level variables considered are the proportion of non-marital cohabiting unions (all types of de facto couples) over all cohabiting couples, and the proportion of elderly people (aged 80+) living in a family dwelling whose household head is his/her son/daughter or son/daughter in law. All community-level variables are computed on the universe of municipalities with more than 20,000 inhabitants, in line with availability of IPUMS-I individual data.

The individual-, provincial- and municipal-level explanatory variables used in the empirical analyses are described in Table 2.2. The final sample is constituted of 410,202 young adults (202,251 women and 207,951 men) aged 17 to 35, grouped into 316 municipalities which in turn are nested into 52 provinces (the 50 Spanish provinces plus the two Autonomous Cities of Ceuta and Melilla).

Table 2.2: Descriptive Statistics for explanatory variables

Variable	Mean	Std. Dev.	Min	Max	Obs.
Individual-level variables, women					
Age: 17-23	0.327	0.469	0	1	202,251
Age: 24-29 (ref.)	0.345	0.475	0	1	202,251
Age: 30-35	0.328	0.470	0	1	202,251
Employment status: Not Employed (ref.)	0.474	0.499	0	1	202,251
Employment status: Employed Part-time	0.083	0.276	0	1	202,251
Employment status: Employed Full-time	0.443	0.497	0	1	202,251
Part-time * Age: 17-23	0.022	0.145	0	1	202,251
Full-time * Age: 17-23	0.082	0.274	0	1	202,251
Part-time * Age: 24-29	0.031	0.173	0	1	202,251
Full-time * Age: 24-29	0.188	0.391	0	1	202,251
Part-time * Age: 30-35	0.031	0.172	0	1	202,251
Full-time * Age: 30-35	0.173	0.378	0	1	202,251
Primary or less education achieved	0.114	0.317	0	1	202,251
Secondary education achieved (ref.)	0.638	0.481	0	1	202,251
Higher education achieved	0.249	0.432	0	1	202,251
School enrolment	0.314	0.464	0	1	202,251
Individual-level variables, men					
Age: 17-23	0.330	0.470	0	1	207,951
Age: 24-29 (ref.)	0.347	0.476	0	1	207,951
Age: 30-35	0.322	0.467	0	1	207,951
Employment status: Not Employed (ref.)	0.322	0.467	0	1	207,951
Employment status: Employed Part-time	0.041	0.198	0	1	207,951
Employment status: Employed Full-time	0.637	0.481	0	1	207,951
Primary or less education achieved	0.151	0.358	0	1	207,951
Secondary education achieved (ref.)	0.672	0.470	0	1	207,951
Higher education achieved	0.177	0.382	0	1	207,951
School enrolment	0.273	0.445	0	1	207,951
Municipal-level variables					
Youth Female Unemployment rate	21.531	8.567	7.527	54.311	316
Youth Male Unemployment rate	12.842	5.296	4.239	40.766	316
% Owner-occupied households	83.199	6.901	54.508	94.791	316
Provincial-level variables					
% Non-marital cohabiting unions	5.467	2.131	2.488	11.810	52
% 80+ whose relation with household head is father/mother (in law)	19.083	4.189	9.831	29.852	52

Source: Own calculations on the IPUMS-I 2001 Spanish sample and 2001 Population and Housing Census (INE)

Multilevel modelling allows investigation of between-group variability while accounting for the association between group-level characteristics and individual outcomes. In the setting of this thesis, multilevel modelling enables a test of the association between young adults' residential choice and structural and cultural factors. The probability of being in one of the three residential statuses (i.e. in the parental home, outside the parental home and in a partnership, the parental home and not in a partnership) is modelled through a three-level random intercept multinomial logistic model, where individuals (level 1) are nested into municipalities (level 2), which are nested into provinces (level 3). The multilevel model allows the intercept of the group regression lines to vary across municipalities and provinces, while assuming the slopes to be constant for each group. The multinomial structure of the model also makes it possible to discern the different effects that individual and regional characteristics might have on the three alternative choices of living arrangements. The model, which belongs to the family of Generalized Linear Mixed Models, can be described by the following multinomial logit link function:

$$\pi_{ijk}^{(s)} = \frac{\exp\left\{\eta_{ijk}^{(s)}\right\}}{1 - \exp\left\{\eta_{ijk}^{(2)}\right\} + \exp\left\{\eta_{ijk}^{(3)}\right\}} \qquad (2.1)$$

and the following linear predictor:

$$\eta_{ijk}^{(s)} = \beta_0^{(s)} + \boldsymbol{\beta}_1^{(s)'}\mathbf{x}_{1ijk} + \boldsymbol{\beta}_2^{(s)'}\mathbf{x}_{2jk} + \boldsymbol{\beta}_3^{(s)'}\mathbf{x}_{3k} + v_k^{(s)} + u_{jk}^{(s)} + \delta_{ijk}^{(s)} \qquad (2.2)$$

where $\pi_{ijk}^{(s)} = \Pr(y_{ijk}^{(s)} = s \mid \mathbf{x}_{1ijk}, \mathbf{x}_{2jk}, \mathbf{x}_{3k}, \mathbf{v}_x, \mathbf{u}_{jk}, \boldsymbol{\delta}_{ijk})$ denotes the probability of being in alternative s $(s = 1,2,3)$ for the categorical response y_{ijk}, for individual $i = 1,\dots,n_{jk}$ in municipality $j = 1,\dots,n_k$ and province $k = 1,\dots,52$. The response variable takes value 2 if the young adult lives independently from parents and in a partnership, and equals 3 if she/he lives independently and not in a partnership, the reference category ($s=1$) being living inside the parental home. The terms $\boldsymbol{x}_{1\boldsymbol{ijk}}$, $\boldsymbol{x}_{2\boldsymbol{jk}}$ and $\boldsymbol{x}_{3\boldsymbol{k}}$ are vectors of individual-, municipal- and provincial-level characteristics, respectively. The random effects $\boldsymbol{v_k}$, $\boldsymbol{u_{jk}}$, $\boldsymbol{\delta}_{ijk}$ are, respectively, provincial-, municipal- and individual-specific effects, assumed to be independent across levels and such that the following assumptions hold:

$$\begin{aligned}
\boldsymbol{v'_k} &= \left(v_k^{(2)}, v_k^{(3)}\right)' \overset{iid}{\sim} N\left(0, \Sigma_{\boldsymbol{v}}\right) \\
\boldsymbol{u'_{jk}} &= \left(u_{jk}^{(2)}, u_{jk}^{(3)}\right)' \overset{iid}{\sim} N\left(0, \Sigma_{\boldsymbol{u}}\right) \\
\boldsymbol{\delta'_{ijk\delta}} &= \left(\delta_{ijk}^{(2)}, \delta_{ijk}^{(3)}\right)' \overset{iid}{\sim} N\left(0, \Sigma\ \right).
\end{aligned} \tag{2.3}$$

All the parameters to be estimated in the models are contrast-specific, in the sense that the choice over the two alternatives (living outside the parental home and in a partnership, living independently but not in a partnership) can be affected differently by different local-specific factors. Conditional independence follows from the assumptions and allows writing the likelihood function (see Grilli et al., 2007: 385). Estimation is carried out using the Iterative Generalized Least Squares procedure with penalized quasi-likelihood (PQL) method approximation including the second order terms of the Taylor series expansion, implemented in MLwiN.

If the estimation procedure might produce random-effects estimates which are biased downwards (see, e.g., et al., 2004), the huge sample size under consideration is expected to compensate for it (see, e.g., Rabe-Hesketh et al., 2002). The adaptive quadrature approximation of the likelihood function available in STATA through GLLAMM was also tried; however, due to the nature of the data which are a 5% extract from the census, the estimation was too slow. The individual level covariance matrix, Σ_{δ}, is empirically not identified. Using the random-utility model specification (see McFadden, 1973), Grilli and Rampichini (2007) show how to compute the Intraclass Correlation Coefficient (ICC) for this model. They show that the ICC is not affected by the individual-level error-term variance (see Appendix 1). In order to assess the significance of the random effects, Likelihood Ratio Tests (LRT) were performed. The significance of the municipal (provincial) random effect's variances in the three-level model is determined comparing with a two-level model with individuals nested into provinces (municipalities). Estimation of the statistical model was carried out using a penalized quasi-likelihood approach. In this circumstance the regression analysis produces an approximation of the likelihood function, on the basis of which the LRT is performed.

A first model (Model 1) provides an estimate of the existing between-area variability, with age as the only control variable, as age is a very strong predictor of home leaving. In a second step (Model 2), the remaining individual-level variables are added to the model, while level-2 and 3 variables are added gradually in subsequent models (Models 3-4 and 5-6 respectively), in order to test the relative changes in the unexplained provincial and municipal variances they give rise to. The provincial and municipal random-effects' variances are expected to decrease after individual covariates are included in the model. Moreover, the variance of the random effect at the provincial (municipal) level is expected to further decrease when provincial- (municipal-) level covariates are included in the model.

2.5 Results

Results from the three-level random-intercept multinomial logistic model for women and men are shown in Table 2.3 and 2.4 respectively. Results for women are discussed first. In Model 1, where only age is controlled for, the initial local variability for the two alternative choices of independent living (outside the parental home with a partner, outside the parental home without a partner) is estimated. As expected, the coefficient estimates for the age variables show that the likelihood to be living outside the parental home increases with age. However, for women in the lower (higher) age category, the probability to live outside the parental home with a partner is much lower (higher) than the probability to live independently without a partner, which is in line with the general idea that age at marriage (or age at stable partnership) is increasing. Initially, the local variability is higher among women living with a partner (between-municipality and between-province variances are estimated as 14% and 12%, respectively) than among women living not in partnership (13% and 9%, respectively) and the variability at the municipal level is higher than the variability at the provincial level. According to the ICC, municipal- and provincial-level random-effect variances explain, respectively, 4% and 3% of the total variance in the equation contrasting women who live independently with a partner with women who live in the parental home; similarly, in the other equation, the ICC equals 4% and 2%, respectively. These estimates suggest that significant unexplained variability across municipalities and provinces exists in the choice of each type of living arrangement outside the parental home. Individual-level variables account for part of the overall provincial and municipal variability (Model 2). The sign of the education variables is the same for the two outcomes. Young women are more likely to co-reside with parents while enrolled in education. The Spanish higher education system is such that universities are distributed all over the country and it is therefore possible to commute from the parental home. The permanence in the parental home while enrolled in higher education is further induced by the low provision of student accommodations by the universities. There is, however, a difference between the two contrasts: school enrolment affects negatively the log odds of living with a partner more than the log odds of the alternative choice of living outside the parental home without a partner. The latter result is probably induced by the fact that those students who leave home for university generally decide to live in apartment-sharing or room-renting accommodation rather than with a partner, although many of them are expected to be registered in the census as still living in the parental home. The probability to live outside the parental home decreases with educational level achieved, probably because lower educated women enter the labour market relatively earlier than women who achieved higher education and therefore have an earlier access to economic independence. The association between actual employment situation of female respondents and independent living varies according to the choice of living arrangements. For all age groups, working women are more likely to live independently and without a partner rather than with their parents, irrespectively of whether they work part-time or full-time. In the same way, being part-time or full-time employed increases the likelihood of living independently with a partner, but only for younger women. The association between female employment and living arrangements is varies depending on the woman's age. Women at older ages who are working are less likely to co-reside with a partner than with parents, and the likelihood is even lower if they are employed full-time rather than part-time. The distinction between part-time and full-time employment is particularly relevant for women in their early thirties. Indeed, women in this age group who are part-time employed are more likely to live with a spouse or an unmarried partner, while those who are full-time employed are more likely to live with their parents. Models 3 and 4 incorporate the two municipal-level variables. In Model 3 the labour market indicator is included and results show that the female youth unemployment rate in the municipality of residence is negatively associated with each choice of the living arrangement outside the parental home. This result confirms the expectations about the relevance of contextual difficulties in entering the local labour market for the living arrangements decisions. Where economical independence is harder to reach due structural difficulties in the labour market, young adults postpone their exit from the parental home. The housing market indicator influences the two contrasts in opposite ways (Model 4). Residing in a municipality where the proportion of owner-occupied households is high increases the likelihood of living with a partner versus living with parents, but it reduces the likelihood of living outside the parental home without a partner. This result shows that shortage of rentable accommodations matters for young people who live independently from their parents and not in a partnership, i.e. for individuals who face the housing cost on their own and are more likely to opt for rented versus owned households.

Hence, the indicators for difficulty in entering the labour and housing markets significantly and negatively affect the log odds of living outside the parental home, with the exception of the housing market indicator in the case of young women co-residing with a partner.

Table 2.3: Three-level random-intercept multinomial logistic model, Women

Outside with Partner VS In PH	MODEL 1		MODEL 2		MODEL 3		MODEL 4		MODEL 5		MODEL 6	
	β	s.e.	β	s.e.	β	s.e.	β	s.e.	β	s.e.	β	s.e.
Individual-level variables												
Age: 17-23	-2.210***	(0.019)	-2.300***	(0.027)	-2.300***	(0.027)	-2.299***	(0.027)	-2.300***	(0.027)	-2.301***	(0.027)
Age: 30-35	1.842***	(0.013)	1.805***	(0.021)	1.805***	(0.021)	1.796***	(0.021)	1.800***	(0.021)	1.802***	(0.021)
Primary or less education achieved			0.328***	(0.017)	0.328***	(0.017)	0.323***	(0.017)	0.325***	(0.017)	0.325***	(0.017)
Higher education achieved			-0.639***	(0.014)	-0.639***	(0.014)	-0.637***	(0.013)	-0.638***	(0.013)	-0.639***	(0.014)
School enrolment			-1.316***	(0.017)	-1.316***	(0.017)	-1.315***	(0.017)	-1.317***	(0.017)	-1.317***	(0.017)
Part-time * Age: 17-23			0.384***	(0.070)	0.384***	(0.070)	0.385***	(0.070)	0.384***	(0.070)	0.384***	(0.070)
Full-time * Age: 17-23			0.457***	(0.040)	0.457***	(0.040)	0.458***	(0.040)	0.458***	(0.040)	0.458***	(0.040)
Part-time * Age: 24-29			-0.182***	(0.032)	-0.183***	(0.032)	-0.184***	(0.032)	-0.184***	(0.032)	-0.184***	(0.032)
Full-time * Age: 24-29			-0.463***	(0.019)	-0.464***	(0.019)	-0.464***	(0.019)	-0.465***	(0.019)	-0.465***	(0.019)
Part-time * Age: 30-35			0.122***	(0.044)	0.127***	(0.044)	0.124***	(0.044)	0.125***	(0.044)	0.125***	(0.044)
Full-time * Age: 30-35			-0.134***	(0.026)	-0.134***	(0.026)	-0.131***	(0.026)	-0.131***	(0.026)	-0.132***	(0.026)
Municipal-level variables												
Youth female unemployment rate					-0.016***	(0.004)	-0.016***	(0.004)	-0.013***	(0.004)	-0.014***	(0.004)
Prop. owner-occupied households							0.010**	(0.004)	0.012***	(0.004)	0.012***	(0.004)
Provincial-level variables												
Prop. non-marital cohabiting unions									0.068***	(0.021)	0.061***	(0.021)
Family Ties											0.014	(0.013)
Random effects												
Intercept	-0.438***	(0.058)	0.233***	(0.053)	0.596***	(0.108)	-0.218	(0.334)	-0.852**	(0.386)	-1.096**	(0.451)
Intercept Variance, Municipal level	0.137***		0.112***		0.109***		0.102***		0.102***		0.101***	
Intercept Variance, Provincial level	0.117***		0.091***		0.074***		0.083***		0.065***		0.064***	

Table 2.3 (Continued)

-2*Loglikelihood IMP	273553	220623	220643	220732	220683	220668
-2*Loglikelihood IM	290034	221137	221126	221182	221152	221145
-2*Loglikelihood IP	294177	230254	230023	228332	228165	228154
ICC MUNICIPALITY	0.039	0.032	0.031	0.029	0.030	0.029
ICC PROVINCE	0.033	0.026	0.021	0.024	0.019	0.019

Outside without Partner VS In PH	**MODEL 1**		**MODEL 2**		**MODEL 3**		**MODEL 4**		**MODEL 5**		**MODEL 6**	
	β	**s.e.**	**β**	**s.e.**	**β**	**s.e.**	**β**	**s.e.**	**β**	**s.e.**	**β**	**s.e.**
Individual-level variables												
Age: 17-23	-0.919***	(0.017)	-1.027***	(0.026)	-1.028***	(0.026)	-1.026***	(0.026)	-1.026***	(0.026)	-1.026 ***	(0.026)
Age: 30-35	1.265***	(0.014)	1.102***	(0.026)	1.103***	(0.026)	1.093***	(0.026)	1.097***	(0.026)	1.096 ***	(0.026)
Primary or less education achieved			0.563***	(0.019)	0.564***	(0.019)	0.556***	(0.019)	0.558***	(0.019)	0.557 ***	(0.019)
Higher education achieved			-0.475***	(0.016)	-0.476***	(0.016)	-0.473***	(0.016)	-0.474***	(0.016)	-0.474 ***	(0.016)
School enrolment			-0.483***	(0.017)	-0.484***	(0.017)	-0.483***	(0.017)	-0.483***	(0.017)	-0.483 ***	(0.017)
Part-time * Age: 17-23			0.207***	(0.062)	0.207***	(0.062)	0.207***	(0.062)	0.206***	(0.062)	0.206 ***	(0.062)
Full-time * Age: 17-23			0.339***	(0.036)	0.339***	(0.036)	0.340***	(0.036)	0.339***	(0.036)	0.339 ***	(0.036)
Part-time * Age: 24-29			0.146***	(0.038)	0.145***	(0.038)	0.144***	(0.038)	0.144***	(0.038)	0.144 ***	(0.038)
Full-time * Age: 24-29			0.040*	(0.023)	0.039*	(0.023)	0.039*	(0.023)	0.038*	(0.023)	0.038 *	(0.023)
Part-time * Age: 30-35			0.302***	(0.053)	0.306***	(0.053)	0.301***	(0.053)	0.302***	(0.053)	0.301 ***	(0.053)
Full-time * Age: 30-35			0.094***	(0.032)	0.093***	(0.032)	0.095***	(0.032)	0.094***	(0.032)	0.094 ***	(0.032)
Municipal-level variables												
Youth female unemployment rate					-0.017***	(0.004)	-0.018***	(0.003)	-0.016***	(0.003)	-0.014 ***	(0.003)
Prop. owner-occupied households							-0.029***	(0.003)	-0.027***	(0.004)	-0.028 ***	(0.003)
Provincial-level variables												
Prop. non-marital cohabiting unions									0.042***	(0.014)	0.045 ***	(0.013)
Family Ties											-0.016 *	(0.009)

Table 2.3 (Continued)

Random effects												
Intercept	-1.261	***	(0.052)	-1.060	***	(0.051)	-0.679	***	(0.108)	1.752	***	(0.300)
Intercept Variance, Municipal level	0.125	***		0.128	***		0.123	***		0.098	***	
Intercept Variance, Provincial level	0.085	***		0.070	***		0.056	***		0.022	***	
-2*Loglikelihood IMP	273553			220623			220643			220732		
-2*Loglikelihood IM	285482			227567			227449			225506		
-2*Loglikelihood IP	278265			225412			224913			223426		
ICC MUNICIPALITY	0.036			0.037			0.035			0.029		
ICC PROVINCE	0.024			0.020			0.016			0.006		

Random effects						
Intercept	1.243	***	(0.355)	1.592	***	(0.398)
Intercept Variance, Municipal level	0.099	***		0.100	***	
Intercept Variance, Provincial level	0.014	***		0.010	***	
-2*Loglikelihood IMP	220683			220668		
-2*Loglikelihood IM	225450			225349		
-2*Loglikelihood IP	223413			223384		
ICC MUNICIPALITY	0.029			0.029		
ICC PROVINCE	0.004			0.003		

Note: “Outside PH with Partner vs In PH” reports coefficient estimates for the model comparing young adults living independently in a partnership (Outside Parental Home with Partner) versus those still co-residing with parents (In Parental Home). “Outside PH without Partner vs In PH” reports coefficient estimates for the model comparing young adults living independently and not in a partnership (Outside Parental Home without Partner) versus those still co-residing with parents (In Parental Home). Significance of the intercept variance terms is determined according to Likelihood Ratio Tests. IMP stands for the 3-level model with individuals, municipalities and provinces; IM and IP for 2-level models with individuals and municipalities and individuals and provinces, respectively.

p-value: *** < 0.001; ** < 0.01; * < 0.05.

Differences in local labour and housing market conditions, measured at the municipal level, together explain part of the random-effect variances across municipalities: the municipal-level variance drops from 0.112 in Model 2, to 0.102 in Model 4 among women who live outside the parental home in a partnership and from 0.128 to 0.098 among the others. In Model 5, the proportion of cohabiting couples in the province of residence is introduced as an indicator of modernity. This causes the provincial-level variance of the random intercept to fall from 0.083 to 0.065 in the equation for women who live with a partner and from 0.022 to 0.014 for women who do not co-reside with a partner outside the parental home. Results show that, when they live in a relatively modern context, young women are more likely to live outside the parental home, especially with a partner. Finally, the indicator chosen for the strength of family ties, i.e. the proportion of individuals aged 80 or more living with son or daughter (in-law), shows no significant association with the probability of living independently in a co-residential union, and a negative association with the probability of living outside the parental home without a partner (Model 6). The effects of individual-level variables are rather constant when municipal- and provincial-level variables are included in the model, indicating the stability of the parameter estimates. In the same way, the estimates of level-2 variables remain quite stable when level-3 variables are included. In the final model (Model 6), the provincial variability is reduced from 0.117 in the null model to 0.064 for women living outside the parental home with a partner and from 0.085 to 0.010 for women living outside the parental home without a partner. The variability across municipalities falls from 0.137 in Model 1 to 0.101 in the first case and from 0.125 to 0.100 in the second case.

Results for men (Table 2.4) differ from those obtained for women in some instances. The initial local variability at both the municipal and provincial levels is higher for young men living outside the parental home without a partner, compared to those living in a partnership, while for women the opposite is true (Model 1). As in the models for women, the initial variation across municipalities is higher than that across provinces. In Model 2, when individual-level variables are added, the status of being employed positively increases the log odds of living independently, especially when one lives with a partner. The effect of employment status on the likelihood of living independently without a partner is the same irrespectively of its part-time or full-time nature, while being full-time rather than part-time employed increases the probability of being in a partnership outside the parental home. The other individual variables behave as expected: the likelihood of living independently from parents increases with age and decreases with educational attainment, and young men who are enrolled in school are less likely to live outside the parental home. Residing in a municipality with a high youth male unemployment rate reduces the log odds of living outside the parental home and this effect is especially negatively associated with living independently with a partner (Model 3). The inclusion of the housing market indicator in Model 4 produces coefficient estimates similar to those obtained for women. As in the case of women, also for men the proportion of cohabiting couples in the province of residence is positively associated with independent living (Model 5). Finally, residing in areas where family ties are stronger is positively associated with the status of living with a partner outside the parental home, but there is no statistical association with the status of living independently in other types of living arrangements (Model 6). The random-effect variances at the municipal and provincial levels among young men living with a partner outside the parental home falls from 0.121 and 0.104, respectively, in the first model, to 0.081 and 0.087 in the final model. Among young men living independently but not in a partnership, level-2 and level-3 variances fall from 0.134 and 0.126, respectively in Model 1, to 0.103 and 0.027 in Model 6.

Table 2.4: Three-level random-intercept multinomial logistic model, Men

Outside with Partner VS In PH	MODEL 1		MODEL 2		MODEL 3		MODEL 4		MODEL 5		MODEL 6	
	β	s.e.	β	s.e.	β	s.e.	β	s.e.	β	s.e.	β	s.e.
Individual-level variables												
Age: 17-23	-2.552***	(0.027)	-2.272***	(0.028)	-2.273***	(0.028)	-2.271***	(0.028)	-2.274***	(0.028)	-2.274***	(0.028)
Age: 30-35	1.855***	(0.013)	1.783***	(0.013)	1.783***	(0.013)	1.778***	(0.013)	1.786***	(0.013)	1.787***	(0.013)
Primary or less education achieved			0.247***	(0.016)	0.249***	(0.016)	0.247***	(0.016)	0.250***	(0.016)	0.250***	(0.016)
Higher education achieved			-0.504***	(0.016)	-0.505***	(0.016)	-0.503***	(0.016)	-0.505***	(0.016)	-0.505***	(0.016)
School enrolment			-0.431***	(0.019)	-0.431***	(0.019)	-0.431***	(0.019)	-0.431***	(0.019)	-0.431***	(0.019)
Part-time			0.894***	(0.034)	0.893***	(0.034)	0.891***	(0.034)	0.894***	(0.034)	0.895***	(0.034)
Full-time			1.276***	(0.018)	1.274***	(0.018)	1.271***	(0.018)	1.276***	(0.018)	1.276***	(0.018)
Municipal-level variables												
Youth male unemployment rate					-0.029***	(0.006)	-0.029***	(0.006)	-0.028***	(0.006)	-0.029***	(0.006)
Prop. owner-occupied households							0.006	(0.004)	0.008**	(0.004)	0.009**	(0.004)
Provincial-level variables												
Prop. non-marital cohabiting unions									0.048**	(0.023)	0.034	(0.023)
Family Ties											0.030**	(0.013)
Random effects												
Intercept	-1.127***	(0.055)	-2.044***	(0.053)	-1.666***	(0.092)	-2.199***	(0.327)	-2.643***	(0.381)	-3.179***	(0.447)
Intercept Variance, Municipal level	0.121***		0.098***		0.085***		0.081***		0.082***		0.081***	
Intercept Variance, Provincial level	0.104***		0.086***		0.096***		0.102***		0.094***		0.087***	
-2*Loglikelihood IMP	224484		171572		171575		171737		171608		171585	
-2*Loglikelihood IM	237345		171905		171862		171948		171903		171892	
-2*Loglikelihood IP	238333		182565		183983		181908		181007		180231	
ICC MUNICIPALITY	0.034		0.028		0.024		0.023		0.024		0.023	
ICC PROVINCE	0.030		0.025		0.028		0.029		0.027		0.025	

Table 2.4 (Continued)

Outside without Partner VS In PH	**MODEL 1**		**MODEL 2**		**MODEL 3**		**MODEL 4**		**MODEL 5**		**MODEL 6**	
	β	**s.e.**	**β**	**s.e.**	**β**	**s.e.**	**β**	**s.e.**	**β**	**s.e.**	**β**	**s.e.**
Individual-level variables												
Age: 17-23	-0.867***	(0.016)	-0.760***	(0.017)	-0.760***	(0.017)	-0.760***	(0.017)	-0.759***	(0.017)	-0.759***	(0.017)
Age: 30-35	0.929***	(0.014)	0.865***	(0.014)	0.865***	(0.014)	0.861***	(0.014)	0.865***	(0.014)	0.866***	(0.014)
Primary or less education achieved			0.496***	(0.016)	0.498***	(0.016)	0.496***	(0.016)	0.498***	(0.016)	0.498***	(0.016)
Higher education achieved			-0.226***	(0.017)	-0.277***	(0.017)	-0.226***	(0.017)	-0.227***	(0.017)	-0.227***	(0.017)
School enrolment			-0.297***	(0.018)	-0.297***	(0.018)	-0.297***	(0.018)	-0.297***	(0.018)	-0.296***	(0.018)
Part-time			0.356***	(0.031)	0.357***	(0.031)	0.356***	(0.031)	0.357***	(0.031)	0.357***	(0.031)
Full-time			0.339***	(0.016)	0.339***	(0.016)	0.338***	(0.016)	0.339***	(0.016)	0.339***	(0.016)
Municipal-level variables												
Youth unemployment rate					-0.019***	(0.006)	-0.031***	(0.006)	-0.028***	(0.005)	-0.028***	(0.005)
Prop. owner-occupied households							-0.032***	(0.004)	-0.028***	(0.004)	-0.028***	(0.004)
Provincial-level variables												
Prop. non-marital cohabiting unions									0.064***	(0.016)	0.063***	(0.016)
Family Ties											-0.006	(0.010)
Random effects												
Intercept	-1.370***	(0.060)	-1.594***	(0.057)	-1.334***	(0.101)	1.475***	(0.337)	0.748**	(0.381)	0.892**	(0.435)
Intercept Variance, Municipal level	0.134***		0.131***		0.130***		0.103***		0.102***		0.103***	
Intercept Variance, Provincial level	0.126***		0.101***		0.086***		0.046***		0.030***		0.027***	
-2*Loglikelihood IMP	224484		171572		171575		171737		171608		171585	
-2*Loglikelihood IM	236076		171713		171729		171806		171732		171724	
-2*Loglikelihood IP	248458		187597		183344		179991		179208		178968	
ICC MUNICIPALITY	0.038		0.037		0.037		0.030		0.030		0.030	
ICC PROVINCE	0.035		0.029		0.025		0.013		0.009		0.008	

Note: "Outside PH with Partner vs In PH" reports coefficient estimates for the model comparing young adults living independently in a partnership (Outside Parental Home with Partner) versus those still co-residing with parents (In Parental Home). "Outside PH without Partner vs In PH" reports coefficient estimates for the model comparing young adults living independently and not in a partnership (Outside Parental Home without Partner) versus those still co-residing with parents (In Parental Home). Significance of the intercept variance terms is determined according to Likelihood Ratio Tests. IMP stands for the 3-level model with individuals, municipalities and provinces; IM and IP refer to 2-level models with individuals and municipalities and individuals and provinces, respectively.
p-value: *** < 0.001; ** < 0.01; * < 0.05.

2.6 Concluding remarks

This study aimed at shedding some new light on the “latest-late” transition to independent living characterizing the Spanish context, by investigating within-Spain differences in young adults’ living arrangements. It has been shown by the means of multilevel modelling that regional variation in the young Spaniards’ living arrangements exists at the provincial, but even more so at a lower level of territorial aggregation, the municipal one.

With regard to young adults’ residential choice outside the parental home, women and men differ in terms of some specific individual characteristics. For young men whose residential status is a cohabiting (marital) union, being employed is a very important predictor; being employed is also important for both women and men residing in other forms of living arrangements outside the parental home. Conversely, a negative (positive) association is found between being full-time (part-time) employed and living in a partnership outside the parental home for women in their early thirties. Results obtained for men confirm the Southern European pattern of leaving the parental home, where economic independence is a necessary condition for nest-leaving. Younger generations of women are closer to men for what concerns the relationship between employment and residential autonomy, while the breadwinner model, weaker for younger women, might explain the difference found for the older generations.

The gender-specific youth unemployment rate in the municipality of residence proved to be strongly and negatively associated with the probability of living independently for both women and men, irrespectively of their choice of living arrangement. As expected, the effect of the proportion of owner-occupied households depends on the choice of destination outside the parental home. A positive association is found between the proportion of owner-occupied households and the probability of living outside the parental home with a partner, and this effect is similar for both women and men, while the association is negative for young adults who opt to live outside the parental home but without a partner. Thus, difficulties in entering the housing market, according to the indicator chosen here, negatively affect the odds of living outside the parental home. However, for young adults who leave home for partnership, this effect could possibly be mediated by the support they receive from parents.

Young adults living in a modern cultural climate are less likely to live with parents, whatever their choice of the living arrangements. Indeed, in such cultural contexts, the traditional association between leaving home and marriage, home-ownership and economical independence might be weaker than in more tradition-oriented environments. Finally, no clear effect is found for strength of family ties using the indicator variable included in this study.

Individual background variables and contextual characteristics are able to explain a large proportion of the unobserved regional variability in the choice of each type of living arrangements outside the parental home among both women and men. However, in the final model the random-effect variances remain significantly different from zero. This result suggests that there need to be other unobserved factors at both the municipal and provincial levels, beyond the individual characteristics and the local structural and cultural aspects included here, which influence the exit from the parental home and the choice of destination.

2.7 References

Aassve, A., Davia, M. A., Iacovou, M. and Mazzuco, S. (2007). Does Leaving Home Make You Poor? Evidence from 13 European Countries. *European Journal of Population*, 23(4): 315–338.

Baizán, P., Michielin, F. and Billari, F. C. (2002). Political Economy and Life Course Patterns: The Heterogeneity of Occupational, Family and Household Trajectories of Young Spaniards. *Demographic Research* 6: 191–240.

Billari, F. C. and Wilson, C. (2001). Convergence Towards Diversity? Cohort Dynamics in the Transition to Adulthood in Contemporary Western Europe. *MPIDR WP* 2001–39.

Billari, F. C., Castiglioni, M., Casto Martìn, T., Michielin, F. and Ongaro, F. (2002). Dynamics of Fertility and Partnership in Europe. In: *Insights and lessons from comparative research*. Klijzing E. and M. Corijn (Eds.) Volume II, United Nations, Geneva/New York, 17–41.

Billari, F. C. and Rosina, A. (2004). Italian "Latest–Late" Transition to Adulthood: An Exploration of its Consequences on Fertility. *Genus*, 60(1): 71–87.

Billari, F. C. and Liefbroer, A. C. (2007). Should I Stay or Should I Go? The Impact of Age Norms on Leaving Home. *Demography*, 44(1): 181–198.

Billari, F. C., Rosina, A., Ranaldi, R. and Romano, M. C. (2008). Young Adults Living Apart and Together (LAT) with Parents: A Three–level Analysis of the Italian Case. *Regional Studies* 42(5): 625–639.

Cavalli, A. and Galland, O. (Eds.) (1995) *Youth in Europe. Social Change in Western Europe*. London: Pinter.

Corijn, M. and Klijzing, E. (2001). *Transitions to Adulthood in Europe*. Dordrecht: Kluwer Academic Publishers.

Elzinga, C. H. and Liefbroer, A. C (2007). De–standardization of Family–Life Trajectories of Young Adults: A Cross-National Comparison Using Sequence Analysis. *European Journal of Population*, 23: 225–250.

Esping-Andersen, G. (1990). *The Three worlds of Welfare Capitalism*. Cambridge: Polity Press.

Fernández Cordón, J. A. (1997). Youth Residential Independence and Autonomy. A Comparative Study. *Journal of Family Issues*, 18(6): 576–607.

Ferrera, M. (1996). The “Southern Model” of Welfare in Social Europe. *Journal of European Social Policy*, 6(1): 17–37.

Fussell, E., Gauthier, A. H. and Evans, A. (2007). Heterogeneity in the Transition to Adulthood: The cases of Australia, Canada, and the United States. *European Journal of Population*, 23: 389–414.

Holdsworth, C. (1998). Leaving Home in Spain: A Regional Analysis. *International Journal of Population Geography*, 4: 341–360.

Holdsworth, C. and Irazoqui Solda, M. (2002). First Housing Moves in Spain: An Analysis of Leaving Home and First Housing Acquisition. *European Journal of Population,* 18: 1–19.

Holdsworth, C., Voast, D. and Tranmer, M. (2002). Leaving Home in Spain: When, Where and Why? *Regional Studies*, 36(9): 989–1004.

Holdsworth, C. (2004). Family Support during the Transition Out of the Parental Home in Britain, Spain and Norway. *Sociology*, 38(5): 909–926.

Iacovou, M. (2002). Regional variations in the transition to adulthood. *Annals of the American Academy of Political and Social Science*, 580(1): 40–69.

Instituto Nacional de Estadistica (INE). 2001 Population and Housing Census. http://www.ine.es/censo/en/inicio.jsp?_IDIOMA=en . Retrieved in May, 2008.

Integrated Public Use Microdata Series – International. Minnesota Population Center, Minneapolis: University of Minnesota. Retrieved in May, 2008. https://international.ipums.org/international/

Kohler, H.-P., Billari, F. C. and Ortega, J. A. (2002). The Emergence of Lowest-Low Fertility in Europe During the 1990s," Population and Development Review 28(4): 641–681.

Lesthaeghe, R. (1983). A century of demographic and cultural change in Western Europe: an exploration of underlying dimensions. *Population and Development Review*, 9(3): 411–435.

Martínez-Granado, M. and Ruiz-Castillo, J. (2002). The decisions of Spanish youth: A cross-section study. *Journal of Population Economics*, 15(2), 305–330.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In: *Frontiers in econometrics*. Zarembka, P. (Ed.) 105–142, Academic Press: New York.

Minnesota Population Center (2009). Integrated Public Use Microdata Series, International, Version 5.0, Minneapolis: University of Minnesota, 2009.

Mulder, C. H. (2006). Home-ownership and family formation. *Journal of Housing and the Built Environment*, 21(3): 281–298.

Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2004). Gllamm Manual. U.C. Berkeley Division of Biostatistics Working Paper Series. WP 16(0. Downloadable from http://www.gllamm.org/docum.html.

Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, 2(1): 1–21.

Reher, D. S. (1998). Family Ties in Western Europe: Persistent Contrasts. *Population and Development Review*, 24: 203–234.

Reher, D. S. (1997). *Perspectives on the family in Spain, past and present*. Oxford: Clarendon Press.

Reher, D. S. (1991). Marriage Patterns in Spain, 1887-1930. *Journal of Family History*, 16(1): 7–30.

Rossi, G. (1997). The Nestlings. Why Young Adults Stay at Home Longer: The Italian Case. *Journal of Family Issues*, 18(6): 627–644.

Vitali, A. (2010). Regional differences in young Spaniards' living arrangement decisions: A multilevel approach. *Advances in Life Course Research*, 15(2-3): 97–108.

Van de Kaa, D. J. (2001). Postmodern fertility preferences: from changing value orientation to new behaviour. *Population and Development Review*, 27: 290–331.

Van de Kaa, D. J. (1987). Europe's Second Demographic Transition. *Population Bulletin*, 42(1): 1–57.

2.8 Appendix

Grilli and Rampichini (2007) show how to compute the Intraclass Correlation Coefficient (ICC) for a two-level multilevel multinomial model. The reasoning can easily be extended to a three-level model. Let $U_{ijk}^{(s)} = \eta_{ijk}^{(s)} + \varepsilon_{ijk}^{(s)}$, where $\eta_{ijk}^{(s)}$ is the linear predictor of the model and $\varepsilon_{ijk}^{(s)}$ is the individual error term in the random utility formulation of the model, independently and identically distributed according to a Gumbel distribution (McFadden, 1973). Setting all parameters of the reference category (s=1) equal to zero, the difference between utilities associated with different choices (the different alternatives of the response variable) can be written as follows: $U_{ijk}^{(s)} - U_{ijk}^{(1)} = \eta_{ijk}^{(s)} + \left[\varepsilon_{ijk}^{(s)} - \varepsilon_{ijk}^{(1)}\right]$.

The term $\varepsilon_{ijk}^{(s)} - \varepsilon_{ijk}^{(1)}$ is the difference between two independent Gumbel distributions and as such, it is therefore distributed according to a Logistic distribution with variance equal to the fixed quantity $\pi^{2/3}$. Substituting the expression for the linear predictor yields:

$$U_{ijk}^{(s)} - U_{ijk}^{(1)} = \beta_0^{(s)} + \boldsymbol{\beta}_1^{(s)'} \mathbf{x}_{1ijk} + \boldsymbol{\beta}_2^{(s)'} \mathbf{x}_{2jk} + \boldsymbol{\beta}_3^{(s)'} \mathbf{x}_{3k} + v_k^{(s)} + u_{jk}^{(s)} + \delta_{ijk}^{(s)} + \left[\varepsilon_{ijk}^{(s)} - \varepsilon_{ijk}^{(1)}\right] \qquad (2.4).$$

It follows that the variance of the difference between utility associated with the choice s and the reference category equals to the sum of three variances: the provincial-level variance $-Var(v_k^{(s)})-$, the municipal-level variance $-Var(u_{jk}^{(s)})-$, the individual-level variance $-Var(\delta_{ijk}^{(s)})-$ and the variance of the difference between the residual error terms associated with the two choices $-Var(\varepsilon_{ijk}^{(s)}-\varepsilon_{ijk}^{(1)})$. If the individual error term δ_{ijk} is not included in the model, its variance is absorbed in the variance of the difference of the errors in the random utility model, and this variance is fixed at $\pi^{2/3}$, thus all other parameters are implicitly rescaled. It follows that the ICC of the *s*-th equation for the province (level three) can be computed as the proportion of total variance due to provincial variability:

$$\mathrm{ICC}_v^{(s)} = \frac{\mathrm{Var}\left(v_k^{(s)}\right)}{\mathrm{Var}\left(v_k^{(s)}\right)+\mathrm{Var}\left(u_{jk}^{(s)}\right)+\pi^2/3} \qquad (2.5).$$

The ICC at the municipal level (level two), instead, cannot be computed as the proportion of total variance due to municipal variability because the ICC measures how similar or different two individuals residing in two different municipalities are. If the two municipalities belong to two different provinces and this is taken into account, the contribution of the municipality will be underestimated because it will also incorporate the provincial effect. Thus, the true ICC at the municipal level can be computed excluding the provincial variance at the denominator, as follows:

$$\mathrm{ICC}_u^{(s)} = \frac{\mathrm{Var}\left(u_{jk}^{(s)}\right)}{\mathrm{Var}\left(u_{jk}^{(s)}\right)+\pi^2/3} \qquad (2.6).$$

3. Living arrangements of second generation immigrants in Spain: A cross-classified multilevel analysis

3.1 Introduction

Young adults who grew up in families of immigrant origin are different from their parents and also from native-born peers. They are in fact bi-cultural, exposed to different normative sets: they learn the culture of the country of origin from their parents and family, while peers and the surrounding social contexts are vehicles for the culture of the country of residence. During their life cycle, they mix the two different cultures they have been exposed to. These two cultures may have different traditions regarding the transition to adulthood, in terms for example of age at leaving the parental home, destination after the first move and conditions that are expected to be met for the move to take place (Wakil et al., 1981;

Nauck, 2001; De Valk and Liefbroer, 2007; Rumbaut et al., 2010). The exposure to two sets of cultural values may eventually give rise to inter-generational conflicts when second generation immigrants start to take life-course decisions in early adulthood (Fuligni et al., 1999; Giguère et al., 2010).

Socio-demographic research on the transition to adulthood of second generation immigrants generally focuses on three life course decisions taken during young adulthood, namely marriage (Kalmijn et al., 2005; Cortina Trilla et al., 2008; Huschek et al., 2010; Kalmijn and Van Tubergen, 2010), employment (Adsera et al., 2007; Algan et al., 2010) and fertility (Roig Vila et al., 2007). Also the analysis of living arrangement decisions of second generation immigrants is receiving increasing attention from social scientists (Fussell et al., 2005; De Valk, 2006; De Valk et al., 2007; Van Hook and Glick, 2007; De Valk and Mencarini, 2009; Beck et al., 2010; Rumbaut et al., 2010).

A number of North American studies document ethnic differences in the living arrangements of the immigrant population in comparison to natives. In particular, evidence is provided that patterns of co-residence between young adults and their immigrant parents vary by origin (Goldscheider et al., 1988, 1989; Burr and Mutchler, 1993; Boyd, 2000; Glick et al., 2002; Mitchell, 2004; Mitchell et al. 2004). Cultural explanations have mainly been proposed to explain the variation in the observed proportions of young adults co-residing with their immigrant parents (Goldscheider et al., 1988; Boyd, 2000; de Valk and Billari, 2007; Giuliano 2007; Boyd and Park, 2010). Analyzing second generation Western-European immigrants in the US, Giuliano (2007) finds that young adults' living arrangement decisions reflect their country of origin more than their present country of residence. Boyd and Park (2010) find that adult children of immigrants who come from countries identified as emphasizing familism (individualism) are more (less) likely to co-reside with parents in Canada.

Less is known about the living arrangements of immigrants' children in Western European countries. This is particularly true for Mediterranean countries which only recently started to attract substantial flows of immigrants. In these countries, the role of immigrants is widely discussed in public debates on issues such as their impact on the labour market and criminality, but their role for future demographic trends is usually overlooked.

Particularly interesting is the case of Spain on which this chapter focuses. First, in Spain, as in other Southern European countries, young adults leave the parental home rather late, so that the label "latest-late" has been introduced to describe the transition to adulthood in these countries (Billari et al, 2002). Second, in the recent past Spain has experienced a steady increase in immigration flows, together with the rapid diversification of immigrant origins. In a “latest-late” context like Spain, the comparison between second generation immigrants and natives’ demographic behaviours is important to better understand possible future dynamics of an increasing portion of the population. Finally, past research and the previous chapter of this thesis have emphasized the non-homogeneity of the transition to adulthood over the different Spanish geographical areas (Reher, 1991; Holdsworth et al, 2001), while the role of the local context for immigrants is less documented in the literature.

This chapter focuses on the analysis of two sources of contextual heterogeneity that characterise second generation immigrants’ living arrangement decisions in Spain. In fact, cross-classified multilevel models allow taking simultaneously into account the influence of the country of origin and that of the province of residence. This modelling approach enables to avoid focusing on a selected number of countries of origin and to exploit the whole heterogeneity of immigrant origins. This chapter assesses if second generation immigrants conform to the latest-late pattern of transition to adulthood which is prevalent among young Spaniards or whether the culture of their country of origin still plays a role with respect to the living arrangement decisions. The chapter further investigates if the province of residence in Spain matters for living arrangement decisions and if it has a different effect for native and immigrant young adults.

3.2 Effect of the immigrants' country of origin

Many authors have recently acknowledged the need for a design that uses data on immigrants from a multitude of countries of origin (Farley et al., 2002; Crul et al., 2003; Van Tubergen, 2004; Levels and Dronkers, 2008; Clark et al., 2009). However, the inclusion into the analysis of all immigrant groups present in a country has to face two obstacles relating to data availability and the method of analysis.

First of all, available data can limit the possibility of considering more than a few immigrant origins. General surveys usually have limited sample sizes and do not allow the implementation of reliable analyses for many immigrant groups. This chapter instead relies on public use micro-census data that allow having a representative sample of all immigrant groups present in Spain together with a sample of natives.

As for the methodology, empirical analyses usually rely on estimating separate statistical models for each immigrant group (e.g.

Boman, 2010) or on using dummy variables to represent heterogeneity across immigrant groups (e.g., Kritz et al., 2000; de Valk and Billari, 2007; Giuliano, 2007; Cortina Trilla et al., 2008). With this approach, however, it is impractical to consider more than a few immigrant origins. Instead, by taking a multilevel approach, heterogeneity among immigrant groups can be modelled through a single random effect. In this way, not only there is no limitation on the number of immigrant origins that can be included, but also small immigrant groups can be considered since they are appropriately weighted in the estimation, depending on the immigrant group sample sizes (Snijders et al., 1999).

In order to assess the role of the cultural heritage on second generation immigrants' living arrangements, it would be interesting to include in the analyses the average (or median) age at leaving home that is typical in each country of origin. This variable would be a direct measure of the normative behaviour in the country of origin. However, data on the timing of leaving home are not available on all the countries of origin considered. As a proxy for mean age at leaving home, I use the mean age at marriage in the country of origin. The idea is that if there is a strong intergenerational transmission of cultural values from immigrants to their children, the probability of co-residing with parents is expected to be higher for those immigrant groups characterised by high mean age at marriage and vice versa. On the contrary, if immigrant children adapt to the behaviours of the native population, no significant association between this variable measured at the countries of origin level and the probability of co-residence with parents is expected to be found.

3.3 Effect of the province of residence

Previous research documented that the timing and quantum of home leaving and marriage in Spain are characterized by substantial regional diversity (Reher, 1991; Holdsworth et al., 2001). Also, in the second chapter a strong effect of the geographical area of residence on young Spaniards' living arrangement decisions was found. Consistently, also in this chapter Spain is not considered as a homogeneous destination for immigrants, but the role of geographical differences is assessed. Therefore, the provincial level is included in the multilevel analyses to investigate if the place of residence matters for the living arrangement decisions and if it has a different effect for natives and immigrant young adults.

A number of socio-demographic studies have called attention to the influence of the residential context on young adults' living arrangements. Structural-economic, institutional and cultural factors have been found to explain much of the existing variability in the transition to independent living experienced by young adults (Billari, 2004; Mulder, 2006; Liefbroer et al., 2010), in particular for young Spaniards, as documented in Holdsworth (2002) and in the second chapter of this thesis. According to this literature, in areas where the entrance into the labour and housing markets is restricted by structural limitations, independence in young adulthood is postponed. It is therefore useful to incorporate in the analysis some measures describing the economic situation of the residential context in order to study young adults' living arrangements. To this aim, following the considerations discussed in the second chapter, the provincial youth unemployment rate and the proportion of owner-occupied households were chosen. It is expected that young adults residing in provinces with high unemployment rate and high proportion of owner-occupied households – i.e. facing difficulties in finding a job and an available dwelling – will postpone their residential independence and thus will more likely live with their parents rather than outside the parental home, compared to young adults living in provinces were economic circumstances are more favourable.

On the other hand, there are also cultural peculiarities of the residential context which need to be considered in the study of regional variation in young adults' living arrangements. Young adults living in traditional contexts are expected to conform to the usual Southern European pattern of leaving the parental home for marriage. Young people living in metropolitan contexts where postmodern values are more widespread instead, are expected to be more open to other forms of independent living – e.g. living alone, in a non-marital cohabitation or co-residing with non-relatives – before, eventually, getting married. These individuals typically leave the parental home earlier than peers who wait the “right moment” for marriage. Again, this cultural trait is measured by the proportion of non-marital cohabiting unions on the total number of co-residing unions in the province of residence. If the cultural climate of the province of residence has an effect on young adults living arrangements, it is expected that young people living in provinces where non-marital cohabitation is more widespread are more likely to live independently from their parents. If youth of immigrant origin replicated behaviours induced by cultural values typical of their countries of origin irrespectively of the Spanish province in which they reside, no significant association between the proportion of cohabiting unions in the province of residence and the probability of co-residence with parents would be found.

For youth of immigrant origin it is also interesting to consider the normative timing of leaving home in the province of residence in Spain. Again, the mean age at marriage of natives in the province of residence is used as a proxy. If young adults of immigrant origin assimilate to natives' behaviours, co-residence probabilities are expected to be higher for immigrants living in provinces characterised by high mean age at marriage and vice versa. If instead they maintained intact the cultural marital trait typical of their country of origin, no significant association between the mean age at marriage in the province of residence and the probability of co-residence with parents would be found. Measuring this indicator for Spanish natives in the province of residence and for stayers in the countries of origin simultaneously allows understanding which of the two the cultural traits – i.e. the country of origin or the province of residence – is more influential on the living arrangement decision.

In this chapter the importance of the residential context at destination for immigrants is stressed. An important point is how to identify the residential context for those individuals who experienced different moves and were therefore exposed to different contexts. Literature on countries with a long-lasting immigration history shows that immigrant populations tend to move within their destination country more often than native populations. Also, some immigrants might move across different nations before eventually settling down. This also applies for Spain (Reher et al., 2009). The census contains information on the internal/international migration status, recoding the previous residence of each member of the household as well as their residence ten years before the census date. The unit of analysis is the province, therefore only movements across different major administrations are of interest. In 2001, roughly 10% of the immigrants in the sample used (694 out of 6761) had ever resided in a different Spanish province than the current one (or abroad).

3.4 Interaction between country of origin and province of residence

The two contextual effects discussed so far, i.e. immigrants' country of origin and province of residence can also interact. The literature on immigrants' assimilation has put attention on the so-called "community effect", i.e., the impact of specific characteristics of immigrants' communities in the area of residence such as the size of the local community on the speed of assimilation (Van Tubergen et al., 2004; Van Tubergen et al., 2005; Levels et al. 2008). For example, in areas where a given immigrant community is large, it is more likely that members of that community keep familiarizing with peers of the same cultural background. As a consequence, the role of the country of origin will be stronger than in areas where the own immigrant community is small. In the context of this thesis, it is of interest to assess if the higher the size of the immigrant group in a given province, the more likely it is that norms regarding living arrangements remain more preserved within immigrants of the same community, and thus the more likely it is that youth of immigrant origin will maintain behaviours typical of their countries of origin.

I test the "community effect" hypothesis by estimating an extended model with an additional random effect allowing the country of origin effect to vary by province of residence in Spain (“interaction” random effect). I will find evidence in support of the “community effect” hypothesis if the variance of the additional random effect will be sizable. In this case, the same country of origin will have a different effect on second generation living arrangements depending on the province-specific immigrant community characteristics. To further test if this “community effect” can be explained by the size of the local community, I also introduce a control variable measuring the relative size of each immigrant community, i.e., the proportion of immigrants from a given sending area represented in the sample in each Spanish province on the total resident population. This variable is measured at the interacted level and is computed using the whole IPUMS-I sample representing the 5% of the Spanish census –thus including all age groups and not just the extract used in the analysis, which refers to young people only. Differently than what was done in chapter two, in this chapter I confine the study of the living arrangements to the dichotomy “inside versus outside the parental home”, and therefore no clear association between the control and the dependent variable is expected. It is not the size of the community per se which increases or decreases the likelihood of co-residing with one’s parents, but the association works through the country of origin effect. This is because for some nationalities the norm is to foster young people to live on their own relatively early, while for other nationalities it is the opposite; since both mechanisms are expected to be at work when the immigrant group size is big, they will compensate each other. However, it is expected that once the control variable measuring the immigrant group size in the province of residence is included in the model, the explanatory power of the mean age at marriage in the country of origin will increase. In fact, in provinces where the own immigrant group size is larger, the norms regarding living arrangements are expected to remain more preserved through a stronger group control. According the same reasoning, it is expected that the inclusion of the control variable will reduce the magnitude of the estimated association between the living arrangements and the provincial mean age at marriage, since in provinces where the own immigrant group size is larger, youth of immigrant origin will be less exposed to the receiving context than they are to the context of their country of origin.

3.5 Data and descriptive analyses

Micro-census data are considered as an opportunity to disaggregate a large sample of immigrants according to their place of residence and country of origin, i.e., the two sources of heterogeneity which are of interest in this chapter. Hence, in order to address the research questions, I again rely on micro-census data from the Spanish 2001 Population and Housing Census from Integrated Public Use Microdata Series, International (IPUMS-I) (Minnesota Population Center, 2009) and provincial level information from the whole census, via the Spanish National Statistical Institute (INE, 2001). Information on the nativity status, country of immigration, and years since immigration took place allow identification of second generation immigrants. Though the country of origin has not been taken into account in the automatic stratification procedure, country of origin groups in the sample are extremely well represented (e.g. Moroccan immigrants represent the 0.76% of the total resident population in both the whole census and in the sample, Ecuadorian 0.53% and 0.54% respectively, Colombian 0.42% and 0.43%).

However, second generation immigrants are identifiable only among those who were born in the country of origin, while those born in Spain from immigrant parents are not identifiable according to census data, unless they co-reside with their parents. This thesis refers to the so called

"1.5 generation" of immigrants, i.e. individuals who were born outside Spain from foreign parents and who immigrated before age 12. In the following, for simplicity the term second generation will be used. Following Cortina Trilla et al. (2008), who use IPUMS-I data for Spain to study marriage patterns of the foreign-born population, immigrants who report a correspondence between their year of birth and the year of immigration are excluded from the sample, due to inconsistencies. The final sample includes individual information for 6,761 young adults of immigrant origins, aged 17 to 35,[4] coming from 70 different countries which are grouped into 35 sending areas, and residing in one of the 50 Spanish provinces (the Autonomous cities of Ceuta and Melilla are excluded from the analyses).

As already mentioned in chapter two, the census is not an ideal source for describing young people's living arrangements. For immigrant youth there are additional cases which cannot be identified through census data, with respect to those already listed. One is that a change in young adults' living arrangements (from dependent to independent living) may be the result of their parents' decision (i.e. the first generation) rather than their own. In particular, parents may come back to their

[4] The minimal age included in the sample is 17 because it is the age marking the end of secondary level of education in Spain.

original country while their children (the 1.5 generation) may stay in Spain.

As shown in Table 3.A in the appendix, countries of origin represented by more than 45 second generation immigrants are considered as separate nationalities (21 in total). The remaining countries are grouped together geographically to form 14 additional clusters, so that, in total, 35 “sending areas” are considered. As shown in Table 3.1, sending area sample sizes range from 15 (Western Europe) to 1,528 (France), the mean size being 193. The number of immigrants in provinces, reported in Table 3.2, ranges from 6 (Cuenca) to 1,036 (Madrid), the mean size being 135. Finally, data on the mean age at marriage for women in each sending area are obtained from the World Marriage Patterns 2000 data produced by the United Nations.

In order to implement comparative analyses, a sample of 562,648 Spanish young adults aged 17 to 35 is also obtained from IPUMS-I. The sample sizes at the provincial level for natives range from 1,029 (Soria) to 75,297 (Madrid), the mean size being 11,253 (Table 3.2).

Table 3.1: Proportion of second generation immigrants outside the parental home (descending order) by sending areas

Sending Area	%	N	Sending Area	%	N
Ecuador	85.6	97	South-Eastern Asia	48.1	27
Western Africa	84.9	53	Uruguay	48.1	79
Portugal	84.4	128	South-Central Asia	44.4	36
Middle Africa	83.7	49	Southern Europe	44.4	63
Northern Africa	76.0	25	United Kingdom	44.2	326
Western Europe	73.3	15	Northern Europe	43.4	189
Eastern Europe	63.9	108	Switzerland	42.5	671
Morocco	62.3	443	South America	41.7	24
Colombia	61.7	128	Venezuela	38.0	474
France	59.4	1,528	Chile	36.5	52
Belgium	57.1	205	Australia	35.9	39
Brazil	54.4	68	Argentina	33.3	288
Americas	54.3	46	Peru	30.9	68
Canada	53.3	45	Western Asia	27.3	22
Germany	51.9	1,141	United States	27.1	70
China	51.6	62	Mexico	24.6	65
Central America	51.1	47	Eastern Asia	23.5	17
Dominican Rep.	50.8	63	**Total**	51.9	6,761

Note: Proportions refer to second generation immigrants residing in Spain, tabulated according their country of origin

Simple descriptive analysis presented in Tables 3.1 and 3.2 clearly show a considerable degree of heterogeneity in young adults' living arrangements with respect to the two dimensions of the country of origin and province of residence. Table 3.1 shows that the proportions of second generation immigrants living outside the parental home vary greatly by origin. At one extreme there are sending areas like Ecuador, Western and Middle Africa, and Portugal with proportions higher than 80 per cent, well above the overall mean (51.9 per cent). At the other extreme of the ranking there are Western and Eastern Asia, United States, and Mexico with percentages below 30 per cent. It should be acknowledged that second generation immigrants who are classified as living independently from their parents, in the case of some selected countries of origin, might co-reside with other relatives in different forms of extended family arrangements; however, in the sample these proportions are low enough to avoid including a separate classification in the living arrangements' category.

Table 3.2 reports the proportions of second generation immigrants and natives living outside the parental home by their province of residence in Spain. Here the heterogeneity is less pronounced but still noteworthy. In the second generation immigrants sample, proportions range from 20 per cent (Soria) to 90 per cent (Huesca), while for natives the variability is lower, from a minimum of 21.4 per cent (Zamora) to a maximum of 49.8 per cent (Baleares Islands).

Table 3.2: Distribution of Spanish provinces according to the proportion of immigrant and native young adults outside the parental home

Province	Immigrants		Natives		Province	Immigrants		Natives	
	%	N	%	N		%	N	%	N
Huesca	90.0	10	35.0	2,411	Cáceres	50.0	28	42.4	5,226
Albacete	72.4	29	39.6	5,069	Cuenca	50.0	6	38.9	2,332
Toledo	70.7	58	40.5	7,173	Segovia	50.0	8	32.1	1,846
Zaragoza	65.7	102	39.1	11,234	Teruel	50.0	10	37.1	1,505
Castellón/Castelló	64.6	65	45.1	6,592	Málaga	49.9	409	42.4	18,245
Jaén	63.3	30	44.2	9,039	Guipúzcoa	48.8	43	38.0	9,425
Huelva	63.3	49	42.6	7,095	Tarragona	48.3	60	44.5	8,061
Lleida	63.0	27	40.3	4,549	León	47.7	111	32.0	5,980
Barcelona	61.1	699	44.9	65,627	Sevilla	46.9	179	41.2	27,208
Rioja (La)	60.6	33	38.8	3,474	Badajoz	46.5	43	41.5	9,083
Murcia	59.9	252	41.2	17,356	Asturias	46.4	224	35.0	13,584
Valencia/València	59.4	404	41.6	31,187	Palmas (Las)	46.4	138	43.9	14,184
Girona	59.3	86	47.1	7,269	S. Cruz de T.	44.7	275	41.1	11,672
Alicante/Alacant	58.5	289	44.7	19,750	Cantabria	44.2	52	34.7	7,405
Cádiz	58.2	110	39.8	17,705	Valladolid	43.2	74	32.8	7,225
Navarra	55.7	70	36.9	7,231	Coruña (A)	42.2	372	35.4	14,641
Ciudad Real	55.0	20	39.3	6,355	Ourense	40.9	193	34.9	3,516
Córdoba	54.3	70	44.3	10,932	Guadalajara	40.9	22	46.3	2,254
Almería	54.0	87	46.2	7,594	Palencia	40.0	20	31.5	2,210
Balears (Illes)	53.2	139	49.8	11,489	Zamora	40.0	40	31.4	2,306
Granada	53.0	149	41.8	11,639	Ávila	39.1	23	33.2	1,890
Madrid	52.6	1,036	40.2	75,297	Pontevedra	38.9	329	35.7	13,063
Álava	52.0	25	38.2	4,084	Lugo	37.5	72	33.9	4,227
Burgos	51.9	27	31.5	4,489	Salamanca	33.8	77	32.2	4,360
Vizcaya	50.6	77	33.7	15,531	Soria	20.0	10	32.0	1,029
					Total	51.9	6,761	40.7	562,648

3.6 Methods

In order to accomplish the objectives of this study, I opted for multilevel models. Multilevel models are widely used in social sciences when data have a hierarchical structure, i.e., units are grouped into clusters. For example, Billari et al. (2008) consider Italian young adults clustered into households and municipalities by using a three-level model. Such a modelling approach enables to take into account the non-independence of units in the same cluster (e.g., municipalities) and to include in the same model variables defined at the different levels. Srholec (2010) clearly discusses the advantages of using multilevel models to analyse hierarchical data. However, research on second generation immigrants utilizes data which are intrinsically characterised by a cross-classified (non-hierarchical) structure. This thesis considers a multilevel structure consisting of second generation immigrants at the first level, clustered into a cross-classification of non-nested second level units defined by place of birth and province of current residence in Spain. This modelling approach allows partitioning the relative importance of the two sources of heterogeneity, while testing the role of macro-level variables measured both at the country of origin and provincial level. For example, the higher variability at the provincial level found for immigrants in Table 3.2 could be simply due to non-homogeneous concentration of immigrants in certain areas. This will be assessed with the cross-classified multilevel analyses, which allows estimating the provincial level variability after the heterogeneity of immigrant origins has been controlled for.

Cross-classified multilevel models have received some interest in studying immigrants' behaviours (e.g., Van Tubergen et al., 2004; Levels and Dronkers, 2008; Levels et al. 2008; Kalmijn et al., 2010), but they should deserve much more consideration in this field of research because of their ability to simultaneously take into account different contextual influences to which immigrants are exposed.

Empirical analyses are based on a cross-classified multilevel logistic model (see, e.g., Rasbash et al., 1994; Van den Noortgate et al., 2003) where the outcome is the probability of living outside the parental home for second generation immigrants. The model, presented in the latent index formulation, takes the form:

$$Y^*_{i(p,s)} = X_{i(p,s)}\beta + Z_p\gamma + W_s\delta + e_{i(p,s)} + u_p + v_s \tag{3.1}$$

where Y^* indicates the (unobserved) propensity to leave the parental home, such that Prob(Y=1) = Prob(Y^*>0). The subscript $i(p,s)$ indicates an immigrant belonging to a generic unit of the cross-classified structure, where $i=1,2,\dots,n_{(p,s)}$; $p=1,2,\dots,50$ indicates the province and $s=1,2,\dots,35$ indicates the sending area. Individual, provincial, and sending area-level variables are identified with X, Z, and W, respectively. The individual error term, $e_{i(p,s)}$, is assumed to follow a standard logistic distribution with mean 0 and variance equal to the fixed fixed at to $\pi^{2/3}$ (i.e. 3.29), while the province (u_p) and the sending area (v_s) error terms are assumed to be normally distributed with zero mean and variance to be estimated (Snijders et al., 1999). These variances are of interest because they measure the importance of the two sources of heterogeneity under analysis.

As individual level covariates the following are considered: gender (reference category: woman), age, educational enrolment (ref. not in school), educational level achieved –primary or less, secondary (ref.), university education achieved–, and employment status (ref. not employed). Gender heterogeneity in the effects of covariates is allowed by including interactions between all individual level variables and the gender indicator. As already mentioned, at the provincial level two indicators of the difficulty to enter the labour and housing markets are considered, the youth unemployment rate and the proportion of owner-occupied households in the province of residence, an indicator of new family models, i.e., the proportion of cohabiting couples, and the mean age at marriage in the province of residence. At the sending area level the mean age at marriage measured in the country of origin is considered.

To contrast immigrant and Spanish young adults' living arrangements, a two-level logistic regression model is also estimated, where natives are nested into provinces:

$$Y^{*} = X_{ip}\beta + Z_{p}\gamma + e_{ip} + u_{p} \qquad (3.2).$$

This model is similar to model (3.1) but here only a random effect for the provincial level is included. The individual and provincial-level covariates that are considered for natives are the same as those illustrated for immigrants.

Finally, we test the "community effect" by introducing an additional random effect in equation (3.1):

$$Y^{*}_{i(p,s)} = X_{i(p,s)}\beta + Z_{p}\gamma + W_{s}\delta + C_{(p,s)} + e_{i(p,s)} + u_{p} + v_{s} + z_{(p,s)} \qquad (3.3).$$

The inclusion of the random effect $z_{(p,s)}$, also assumed to be normally distributed with zero mean and variance to be estimated, allows the country effect to vary by province. In fact, the total contribution of the country of origin to the linear predictor will be equal to $v_s + z_{(p,s)}$ and the same country of origin in different provinces will have a different effect. This "community effect" can be explained by community specific characteristics $C_{(p,s)}$, such as the relative size of any given immigrant community in each province.

For both cross-classified and nested multilevel models, I use maximum likelihood estimation with Laplacian approximation as implemented by *xtmelogit* in Stata11 (StataCorp, 2009). Descriptive statistics for all covariates are presented in Table 3.3, separately for the two samples considered.

Table 3.3: Descriptive statistics for the individual, provincial and sending area-level covariates

	Mean	Std. dev.	Min	Max
Individual-level variables - Natives' sample (*N* 562,648)				
Male	0.51	0.50	0	1
Age	26.30	5.32	17	35
Still in education	0.28	0.45	0	1
Male * Still in education	0.26	0.44	0	1
Employed	0.53	0.50	0	1
Male * Employed	0.62	0.48	0	1
Primary or less	0.14	0.34	0	1
Male * Primary or less	0.16	0.36	0	1
Higher education	0.19	0.39	0	1
Male * Higher education	0.16	0.36	0	1
Individual-level variables - Immigrants' sample (*N* 6,761)				
Male	0.50	0.50	0	1
Age	26.81	5.53	17	35
Still in education	0.28	0.45	0	1
Male * Still in education	0.27	0.44	0	1
Employed	0.52	0.50	0	1
Male * Employed	0.59	0.49	0	1
Primary or less	0.14	0.35	0	1
Male * Primary or less	0.17	0.38	0	1
Higher education	0.18	0.39	0	1
Male * Higher education	0.14	0.35	0	1
Provincial-level variables (*N* 50)				
Youth unemployment rate	16.63	5.66	8.10	33.99
% Owner-occupied households	83.32	4.43	69.54	89.46
% Non-marital cohabiting unions	4.80	2.28	1.92	11.65
Mean age at marriage	27.90	0.83	26.34	29.76
Sending area-level variables (*N* 35)				
Mean age at marriage	23.88	2.36	20.41	29.00
Community-level variables (*N* 792)				
Size of community	2.21	3.47	0	32

3.7 Results

3.7.1 Fixed effect estimates

Table 3.4 presents the coefficient estimates of the cross-classified logistic regression model for the immigrants' sample and Table 3.B in the Appendix those of the two-level logistic regression model estimated on the natives' sample. Estimates are not strictly comparable across the two samples of natives and immigrants given their different scale; however, their sign and significance can be compared.

The effect of individual-level covariates is similar for second generation immigrants and natives. As expected, the likelihood of living independently from parents is higher for women, because they tend to marry older partners, and increases with age. Women who are still enrolled in (higher) education are more likely to co-reside with their parents, while the opposite is found for men. In both samples, employment status decreases the likelihood to live independently for women, while the association is positive for men. The educational level achieved does not influence the living arrangement decision for young adults of immigrant origins, the only remarkable exception is for women with higher education, who are more likely to co-reside with their parents. For native women a similar effect is found: the higher the educational level achieved, the lower the likelihood of living independently. The same holds also for immigrant and native men, although differences in educational achievement have a significantly lower impact on the choice of the living arrangements, if compared to women. The result can be explained by the fact that low-educated individuals enter the job market relatively earlier than peers who enrolled in higher education.

Also the effect of provincial-level covariates among the two samples shows similarities. Youth unemployment rate in the province of residence is positively associated with home-leaving for natives but of low significance, while the association is not statistically significant for immigrants. The proportion of owner-occupied households in the province of residence is positively associated with the probability of living independently for youth of immigrant origin, while it is not significant for natives. The poor significance found for unemployment rates and housing conditions should not, however, be interpreted as a lack of importance of the economic context. Also, the positive (though poorly significant) estimate for provincial unemployment rate is inconsistent with expectations.

Table 3.4: Regression coefficient estimates from the cross-classified multilevel logistic model, immigrants' sample, and 2-level logistic model, natives' sample (standard errors in parenthesis)

	Immigrants' sample			**Natives' sample**		
Individual-level variables:						
Male	-1.535	(0.15)	***	-1.651	(0.018)	***
Age	0.215	(0.01)	***	0.267	(0.001)	***
Still in education	-0.957	(0.12)	***	-1.183	(0.015)	***
Male * Still in education	1.404	(0.18)	***	1.341	(0.024)	***
Employed	-0.255	(0.11)	**	-0.502	(0.012)	***
Male * Employed	1.008	(0.16)	***	1.204	(0.019)	***
Educational level achieved:						
Primary or less	-0.037	(0.14)		0.303	(0.016)	***
Male * Primary or less	0.055	(0.18)		-0.104	(0.021)	***
Higher education	-0.898	(0.10)	***	-0.630	(0.011)	***
Male * Higher education	0.277	(0.15)	*	0.110	(0.017)	***
Provincial-level variables:						
Youth Unemployment Rate	0.002	(0.01)		0.012	(0.007)	*
Owner-occupied Households	0.030	(0.01)	**	0.004	(0.010)	
Cohabiting couples	0.091	(0.03)	***	0.072	(0.019)	***
Mean age at marriage	-0.257	(0.07)	***	---		
Sending area-level variable:						
Mean age at marriage	-0.163	(0.06)	**	---		
Log Likelihood	-3611.6548			-269190.45		
N	6,761			562,648		

These two results are due to the fact that the level of measurement used, i.e., the provincial level, could not be the most appropriate to capture the impact of structural variables which show some degree of heterogeneity also across different municipalities of the same province, as shown in chapter two. The introduction of a finer level of analysis, however, would have been problematic because micro-census data are available at the municipal level only for municipalities with more than 20,000 inhabitants and this would have cause the loss of many immigrants from the sample.

The proportion of cohabiting couples, instead, shows a positive association with the probability of living outside the parental home both for immigrants and natives. This indicates that where non-traditional family models are more common and socially accepted, young adults are encouraged to leave the nest without waiting the “right moment” for marriage. From these results it seems that the local culture matters more than the contextual economic constraints in influencing young adults’ living arrangements.

Finally, in the model for immigrants, the mean age at marriage in the province of residence and country of origin are negatively and significantly associated with the probability of co-residing with parents. Second generation immigrants coming from countries where the age at marriage is lower are less likely to co-reside with their parents. This corroborates the idea that the cultural heritage of second generation immigrants still plays a role in their transition to adulthood. However, letting the marital age in the country of origin “compete” in the model with the same indicator measured at the province of residence in Spain, the provincial level gets the cultural trait more strongly pronounced in the immigrant's behaviour. Thus, while maintaining a connection with behaviours typical of their countries of origin, youth of immigrant origin who live in a province where natives marry later, will more likely co-reside with their parents.

When I exclude from the analysis individuals who resided in a different province or abroad either in 1991 (i.e. 10 years before the 2001 census) or identify their previous residence in a different province or abroad, results are fairly stable (not shown), therefore these cases are not treated as a separate category.

3.7.2 Random effect estimates

Table 3.5 presents the variance component estimates for different types of multilevel models. For the immigrant sample three types of models are considered: two-level hierarchical models with individuals nested in provinces (IP), individuals nested in sending areas (IS), and cross-classified model (IPS). Of course, for the Spanish sample only two-level models (IP) are considered. The magnitude of the province and sending area random effects is assessed using the Intra-class Correlation Coefficient (ICC). The ICC for a given dimension of the multilevel structure (i.e., province or sending) is calculated as the proportion of the estimated variance at that level to the total variance, obtained as the sum of the variances of province, sending and individual effects, the latter being the fixed quantity 3.29 (Snijders et al., 1999: 224).

The null model includes only age and gender as covariates. Comparing the three types of models estimated on the immigrant sample, it emerges that ignoring the cross-classified structure of the data would lead to overestimating the provincial variability. In fact, the ICC at the provincial level decreases from 4.11 per cent (IP) to 2 per cent when the sending area effect is introduced together with the provincial one (IPS). The relative weight of the residual variability at the provincial level is slightly higher for the Spanish sample (ICC = 2.7 per cent) than it is for the immigrants' sample (ICC=2 per cent). This confirms that the higher variability across provinces that resulted from the descriptive statistics in Table 3.2 is due to the non-homogeneous geographical displacement of immigrants.

From the variance decomposition of the cross-classified model it is evident that the country of origin contributes more to explain variability in home-leaving (21 per cent of the total variance) among immigrants than the province of residence does (2 per cent of the total variance). Although the provincial variance is small compared to the country of origin effect, both effects are significant at the 1 per cent level, according to Likelihood Ratio Tests (Snijders et al., 1999). While it is true that second generations navigate between the two cultures they are exposed to, this chapter shows that norms associated with the sending country are more important for determining home leaving than characteristics of the environment in the country of residence. Giuliano (2007) finds similar results for second generation immigrants from Europe in the US. She explains the strong intergenerational transmission of cultural values from immigrants to their children using the cultural interpretation based on strength of family ties originally proposed by Reher (1998). According to this theoretical framework, differences in family structures, marital patterns and living arrangements across European societies have historical roots and are embedded into the cultural heritage of the European populations. For this reason, individual and contextual economic factors of the destination country are not enough to explain differences in living arrangement decisions across immigrant groups. In the case of Spain the importance of the country of origin might be further emphasized by the fact that Spain is a country of recent immigration, and therefore the integration of immigrants is expected to be slower than in countries which have been receiving immigrant populations for a long time.

The high intra-class correlation coefficient at the sending area level suggests the existence of a strong heterogeneity across immigrant groups. This reinforces the importance of not considering immigrants as a unique category and increases the interest for a comprehensive analysis of all immigrant groups.

Residual variance at both provincial and sending area levels remains significant also after controlling for the other individual covariates (Model IPS+X). This is also the case for the provincial effect in the Spanish sample (Model IP+X). This residual variability is attempted to be “explained” by introducing macro-level variables. This is the exercise which is conducted in the remaining rows of Table 3.5.

Introducing all provincial-level variables contributes to explain 60 and 30 per cent of the residual provincial-level variance for the immigrants and natives’ samples, respectively. In both samples most of this explanatory power is attributable to the proportion of cohabiting couples in the province of residence, followed by the mean age at marriage, which respectively explains, if included alone, 22 and 14 per cent of the provincial effect’s variance. The mean age at marriage in the sending area alone explains 17 per cent of the residual variability across sending areas, thus confirming that norms and behaviours which are typical of the country of origin still play a role for second generation immigrants, when the co-residence with parents is concerned.

Table 3.5: Random effects estimates from the multilevel logistic model

	Immigrants' sample: Province						**Immigrants' sample: Sending area**					
Model	Var.		Δ% Var.	ICC (%)	LRT		Var.		Δ% Var.	ICC (%)	LRT	
IP + Age + Gender	0.141	(0.044)		4.11	111.36	***	-	-		-	-	
IS + Age + Gender	-			-	-		0.947	(0.255)		22.35	502.84	***
IPS + Age + Gender	0.086	(0.032)		2.00	44.70	***	0.909	(0.246)		21.21	436.19	***
IPS + X (*baseline*)	0.080	(0.031)		1.90	39.94	***	0.844	(0.231)		20.03	373.33	***
IPS + X + Youth UR	0.077	(0.030)	-3.38	1.84	35.14	***	0.842	(0.230)	-0.25	20.00	371.68	***
IPS + X + Owner-occupied HH	0.080	(0.031)	0.10	1.90	39.95	***	0.844	(0.231)	0.04	20.04	373.30	***
IPS + X + Cohabiting Couples	0.062	(0.027)	-22.20	1.49	22.24	***	0.843	(0.231)	-0.16	20.09	371.44	***
IPS + X + Age at Marriage (Prov)	0.069	(0.027)	-13.91	1.64	39.03	***	0.848	(0.232)	0.47	20.16	376.29	***
IPS + X + Z	0.032	(0.019)	-60.07	0.77	8.97	***	0.836	(0.229)	-0.96	20.11	370.61	***
IPS + X + Age at marriage (Send)	0.080	(0.031)	-0.66	1.96	39.55	***	0.698	(0.194)	-17.32	17.16	344.93	***
IPS + X + Z + W	0.032	(0.018)	-60.07	0.80	8.90	***	0.690	(0.192)	-18.25	17.20	342.36	***
	Natives' sample											
	Var.		Δ% Var.	ICC (%)	LRT							
IP + Age + Gender	0.091	(0. 019)		2.70	5270.10	***						
IP + X (*baseline*)	0.078	(0 .016)		2.33	4425.34	***						
IP + X + Youth UR	0.078	(0 .016)	-0.80	2.31	4350.28	***						
IP + X + Owner-occupied HH	0.071	(0.015)	-9.03	2.12	4103.76	***						
IP + X + Cohabiting Couples	0.059	(0 .012)	-25.16	1.75	3208.10	***						
IP + X + Z	0.055	(0 .011)	-29.62	1.65	3142.82	***						

Note: I: Individual; P: Province; S: Sending Area; X, Z, W: Individual- Provincial- and Sending Area level variables, respectively.
IP: 2-level model with individuals nested within provinces; IS: 2-level model with individuals nested within sending areas; IPS: cross-classified multilevel model. The columns "Immigrants: Province" and "Immigrants: Sending area" refer, respectively, to the provincial and sending area random effects of the cross-classified multilevel models applied to the immigrant sample. The column, "Spain", refers to the provincial random effects of two-level models applied to the Spanish sample. Δ% Var.: percent variation of the random effect variance with respect to the baseline model; ICC: Intraclass Correlation Coefficient; LRT: Likelihood Ratio Test, the test statistics is calculated as -2 multiplied by the difference between the loglikelihood of the model without the random effect whose variance is being tested and the loglikelihood of the model including the random effect.
p-value: ***<0.01;** <0.05;*<0.10

Table 3.6 reports the random effects estimates of the full multilevel model for immigrants, where the interaction between immigrants' country of origin and province of residence effects is included as an additional random effect in the model. In this case, there is no particular interest in the fixed effects estimates from this model, which are, as expected, very similar to those of the model without the interaction random effect. Instead, the interest lies in estimating the variance of this random effect. In the model (not reported for brevity) that excludes the size of the immigrant community in the province, the estimated variance for this random effect is rather small. A similar result is obtained introducing the size variable, which is not statistically significant, as expected. Therefore, these results do not support the "community effect" hypothesis. For second generation immigrants in Spain, cultural background of the country of origin plays a key role in shaping living arrangements independently on the specific characteristics of the local immigrants' communities. This may be due to the fact that immigration in Spain in 2001 was still a recent phenomenon and thus the different immigrant communities did not have sufficient time to develop specific cultural traits in the different provinces where they installed.

Table 3.6: Regression coefficient estimates from the multilevel logistic models with community effect (standard errors in parenthesis)

Fixed effects estimates				
Individual-level variables:				
Male	-1.547***	(0.15)		
Age	0.217***	(0.01)		
Still in education	-0.966***	(0.12)		
Male * Still in education	1.414***	(0.19)		
Employed	-0.252**	(0.11)		
Male * Employed	1.010***	(0.16)		
Educational level achieved:				
Primary or less	-0.031	(0.14)		
Male * Primary or less	0.042	(0.18)		
Higher education	-0.909***	(0.10)		
Male * Higher education	0.282*	(0.15)		
Provincial-level variables:				
Youth UR	-0.0003	(0.01)		
Owner-occupied HH	0.028**	(0.01)		
Cohabiting couples	0.090***	(0.03)		
Mean age at marriage	-0.274***	(0.07)		
Sending area-level variables:				
Mean age at marriage	-0.167**	(0.07)		
Community-level variables:				
Prop. Immigrants	-0.007	(0.01)		
Log Likelihood	-3607.8675			
N	6,761			
Random effects estimates				
	Var.		ICC (%)	LRT
Province	0.083	(0.044)	2.02	
Sending area	0.720	(0.203)	17.48	
Community	0.027	(0.020)	0.66	7.57***

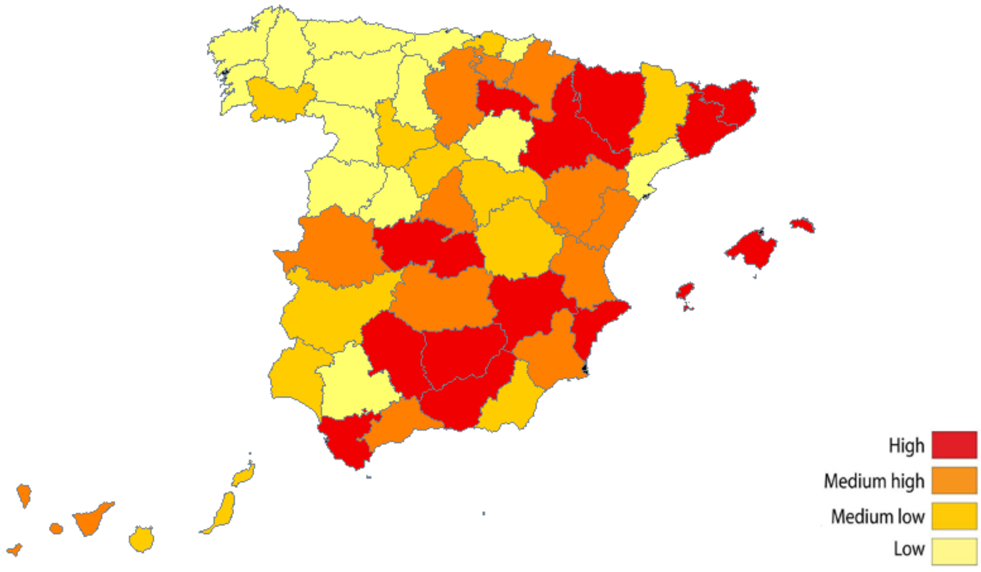
p-value: ***<0.01;** <0.05;*<0.10

3.7.3 Mapping provincial and sending area effects

To better highlight interesting aspects of the sources of heterogeneity under study, Empirical Bayes predictions (Rabe-Hesketh et al., 2008) of provincial and sending area errors are calculated. For the immigrants' sample a prediction of the error terms for each province (u_p) and for each sending area (v_s) is obtained. Similarly, for the Spanish sample predictions of each provincial error are obtained. Groups with positive (negative) predictions tend to have below (above) the mean proportions of young adults co-residing with parents. The higher the predicted error, the stronger is the deviation from the mean. Provinces and sending areas are classified in four groups, according to the quartiles of the predicted error term distribution. For example, areas with predicted error below the first quartile fall in the first group, which are labelled as “low”. These areas, identified by the light yellow colour in Figures 3.1 to 3.3, are those where young adults show the lowest rates of independent living. At the other extreme, areas above the third quartile are labelled as “high”, because they are characterised by the highest proportions of young adults living outside the parental home.

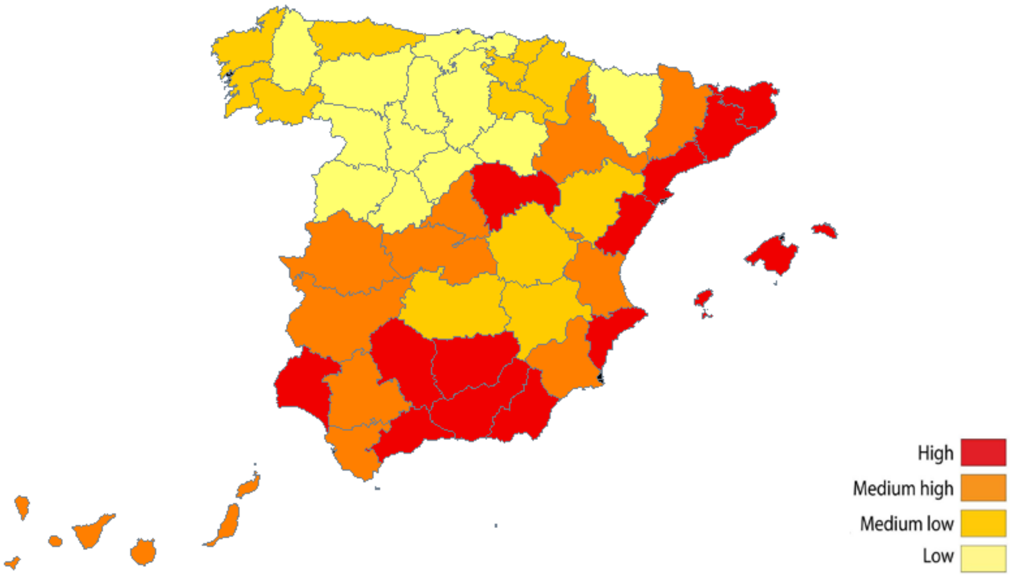
Figures 3.1 and 3.2 display provincial error predictions for the immigrants' and natives' samples, respectively. In both cases the model with individual-level covariates (i.e., Model IPS+X) is used only to show the “gross” provincial heterogeneity. Comparing the two figures it emerges that, overall, the provincial effects are similar in the two samples. For example, the highest propensities to live outside the parental home are found, both for immigrants and natives, in the autonomous communities of Andalucía (provinces of Granada, Jaén and Córdoba), Cataluña (provinces of Girona and Barcelona) and of the Balearic Islands. In the same way, the highest prevalence of young adults-parents co-residence is found, both for immigrants and natives, in the Autonomous Community of Castilla y León (provinces of León, Zamora, Salamanca, Palencia and Ávila). However, there are also provinces that show different patterns for the two samples. An example is represented by the province of Huesca, which falls in the “high” category for immigrants and in the “low” category for natives, while the opposite is found in the province of Tarragona.

Figure 3.1: Empirical Bayes predictions of province effects, immigrants' sample



Note: Predictions are obtained from the cross-classified model estimated on the immigrants sample, using individual-level variables only (i.e., Model IPS+X).

Figure 3.2: Empirical Bayes predictions of province effects, natives' sample

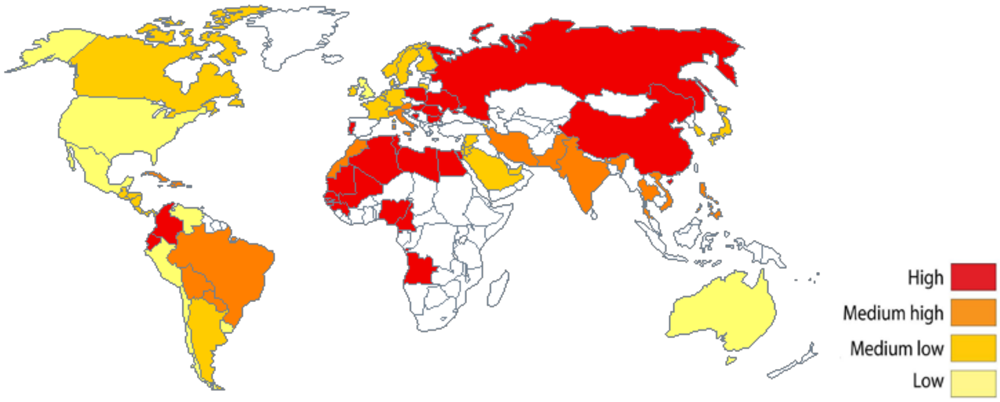


Note: Predictions are obtained from the 2-level model estimated on the natives' sample, using individual-level variables only (i.e., Model IP+X)

From Table 3.5 it emerges that the provincial effect remains more important for natives than for immigrants also after controlling for individual covariates (the ICC at provincial level is equal to 2.33 and 1.90 per cent, respectively). This difference can be quantified by computing predicted probabilities of living independently for a typical individual residing in two “extreme” provinces. For example, a 30 year old employed woman, residing in the province of Barcelona (the province with the highest error) and who achieved secondary education, has a predicted probability of living independently equal to 70 per cent if she is native and 82 per cent if she has immigrant origin. If the same woman resides in the Northern province of Léon (the province with the lowest error), the predicted probabilities fall to 47 and 70 per cent, respectively. These results show that, despite the relative low provincial variance, the heterogeneity in the living arrangements across the Spanish provinces is not negligible, especially for native young adults.

Figure 3.3 displays the predictions of sending area errors obtained from the cross-classified model, estimated on the immigrants’ sample using individual-level variables only (i.e., Model IPS+X). This figure clearly suggests that, despite the high heterogeneity in second generation immigrants’ living arrangements shown by the ICC calculations reported in Table 3.5, sending areas can be geographically clustered. For example, immigrants who were born in the countries of the Maghreb area show similar behaviours in terms of living arrangements: high probabilities of living independently due to low mean age at marriage and early transition to adulthood with respect to other origins.

Figure 3.3: Empirical Bayes predictions of sending area effects



Note: Predictions are obtained from the cross-classified model estimated on the immigrants sample, using individual-level variables only (i.e., Model IPS+X)

To gain insight into the magnitude of the heterogeneity in living arrangements, the predicted probability of living outside the parental home has been calculated for each immigrant origin and for a typical individual (woman aged 25, employed, and with secondary education achieved). These probabilities vary from 92 per cent if her country of origin is Ecuador (highest country error) to 32 per cent if it is Australia (lowest country error), showing a strong degree of heterogeneity. Moreover, second generation immigrants from Venezuela, Australia, Switzerland, Mexico, Uruguay, and Peru show the most similar predicted probabilities with respect to Spanish peers. This conclusion is confirmed by the results of a cross-classified model (not showed here but available upon request) where Spain is included as an additional “sending area” together with the 35 immigrant origins described in Table 3.A in the appendix. Since the original Spanish sample is much bigger than all immigrant groups samples, a random sub-sample of native young adults was used. The model contribution to mapping can be appreciated comparing the rank of Empirical Bayes predictions in the model with individual-level covariates to those obtained in a null model. This rank is reported for province and sending area effects in the Appendix in Figures 3.C and 3.D, respectively.

3.8 Concluding Remarks

The literature on the transition to adulthood for second generation immigrants in Southern European countries which recently experienced a tremendous change in immigration dynamics is not well developed yet. This chapter contributes to this field by studying young adults-parents co-residence among second generation immigrants in Spain in comparison to natives.

Considering the need, recently stressed by many authors, for studies that fully describe the heterogeneity characterising immigrants' behaviours, all immigrant groups present in Spain at the time of the 2001 census and represented in the 5 per cent sample extracted from the IPUMS-I database are included in the sample.

The heterogeneity of the country of origin is studied together with the effects due to the place of residence, namely the province, which in past studies has been found to be important for the transition to adulthood in Spain. This is made possible by using cross-classified multilevel models which offered the opportunity to disentangle the two sources of variability, and to introduce variables measured both at the country of origin and province of residence level.

Second generation immigrants are shown to be extremely heterogeneous with respect to their country of origin, even though a geographical clustering is evident. Cultural and structural explanations of different living arrangements are linked and therefore difficult to separate. Young adults who grew up in families of immigrant origin are exposed to two different normative sets: the culture of the country of origin and the environment in the country of residence. This chapter shows that for second generation immigrants in Spain, the country of origin contributes more to explaining the existing variability in independent living than the province of residence. However, even if the heterogeneity due to provincial effect is lower, it is not negligible. Moreover, the effect due to the province of residence is slightly higher for Spanish natives than for immigrants. Results do not provide evidence in support of the “community effect” hypothesis probably because, being Spain a country of recent immigration, immigrant communities were yet not established in 2001. In the future, with more recent data, it will be interesting to test again the “community effect” hypothesis to assess if the assimilation of immigrants had developed differently across Spain.

Undoubtedly there are other structural factors –with respect to the ones considered here– which are expected to be responsible for the living arrangement decision, and also for the age at living the parental home. For example, one would expect that availability of certain occupations matters. Smilingly, house prices would be a better predictor of difficulty to enter the housing market than the indicator used in this thesis. To study the existence of different housing preferences between natives and second generation immigrants is intriguing, though not possible with census data. Also, housing preferences might incorporate cultural preferences: immigrants from Northern Africa might choose to live in extended families because they maintain the preference for this form of living arrangement, typical of their country of origin. Further research should therefore be directed to consider more possible choices of living arrangements as the object of investigation and not just the probability to live with parents versus living independently from parents.

An interesting result of this chapter is the strong negative association found between the mean age at marriage measured in the country if origin and the probability to reside outside the parental home for second generation immigrants. This corroborates the idea that the cultural heritage of second generation immigrants still plays a role in influencing their demographic behaviours.

The findings showed in this chapter, however, are based on the 2001 Spanish census and for this reason they should not be generalized to draw ultimate conclusions about co-residence patterns between young people of immigrant origin and their parents. Spain, in fact, is a country of recent immigration which has been experiencing unprecedented changes in immigration flows during the last decade. Moreover, past studies for other countries found that third generation immigrants tend to be less influenced by the culture of their country of origin. Future work using the next Spanish census could give a more precise view of the phenomenon under study.

From a methodological point of view, this chapter shows that ignoring the cross-classified structure of the data leads to overestimating the provincial variability. The adoption of cross-classified multilevel models has great potential in the study of demographic behaviours of immigrants: if a comprehensive perspective of migration movements is adopted and simultaneous sources of heterogeneity at sending and receiving contexts (i.e., countries, regions, provinces) are to be considered, cross-classified multilevel modelling proves to be a useful tool of analysis.

3.9 References

Adsera, A. and Chiswick, B. R. (2007). Are there gender and country of origin differences in immigrant labor market outcomes across European destinations? *Journal of Population Economics*, 20(3), 495–526.

Algan, Y., Dustmann, C., Glitz, A. and Manning, A. (2010). The economic situation of first and second-generation immigrants in France, Germany and the United Kingdom. *Economic Journal*, 120(542): F4–F30.

Beck, A. N. and Tienda, M. (2010). Better Fortunes? Living Arrangements and Wellbeing of Migrant Youth in Six OECD Countries. Paper presented at the 2010 Annual Meeting of The Population Association of America, Dallas, Texas.

Billari, F. C., Castiglioni, M., Castro Martín, T., Michielin, F. and Ongaro, F. (2002). Household and union formation in a Mediterranean fashion: Italy and Spain. In Klijzing, E. and Corijn, M. (Eds.) *Fertility and Partnership in Europe: Findings and Lessons from Comparative Research* (Vol. II): 17–41. New York: United Nations.

Billari, F. C. (2004). Becoming an Adult in Europe: A Macro(/Micro)-Demographic Perspective. *Demographic Research* Special Collection, 3: 15–44.

Billari, F. C., Rosina, A., Ranaldi, R., Romano, M. C. (2008). Young Adults Living Apart and Together (LAT) with Parents: A Three level Analysis of the Italian Case. *Regional Studies*, 42(5): 625–639.

Boman, A. (2011). The Mobility of Immigrants and Natives: Evidence from Internal Migration Following Job Displacement. *Regional Studies*, 45(3): 283–297.

Boyd, M. (2000). Ethnic variations in young adults living at home. *Canadian Studies in Population*, 27(1): 135–158.

Boyd, M. and Park, S. (2010). Generational change and cultural preferences: 1.5 and second generation adults living with parents, Paper presented at the 2010 Annual Meeting of The Population Association of America, Dallas, Texas.

Burr, J. A. and Mutchler, J. E. (1993). Ethnic living arrangements: cultural convergence or cultural manifestation? *Social Forces*, 72(1): 169–179.

Clark, R. L., Glick, J. E. and Bures, R. M. (2009). Immigrant families over the life course: Research directions and needs. *Journal of Family Issues*, 30(6): 852–872.

Corijn, M. and Klijzing, E. (2001). *Transitions to adulthood in Europe*. Dordrecht: Kluwer Academic Publishers.

Cortina Trilla, C., Esteve, A. and Domingo, A. (2008). Marriage patterns of the foreign-born population in a new Country of immigration: The case of Spain. *International Migration Review*, 42(4): 877–902.

Crul, M. and Vermeulen, H. (2003). The second generation in Europe. *International Migration Review*, 37(4): 965–986.

De Valk, H. A. G. (2006). Pathways into adulthood. A comparative study on family life transitions among migrant and Dutch youths. Utrecht, The Netherlands, Dissertation.

De Valk, H. A. G. and Billari, F. C. (2007). Living arrangements of migrant and Dutch young adults: the family influence disentangled. *Population Studies*, 61(2): 201–17.

De Valk, H. A. G. and Liefbroer, A. C. (2007). Timing preferences for women's family-life transitions: Intergenerational transmission among migrants and Dutch. *Journal of Marriage and the Family*, 69(1): 190–206.

De Valk, H. A. G. and Mencarini, L. (2009). Growing up in an immigrant family: the position of children of immigrants in Italy and the Netherlands, Paper presented at the XXVI IUSSP conference, Marrakech.

Elzinga, C. H., and Liefbroer, A. C. (2007). De-standardization of family-life trajectories of young adults: A cross-national comparison using sequence analysis. *European Journal of Population*, 23(3): 225–250.

Farley, R. and Alba, R. (2002). The new second generation in the United States. *International Migration Review*, 36(3): 669–701.

Fuligni, A. J., Tseng, V. and Lam, M. (1999). Attitudes toward family obligations among American adolescents with Asian, Latin American, and European backgrounds. *Child Development*, 70(4): 1030–1044.

Fussell, E. and Furstenberg, F. F. (2005). The transition to adulthood during the twentieth century: Race, nativity, and gender. In Settersten, R. A., Furstenberg, F. F. and Rumbaut, R. G. (Eds.) *On the frontier of adulthood: Theory, research and public policy*, pp. 29–75. Chicago: University of Chicago Press.

Giuliano, P. (2007). Living arrangements in Western Europe: Does cultural origin matter? *Journal of the European Economic Association*, 5(5): 927–952.

Glick, J. E. and Van Hook, J. (2002). Parents' coresidence with adult children: can immigration explain racial and ethnic variation? *Journal of Marriage and Family*, 64(1): 240–253.

Giguère, B., Lalonde, E. and Lou, E. (2010). Living at the crossroads of cultural worlds: The experience of normative conflicts by second generation immigrant youth. *Social and Personality Psychology Compass*, 4(1): 14–29.

Goldscheider, C. and Goldscheider, F. K. (1988). Ethnicity, religiosity and leaving home: the structural and cultural bases of traditional family values. *Sociological Forum*, 3(4): 525–547.

Goldscheider, F. K. and Goldscheider, C. (1989). *Ethnicity and the new family economy: Living arrangements and intergenerational financial flows*. Boulder: Westview Press.

Holdsworth, C. (1998). Leaving home in Spain: A regional analysis. *International Journal of Population Geography*, 4(4): 341–360.

Holdsworth, C. and Irazoqui Solda, M. (2002). First housing moves in Spain: An analysis of leaving home and first housing acquisition. *European Journal of Population*, 18(1): 1–19.

Holdsworth, C., Voast, D., and Tranmer, M. (2002). Leaving Home in Spain: When, where and why? *Regional Studies*, 36(9): 989–1004.

Huschek, D., Liefbroer, A. C. and De Valk, H. A. G. (2010). Timing of first union among second-generation Turks in Europe: The role of parents, peers and institutional context. *Demographic Research*, 22(16): 473–504.

Instituto Nacional de Estadistica (INE) (2001). Population and Housing Census. INE, Madrid.

Kalmijn, M., de Graaf, P.M. and Janssen, J. P. G. (2005). Intermarriage and the risk of divorce in the Netherlands: The effects of differences in religion and in nationality, 1974/94. *Population Studies*, 59(1): 71–85.

Kalmijn, M. and Van Tubergen, F. (2010). A comparative perspective on intermarriage: Explaining differences in marriage choices among national origin groups in the United States. *Demography*, 47(2): 459–479.

Kritz, M., Gurak, D. T. and Chen, L. W. (2000). Elderly immigrants: Their composition and living arrangements. *Journal of Sociology and Social welfare*, 27(1): 85–114.

Levels, M. and Dronkers, J. (2008). Educational performance of native and immigrant children from various countries of origin. *Ethnic and Racial Studies*, 31(8): 1404–1425.

Levels, M., Dronkers, J. and Kraaykamp, G. (2008). Immigrant children's educational achievement in Western countries: Origin, destination, and community effects on mathematical performance. *American Sociological Review*, 73(5): 835–853.

Liefbroer, A. C. and Toulemon, L. (2010). Demographic perspectives on the transition to adulthood: An introduction. *Advances in Life Course Research*, 15(2–3): 53–58.

Mitchell, B. A. (2004). Making the move: Cultural and parental influences on Canadian young adults' homeleaving decisions. *Journal of comparative Family Studies*, 35(3): 423–441.

Mitchell, B. A., Wister, A. V. and Gee, E. M. (2004). The ethnic and family nexus of homeleaving and returning among Canadian young adults. *Canadian Journal of Sociology*, 29(4): 543–575.

Mulder, C. H. (2006). Home-ownership and family formation. *Journal of Housing and the Built Environment*, 21(3): 281–298.

Minnesota Population Center (2009). *Integrated Public Use Microdata Series, International*, Version 5.0, Minneapolis: University of Minnesota, 2009.

Nauck, B. (2001) Social capital, intergenerational transmission and intercultural contact in immigrant families. *Journal of Comparative Family Research*, 32(2): 159–173.

Rabe-Hesketh, S. and Skrondal, A. (2008). *Multilevel and longitudinal modeling using Stata*, 2nd ed. College Station, TX: Stata Press.

Rasbash, J. and Goldstein, H. (1994). Efficient analysis of mixed hierarchical and cross-classified random structures using a multilevel model. *Journal of Educational and Behavioral Statistics*, 19(4): 337–350.

Reher, D.S. (1991). Marriage patterns in Spain, 1887-1930. *Journal of Family History*, 16(1), 7–30.

Reher, D.S. and Silvestre, J. (2009). Internal Migration Patterns of Foreign-Born Immigrants in a Country of Recent Mass Immigration: Evidence from New Micro Data for Spain. *International Migration Review*, 43(4): 815–849.

Roig Vila, M. and Castro Martín, T. (2007). Childbearing patterns of foreign women in a new immigration country: The case of Spain. *Population*, 62(3): 351–379.

Rumbaut, R. G. and Komaie, G. (2010). Immigration and adult transitions. *The Future of Children*, 20(1): 43–66.

Snijders, T. and Bosker, R. (1999) *An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage.

Srholec, M. (2010). A Multilevel Approach to Geography of Innovation. *Regional Studies*, 44(9): 1207–1220.

StataCorp (2009). *Stata 11 Longitudinal Data/ Panel Data Reference Manual,* College Station, TX: Stata Press.

Van den Noortgate, W., De Boeck, P. and Meulders, M. (2003). Cross-classification multilevel logistic models in psychometrics. *Journal of Educational and Behavioral Statistics*, 28(4): 369–386.

Van Tubergen, F., Maas, I. and Flap, H. (2004). The economic incorporation of immigrants in 18 Western societies: Origin, destination, and community effects. *American Sociological Review*, 69(5): 704-727.

Van Tubergen, F. and Kalmijn, M. (2005). Destination-language proficiency in cross-national perspective: a study of immigrant groups in nine Western countries. *American Journal of Sociology*, 110(5): 1412–1457.

Vitali, A. (2010). Regional differences in young Spaniards' living arrangement decisions: A multilevel approach. *Advances in Life Course Research*, 15(2-3): 97–108.

Wakil, S. P., Siddique, C. M. and Wakil, F. A. (1981). Between two cultures: a study of socialisation of children of immigrants. *Journal of Marriage and the Family*, 43(4): 929–940.

3.10 Appendix

Table 3.A: List of sending countries and grouping of countries into sending areas

Sending areas	Subtotal	Total	Sending areas	Subtotal	Total
Argentina		288	India	18	
Australia		39	Iran	7	
Belgium		205	Pakistan	11	
Brazil		68	South-Central Asia		36
Canada		45	Philippines	20	
Chile		52	Thailand	2	
China		62	Vietnam	5	
Colombia		128	South-Eastern Asia		27
Dominican Republic		63	Armenia	2	
Ecuador		97	Israel	2	
France		1,528	Jordan	3	
Germany		1,141	Lebanon	4	
Mexico		65	Saudi Arabia	2	
Morocco		443	Syria	6	
Peru		68	United Arab Emirates	3	
Portugal		128	Western Asia		22
Switzerland		671	Denmark	14	
United Kingdom		326	Finland	2	
United States		70	Ireland	5	
Uruguay		79	Lithuania	5	
Venezuela		474	Norway	4	
Angola	6		Sweden	26	
Cameroon	4		Northern Europe, n.s.	133	
Equatorial Guinea	39		Northern Europe (others)		189
Middle Africa		49	Andorra	25	
Algeria	19		Italy	38	
Egypt	2		Southern Europe (others)		63
Libya	2		Austria	4	
Tunisia	2		Luxembourg	11	
Northern Africa (others)		25	Western Europe (others)		15

Table 3.A (Continued)

	Cape Verde	5			Cuba	44	
	Gambia	10			Haiti	2	
	Guinea	14		Americas			46
	Guinea-Bissau	2			Costa Rica	4	
	Mali	5			El Salvador	6	
	Mauritania	4			Guatemala	6	
	Nigeria	5			Honduras	10	
	Senegal	8			Nicaragua	3	
Western Africa			53		Panama		18
	Japan	4		Central America (others)			47
	Korea, RO (South)	13			Bolivia		19
Eastern Asia (others)			17		Paraguay		5
				South America (others)			24
				Total			6,761

Note*:* The 21 countries of origin that form a separate group are listed first, followed by immigrant groups with small sizes that are aggregated geographically. “n.s.” means not specified.

Table 3.B: Regression coefficient estimates from the 2-level logistic models (standard errors in parenthesis)

	2-level model – Province			2-level model – Sending area		
Individual-level variables:						
Male	-1.41	(0.14)	***	-1.520	(0.15)	***
Age	0.183	(0.01)	***	0.214	(0.01)	***
Still in education	-1	(0.12)	***	-0.946	(0.12)	***
Male * Still in education	1.267	(0.17)	***	1.382	(0.18)	***
Employed	-0.22	(0.10)	**	-0.254	(0.10)	**
Male * Employed	0.881	(0.15)	***	0.996	(0.16)	***
Educational level achieved:						
Primary or less	0.273	(0.14)	**	-0.037	(0.14)	
Male * Primary or less	0.044	(0.17)		0.058	(0.18)	
Higher education	-0.91	(0.10)	***	-0.895	(0.10)	***
Male * Higher education	0.266	(0.15)	*	0.280	(0.15)	*
Provincial-level variables:						
Youth Unemployment Rate	0.0004	(0.01)		-0.001	(0.01)	
Owner-occupied Households	0.05	(0.01)	***	0.026	(0.01)	***
Cohabiting couples	0.118	(0.03)	***	0.099	(0.02)	***
Mean age at marriage	-0.27	(0.07)	***	-0.235	(0.05)	***
Sending area-level variable:						
Mean age at marriage	-0.07	(0.01)	***	-0.16	(0.06)	**
Log Likelihood	-3782.8358			-3616.1027		
N	6,761			562,648		

Note: p-value: *** < 0.01;** < 0.05;* < 0.10

Figure 3.C: Rank of Empirical Bayes predictions of province effects

Sending areas	Null Model	Sending areas	Model IPS+X
United States	-0.95	Australia	-1.34
Mexico	-0.92	Venezuela	-0.87
Peru	-0.83	Switzerland	-0.80
Western Asia	-0.78	Peru	-0.79
Argentina	-0.76	Mexico	-0.77
Eastern Asia	-0.73	Uruguay	-0.76
Chile	-0.58	United Kingdom	-0.72
Australia	-0.54	United States	-0.71
Venezuela	-0.48	Chile	-0.71
South America	-0.34	Western Asia	-0.71
South-Central Asia	-0.33	Northern Europe	-0.68
Northern Europe	-0.30	France	-0.65
Southern Europe	-0.29	Eastern Asia	-0.64
Switzerland	-0.28	Germany	-0.60
United Kingdom	-0.27	Argentina	-0.57
South-Eastern Asia	-0.15	Central America	-0.38
Uruguay	-0.13	Belgium	-0.33
Dominican Republic	-0.09	Canada	-0.19
China	-0.06	Southern Europe	-0.13
Central America	-0.04	South-Eastern Asia	-0.09
Germany	0.04	Brazil	-0.02
Americas	0.08	South America	0.00
Brazil	0.10	South-Central Asia	0.09
Canada	0.11	Americas	0.20
Belgium	0.24	Western Europe	0.26
France	0.30	Dominican Republic	0.47
Colombia	0.33	Morocco	0.71
Morocco	0.35	China	0.72
Eastern Europe	0.44	Colombia	0.92
Western Europe	0.63	Eastern Europe	1.15
Northern Africa	0.70	Northern Africa	1.21
Middle Africa	1.22	Middle Africa	1.36
Western Africa	1.28	Portugal	1.72
Ecuador	1.43	Western Africa	1.80
Portugal	1.57	Ecuador	1.84

Note: Predictions are obtained from the cross-classified model estimated on the immigrants sample (i.e., Model IPS+X)

Figure 3.D: Rank of Empirical Bayes predictions of sending area effects

Provinces	Null Model	Provinces	Model IPS+X
Salamanca	-0.41	Pontevedra	-0.41
Pontevedra	-0.33	León	-0.31
Lugo	-0.27	Sevilla	-0.30
León	-0.22	Asturias	-0.30
Ourense	-0.21	Cantabria	-0.28
Zamora	-0.19	Salamanca	-0.26
Ávila	-0.19	Lugo	-0.24
Valladolid	-0.18	Palencia	-0.22
Coruña (A)	-0.15	Soria	-0.21
Soria	-0.14	Zamora	-0.19
Asturias	-0.11	Ávila	-0.19
Palencia	-0.10	Coruña (A)	-0.16
Sevilla	-0.10	Tarragona	-0.16
Guadalajara	-0.09	Valladolid	-0.14
Cantabria	-0.09	Guipúzcoa	-0.12
Badajoz	-0.07	Ourense	-0.11
Tarragona	-0.06	Palmas (Las)	-0.09
Cáceres	-0.05	Badajoz	-0.08
Guipúzcoa	-0.05	Almería	-0.08
Palmas (Las)	-0.02	Vizcaya	-0.07
Vizcaya	-0.02	Guadalajara	-0.07
Segovia	-0.02	Cuenca	-0.02
Burgos	0.00	Lleida	-0.01
Almería	0.00	Segovia	0.00
Cuenca	0.00	Huelva	0.00
Teruel	0.00	Teruel	0.01
Álava	0.01	Álava	0.01
Málaga	0.01	Burgos	0.03
Granada	0.02	Navarra	0.04
Madrid	0.03	Madrid	0.04
Ciudad Real	0.04	Valencia/València	0.05
Córdoba	0.05	Murcia	0.05
Santa Cruz de T.	0.07	Cáceres	0.07
Navarra	0.07	Ciudad Real	0.08
Lleida	0.08	Málaga	0.08
Balears (Illes)	0.08	Santa Cruz de T.	0.10
Jaén	0.10	Castellón/Castelló	0.13
Girona	0.10	Córdoba	0.15
Rioja (La)	0.12	Huesca	0.17
Murcia	0.12	Rioja (La)	0.18
Huelva	0.14	Cádiz	0.19
Cádiz	0.15	Jaén	0.22
Valencia/València	0.16	Toledo	0.24
Huesca	0.17	Girona	0.24
Albacete	0.18	Albacete	0.26
Castellón/Castelló	0.20	Granada	0.26
Alicante/Alacant	0.20	Zaragoza	0.30
Toledo	0.29	Balears (Illes)	0.31
Zaragoza	0.29	Barcelona	0.39
Barcelona	0.38	Alicante/Alacant	0.41

Note: Predictions are obtained from the cross-classified model estimated on the immigrants sample (i.e., Model IPS+X)

4. A Spatial Panel Analysis of Italian Regional Fertility

4.1 Introduction

In this chapter I face the challenge of re-introducing space into fertility analysis. I make use of very recent spatial econometric techniques to unravel the spatial and temporal mechanisms behind recent fertility change in Italian provinces.

The European Fertility Project (Coale et al., 1986; Watkins, 1987; Bongaarts et al., 1996) studied the fertility decline occurred in the 18th and 19th century in several European provinces. Results showed that fertility decline spread beyond the logic predicted by socio-economic differentials across provinces. Rather, provinces which shared same language, ethnicity and religion –that is, same cultural characteristics (Anderson, 1986; Knodel et al., 1979)– experienced similar fertility transitions. These considerations directly link to the diffusionist perspective of fertility decline. Fertility decline results from the diffusion of new attitudes and ideas towards the value and cost of children and new behaviours due to acquired knowledge of and information regarding birth control techniques, spreading among people and places (Cleland et al., 1987; Casterline, 2001). Attitudes and behaviours are “new” because they were absent or rare in the past. For this reason, the spread of new ideas and behaviours are respectively referred to as “innovation diffusion” and “behavioural innovation”, as opposed to “adjustment processes” to changed economic circumstances (Carlson, 1966). The spread of new ideas and knowledge is dynamic in its essence and acts as a social influence or a social learning process at the individual level through kinship, communication networks and the mass media (Retherford et al., 1983; Montgomery et al., 1996; Kohler, 2000; Carter, 2001; Kohler, 2001). Over time, this process results in a diffusion mechanism across space, leading to collective outcomes at the population level. At the aggregate level, provinces sharing a homogeneous cultural context were found to have similar transition onset and pace. In Bongaarts et al.’s words (1996), “diffusion refers to the process by which innovation spreads among regions, social groups, or individuals, often apparently independently of social and economic circumstances”, while Cleland et al. (1987) explain the fertility transition with ideational rather than structural factors.

The diffusionist perspective on fertility has been mainly applied to study historical settings (Tolnay, 1995; Bocquet-Appel et al., 1998;

Murphy, 2010). Based on the principle that the same diffusive process applies to contemporary settings where fertility is in transition, some scholars have tested the validity of this approach using spatial statistic and econometric techniques on contemporary fertility decline in developing regions of the world (Watkins, 1987; Weeks et al., 2000; Bocquet-Appel et al., 2002). However, diffusion mechanisms are in place whenever there is an innovation. Whenever there is a fertility transition, one could in principle find spatial patterns. For instance, Lesthaeghe et al. (2002) find a common spatial pattern between the First and the Second Demographic Transitions in Belgium, Switzerland and France. Of course in contemporary developed societies, contraception techniques are widely known and used and thus the concept of diffusion is necessarily confined to ideational change and the spread of new ideas more generally.

It is easy to recognize spatial patterns in demographic behaviours when the relative demographic indicator is plotted on a geographical map, for example a choropleth map like the one presented in Figure 4.1 showing the Period Total Fertility Rate (TFR) of Italian regions for 1952 and 2009. The next step is to understand how to incorporate a diffusion process into statistical models. Recent advances in spatial statistics and econometrics allow a set of different tools for incorporating spatial effects into regression analyses. It is up to the researcher to choose the tool that best suits the theoretical framework chosen. However, there is no such thing as a "diffusion theory of fertility" and for this reason it is not straightforward to set up a model of diffusion.

In this chapter I shall use the diffusionist perspective to fertility transition for describing the current temporal and spatial trends in Italian provincial fertility in association with a series of indicators of marital behaviours, female occupation, contribution of foreign fertility and economic development. Following the spatial econometric literature, I explicitly take into account spatial dependence among Italian provinces. By including a spatially lagged dependent variable on the right-hand side of the regression equation, spatial lag models allow the estimation of a spatial autocorrelation coefficient which captures the correlation between the dependent variable measured in a given area and the same variable measured in neighbouring areas. The evolution of sub-regional fertility over time is therefore modelled taking into account fertility levels recorded in neighbouring places. This ability to explicitly take into account the influence that fertility in neighbouring areas has on fertility in a given area makes spatial lag models the appropriate tool for studying spatial diffusion and hence to test the diffusionist perspective to fertility transition.

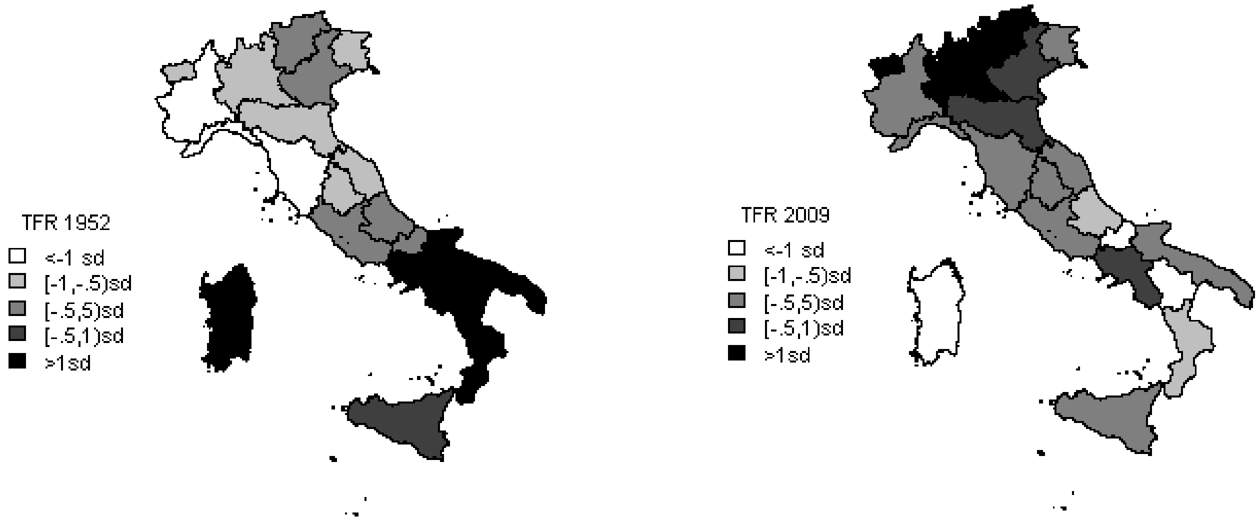
4.2 How many Italies are there? A look at sub-national fertility differentials

Scholars have often studied fertility in Italy in cross-national comparisons with other low-fertility settings, namely Mediterranean and sometimes Central and Easter European countries. What is frequently overlooked in these studies is that Italy presents great intra-country variations (Rallu, 1983; Kertzer et al., 2009). The story of Italian regional heterogeneity dates back in history and is not confined to a North-South divide. Regional heterogeneity embraces many aspects of the society as a whole. There are, in fact, different local economies, different labour markets and housing conditions, different levels of poverty, and different cultures. These differences in turns are reflected in different demographic behaviours among which, fertility behaviours. For these reasons, exploring regional differentials is likely to contribute to the study of determinants of lowest-low fertility. Livi-Bacci (1977), Watkins (1990) and Franklin (2003) showed that regional fertility differentials existed in Italy even before the (First) Demographic Transition, which in Italy started at the end of the 19^{th} century, and persisted after the political unification in 1861. Historically, fertility was considerably higher in the South of Italy than in the Centre and North. During the economic recovery following the Second

World War Italy witnessed a baby boom like all other industrialized countries. In this period, fertility increased in most regions, but not in all of them. In Southern Italy, fertility was already high during the 1950s, and therefore it remained quite stable, while in most Northern and Central regions it experienced a steady increase (Terra Abrami et al., 1993). The fertility trend reversed during the mid-1960s. A fertility decline came to a halt in 1995, when a TFR of 1.19 was recorded. From 1995 onwards, fertility has been increasing slightly at the national level, and regional differences have emerged again: during the 2000s, fertility increased in most regions while in a few others it continued to decline. In very recent years there has been a reversal in that it is the North which now shows the highest regional fertility, something that used to characterize the South. Regional fertility also appears more heterogeneous than it was in the past because we no longer observe a clear divide between Northern and Southern regions. For instance, fertility levels in the Southern region of Campania are more similar to those observed in North-Eastern regions than to other Southern regions.

Figure 4.1 maps the period total fertility rate (TFR) in the twenty Italian regions for the two years marking the beginning and the end of the period for which the National Statistics Institute (Istat) provides available data at the regional level. The figure shows two main features of Italian regional fertility: first, there is sub-national variation (*spatial heterogeneity*) and second, there is spatial clustering (*spatial dependence*).

Figure 4.1: TFR in the 20 Italian regions in year 1952 and 2009



Note: The legend is to be read in terms of standard deviations (sd) from the mean: ">1 sd" indicates regions whose TFR is one sd above the mean; "[.5;1)" between .5 and 1 sd above the mean; [-.5;.5) .5 sd around the mean; [-1;-.5) between .5 and 1 sb below the mean; "<-1" 1 sd below the mean. Mean and standard deviations were 2.39 and 0.74, respectively in 1952 (a) and 1.37 and 0.13 in 2009 (b). Panel (a) considers Molise and Abruzzi as a unique region since Molise became an autonomous region only in 1964.
Source: Istat, "Tavole di Fecondità della popolazione italiana per regione di residenza" for the period 1952–2004, and Survey on Live Births after 2004.

If in the early 1950s there was a (although not perfect) core-periphery pattern with high levels of TFR observed in Southern regions and low values observed in Northern regions, in 2009 the picture is completely different. In 2009 all Southern regions show a TFR below the national average, Campania and Sicily being the only exceptions;

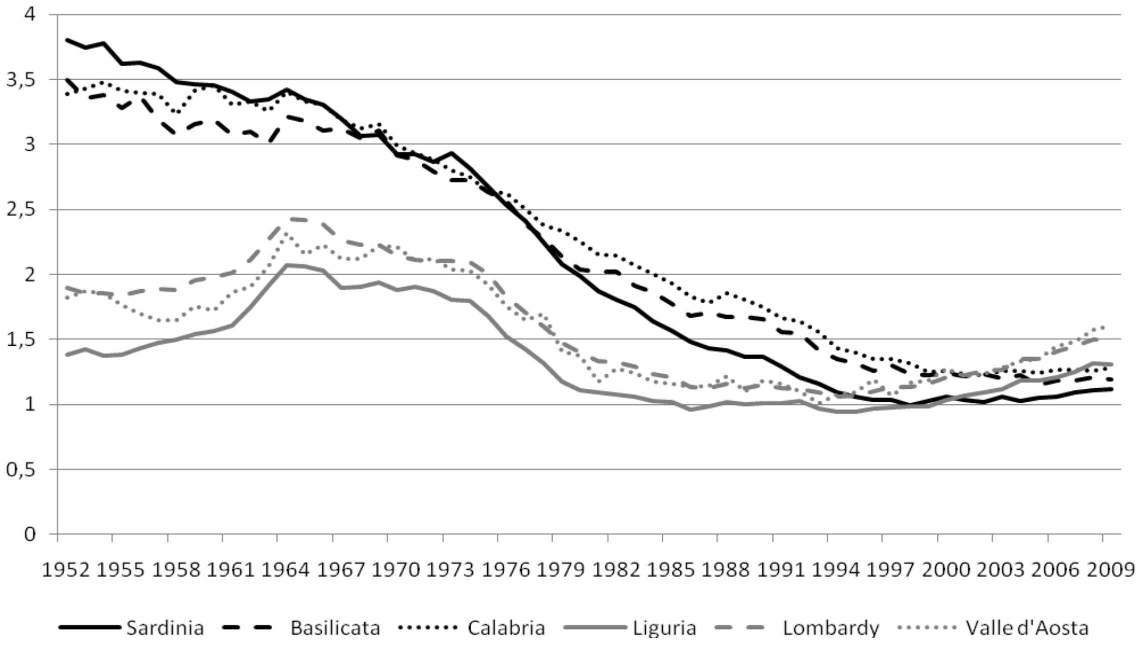
conversely, all Northern regions have a TFR above the national average, with the exception of Liguria. Of course regional differentials in fertility levels in 1952 were not the same as they are today. At the beginning of the period of analysis, in fact, Italian TFR was equal to 2.34 children per woman, with huge regional variations ranging from a maximum of 3.8 in Sardinia to a minimum of 1.39 in Liguria. In 2009, when the national TFR was 1.41, variations around the mean were very moderate, ranging from 1.12 in Sardinia and Molise to 1.61 in Valle d'Aosta.

Figure 4.2 shows the evolution of the TFR over the period 1952-2009 for three selected Southern regions, Sardinia, Basilicata, Calabria and three selected North-Western regions, Lombardy, Liguria and Valle d'Aosta. Liguria held the lowest regional TFR in Italy for almost the whole period, with a value as low as 1.39 already in 1952. A very low fertility level was observed also for the North-Western region of Piedmont where the TFR was 1.49 children per woman in 1952. In the same year, the TFR in Sardinia (South) was 3.8 children per woman – almost three times that of Liguria. The TFR was above 3 children per woman also in the Southern regions of Basilicata, Calabria, Apulia and Campania (3.49, 3.39, 3.38 and 3.18, respectively). Liguria (North-West) and Emilia Romagna (North-East) were the first two regions to cross the lowest-low fertility threshold of 1.3 in 1979 (with a TFR of 1.18 and 1.28, respectively), followed by

Friuli-Venezia-Giulia (North-East) in 1980 (1.25), and Piedmont (North-West), Tuscany (Centre) and Valle d'Aosta (North-West) in 1981 (1.27, 1.25 and 1.18). The same threshold was crossed more than 10 years later in Southern regions, starting in 1991 with Sardinia (1.29) followed in 1993 by Abruzzi (1.3), while Calabria (1.25) and Apulia (1.3) reached below replacement fertility in 1999 and 2003, respectively. Campania and Sicily always remained above the 1.3 threshold throughout the whole period. Also the North-Eastern region of Trentino-Alto-Adige remained always above the threshold.

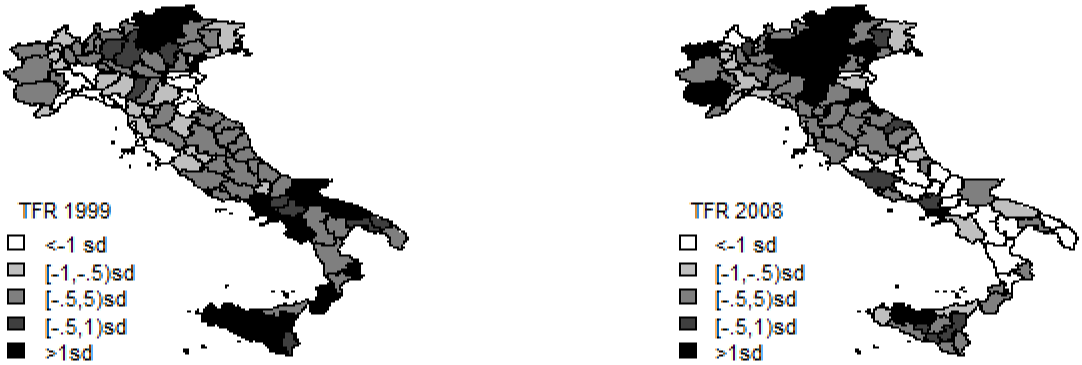
In the same way as in the early 1980s they were the forerunners of lowest-low fertility, in the 2000s the Northern regions were the forerunners of fertility recuperation. By 2008, in fact, all Northern and Central regions, with the exception of Trentino-Alto-Adige, had exited from lowest-low fertility, the forerunner regions being Veneto (North-East), Lombardy (North-East), Valle d'Aosta (North-East), Emilia-Romagna (North-West) and Umbria (Centre) in 2004 (with TFR equal to, respectively, 1.36, 1.35, 1.33, 1.32 and 1.31). All Southern regions instead registered lowest-low fertility levels still in 2009, the only exception being Apulia, with a TFR of 1.34 children per woman. Particularly noteworthy is the case of Sardinia, which, during the 1950s was the region with the highest fertility, above 3.5 children per woman, and then, during the 1970s and 1980s experienced the fastest reduction in fertility among Italian regions until the 2000s, when it became the region with the lowest fertility with 1.12 children per woman in 2009. Italian Regional data therefore suggest that the aggregate level hides great intra country variation and its use induces loss of information.

Figure 4.2: Total Fertility Rate in four selected Italian regions over the period 1952–2009: Sardinia, Basilicata, Calabria (South), Lombardy, Liguria and Valle d’Aosta (North-West)



Source: Istat, “Tavole di Fecondità della popolazione italiana per regione di residenza” for the period 1952–2004, and Survey on Live Births after 2004.

Figure 4.3: TFR in 99 Italian provinces in year 1999 and 2008



Note: The legend is to be read in terms of standard deviations (sd) from the mean: ">1 sd" indicates provinces whose TFR is one sd above the mean; "[.5;1)" between .5 and 1 sd above the mean; [-.5;.5) .5 sd around the mean; [-1;-.5) between .5 and 1 sb below the mean; "<-1" 1 sd below the mean. Mean and standard deviations were 1.18 and 0.15, respectively in 1999 (left panel) and 1.39 and 0.11 in 2008 (right panel). In 2006 four provinces came to exist in Sardinia. In lack of population data for all municipalities, we cannot reconstruct backward the TFR of all provinces in Sardinia. Thus, for comparability over time I excluded Sardinia and refer to the remaining 99 Italian provinces.
Source: Istat, Survey on Live Births.

4.3 Correlates of Fertility over time

Existing literature suggests a wide range of indicators which should, to some extent, help explaining cross-country fertility differentials.

One of the most cited indicators is female employment. At the country level it has been shown that female participation in the labour market correlates negatively with fertility in a variety of European countries (see, e.g., Brewster et al., 2000): countries where female employment rate is high tend to have low fertility. However, in some countries, the negative correlation between female employment and fertility has reversed its sign by the late 1980's, while in some others it still persists. Consequently, at a cross-sectional level, two distinct equilibria can be discerned: Northern-European countries are characterized by both high female employment and high fertility while Southern-European countries are characterized by both low female employment and low fertility (Ahn et al., 2002; Engelhardt et al., 2004a; Engelhardt et al., 2004b; Boeri et al., 2005).

A sign reversal in the cross-country correlation with fertility has been observed also for other indicators such as marriage propensity, cohabitation, divorce, extramarital births (Billari et al., 2004; Prskawetz et al., 2010) and GDP (Bryant, 2007; Myrskylä et al., 2009). The contribution of fertility of foreigners on total national fertility is another crucial correlate of fertility in low-fertility contexts (Coleman, 2006; Billari, 2008; Billari et al.; 2008; Sobotka, 2008).

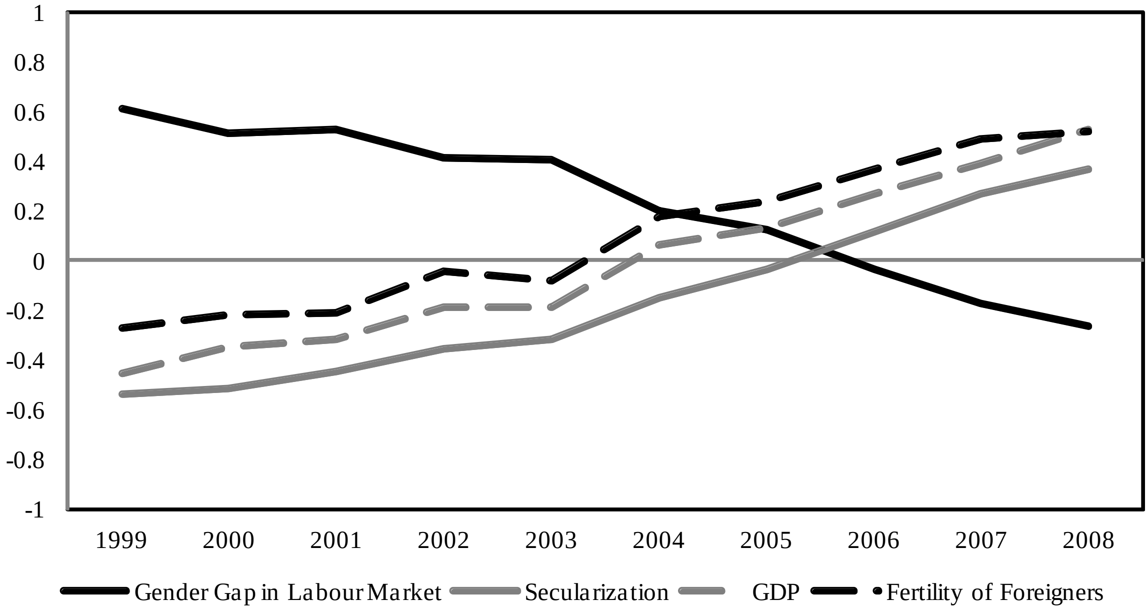
For what concerns sub-regional fertility differentials in Italy, Castiglioni et al. (2009) find that the fertility increase in Central and Northern Italian provinces in the late 1990s is positively associated with the fertility of foreigners, the spread of new marital behaviours and income. Also Billari (2008) explains the recent fertility recuperation of North-Western regions in terms of earlier spread of new marital behaviours –the “new demographic spring” for Italy (Dalla Zuanna, 2005)– which include non-marital cohabitation, extramarital births and marital instability. Dalla Zuanna et al. (1999) provide an overview of sub-regional differences in fertility behaviours observed at the beginning of the 1990s, showing that Italian provinces can be grouped into six clusters on the basis of a selection of indicators measuring reproductive and marital behaviours and economic circumstances (marital and extramarital fertility, voluntary abortions, shotgun marriages, degree of industrialization, unemployment rate, and secularization). Franklin et al. (2004) show that changes in Italian fertility in the period 1952–1991 are mainly explained by regional age-specific fertility differentials, while Waldorf et al. (2002) find that the Easterlin hypothesis is confirmed in most Italian regions over the period 1952–1995.

Figure 4.4 shows the evolution over time of the cross-province correlation coefficient between the period TFR with four indicators which will then be used in regression analyses: gender gap in the labour market, proportion of civil marriages on total marriages –considered as an indicator of secularization–, GDP and contribution of fertility of foreigners.

Figure 4.4 shows that between 2002 and 2004 the correlation of GDP and fertility of foreigners with the TFR gets close to zero. The same happens between 2004 and 2007 for gender gap in the labour market and secularization. One would then conclude that in those years none of the indictors is correlated with fertility. A change in the correlation coefficient between fertility and each of its correlates over time is an indication that something is changing. Emerging values, norms, ideas and change in the socio-economic context might be driving the change, while fertility levels themselves are changing over time. Of course, such changes do not occur uniformly in all sub-national areas of the country. As it is clear from Figure 4.2, at the beginning of the 2000s some provinces were experiencing increasing fertility and some others declining fertility. As explained by the diffusionist perspective on fertility transition, changes in fertility are the result of innovation diffusion and behavioural innovation spreading through social interaction processes and as such, these changes diffuse among the population (and hence across different areas of the country) over time. Whenever there is a sign reversal in the cross-country correlation between two variables, obviously there is a time period when the correlation crosses zero. Intuitively, when the cross-sectional correlation gets close to zero, the co-existence of opposite trends at the local level might well lead to a lack of global association. Thus, the cross-sectional correlation between fertility and one of its correlates starts to lower when a change in the association occurs in given areas. As the new ideas and behaviours that have caused such a change diffuse across space, the cross-country correlation lowers even further and reaches zero at a stage when the change has reached approximately half of the areas. The correlation then changes its sign when other areas assimilate to the change.

In the rest of the chapter, I shall try to show that the relationship between the TFR and these commonly used indicators is overestimated in regression models where spatial dependence is not taken into account. In fact, the relationship between TFR and each indicator might mask a spatial relationship. In order to shed light on this, spatial regression methods are required. In the following I shall use Geographically Weighted Regression techniques to show that the association between each indicator and fertility varies spatially and that there are, in fact, different local associations.

Figure 4.4: Correlation between TFR and four indicators, Italian provinces, 1999-2008.



Note: In 2006 four provinces came to exist in Sardinia. In lack of population data for all municipalities, we cannot reconstruct backward the TFR of all provinces in Sardinia. Thus, for comparability over time I excluded Sardinia and refer to the remaining 99 Italian provinces.

4.4 Data

The main dependent variable is the period TFR measured on a panel of 99 Italian provinces (Sardinia is excluded from the analyses, see note 2) over the period 1999-2008 and is obtained from the Survey on Live Births, which Istat produces annually, starting from 1999. The survey provides territorial data referring to geographical macro-areas, regions, provinces and regional and provincial capitals, covers the whole population of newborns[5] and collects information on births disaggregated by sex, citizenship, date and place of birth of the newborn, together with age, marital status and citizenship of both parents.

As independent variables I use GDP and GDP^2, fertility of immigrants, the gender gap in labour market participation and an indicator of secularization. The GDP is expressed in Euros per inhabitants and is calculated at current market prices (Source: Eurostat, Regional Statistics); I also consider GDP^2 in order to capture the cross-sectional nonlinear relationship between TFR and GDP. Fertility of immigrants is measured by the proportion of children with at least one foreign parent on the total number of children born in a given year (Source: Istat, Migration and calculation of yearly resident population). The gender gap in labour market participation equals to one minus the proportion of working women on the total female population aged 15 and over, relative to the same proportion calculated for the male population (Source: Istat, Labour Force Quarterly Survey data for the period 1999–2003 and Labour Force Survey data after 2003).[6] This indicator varies between zero (no gender gap, i.e. the labour participation of women equals that of men) and one (greatest gender gap, i.e. no female labour participation). Finally, I use

[5] Coverage passed from 95,8% of total babies born in 1999 to 98,9% in 2008.
[6] Labour Force Quarterly Survey data should not be compared with Labour Force Survey data, due to the reorganization carried out to meet European harmonization criteria. However, in lack of other comparable data at provincial level for the whole time series, this indicator was the best we could find.

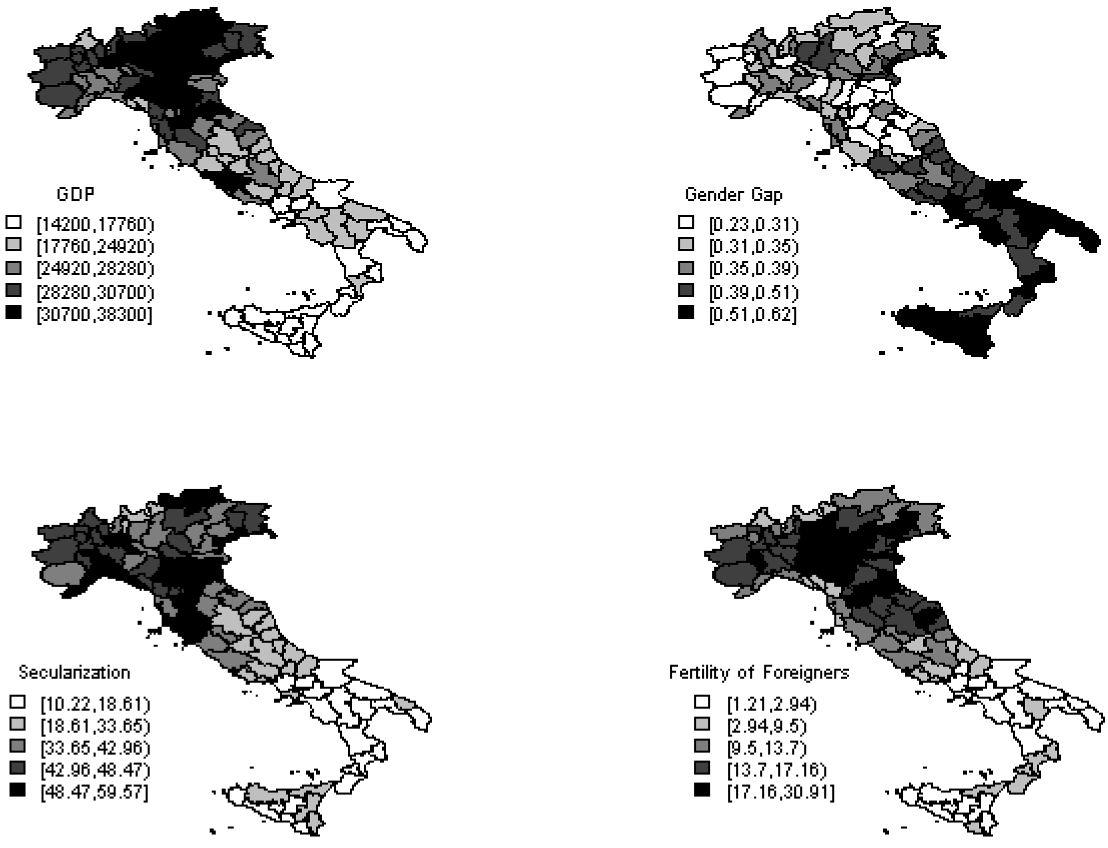
the proportion of civil marriages on the total number of celebrated marriages as an indicator of secularization (Source: Istat, Marriages).

Figure 4.5 reports the spatial distribution of the selection of indicators which will be then used in regression analyses measured in the year 2007:[7] GDP, gender gap in the labour market, an indicator of secularization and fertility of foreigners.[8] The figure shows that all indicators assume the highest values in Northern provinces, intermediate values in the center and the lowest values in the South, with the exception of gender gap which shows the opposite distribution. Northern-Eastern provinces are the richest, with GDP values in the top quintile of the distribution. Instead GDP ranges in the first quintile for all the Sicilian provinces plus other Southern provinces. Gender gap in the labour market is below 0.35 in most Northern provinces, meaning that female participation in the labour market is close to male participation. This same indicator is above 0.39 in all Southern provinces, in particular in Sicily and Apulia where it is above 0.5, meaning that the proportion of women in the labour market is less than half the proportion of men. The proportion of civil marriages on total marriages (secularization) is above 34% in all Northern provinces, while it is below 20% in most Southern provinces,

[7] Although the provincial panel is available up until 2008, the year 2007 is chosen for comparability with the Adjusted TFR whose time series refers to the period 2000-2007.
[8] GDP^2 is not reported because of its information redundancy, albeit it is included in regression models.

with Central regions ranging in between. The contribution of fertility of foreigners to the total fertility is quite low in Southern provinces (ranging from 1 to 10%) while it becomes more important in some Central and Northern provinces, which are also more attractive for immigrants. In regression analyses all variables are standardized to ease comparisons.

Figure 4.5: Observed values of indicators by quantile ranges, year 2007



In the demographic literature it is well known that period fertility measures are distorted when women postpone births (i.e. when the mean age at birth increases), because the period TFR is deflated i.e. period TFR is lower than total cohort fertility rate. The opposite happens when births are being anticipated (i.e. when the mean age at birth decreases), leading the period TFR to be higher than the total cohort fertility rate (Ryder, 1956; Bongaarts et al., 2010). When the Period TFR is biased, Bongaarts-Feeney adjustments (Bongaarts et al., 1998; 2000) are needed to correct distortions due to year-to-year tempo changes. A discussion on which tempo effects should be regarded as distortions –and therefore should be corrected– can be found in Bongaarts et al. (2010) and Ní Bhrolcháin (2011). The fertility measure used in this thesis refers to Italian provinces in the period 1999-2008. This setting certainly represents a situation of postponement. Ideally, I would like to model age-specific completed cohort fertility over time in Italian provinces. However, this measure is unavailable at the provincial level for the time period I am interested in analyzing. A second best strategy therefore would be to use Bongaarts-Feeney adjustments to correct for the tempo distortion of period TFR. However, TFR and mean age at childbearing disaggregated by birth order are not available at the provincial level and therefore the Bongaarts-Feeney adjustment formula cannot be applied in its original form, i.e.

summing the parity-specific TFRs corrected by rate of increase in parity-specific mean age at childbirth. Using the TFR and the mean age at childbearing (MAC) not disaggregated by birth order does not compensate for tempo distortions and therefore statistical analyses are carried out using the period TFR. As a robustness check I also perform the analyses on a simplified version of the adjusted TFR, computed as $TFR'_t = TFR_t/[1-r_t]$, i.e. I correct the period TFR in a given year by the annual rate of increase in mean age $r_t = [MAC_{t+1}-MAC_{t-1}]/2$ using the mean age at birth. Empirical analyses using as dependent variable the adjusted TFR will be based on the period 2000-2007 because, due to the way it is defined, this measure cannot be calculated for the first and last year of the time series.

4.5 Incorporating space into regression analyses

The importance of *spatial heterogeneity* is recognized in cross-national studies on fertility in which cross-country differences are alternatively modelled through separate analyses by country (Engelhardt et al., 2004a) or through dummy variables identifying groups of countries (Engelhardt et al., 2004b), country fixed effects or random effects (Prskawetz et al., 2010). Spatial heterogeneity is frequently considered also in sub-national studies of fertility in Italy. For instance, Castiglioni et al. (2009) focus their analyses only on Northern regions; Caltabiano (2008) compares cohort age-specific fertility between one Northern (Lombardy) and one Southern region (Campania), while Caltabiano et al. (2009) does the same comparison between the North and the South.

The concept of *spatial dependence* or *spatial autocorrelation*, instead, is less commonly considered, although spatial contiguity generally induces dependence in demographic behaviours. A number of studies called the attention on the existence of spatial patterns and the need to take these into account when studying demographic behaviours (Boyle, 2003; Goodchild et al., 2004; Weeks, 2004; Castro, 2007; Voss, 2007; Chi et al., 2008; Lesthaeghe, 2010). Although geographically referenced data have become increasing available, it is still uncommon for demographers to explicitly account for spatial dependence. In particular, very few studies model spatial dependence in fertility (Weeks et al., 2000; Waldorf et al., 2002; Işik et al., 2006; Muniz, 2009; Goldstein et al., 2010; Murphy, 2010; Potter et al., 2010).

Figure 4.1 shows that Tobler's "first law of geography" (Tobler, 1970) applies also for the Italian regional TFR: closer regions have more similar TFRs than regions which are far apart, and this is true for all years in the time series. Figure 4.3 shows that also at the provincial level spatial contiguity implies dependence in fertility measures.

The existence of spatial autocorrelation in the variable(s) of interest can be tested using the Moran's *I* index. The Moran's *I* is a global diagnostics tool for exploratory spatial data analysis which tests whether the value of a variable observed in a given location is independent of the values observed in neighbouring locations.

The Moran's *I* index (Moran, 1950) calculates the ratio between the product of the variable of interest and its spatial lag, with the product of the variable of interest, adjusted for the spatial weights used:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j\neq i}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j\neq i}^{n} w_{ij}\left(y_i - \bar{y}\right)\left(y_j - \bar{y}\right)}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \qquad (4.1)$$

where y_i is the value assumed by the variable in the *i*-th location, $\bar{y}$ is the sample mean and w_{ij} is the spatial weight assigned to the *j*-th location. Weights are specified as follows:

$$w_{ij} = \begin{cases} \dfrac{1}{\eta_i} & \text{if } j \in N(i) \\ 0 & \text{otherwise} \end{cases} \qquad (4.2)$$

where $N(i)$ defines the set of all neighbours to the spatial unit *i* and η_i is the cardinality of $N(i)$ (i.e. the number of neighbours to spatial unit *i*) and it is assumed that a unit cannot be its own neighbour i.e., $w_{ii}=0$.[9] In this case neighbours are defined on the basis of a contiguity criterion, according to which two locations are neighbours if they share a border or an edge (queen criterion).[10]

The Moran's *I* index ranges between -1 (perfect negative spatial autocorrelation, i.e. a location in which a high value of the variable is observed is surrounded by locations with low values of the variable) and 1 (perfect positive spatial autocorrelation, i.e. similar values are clustered together in space). When the index is close to 0 there is a random spatial distribution, i.e. no spatial autocorrelation.

Figure 4.6 reports the evolution over time of the Moran's *I* index calculated on provincial TFRs. Although the Moran's *I* index declines over time, all indexes are positive and significant at the 1% level per random permutation. We can reject the null hypothesis that there is zero spatial

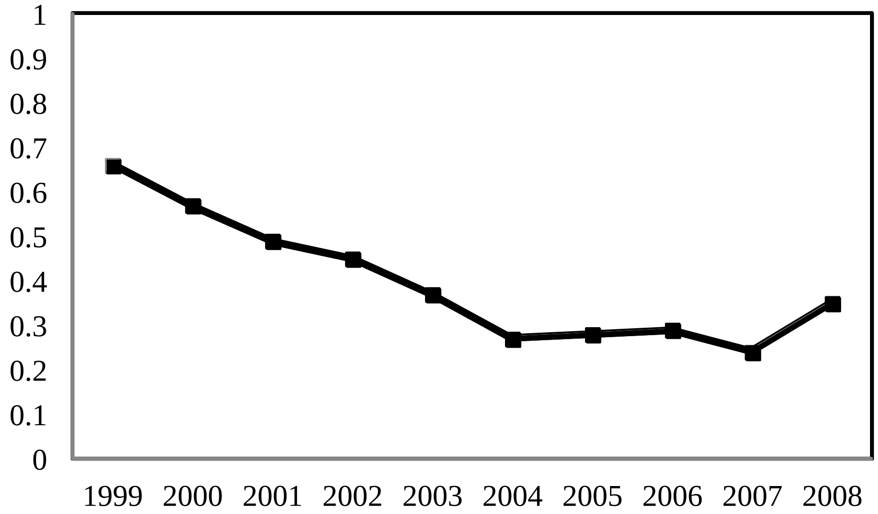
---

[9] w_{ij} are the entries of a spatial weight matrix. A spatial weight matrix is a matrix that selects neighbors. Suppose that the spatial unit *i* has two neighbors, say the spatial units *j* and *k*; then, the i^{th} row of the *W* matrix will have two non-zero elements i.e., the entries w_{ij} and w_{ik} will be different from zero. The matrix *W* is row standardized because the following condition holds: $\sum_{j=1}^{\eta_i} w_{ij} = 1$.

[10] In the spatial econometric and statistical literature the choice of an appropriate neighbouring structure is of paramount importance. In this chapter the spatial units are provinces, i.e. irregular polygons, therefore the neighbouring structure can be chosen among the contiguity, k-nearest neighbours or distance criterion. The latter is based on a measure of distance computed between the centroids of the polygons, and it can be Euclidean, Great Circle, Canberra, etc. or it can be specified by the user basing on a given dissimilarity criterion (e.g. social or economic distance or such that the morphology of the area, the actual time or cost it takes to travel the distance, etc. are taken into account. For an extensive review on the selection of the neighbouring and weighting schemes see Anselin (2002), Leenders (2002) and Chi et al. (2008).

autocorrelation present in the provincial TFR. Provinces, therefore, cannot be modelled as independent units. Indeed, provinces are spatially dependent; in other words, it cannot be assumed that fertility observed in a given province is independent from fertility observed in a neighbouring province. However, independence among observations is the main assumption of traditional regression models. In this chapter I do not superimpose a geographical structure which would a priori generate clusters of regions, as it would, for example, the inclusion of control dummy variables identifying the three macro regions of South, North and Centre. Therefore I explicitly take into account spatial dependence among provinces by the means of spatial regression models.

Figure 4.6: Dynamic change of global Moran's *I* of TFR, 1999-2008



Spatial modelling allows for the introduction into regression models of spatial (and social) interactions among neighbouring observations in space. The idea is to include in the statistical model a function of neighbouring observations through a spatial lag operator generating a new variable which is a weighted average of the neighbouring observations. Spatial dependence can be modelled applying the spatial lag operator to the dependent variable, to independent variables or to the error term, yielding the spatial lag model, the spatial Durbin model and spatial error model, respectively. Besides the features of cross-sectional spatial regressions, panel data with spatial interaction allows accounting also for the dynamics of the process being studied.

The spatial lag and spatial error model specifications might induce very different spatial correlation patterns among the observations and sometimes the use one specification might be inappropriate according to the theoretical framework of reference (Aselin, 2002, p. 248). Anselin et al. (2008, p. 6) explain that the “spatial lag model is typically considered as the formal specification for the equilibrium outcome of a spatial or social interaction process, in which the value of the dependent variable for one agent is jointly determined with that of the neighboring agents”. The spatial error model instead “does not require a theoretical model of spatial/social integration, but, instead, is a special case of a nonspherical error covariance matrix” (Anselin et al., 2008, p. 8) and such that spatial dependence in the dependent variable is the result of a spatial clustering of unobserved independent variables omitted from the model. A model specification including both a spatial lag and a spatial error term is possible and it is advisable when the assumptions underlying the two models are thought of being satisfied.

Diffusion processes during fertility transitions have been studied empirically in only few contributions. While some have modelled diffusion with an autocorrelation coefficient on the error term using a spatial error model (Loftin et al., 1983), most agree on using an autocorrelation coefficient on fertility –the dependent variable– using a spatial lag model

(Montgomery et al., 1993; Tolnay, 1995; Casterline, 2001; Palloni, 2001; Muniz, 2009; Murphy, 2010). The spatial lag model represents a diffusive process in the dependent variable, and as such it is appropriate for modelling social network as well as diffusion processes, including the diffusion of behavioural innovations and diffusion of new ideas, because such processes spread among individuals over space. Since this is exactly the idea behind the diffusionist perspective about fertility transitions, the spatial lag model should be the correct choice for modelling diffusion in fertility.[11]

4.5.1 A Spatial Cross-Sectional Perspective

Geographically Weighted Regression (GWR) techniques are local regressions which allow for the estimation of heterogeneous relationships between the dependent and independent variables when the observations are measured at different locations (Brunsdon et al., 1998; Brunsdon et al., 1999; Fotheringham et al., 2002). This technique is particularly useful when the relationship among variables differ from location to location (non-stationarity).

[11] In order to check the appropriateness of the spatial lag model for modelling diffusion in fertility, the Lagrange Multiplier test was also performed and, as expected, it is in favour of the spatial lag model.

For a given cross-section and for each location, GWR fits a single linear regression equation of the form:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \tag{4.3}$$

where y_i denotes the response variable in province $i=1,...,N$, x_{ik} the k-th independent variable measured in province i, (u_i,v_i) the coordinates (longitude and latitude) of the centroid of the i-th province, $\beta_k(u_i,v_i)$ the parameter associated to the k-th variable in the i-th province and ε_i the error term (Fotheringham et al., 2002). For each observation (i.e. province) i, GWR estimates an intercept term and a vector of parameter estimates using a modification of the weighted least squares model. Each regression equation (one for each province) is calibrated using a different weighting scheme on the basis of spatial dependence among neighbouring provinces. Provinces can be thought of as (irregular) spatial polygons and it is possible to calculate their centroids' geographic coordinates (longitude and latitude) on the basis of which geographical distances can be computed. Weights are inversely proportional to the distance between provinces' centroids. The vector of parameter estimates for a given location i is obtained using a weighting scheme:

$$\hat{\boldsymbol{\beta}}(u_i,v_i) = (\boldsymbol{X}^{\mathrm{T}} \mathbf{W}(u_i,v_i) \boldsymbol{X})^{-1} \boldsymbol{X}^{\mathrm{T}} \mathbf{W}(u_i,v_i) \boldsymbol{y} \tag{4.4}$$

where $\mathbf{W}(u_i,v_i)$ is an nxn diagonal spatial weight matrix of the form:

$$\mathbf{W}(u_i,v_i) = \begin{pmatrix} w_{i1} & 0 & \dots & 0 \\ 0 & w_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{iN} \end{pmatrix}. \qquad (4.5)$$

The generic entry w_{ij} is the weight given to location j for the estimate of the local parameters at location i. This weight follows a Gaussian distance decay and is equal to $w_{ij} = \exp[-1/2(d_{ij}/h)^2]$ where d_{ij} is the Euclidean distance between locations i and j. The term h is the bandwidth which determines the number of locations to be included in each local regression. As the bandwidth increases, the gradient of the kernel becomes less steep and more locations will be included in the local calibration. In order to choose the optimal value for the bandwidth, I use the Akaike Information Criterion (AIC) which minimizes the root mean square prediction error for the geographically weighted regression.

Results from GWRs can be plotted on a map so that it becomes fairly easy to detect spatial non-stationarity in the association between variables. In order to evaluate the significance of parameter estimates I use t-values. Estimation is carried out using the “spgwr” library (Geographically Weighted Regression) in R.

GWRs account for spatial heterogeneity, allowing the effect of covariates to vary spatially, and for spatial dependence, allowing the effect of explanatory variables in neighbouring locations to count more than locations that are far apart. However, spatial dependence only works though the association between the explanatory variables and the dependent variable. Also, GWR methodologies are available only to study cross-sectional dataset and therefore I limit the analysis to the year 2007. Işik et al. (2006) and Muniz (2009) have used GWRs to explain fertility differentials in Turkey and Brazil, respectively.

In the context of this chapter locations are the 99 Italian provinces,[12] and thus GWRs estimate a regression equation for each province while taking into account spatial dependence in the model. In order to compare the marginal effect of the different indicators on fertility (TFR), i.e., in order to assess which indicator has more explanatory power on fertility, I run a regression model which simultaneously includes GDP and its square, fertility of foreigners, gender gap in labour market participation and secularization. All variables are standardized according to their mean and standard deviation in year 2007. As an additional robustness check, GWRs are also estimated using the Adjusted TFR as the dependent variable. The year 2007 is chosen for comparability with regressions using the adjusted fertility measure as dependent variable.

[12] I exclude the 8 provinces of Sardinia because 4 of them came to exist in 2006, making it impossible, in lack of municipal data, to reconstruct backward the time series of all variables used in the analyses.

Figure 4.B in the Appendix reports the spatial distribution of the provincial TFR and the Adjusted TFR in 2007.

4.5.2 A Spatial Panel Perspective

The next step in the analysis is the inclusion of the time dimension in the study of diffusion of fertility in Italian provinces. The interaction between provinces is modelled including a spatially lagged dependent variable (a spatial lag). The spatial lag allows the TFR to depend on the TFR observed in neighbouring provinces.

Spatial panel data are one of the most promising tool to analyze the spatial and the temporal dimensions simultaneously (Anselin, 1988; Elhorst, 2003; Anselin et al., 2008; Elhorst, 2010). I employ two different spatial panel data regression models: the fixed and random effects models and compare coefficient estimates with those obtained through a traditional panel model with fixed effects.

The first model I estimate (Model 1) is the traditional panel model with spatial, i.e. provincial, and time-period fixed effects which can be expressed as follows:

$$y_{it} = x_{it}\beta + \mu_i + \nu_t + \varepsilon_{it}\,, \quad i = 1,\ldots,N; \quad t = 1,\ldots,T \qquad (4.6)$$

where *i* indexes the provinces and *t* the time periods. The dependent variable y_{it} is the TFR observed in location *i* at time *t*, x_{it} is the vector of independent variables of dimension $1 \times k$, β a matching vector of fixed unknown parameters, while μ_i and ν_t denotes province-specific and time-period fixed effects, assumed to be constant over time and independent of the error term ε_{it}. The error term ε_{it} is independently and identically normally distributed with mean 0 and variance to be estimated. Fixed effects control for all fixed (i.e. time invariant) provincial-specific characteristics.

The fixed effects panel model described in (4.6) can be extended to include a spatially lagged dependent variable (spatial lag) on the right-hand side of the regression equation (Anselin, 1988). The introduction of the spatial lag allows the TFR in a given province (y_{it}) to depend on the TFR observed in neighbouring provinces (and on the included independent variables). The resulting model is the spatial panel model with provincial and time-period fixed effects, which is the second model that I estimate (Model 2). It is formally described as follows:

$$y_{it} = \rho \sum_{j=1}^{N} w_{ij} y_{jt} + x_{it}\beta + \mu_i + \nu_t + \varepsilon_{it}\ , \quad i = 1,\ldots,N;\ \ t = 1,\ldots,T \qquad (4.7).$$

The difference with respect to the fixed effects panel model (Model 1) in (4.6) is the term $\sum_{j=1}^{N} w_{ij} y_{jt}$, which is the spatial lag of the dependent variable (Cliff et al., 1973), with w_{ij} equal to the weight assigned to province j. The coefficient ρ measures the spatial autocorrelation in the dependent variable. A positive and statistically significant estimate of ρ

has to be interpreted as spatial autocorrelation in the TFR or, in other words, that provinces with similar values of the TFR tend to cluster together in space, which is evidence in favour of spatial diffusion of fertility. Spatial dependence operates through a pre-defined, user-specified weight matrix (***W***). The weight matrix is constant over time, has dimension *N*x*N* and is a non-stochastic row-standardized spatial weight matrix which takes into account the neighbouring structure of the spatial units such that its entries satisfy the formula in (4.2).

In order to check the robustness of spatial dependence in the TFR, I estimate the spatial panel model using two additional specifications which will be referred to as Model 3 and 4, both using random rather fixed effects for the provinces. Both models are particular cases of the general spatial panel model proposed by Baltagi et al. (2007), which accounts for spatial dependence between provinces at each time period, for serial correlation on each province over time, and allows for heterogeneity across provinces using a random effect. The spatial panel model with provincial random effects is described as follows:

$$y_{it} = \rho \sum_{j=1}^{N} w_{ij} y_{jt} + x_{it}\beta + \mu_i + \varepsilon_{it} \text{ , } i = 1,\dots,N; \quad t = 1,\dots,T \qquad (4.8).$$

In the specification in (4.8), the term μ_i denotes the province-specific random effect, assumed to be constant over time and independent of the error term ε_{it}. Both μ and ε are independently and identically normally distributed with mean 0 and variance to be estimated. The ratio between their variances, defined as $\varphi = \sigma^2_{\mu}/\sigma^2_{\varepsilon}$, gives the contribution of provincial-specific variation in the TFR relative to the variation due to unobserved factors. The spatial panel model with provincial random effects will be referred as Model 3.

The last model specification that I consider extends Model 3 to incorporate (time) serial correlation. The resulting model (Model 4) is a spatial panel model with provincial random effects and serial correlation and it is expressed as follows:

$$y_{it} = \rho \sum_{j=1}^{N} w_{ij} y_{jt} + x_{it}\beta + \mu_i + \varepsilon_{it} \text{ , } i = 1,\dots,N; \quad t = 1,\dots,T \qquad (4.9)$$

$$\varepsilon_t = \psi \, \varepsilon_{t-1} + e_t \qquad (4.10).$$

For each cross-section in (4.9) the error component ε_t is further decomposed in (4.10) to isolate a first-order autoregressive component and the coefficient ψ measures its serial correlation. The general model proposed by Baltagi et al. (2007) further allows for spatial dependence in the error term, which I do not include in my model.

I rely on a panel constituted of repeated observations for provinces over the period 2000-2007. The spatial distribution of the provincial TFR and the Adjusted TFR at the beginning and at the end of the panel are mapped in Figure 4.B in the Appendix. As an additional robustness check, spatial panel models are also estimated using the Adjusted TFR as the dependent variable. The models are estimated by a two-step Maximum Likelihood procedure using the “splm” library (Econometric Models for Spatial Panel Data) in R.

4.6 Results

4.6.1 Cross-section: Geographically Weighted Regressions

Results from GWRs are a series of local parameter estimates which measure the association between each of the independent variables and the TFR for each province, controlling for the other independent variables included in the model. To ease interpretation, parameter estimates are reported on the map of Italian provinces. In this way it is possible to interpret the spatial distribution and the magnitude of the local associations.

Figure 4.7 reports a map plotting the local parameter estimates by quintile ranges together with their relative t-values. Provincial fertility results negatively associated with GDP. This negative relationship is particularly strong in the Central provinces, where the local parameter estimates are lower than -0.5. As one moves from the Centre to the North, this relationship becomes weaker, i.e. closer to zero. In Southern provinces, instead, the relationship remains negative and statistically different from zero. The non linearity in the relationship between fertility and GDP is captured by the positive parameter estimates for GDP^2. Therefore fertility is higher in the most economically developed areas of the country (Northern provinces). This result confirms the recent findings that “advances in development reverse fertility declines” (Myrskylä et al., 2009). Fertility of foreigners is positively and significantly associated with fertility in all Italian provinces. The strength of this association ranges from 0.55 to 1.13 and follows a perfect core-periphery pattern with low parameter estimates in the North and high estimates in the South. The spatial distribution of fertility of foreigners mapped in Figure 4.5 shows a great variability across Italian provinces, with high contribution in the North and very low in the South. This is due to the fact that foreigners are concentrated in Northern and, to a lesser extent, Central provinces. For most immigrants, the South of Italy represent a transit place for reaching their final destinations in Northern Italy or continental Europe, and therefore foreign women in Southern provinces tend to have lower fertility than foreign women in the rest of Italy. Results show that an increase in fertility of foreigners in Southern regions would lead to an increase in fertility. In particular, a standard deviation increase in fertility of foreigners would imply an increase in fertility in the range of 0.91 to 1.13 standard deviations. The association between secularization and fertility is positive in all Southern provinces, while it is not statistically significant in Northern and Central provinces. The indicator here chosen for secularization is the proportion of civil marriages on the total number of marriages, therefore provinces where such indicator assumes higher values are provinces with low religiosity as well as higher diffusion of new family models. It is expected that these provinces will also show high rates of divorces and legal separations, non-marital cohabitation, and out-of-wedlock births. The highest values of secularization are found in the North of Italy and so it is for the other above-mentioned indicators which correlate with secularization, while Southern provinces are more traditional in this respect. Results from GWRs show that an increase in secularization would increase fertility in Southern provinces. Similarly, provincial fertility is found to be positively associated with the gender gap in the labour market in all Southern provinces, while the association is not statistically significant in Northern provinces. An increase in female labour force participation in Southern provinces would lead to a further fertility decline. In other words, in the South of Italy there is an inverse relationship between female employment and fertility. Previous studies document that in a cross-country perspective, Italy together with the other Mediterranean countries, maintains a negative association between fertility and female employment (Brewster et al., 2000; Ahn et al., 2002). However, in this chapter it is shown that such a negative association holds only for the South of Italy. Therefore, for Southern provinces in 2007 we observe the traditional pattern between fertility and female employment which used to be universal up until the late 80s in all advanced countries. Such association instead is not statistically significant in the North of Italy, once the other important variables are controlled for. Therefore Northern regions are in between the traditional association observed in the rest of Italy and Northern European countries, where high female employment is associated with high fertility levels. Though far from Northern European standards, Northern Italian provinces allow an easier combination of work and children with respect to other areas of the country. Female labour force participation is in fact higher than in Southern regions and part-time work and childcare facilities are more widespread.

Results are robust to tempo distortions as accounted in the Adjusted TFR (see Figure 4.A in the Appendix). When the dependent variable is the Adjusted TFR, local parameter estimates and their significance are similar to those obtained on the TFR. The range of values of local parameters is slightly smaller, indicating a smaller variation in the association between variables across provinces.

Spatial heterogeneity is not just a matter of different levels in fertility observed across the country. Results from GWRs show that there is substantial heterogeneity also in the association between fertility and its correlates across provinces. In other words, the association between each indicator and fertility varies locally from being statistically insignificant in some provinces to being significant in other provinces. Also, among provinces for which the association is significant, the magnitude of the association varies considerably. A cross-sectional approach which does not take into account spatial dependence across provinces fails to incorporate the existing local heterogeneity in both the dependent and independent variables. This chapter focuses on Italian provinces but the same considerations on the appropriateness of geographically weighted regressions are applicable to other national contexts characterized by a high internal heterogeneity in both fertility and its correlates.

Figure 4.7: Results from GWRs on TFR: local parameter estimates by quintile ranges and t-values, year 2007

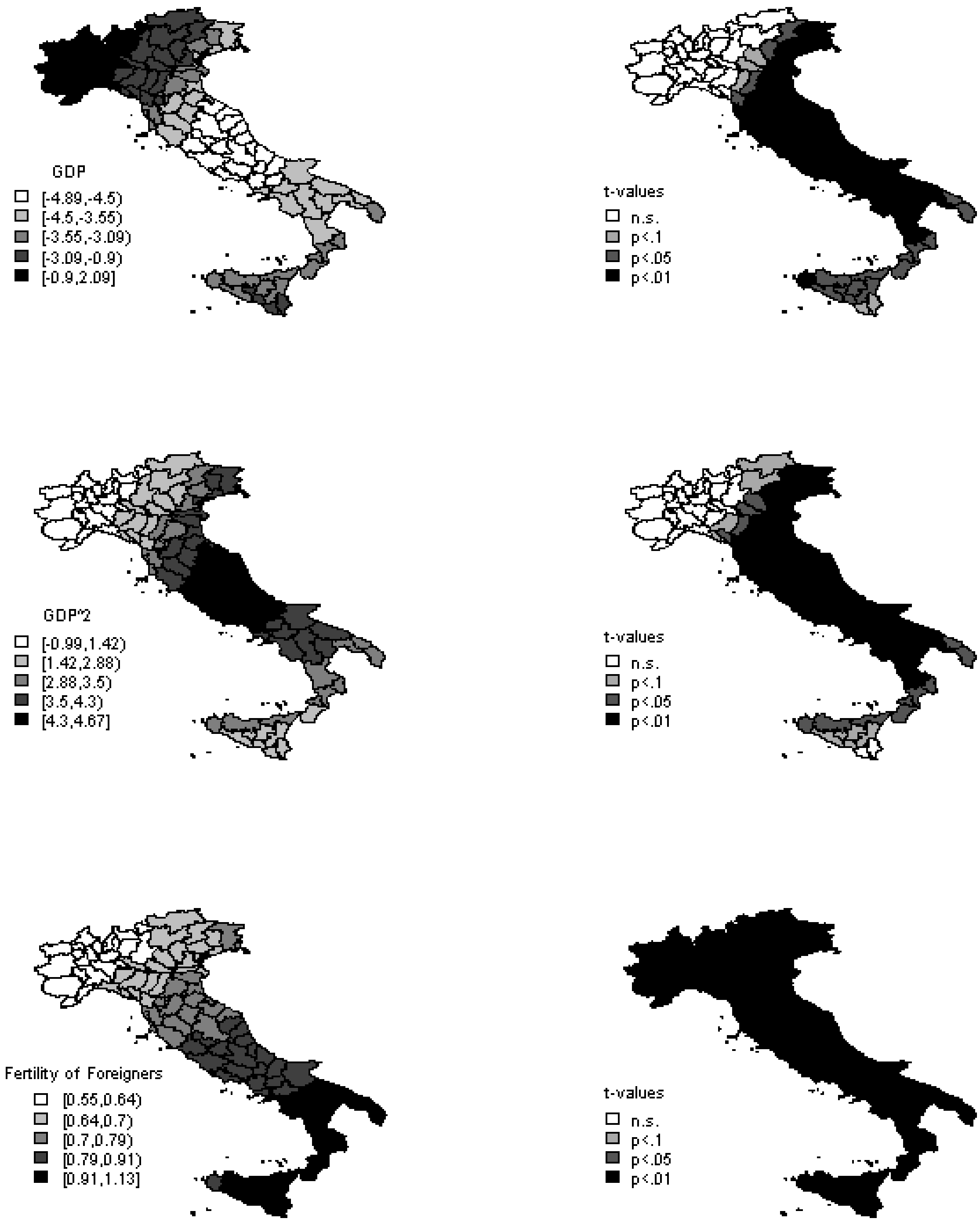
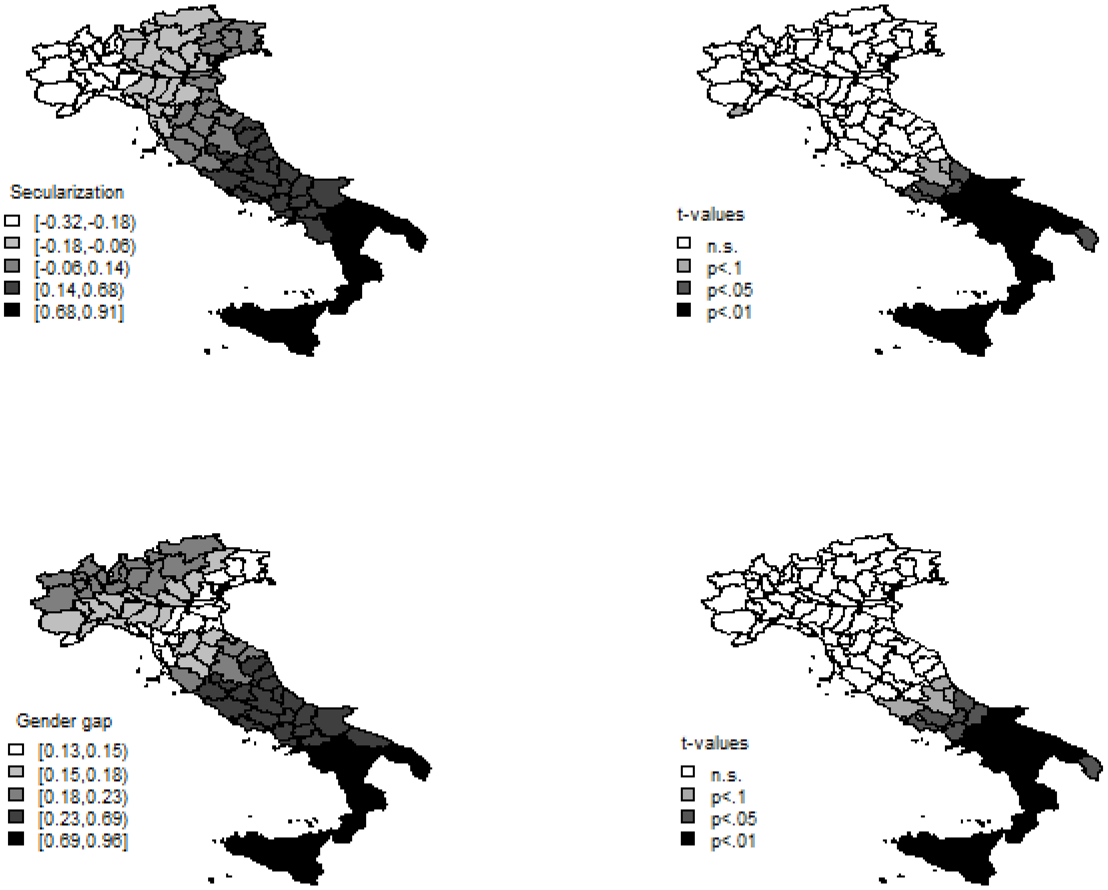


Figure 4.7: (Continued)



4.6.2 Spatial Panel Models

Table 4.1 reports coefficient estimates for the traditional panel model with provincial and time-period fixed effects (Model 1) and the spatial panel model with provincial and time-period fixed effects (Model 2).

My investigation shows that disregarding spatial dependence leads to overestimating the effect of all indicators chosen to explain fertility. For instance, if I estimate the effect of provincial GDP on the provincial TFR using the traditional fixed effects panel model (Model 1), I find that a one standard deviation increase in the indicator leads to a 1.02 standard deviation reduction in the TFR. This effect weakens to -0.71 when I account for spatial dependence across neighbouring provinces, using the spatial panel fixed effects model (Model 2). A similar reduction in the estimated coefficient exists for fertility of foreigners (from 0.60 in the traditional panel model to 0.45 in the spatial panel model), gender gap in the labour market remains (from 0.08 to 0.07) and secularization (from 0.21 to 0.1).

Table 4.1: Estimates of the regression of fertility (TFR) on selected indicators, panel and spatial panel models with provincial fixed effects

	Model 1			**Model 2**		
	β		s.e.	β		s.e.
GDP	-1.024	***	0.233	-0.707	***	0.208
GDP^2	1.320	***	0.189	0.903	***	0.173
Fertility of Foreigners	0.597	***	0.045	0.446	***	0.042
Gender Gap	0.079	*	0.036	0.074	*	0.032
Secularization	0.214	***	0.057	0.109	*	0.051
ρ				0.329	***	0.037

Note: Model 1 refers to the traditional panel model with provincial fixed effects and Model 2 to the spatial panel model with provincial fixed effects. All variables are standardized. ρ: spatial autocorrelation in the dependent variable.
p-value: *** < 0.001; ** < 0.01; * < 0.05; **.** < 0.1.

The estimated spatial autocorrelation coefficient of the TFR (ρ) is equal to 0.33, indicating a positive spatial dependence of fertility across provinces. Basing on the selection of indicators chosen, GDP is the most important predictor of fertility in Italian provinces, followed by fertility of immigrants. If the contribution of fertility of immigrant increases by one standardized unit, the provincial TFR would increase by 0.45 standardized units. It should be noted that the effect of foreign fertility is probably underestimated as the data refer to births with at least one foreign parent legally resident in one of the Italian provinces. As a result, if we consider that illegal immigration is a widespread phenomenon in Italy, the true contribution of fertility of foreigners on total fertility is expected to be more important than what we are actually estimating. On average, if the gender gap in the labour market is increased by one standardized unit, the TFR would increase by 0.07 standard units. Provinces where secularization is more widespread tend to have higher fertility with respect to provinces where it is less widespread. From Geographically Weighted Regressions we know that these average effects mask different provincial patterns which cannot be taken into account in the panel model.

Results from the spatial panel models with provincial random effects (Model 3) and provincial random effects plus serial autocorrelation in the disturbance (Model 4) are presented in Table 4.2. These additional specifications confirm that, once independent variables are controlled for, the spatial autocorrelation coefficient of the TFR (ρ) remains significant and equal to 0.4. In Model 4, the estimated coefficient measuring serial correlation (ψ) is positive and significant indicating that the TFR in a given year is correlated with the TFR in the previous year. The contribution of provincial-specific variation in the TFR is significantly higher than the contribution of the variation due to unobserved factors (φ). Parameter estimates obtained in random effects models are slightly different than those obtained with fixed effects. An explanation of this issue can be found in Elhorst (2010).

Table 4.2: Estimates of the regression of fertility (TFR) on selected indicators, spatial panel models with provincial random effects

	Model 3			Model 4		
	β		s.e.	β		s.e.
GDP	-0.821	***	0.149	-0.949	***	0.177
GDP^2	0.920	***	0.143	1.049	***	0.169
Fertility of Foreigners	0.418	***	0.035	0.403	***	0.037
Gender Gap	0.100	**	0.033	0.097	**	0.034
Secularization	0.013		0.044	-0.011		0.047
ρ	0.448	***	0.033	0.442	***	0.034
φ	8.783	***	1.466	7.934	***	1.383
ψ				0.354	***	0.051

Note: Model 3 refers to the spatial panel model with provincial fixed effects and Model 2 to the spatial panel model with provincial fixed effects. All variables are standardized. Significance of spatial and temporal parameters (φ, ψ, λ) is tested by the means of one-dimensional conditional tests developed in Baltagi et al. (2007). ρ: spatial autocorrelation in the dependent variable; φ: $(\sigma_\mu/\sigma_\varepsilon)$, contribution of provincial-specific variation in the dependent variable relative to the variation due to unobserved factors; ψ: serial correlation.
p-value: *** < 0.001; ** < 0.01; * < 0.05; **.** < 0.1.

What matters in this chapter is that the positive and significant estimate of the spatial autocorrelation coefficient is robust across different model specifications. Results of the spatial panel models estimated using the Adjusted TFR as the dependent variable (Table 4.A), confirm the existence of a significant and positive spatial autocorrelation in fertility. However, the estimated spatial dependence in tempo-adjusted fertility is lower than the spatial dependence in period fertility.

4.7 Conclusion

Starting from the mid-1990s, most regions located in the North and Centre of Italy showed an increasing trend in fertility levels, while most regions in the South experienced a continued fertility decline. The coexistence of the two fertility trends implied a progressive convergence of provincial fertility levels. Of course, if the dual dynamics had to remain, convergence would be a transitory phenomenon. There are, however, reasons to believe that the South will adjust to the national trend in the future. For instance, Lesthaeghe et al., 2002 have shown that behavioural innovations in the context of the First and Second Demographic Transitions spread following a spatial pattern. According to the spatial diffusion approach, we could expect that as Southern regions acted as followers in the fertility decline of Northern regions, they will also follow in the fertility recuperation, leading to another period of fertility convergence towards higher levels.

This chapter contributes to the demographic literature on diffusionist perspective to fertility transition by studying the temporal and spatial dimensions of Italian provincial fertility trends simultaneously.

Results from Geographically Weighted Regressions showed that the association between fertility and GDP, secularization, fertility of foreigners and gender gap in the labour market is not homogeneous across provinces. The strength and in some cases also the sign of such associations vary spatially.

Results from spatial panel regressions show that spatial dependence (or autocorrelation) in provincial fertility persists even after controlling for the usual correlates of fertility. The spatial lag coefficient always results positive and statistically significant, which is interpreted as a confirmation of spatial diffusion in fertility. Results are robust to different model specifications and the significance of the spatial lag persists even when correcting for tempo distortions in fertility.

The interpretation of the autocorrelation effect as evidence in favour of diffusion is reinforced by the fixed-effects design employed in the regression models. The fixed-effects model controls for all time-constant provincial-specific unobservable characteristics and therefore it can safely be stated that the spatial effect is not the result of spatially clustered time-constant explanatory variables.

To shed light on the role of tempo effects on diffusion of fertility it is required to use a tempo-adjusted measure of fertility which is properly calculated taking into account parity-specific TFR and mean age at childbearing. Such information is not available at the provincial level, therefore a further step for future research could be to inspect diffusion mechanisms at the regional level, for which fertility measures by birth order are available. Of course this would come at the price of losing the smaller homogeneous contextual dimension of the provincial level.

4.8 References

Ahn, N. and Mira, P. (2002). A note on the changing relationship between fertility and female employment rates in developed countries. *Journal of Population Economics*, 15: 667-682.

Anderson, B. A. (1986). Regional and cultural factors in the decline of marital fertility in Western Europe. In Coale, A. J. and Watkins, S. C., *The Decline of Fertility in Europe*. Princeton: Princeton University Press.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.

Anselin, L. (2002). Under the hood Issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27(3): 247–267.

Anselin, L., Le Gallo, J. and Jayet, H. (2008). Spatial panel econometrics. In: Matyas, L. and Sevestre, P. (Eds) *The econometrics of panel data, fundamentals and recent developments in theory and practice (3rd edition)*. Dordrecht: Kluwer Academic Publishers.

Baltagi, H.B.; Song, S.H.; Jung, B.C. and Koh, W. (2007). Testing for Serial Correlation, Spatial Autocorrelation and Random Effects Using Panel Data. *Journal of Econometrics*, 140(1): 5–51.

Billari, F.C. and Kohler, H.-P. (2004). Patterns of Lowest-low Fertility in Europe. *Population Studies*, 58, 161–176.

Billari, F. C. and Dalla Zuanna, G. (2008). *La Rivoluzione nella Culla: Il Declino che non c'è*. Milan: Egea.

Billari, F. C. (2008). Lowest-low Fertility in Europe: Exploring the Causes and Finding Some Surprises. *Japanese Journal of Population*, 6(1): 2–18.

Bocquet-Appel, J.-P. and Jakobi, L. (1998). Evidence for a Spatial Diffusion of Contraception at the Onset of the Fertility Transition in Victorian Britain. *Population*-E, 10(1): 181–204.

Bocquet-Appel, J.-P., Rajan, S. I., Bacro, J. N. and Lajaunie, C.. (2002). The Onset of India's Fertility Transition. *European Journal of Population*, 18(3), 211-232.

Boeri, T., Del Boca, D. and Pissarides, C.A. (2005). *Women at work, an economic perspective*. Oxford: Oxford University Press.

Bongaarts, J. and Watkins, S. C. (1996). Social interactions and Contemporary Fertility Transitions. *Population and Development Review*, 22(4): 639–682.

Bongaarts, J. and Feeney, G. (1998). On the Quantum and Tempo of Fertility. *Population and Development Review*, 24(2): 271–291.

Bongaarts, J. and Feeney, G. (2000). On the Quantum and Tempo of Fertility: Reply. *Population and Development Review*, 26(3): 560–564.

Bongaarts, J. and Feeney, G. (2010). When is a tempo effect a tempo distortion? *Genus*, 66(2): 1–14.

Boyle, P. (2003). Population Geography: Does Geography Matter in Fertility Research? *Progress in Human Geography*, 27(5): 615–626.

Brewster, K. L. and Rindfuss, R.R. (2000). Fertility and Women's Employment in Industrialized Nations. *Annual Review of Sociology*, 26: 271–296.

Brunsdon, C., Fotheringham, A. S. and Charlton, M. E. (1998). Geographically Weighted Regression-Modelling Spatial Non-Stationarity. *Journal of the Royal Statistical Society, D*, 47(3): 431–443.

Brunsdon, C., Fotheringham, A. S. and Charlton, M. E. (1999). Some notes on parametric signficance tests for geographically weighted regression. *Journal of Regional Science*, 39(3): 497-524.

Bryant, J. (2007). Theories of Fertility Decline and the Evidence from Development Indicators. *Population and Development Review*, 33(1): 101–127.

Caltabiano, M. (2008). Has the Fertility Decline Come to an End in the Different Regions of Italy? New Insights from a Cohort Approach. *Population*, 63(1): 157–171.

Caltabiano, M., Castiglioni, M. and Rosina, A. (2009). Lowest-Low Fertility: Signs of a Recovery in Italy? *Demographic Research*, 21(23): 681–718.

Carlson, G. (1966). The Decline of Fertility: Innovation or Adjustment Process. *Population Studies*, 20(2): 149–174.

Carter, A. T. (2001). Social Processes and Fertility Change: Anthropological Perspectives. In Casterline, J. (Ed.) *Diffusion Processes and Fertility Transition, Selected Perspectives*. Washington, D.C.: National Research Council.

Casterline, J. B. (2001). Diffusion Processes and Fertility Transition: Introduction. In Casterline, J. (Ed.) *Diffusion Processes and Fertility Transition, Selected Perspectives*. Washington, D.C.: National Research Council.

Castiglioni, M. and Dalla Zuanna, G. (2009). Marital and Reproductive Behavior in Italy after 1995: Bridging the Gap with Western Europe? *European Journal of Population*, 25(1): 1–26.

Castro, M. C. (2007). Spatial demography: An opportunity to improve policy making at diverse decision levels. *Population Research and Policy Review*, 26(5–6): 477–509.

Chi, G. and Zhu, J. (2008). Spatial Regression Models for Demographic Analysis. *Population Research and Policy Review*, 27(1): 17–42.

Cleland, J. and Wilson, C. (1987). Demand Theories of the Fertility Transition: An Iconoclastic View. *Population Studies*, 41(1): 5–30.

Cliff, A. D. and Ord, J.K. (1973). *Spatial Autocorrelation*. London: Pion Limited.

Coale, A. J. and Watkins, S. C. (1986). *The Decline of Fertility in Europe*. Princeton: Princeton University Press.

Coleman D. (2006). Immigration and Ethnic Change in Low-fertility Countries: A Third Demographic Transition. *Population and Development Review*, 32(3): 401–446.

Dalla Zuanna, G. and Righi, A. (1999). *Nascere nelle Cento Italie. Analisi Territoriale del Comportamento Riproduttivo nelle Province Italiane*. Collana Argomenti. Roma: Istat.

Dalla Zuanna, G. (2005). Una nuova primavera demografica. *Il Mulino*, 6: 1061–1071.

Elhorst, J. P. (2003). Specification and estimation of spatial panel data models. *International Regional Science Review*, 26(3): 244–268.

Elhorst, J. P. (2010). Spatial panel data models. In Fischer, M. M. and Getis, A. (Eds.) *Handbook of Applied Spatial Analysis*. Berlin Heidelberg: Springer.

Engelhardt, H., Kogel, T. and Prskawetz, A. (2004a). Fertility and Women's Employment Reconsidered: A Macro-level Time-series Analysis for Developed Countries, 1960-2000. *Population Studies*, 58(1): 109–120.

Engelhardt, H. and Prskawetz, A. (2004b). On the Changing Correlation Between Fertility and Female Employment over Space and Time. *European Journal of Population*, 20(1): 35–62.

Fotheringham, A. S., Brunsdon, C. and Charlton, M. E. (2002). *Geographically Weighted Regression: the Analysis of Spatially Varying Relationships*. John Wiley & Sons Ltd., West Sussex, England.

Franklin, R. (2003). Italian Fertility, 1864 to 1961: An Analysis of Regional Trends. Working Paper.

Franklin, R. and Plane, D. A. (2004). A Shift-Share Method for the Analysis of Regional Fertility Change: An Application to the Decline in Childbearing in Italy, 1952-1991. *Geographical Analysis*, 36(1): 1–20.

Goldstein, J. R. and Klüsener, S. (2010). Culture Resurfaces - A Geographic Analysis of Fertility Decline in Prussia. *MPIDR WP* No: WP-2010-012.

Goodchild, M. F. and Janelle, D.G. (2004). Thinking Spatially in the Social Sciences. In Goodchild, M.F. and Janelle, D.G. (Eds.) *Spatially Integrated Social Science*. New York: Oxford University Press, 3–21.

Işik, O. and Pinarcioglu, M. M. (2006). Geographies of a Silent Transition: A Geographically Weighted Regression Approach to Regional Fertility Differences in Turkey. *European Journal of Population*, 22(4): 399–421.

Kertzer, D., White, M., Bernardi, L. and Gabrielli, G. (2009). Italy's Path to Very Low Fertility. The Adequacy of Economic and Second Demographic Transition Theories, *European Journal of Population*, 25(1): 89–115.

Knodel, J. and van de Walle, E. (1979). Lessons from the past: Policy implications of historical fertility studies. *Population and Development Review*, 5(2): 217–245.

Kohler, H.-P. (2000). Social interaction and fluctuations in birth rates. *Population Studies*, 54(2): 223–232.

Kohler, H.-P. (2001). *Fertility and social interaction: an economic perspective*. Oxford: Oxford University Press.

Leenders, R. T. A. J. (2002). Modeling Social Influence through Network Autocorrelation: Constructing the Weight Matrix. *Social Networks*, 24(1): 21–47.

Lesthaeghe, R. and Neels, K. (2002). From the First to the Second Demographic Transition - An Interpretation of the Spatial Continuity of Demographic Innovation in France, Belgium and Switzerland. *European Journal of Population*, 18(4): 225–260.

Lesthaeghe, R. J. (2010). Long-Term Spatial Continuities in Demographic Innovations: Insights from the Belgian Example, 1846-2000. *PSC Research Report* No. 10–695.

Livi-Bacci, M. (1977). *A History of Italian Fertility During the Last Two Centuries*. Princeton: Princeton University Press.

Loftin, C. and Ward, S. K.. (1983). A Spatial Autocorrelation Model of the Effects of Population Density on Fertility. *American Sociological Review*, 48(1): 121–128.

Michielin, F. and Mulder, C.H. (2008). Family events and the residential mobility of couples. *Environment and Planning A*, 40(11): 2770–2790.

Montgomery, M. R. and Casterline, J. B. (1993). The Diffusion of Fertility Control in Taiwan: Evidence from Pooled Cross-Section Time-Series Models. *Population Studies*, 47(3): 457–479.

Montgomery, M. R. and Casterline, J. B. (1996). Social Learning, Social Influence, and New Models of Fertility. *Population and Development Review*, 22: 151–175.

Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, 37: 17–23.

Muniz, J. O. (2009). Spatial Dependence and Heterogeneity in Ten Years of Fertility Decline in Brazil. Population Review, 48(2): 32–65.

Murphy, T. E. (2010). Old Habits Die Hard (Sometimes) Can département heterogeneity tell us something about the French fertility decline? *IGIER WP* n. 364: 1–38.

Myrskylä, M., Kohler, H.-P., and Billari, F.C. (2009). Advances in development reverse fertility declines. *Nature*, 460(6): 741–743.

Ní Bhrolcháin, M. (2011). Tempo and the TFR. *Demography*, 48(3): 841–861.

Palloni, A. (2001). Diffusion in sociological analysis. In Casterline, J. (Ed.) *Diffusion Processes and Fertility Transition, Selected Perspectives*. National Research Council, Washington, D.C.

Potter, J. E., Schmertmann, C. P., Assunção, R. M. and Cavenaghi, S. M. (2010). Mapping the Timing, Pace, and Scale of the Fertility Transition in Brazil. *Population and Development Review*, 36 (2): 283–307.

Prskawetz, A., Mamolo, M. and Engelhardt, H. (2010). On the Relation Between Fertility, Natality, and Nuptiality. *European Sociological Review*, 26(6): 675–689.

Rallu, J.-L. (1983). Permanence des Disparités Régionales de la Fécondité en Italie? *Population*, 38(1): 29–59.

Retherford, R. D. and Palmore. J. A. (1983). Diffusion processes affecting fertility regulation. In Bulatao, R. A. and Lee, R. D. (Eds.) *Determinants of Fertility in Developing Countries*. New York: Academic Press.

Ryder, N. B. (1956). Problems of trend determination during a transition in fertility. *Milbank Memorial Fund Quarterly*, 34: 5–21.

Sobotka T. (2008). The rising importance of migrants for childbearing in Europe. *Demographic Research*, 19(9): 225–248.

Terra Abrami, V. and Sorvillo, M. P. (1993). La Fécondité en Italie et dans ses Régions: Analyse par Période et par Génération. *Population*, 48(3): 735–751.

Tobler, W. (1970). Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46 (2): 234–240.

Tolnay, S. E. (1995). The spatial diffusion of fertility: a cross-sectional analysis of counties in the American South, 1940. *American Sociological Review*, 60(2): 299–308.

Voss, P. R. (2007). Demography as a Spatial Social Science. *Population Research and Policy Review*, 26(5): 457–476.

Weeks, J. R., Gadalla, M. S., Rashed, T., Stanforth, J. and Hill, A. G. (2000). Spatial variability in fertility in Menoufia, Egypt, assessed through the application of remote-sensing and GIS technologies. *Environment and Planning A*, 32(4): 695–714.

Weeks, J. R. (2004). The Role of Spatial Analysis in Demographic Research. In Goodchild, M.F. and Janelle, D.G (Eds.) *Spatially Integrated Social Science*. New York: Oxford University Press, 3–21.

Waldorf, B., Franklin R. (2002). Spatial Dimensions of the Easterlin Hypothesis: Fertility Variations in Italy. *Journal of Regional Science*, 42: 549–578.

Watkins, S. C. (1987). The fertility transition: Europe and the third world compared. *Sociological Forum*, 2(4): 645–75.

Watkins, S. C. (1990). From Local to National Communities: The Transformation of Demographic Regimes in Western Europe, 1870-1960. *Population and Development Review*, 16: 241–272.

4.9 Appendix

Figure 4.A: Results from GWRs on Adjusted TFR: local parameter estimates by quintile ranges and t-values, year 2007

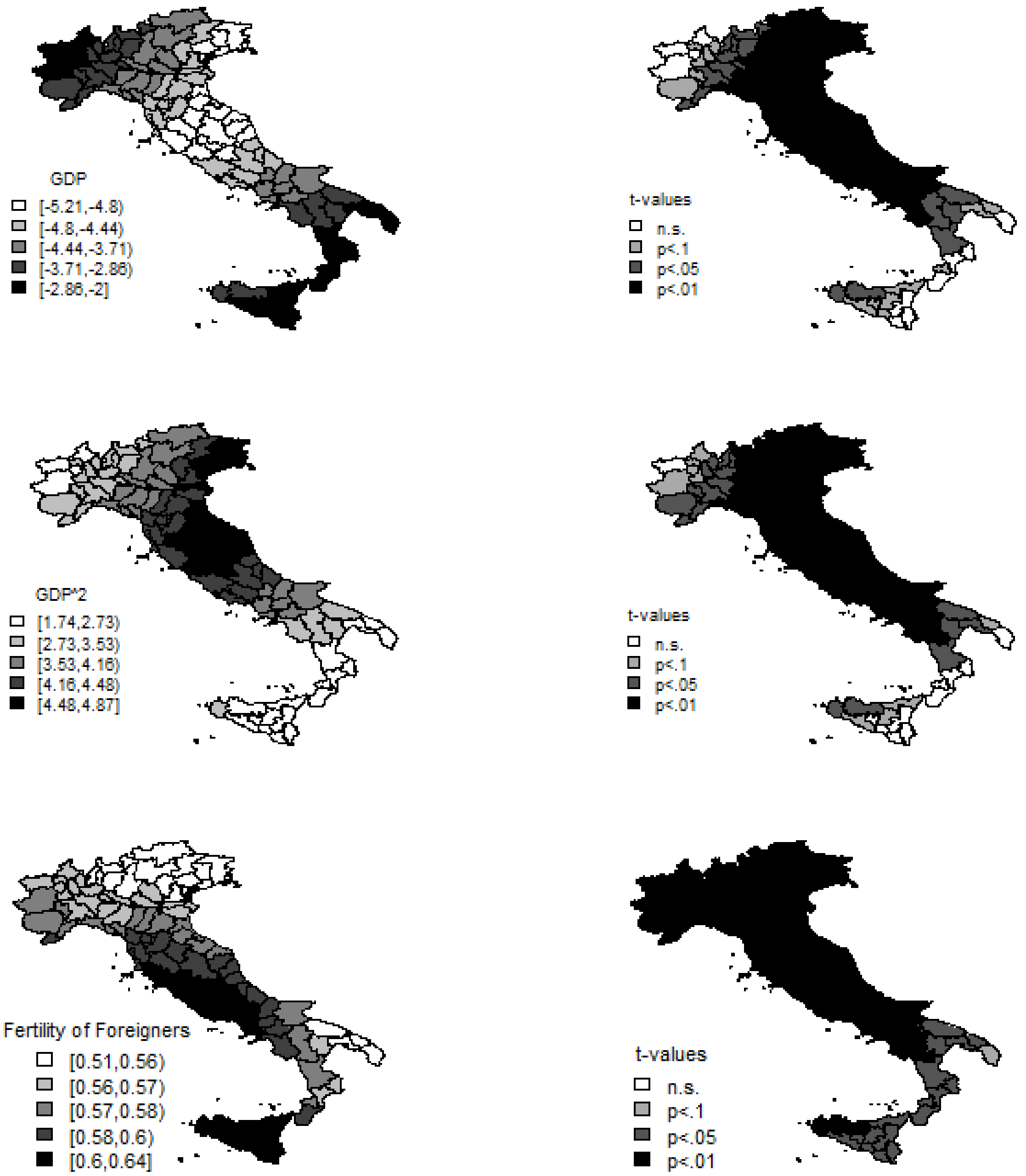


Figure 4.A: (Continued)

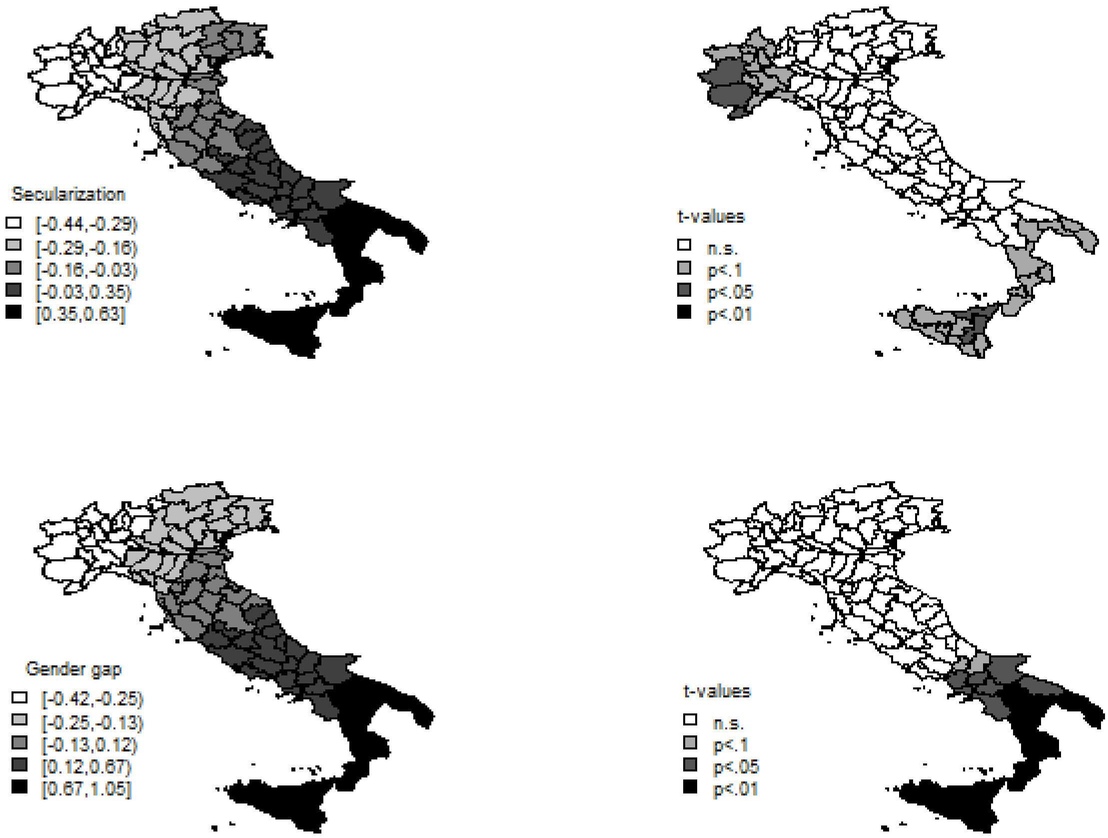
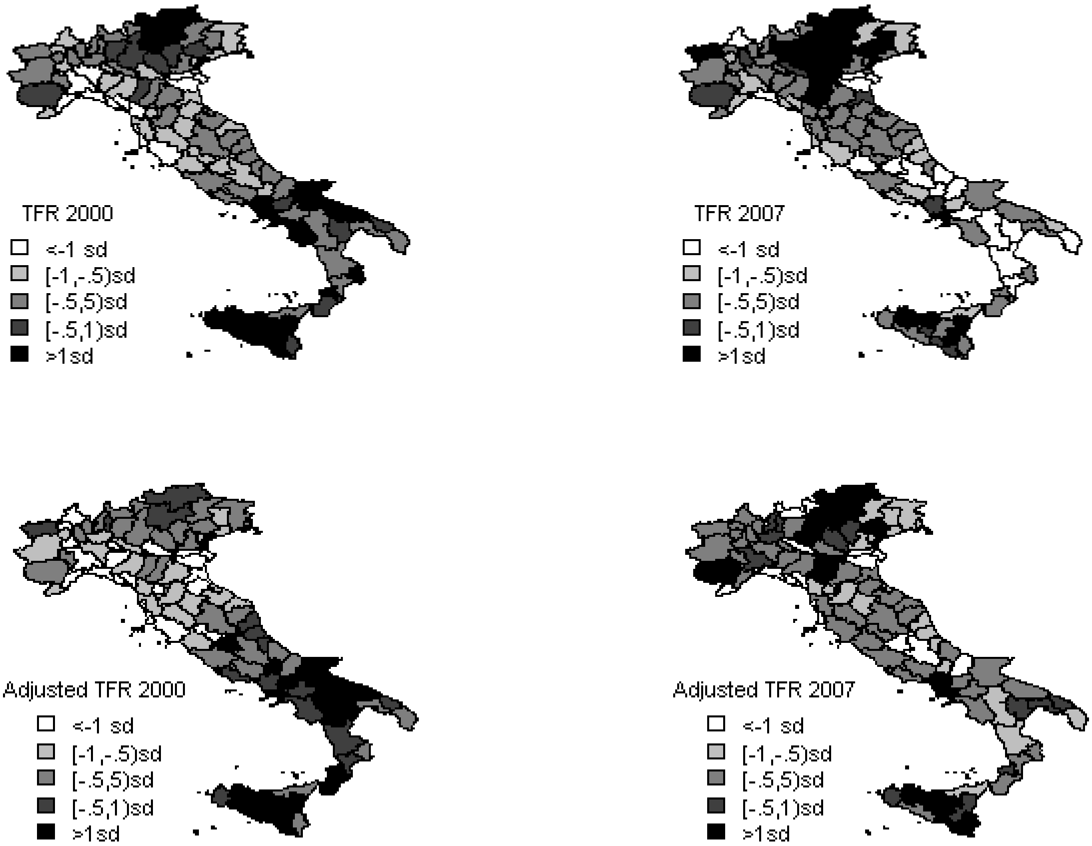


Table 4.A: Estimates of the regression of fertility (Adjusted TFR) on selected indicators, panel and spatial panel models with provincial fixed effects

	Model 1			Model 2			Model 3			Model 4		
	β		s.e.	β		s.e.	β		s.e.	β		s.e.
GDP	-2.750	***	0.642	-2.675	***	0.596	-2.432	***	0.332	-2.512	***	0.353
GDP^2	2.128	***	0.521	2.030	***	0.484	2.207	***	0.305	2.295	***	0.323
Fertility of Foreigners	0.499	***	0.124	0.467	***	0.115	0.300	***	0.066	0.276	***	0.069
Gender Gap	-0.010		0.156	-0.024		0.145	0.176	*	0.075	0.156	*	0.078
Secularization	0.058		0.100	0.060		0.093	-0.141	*	0.071	-0.148	*	0.073
ρ				0.083	.	0.048	0.163	***	0.045	0.147	**	0.046
φ							0.396	***	0.083	0.342	***	0.085
ψ										0.189	***	0.045

Note: Model 1 refers to the traditional panel model with provincial fixed effects and Model 2 to the spatial panel model with provincial fixed effects. All variables are standardized. ρ: spatial autocorrelation in the dependent variable.
p-value: *** < 0.001; ** < 0.01; * < 0.05; **.** < 0.1.

Figure 4.B: TFR and Adjusted TFR in 99 Italian provinces in year 2000 and 2007



5. Concluding Remarks

This thesis studies the role of context on demographic behaviours exploring possibilities and limitations offered by multilevel and spatial approaches. The first two chapters focus on living arrangement decisions undertaking an individual perspective and evaluating the influence of contextual effects by the means of multilevel modelling. The fourth chapter adopts a spatial approach to the study of diffusion in fertility with a macro perspective.

In the second chapter I study living arrangement decisions of young adults in Spain, exploring sub-national contextual heterogeneity at the provincial and at a smaller, more homogeneous contextual level, the municipality of residence by applying multilevel modelling. I find that both structural (labour and housing markets) and cultural (new family models) factors have an influence on living arrangement decisions. In particular, in those contexts where structural-economical factors are less favourable young people are more likely to co-reside with parents, while in places where new family models more widespread, they are more likely to live independently from parents.

In the third chapter I extend the framework developed in chapter 2 to study living arrangements of second generation immigrants in Spain. The heterogeneity of the country of origin is studied together with the effects due to the place of residence, namely the province. Results show that the cultural heritage of the country of origin plays an important role in living arrangement decisions of second generation immigrants. Cross-classified multilevel modelling offers the opportunity to disentangle the two sources of variability while evaluating the relative influence of variables measured both at the country of origin and province of residence level on the outcome of interest. This chapter demonstrates how research on immigrants can benefit from multilevel cross-classified modelling.

The fourth chapter explicitly incorporates the notion of space and contiguity into the statistical model of fertility. Following the literature on diffusion processes in fertility studies, I confirm that areas with a similar cultural heritage and economical circumstances show similar fertility, analysing fertility transitions in contemporary Italian provinces. I show that disregarding spatial dependence leads to overestimating the association between fertility and a series of other indicators. As geographically weighted regression show, associations vary spatially.

Multilevel models offer an exceptional toolkit for analysing the influence of both cultural and economic contexts on individual outcomes.

Their use, however, is not without limitations. One drawback is the identification of the correct context the individuals belong to, which generally is dependent on data availability. In the words of Teachman and Crowder (2002) “there is one guiding principle that applies when considering the impact of context. Specifically, the boundaries and characteristics of contextual units should be defined in such a way that all individuals sharing the context are subject to the same contextual conditions”. The analyses carried out in this thesis show, for example, that when structural-economic factors need to be accounted for, they should be measured at the level of territorial aggregation which affect the individual outcomes. Labour and housing markets serve as an example. While in chapter 2 I find that structural factors influence the living arrangements decisions of young adults, and in chapter 3 the same factors do not result associated with living arrangements. This is because in the former case factors are correctly measured at the municipality level, while in the latter they are measured at the provincial level. If we consider that structural variables could show some degree of heterogeneity also across different municipalities of the same province, it is understandable that the provincial level, in this case, is not appropriate to capture the impact of such variables. In other words, the correlation between indicators measured at a broad geographical level (country, macro-regions or regions) might differ when the same indicators are measured at lower geographical levels –what is referred to as the *modifiable areal unit problem* (MAUP) (Openshaw, 1984; Fotheringham et al., 1991).

Of course, results obtained from the multilevel models performed in this thesis cannot be interpreted in terms of causal effects. The issue of endogeneity arises when observed independent variables included in a regression model are correlated with other omitted unobserved variables, which in turn are associated with the dependent variable. In this circumstance, the unobserved variables are captured by the error term, which then results to be correlated with the independent variables, leading to violation of the model assumptions. In multilevel models the problem of endogeneity can arise from the omission of relevant covariates at any level of the hierarchy considered and therefore it can affect the error terms at all levels. For example, endogeneity can arise from the correlation between the error term at the second (or higher order) level and the level-one variables. This situation is referred to as “cluster-level omitted variable bias” (Rabe-Hesketh et al., 2008). If endogeneity is a concern in the analyses, the level-one variables in question can be replaced by deviations from the cluster mean (i.e. the mean of the original variable across level-two clusters). The latter are instrumental variables for the original level-one variables as they are correlated with the original variables, but not with the error term at the cluster level (level two).

Another limitation of multilevel modelling is that while relaxing the hypothesis of independence among observations belonging to the same cluster, they assume independence between observations across different clusters. In other words, individuals in the same cluster are allowed to be correlated (in the setting of chapter 2, these will be individual residing in the same municipalities, in chapter 3 immigrants residing in the same Spanish province or coming from the same country of origin), but it is not so for individuals belonging to different clusters. While this assumption is perfectly acceptable in a variety of settings, in some circumstances it might not be realistic. Indeed, from a geographic point of view "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). So, it might be the case that individuals, besides being correlated with other individuals in the cluster, are correlated also with individuals belonging to neighbouring clusters (see e.g. Chaix et al., 2005). From a sociological point of view, individual decisions are influenced by social interactions in place within their networks. These considerations directly suggest thinking spatially.

Spatial modelling is preferable in those cases where there are reasons to believe that neighbouring contexts exert some form of influence to the context in which the individual is located. The influence can work through social networks, mobility, the media or spillover effects. Neighbouring contexts defined in terms of geographical distance or adjacency, are generally assumed to have a paramount influence, but in principle, also contexts which are very far apart can be influential, depending on between-contexts connections. Spatial modelling allows units to be correlated on the basis of their “closeness”, i.e. the correlation between individuals in different clusters is allowed and depends on their distance. In principle, distance need not necessarily be geographical, but could be social distance or any other measure of dissimilarity between individuals or areas.

Obviously, the choice of how to model the demographic behaviour, or its timing, is interconnected with the theoretical framework the researcher is referring to. Demography studies populations, and the population is composed of individuals who act in interaction with others. The definition of the context of influence varies depending on the event being studied. However, contexts and spatial effects are embedded in individual decisions. Individuals shape and are shaped by the context in which they live. Disregarding contextual influences and spatial effects leads to biased estimates, hence inaccurate conclusions about the outcome being studied. Most importantly, contextual and spatial effects need to be incorporated into a comprehensive theoretical framework for demographic and sociological research. In fact, besides the methodological challenges that demographic research is facing, there are also substantive challenges which need to be dealt with.

5.1 References

Chaix, B., Merlo, J., Subramanian, S. V., Lynch, J., and Chauvin, P. (2005). Comparison of a spatial perspective with the multilevel analytic approach in neighborhood studies: The case of mental and behavioral disorders due to psychoactive substance use in Malmö, Sweden, 2001. *American Journal of Epidemiology*, 162(2): 171–182.

Fotheringham A. S. and Wong, D. W. S. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning* A, 23(7): 1025–1044 .

Openshaw, S. (1984). *The modifiable areal unit problem*. Concepts and techniques in modern geography n. 38. Norwich: GeoBooks.

Rabe-Hesketh, S. and Skrondal, A. (2008). *Multilevel and longitudinal modeling using Stata*, 2nd ed. College Station, TX: Stata Press.

Tobler, W. (1970). Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46 (2): 234–240.

Teachman, J. and Crowder, K. (2002). Multilevel models in family research: Some conceptual and methodology issues. *Journal of Marriage and the Family*, 64(2): 280-94.