



Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

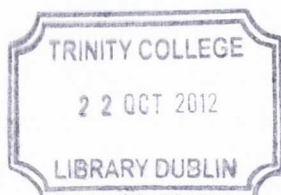
Essays on Environmental Amenities and on Taxation in Ireland

Karen Danielle Mayor

Department of Economics
Trinity College
University of Dublin

Thesis submitted to Trinity College, University of Dublin in
fulfilment of the requirements for the degree of Doctor of
Philosophy (Ph.D.)

2012



Thesis 9731

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and is entirely my own work. All research contained herein that is not entirely my own but is based on research that has been carried out jointly with others is duly acknowledged in the text wherever included.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Summary

This thesis is a collection of essays on household-level tax and environmental issues in an Irish context. The main motivation for this work was to address a number of relatively unexplored policy issues that have arisen within the country at a micro level in the last few years. Four different areas were investigated using national datasets and surveys and different econometric techniques.

Chapter 2 is an empirical investigation into the incidence of income under-reporting by self-employed households in Ireland before and during the Celtic Tiger growth years. The focus is on the level of income under-reporting and particularly which household and occupational groups are most likely to try and evade tax. In the first part of the chapter, a more general group (the self-employed) is scrutinised and in the second part more specific households groups and occupations are looked into. The results indicate that the level of under-reporting in Ireland followed a downward trend over the years under examination. The fall in the level of under-reporting is more pronounced for manual and low professional occupational groups and households living in Dublin. Farming households are found to have higher levels of under-reporting compared to other types of workers. The results support the assumption that tax evasion at a household level tends to occur in industries where cash based transactions dominate.

Chapter 3 explores the feasibility and consequences of creating a national house price valuation system in Ireland for the purpose of the introduction of a property tax. Specifically, a practical system and tax design criteria are set out and key features of the system are tested using available survey data on house characteristics and prices. Using factors based on simple hedonic regressions to estimate house prices yields mixed results due to the imposed simplicity of the model. The scale of the geographic dummies used in the analysis is important. Despite 80% of households being allocated to the correct tax band, large inequalities persist and the remaining 20% are mostly lower income groups. Size variables produce more accurate estimates than other characteristics. Policy solutions are suggested to take account of these results.

Chapter 4 examines the premium paid by homeowners in Dublin for the op-

portunity of living close to parks and green spaces. Urban sprawl and heightened construction during the property boom have led to pressure on urban green spaces. The study measures the value of these spaces to homeowners using similar regressions to the ones in Chapter 3. The results indicate that homeowners will pay a significant premium to live within certain distances of urban green spaces but also that the type of space itself is important.

Chapter 5 investigates the role of maternal and child perceptions in the use of neighbourhood play facilities by children and their effect on childhood health outcomes. Specifically, the determinants of outdoor play by children is examined using maternal and child perceptions of neighbourhood quality as explanatory variables. The results indicate that the availability of play facilities is an important determinant of play in children but maternal perceptions of overall neighbourhood quality also have a significant effect on the probability of a child playing outside. Children are less concerned about their surroundings. The results of a two-stage analysis indicate that there is a direct link between children's health and active playtime. This has important implications for urban planning and obesity policies.

Acknowledgements

I wish to express my deep gratitude to several people. Without their guidance, help and encouragement I would not have been able to finish this work on time, and my experiences as a graduate student would have been less productive and a lot less enjoyable.

The work on this thesis would not have been possible without the help and support of my supervisors. I would like to thank Dr. Seán Lyons at the ESRI for his constant motivation, patience and positive attitude. His efforts have exceeded those expected of a supervisor and I am very grateful. I would also like to thank Dr. Gaia Narciso for her motivation and encouragement and for taking me on as a student mid-way through my thesis. I am also in debt to Professor Patrick Honohan for his supervision of Chapter 2 of this thesis. I would like to thank Dr. Pedro Vicente, Dr. Francis O'Toole and Dr. Iulia Siedschlag for their helpful comments, suggestions and support along the way. Special thanks are also due to Dr. Andrew Somerville for giving me the opportunity to teach on his course.

I am so grateful to Professor Frances Ruane, Professor John FitzGerald, and Professor Richard S.J. Tol at the ESRI for motivating me to pursue the PhD and for giving me the research experience that allowed me to complete it. Working as an RA at the Institute was a fantastic opportunity as a new graduate and I am really thankful. The financial support of the Government of Ireland Research Scholarship in the Humanities and Social Sciences through Research Assistance on the "Turning Globalization to National Advantage: Economic Policy Lessons from Ireland's Experience" project as well as a Research Scholarship from the Economic and Social Research Institute were a great source of help throughout my research. I would also like to thank the Central Statistics Office and the Office of the Minister for Children and Youth Affairs for kindly providing access to the datasets used in this thesis.

To all my friends and staff in the Economics Department at Trinity College and in particular Colette and Trish in the Economics Department and Colette in the IIIS, thank you for all your help over the last three years.

The PhD students in Trinity were a great source of support and coffee breaks over the last three years and kept me sane through this process! Finally, I owe so much to my family and Donncha who believed in me and for giving me the support to continue, thank you.

Contents

1	Introduction to the Thesis	11
1.1	Context and background	11
1.2	Overview of the Thesis	12
1	Taxation in Ireland	17
2	Tax Evasion in Ireland before and during the Celtic Tiger	19
2.1	Introduction and Motivation	20
2.2	Literature Review	22
2.2.1	Definitions of the black economy	22
2.2.2	Techniques to measure the black economy	23
2.2.3	Estimates of the black economy and tax evasion in the literature	26
2.2.4	Measures for Ireland	29
2.3	Methodology	30
2.4	Data	33
2.4.1	Sample selection	34
2.4.2	Specification of variables used	36
2.5	Results and discussion	39
2.5.1	Non-parametric estimates	39
2.5.2	Estimates of under-reporting for the self-employed	40
2.5.3	Estimates of under-reporting for other occupational groups	43
2.6	Conclusions	44
2.7	Tables	47
2.8	Figures	57
2.9	Appendix	61
2.9.1	Formula derivations	61
2.9.2	Summary statistics explanatory variables, by year	64
2.9.3	Summary statistics instrumental variables, by year	68
2.9.4	Farming regression results	72

2.9.5	Older CES regression results	76
2.9.6	Manual CES regression results	87
2.9.7	Dublin CES regression results	92
2.9.8	Low Professional CES regression results	96
2.9.9	Construction CES regression results	101

3 **Designing a Property Tax without Property Values: Analysis in the case of Ireland** **103**

3.1	Introduction	104
3.2	The Irish housing market and fiscal system	106
3.3	Commission on Taxation recommendations and the old property tax . .	107
3.3.1	Commission on Taxation	108
3.3.2	Historical property tax in Ireland	113
3.4	Literature	114
3.4.1	Equity and Fairness in Taxation	114
3.4.2	Self-assessment and the tax system	115
3.4.3	Thin markets and property tax assessment	117
3.4.4	The case of Northern Ireland	120
3.5	Methodology	121
3.5.1	Predicting house prices	122
3.5.2	Evaluating the house price assessments	123
3.5.3	The assessment to sales price ratio	124
3.6	Data	125
3.6.1	The survey, sample and variables	125
3.6.2	The house value	128
3.7	Results	129
3.7.1	The importance of geographic dummies in hedonic regressions .	129
3.7.2	Absolute Value Errors	130
3.7.3	A/S ratios and assessment errors	131
3.7.4	Horizontal inequity across groups	132
3.7.5	Predicting the tax band and difficulties in self-assessment	133
3.8	Summary and Conclusions	134
3.9	Tables	138
3.10	Figures	147
3.11	Appendix	158

II Environmental amenities in Ireland 161

4 A Hedonic Analysis of the Value of Parks and Green Spaces in the Dublin Area 163

4.1	Introduction	164
4.2	Literature	165
4.3	Model	168
4.4	Data	170
4.4.1	Housing data	170
4.4.2	Control variables	170
4.4.3	Green space data	171
4.5	Results	172
4.5.1	Transport, neighbourhood and house characteristics	172
4.5.2	Density of green spaces	173
4.5.3	Distance to named public access parks in Dublin	175
4.6	Discussion and Conclusions	176
4.7	Tables	177
4.8	Figures	182
4.9	Appendix	184

5 Lazy children or helicopter mothers? Evidence from Ireland on outdoor play, overweight in children and perceived neighbourhood characteristics 189

5.1	Introduction	190
5.2	Literature	192
5.2.1	Physical activity, neighbourhood and green space availability	192
5.2.2	Neighbourhood perceptions	194
5.2.3	Obesity in Ireland	195
5.3	Model	195
5.3.1	Principle Component Analysis	196
5.3.2	Statistical analysis	197
5.4	Data	198
5.4.1	Survey and sample restrictions	199
5.4.2	Summary statistics	200

5.5	Results	204
5.5.1	The determinants of play	204
5.5.2	Neighbourhood perception index	206
5.5.3	Children’s health	207
5.6	Discussion and Conclusions	211
5.7	Tables	214
5.8	Figures	224
5.9	Appendix	228
6	Conclusions	237
6.1	Thesis conclusions	237
6.2	Future Research Directions	240
	Bibliography	242

List of Tables

2.1	Food and income across years, full sample	47
2.2	Food and income across years, Self-Employed	47
2.3	Food and income across years, Employees	48
2.4	Food and income across years, Farmers	48
2.5	Results 1987	49
2.6	Results 1994	50
2.7	Results 1999	51
2.8	Results 2004	52
2.9	Under-reporting bounds SE HH (1987-2004)	53
2.10	Under-reporting bounds SE Dublin based CES (1987-2004)	53
2.11	Under-reporting bounds SE Older CES (1987-2004)	54
2.12	Under-reporting bounds SE Low Professional CES (1987-2004)	54
2.13	Under-reporting bounds SE Manual CES (1987-2004)	55
2.14	Under-reporting bounds Farm HH (1987-2004)	55
2.15	Under-reporting bounds Farm HH inc. own produce (1987-2004)	56
2.16	Under-reporting bounds, construction industry (2004)	56
2.17	Summary Statistics Explanatory Variables 1987	64
2.18	Summary Statistics Explanatory Variables 1994	65
2.19	Summary Statistics Explanatory Variables 1999	66
2.20	Summary Statistics Explanatory Variables 2004	67
2.21	Summary Statistics Instrumental Variables 1987	68
2.22	Summary Statistics Instrumental Variables 1994	69
2.23	Summary Statistics Instrumental Variables 1999	70
2.24	Summary Statistics Instrumental Variables 2004	71
2.25	Results 1987, Farmers	73
2.26	Results 1994, Farmers	74
2.27	Results 1999, Farmers	75
2.28	Results 1987, Farmers including own produce	77
2.29	Results 1994, Farmers including own produce	78
2.30	Results 1999, Farmers including own produce	79
2.31	First stage regressions - main sample	80
2.32	First stage regressions - farmers	81

2.33	First stage regressions - farmers own produce	82
2.34	Results 1987 Older CES subsample	83
2.35	Results 1994 Older CES subsample	84
2.36	Results 1999 Older CES subsample	85
2.37	Results 2004 Older CES subsample	86
2.38	Results 1987 Manual subsample	88
2.39	Results 1994 Manual subsample	89
2.40	Results 1999 Manual subsample	90
2.41	Results 2004 Manual subsample	91
2.42	Results 1987 Dublin subsample	92
2.43	Results 1994 Dublin subsample	93
2.44	Results 1999 Dublin subsample	94
2.45	Results 2004 Dublin subsample	95
2.46	Results 1987 Low professional subsample	97
2.47	Results 1994 Low professional subsample	98
2.48	Results 1999 Low professional subsample	99
2.49	Results 2004 Low professional subsample	100
2.50	Results 2004, Construction subsample	102
3.1	Recommendations on the Taxation of Property by the Commission on Taxation	138
3.2	Valuation bands as proposed by the Commission on Taxation	138
3.3	Summary Statistics for the main variables in the analysis	139
3.4	House values by region, full sample	139
3.5	Regression results: house value regressed on square metres, with (bot- tom panel) and without (top panel) geographical dummies and interac- tions	140
3.6	ABV Error statistics depending on which individual explanatory vari- able is used	141
3.7	ABV Error statistics depending on which combinations of explanatory variable are used	141
3.8	A/S ratio summary statistics for entire sample depending on explana- tory variables used in the hedonic regression	141
3.9	Assessment ratios by region, using just square metres (top), just number of rooms (middle) and using all variables (bottom)	142
3.10	Assessment ratios by income quintile, using just square metres (top), just number of rooms (middle) and using all variables (bottom)	143

3.11	Assessment ratios by age of reference person, using just square metres (top), just number of rooms (middle) and using all variables (bottom) . .	143
3.12	Assessment ratios by employment status of reference person, using just square metres (top), just number of rooms (middle) and using all variables (bottom)	144
3.13	Assessment ratios by employment class of reference person, using just square metres (top), just number of rooms (middle) and using all variables (bottom)	145
3.14	Assessment ratios by rural versus urban, using just square metres (top), just number of rooms (middle) and using all variables (bottom)	145
3.15	Proportion of households being put in the incorrect estimation band by explanatory variable used	146
3.16	Full regression results	159
4.1	Transport, Environmental and Structural Variables	178
4.2	Buffer zones and their related area size	179
4.3	Results - park and green space density regressions	179
4.4	Sensitivity of Model 1 results to locational dummies used	180
4.5	Results - Dummy park is within 2500m of property	181
4.6	Results - Logged distance to individual parks	185
4.7	Results Locality Dummies	186
4.8	Results Locality Dummies	187
5.1	Descriptive statistics for neighbourhood perceptions - mother	214
5.2	Descriptive statistics for neighbourhood perceptions - child	214
5.3	Summary statistics of health variables	214
5.4	International cut-off points for BMI for overweight and obesity by sex .	215
5.5	BMI at or above overweight and obesity cut-offs by sex	215
5.6	BMI at or above overweight and obesity by sex	215
5.7	Descriptive statistics for genetics	216
5.8	Descriptive statistics for child's activity variables	216
5.9	Description of household location	216
5.10	Determinants of play	217
5.11	Determinants of play - including activities	218
5.12	Determinants of play including SAPS	219
5.13	Including principle component	220
5.14	BMI basic regressions	221
5.15	BMI regressions full	222

5.16	Two stage - including outdoor play	223
5.17	List of questionnaires in the Child Cohort Survey, Wave 1 (GUI, 2010) . .	228
5.18	Cronbach's alpha results and eigenvalues	229
5.19	Descriptive statistics for household level variables	229
5.20	Play analysis Urban/Rural samples	230
5.21	PC Analysis Urban/Rural Samples	231
5.22	Play analysis - Gender subsamples	232
5.23	PC analysis - gender subsamples	233
5.24	BMI analysis - Gender subsamples	234
5.25	Diet regression	235
5.26	First stage regressions	235

List of Figures

2.1	Non-parametric SE, EE, Farm plots	58
2.2	Non-parametric SE, EE, Farm smoothed	59
2.3	Lost income due to tax evasion as a proportion of total taxable income .	60
3.1	Average national house prices (1975-2008, nominal terms) in Ireland for new and second hand houses. Source: Department of the Environment, Heritage and Local Government.	147
3.2	Quarterly house price growth (1996-2010) in Ireland for national and Dublin prices. Source: ESRI/PTSB House price index, Base Q3 2003	148
3.3	Net receipts of stamp duties (2002-2007) in Ireland (€millions). Source: CSO <i>Construction and Housing in Ireland Report</i> , 2008 Edition. Based on figures in Table 5.4.	148
3.4	Distribution of dwelling area size (in square metres) for respondents who knew the size of their home with a normal distribution plot.	149
3.5	Plot of interviewer estimate minus owner estimate of house value. The top panel is the original sample. The bottom panel is the reduced sample.	150
3.6	Histogram plots of house value by region (from left to right, top to bottom): Dublin, Border, Mideast, Midlands.	151
3.7	Histogram plots of house value by region (from left to right, top to bottom): Midwest, Southeast, Southwest, West.	152
3.8	ABV errors from a regression using square metre dummies without geographical dummies (top panel) and with geographical dummies (bottom panel).	153
3.9	ABV errors from regressions using dwelling type, number of rooms, number of bedrooms and year built as individual explanatory variables with local authority dummy variables.	154
3.10	Assessment ratios against property price calculated using just size in square metres (top), just number of rooms (middle) and all five explanatory variables (bottom).	155
3.11	Summary of A/S ratios by household characteristics.	156

3.12	Mean assessment error (for self-assessment and for hedonic house price model results using just square metres) as a proportion of the property value, by income band.	157
4.1	Map of Dublin with location of sample houses used in the analysis (red dots). Source: Map by authors. Data from Sherry FitzGerald.	182
4.2	Map of Dublin with green spaces used in analysis (green areas)	183
5.1	Summary structure of instruments used in Child Cohort Survey, Wave 1 Source: GUI, 2010	224
5.2	Histograms of BMIs for full sample (top), just girls (middle) and just boys (bottom)	225
5.3	BMI percentiles for age for boys aged 2 to 20. Source: Krebs, N. F. et al. <i>Pediatrics</i> 2007;120:S193-S228	226
5.4	ScreepLOTS	227

Chapter 1

Introduction to the Thesis

1.1 Context and background

The Republic of Ireland underwent massive changes between the beginning of the 1980s and the late 2000s. These changes were not only economic and financial but also cultural and social. The first and main aspect of change in Ireland over these three decades was an increase in incomes and a convergence of living standards with other Western countries. Indeed, in just a few years Ireland went from a deep recession to a shining example of economic growth and became known around the world as the “Celtic Tiger”. Gross Domestic Product grew by 282% between 1980 and 2007 and average incomes by 200% in the same period. These prosperous years have now again ended and Ireland is facing another deep recession and an unprecedented financial and banking crisis.

Ireland’s trade with Europe and the world as well as its booming construction industry resulted in the country becoming more “globalized”. However, this globalization led to changes in the Irish economy as well as Irish society. On the one hand, economic growth spurred a reduction in income taxes and a focus on so called transaction taxes in the booming property sector. On the other hand, incoming migration and increased urbanisation saw more and more households living in city centres. Diets also changed with increasing incomes and a larger availability of international produce. Recent international and national economic events have resulted in a massive deficit and culminated in the EU-IMF Bailout for Ireland in 2010.

Extensive change has occurred and policy issues have been raised in the aftermath of the boom. All this change has had profound effects on Ireland’s current situation and which policy options are open to future governments. Some of these changes are already visible (for instance the gaping hole in property tax takes) while others will only become apparent in years to come (the effect of changes in living conditions on the long-term health of Irish children). This thesis investigates a number of different,

albeit related, issues of taxation, environment and health in Ireland to see what policy options are available to the country and what possible mistakes can be avoided in the future.

The location of additional tax revenue has become a highly topical issue in Irish policy-making and economics in the last two years. Some issues on this subject have yet to be investigated and this thesis sets out to specifically address a number of topics on tax location and new taxation that have been in the media since the beginning of the recession. The second half of this thesis deals with other post-Celtic Tiger issues of the urban environment and obesity and health. The latter is a newly emerging public health issue and a long term economic issue not only in Ireland but in most of the western world. The specific subjects of the different chapters are described below.

1.2 Overview of the Thesis

Part 1 of this thesis consists of two chapters that were directly motivated by the current fiscal problems faced by the Irish State. There has been a lot of debate since the economic downturn on the fall in fiscal revenues and how this deficit should be managed. Chapter 2 examines a potential source of lost revenue that should be accruing to the State from households through tax evasion and Chapter 3 looks at the introduction of a new source of revenue in the form of a property tax.

Chapter 2 deals with the issue of tax evasion. Tax evasion is a part of the black economy which is defined by the OECD (2010:11) as "economic activity which is not fully compliant with laws and regulations, or, in some cases it can refer to economic activity which is itself illegal such as smuggling or selling illegal drugs". Tax evasion is then a subset of the black economy and Chapter 2 investigates a subset of tax evasion; evasion by self-employed households in Ireland. Obviously, tax evasion of businesses and of large international corporations also contributes to losses in tax revenues, however the focus of this paper is on households and a more micro level of research. Research on this area in Ireland is rare and mostly dates back to the 1980s coinciding with the last recession, a first indication that tax evasion may be related to economic growth. Research in this area is also scarce due to the difficulties involved in estimating illegal activity and most of the papers relating to Ireland are macroeconomic analyses. Chapter 2 uses a technique developed by Pissarides and Weber (1989) to estimate under-reporting bounds for the self-employed in four years, 1987, 1994, 1999 and 2004 corresponding to very different periods in Ireland's economic growth. The

results indicate that the level of under-reporting by self-employed households fell over the course of the period under examination but lost income from under-reporting as a proportion of total taxable income increased. Moreover, the level of under-reporting for occupational groups such as farming households was also found to be high.

Chapter 3 is also motivated by current fiscal problems and in particular the Irish government's wish to restructure revenues emanating from the property sector. Ireland's Celtic Tiger years were fuelled by a property boom that saw prices increase by approximately 313% over the course of 11 years. Property related tax revenues during this boom were extremely high due to revenues from housing stamp duty. Stamp duty is a transactions based tax which is paid on completion of a house sale. Consequently, when the property bubble burst, tax revenues directly from the property sector crashed. Consideration is now being given to introducing a tax on the *ownership* of property which would be a more stable source of fiscal revenue whilst also helping to dampen property speculation in the future. However, in order for any type of new tax to be accepted by the public, it will need to be fair. Consequently introducing a flat rate tax is probably not a viable option and a value based tax will be required. Unfortunately, no database of housing prices exists in Ireland. Chapter 3 is co-authored with Dr. Seán Lyons and Professor Richard S.J. Tol at the Economic and Social Research Institute in Dublin. It uses simple hedonic house price regressions to estimate how the impact of using very little information on house characteristics in order to estimate house prices for the purpose of property taxation. Based on a small number of dwelling characteristics such as house size and number of rooms as well as geographical location variables it is possible to estimate house prices to within certain error bounds. However, many inequalities arise from such a simple system. Dwellings that are mis-estimated tend to be lower income household dwellings. These house prices tend to be over-estimated leading to an estimation system which is regressive. However, these problems can be dealt with through the introduction of income and price exemptions.

Part 2 of this thesis looks at a different aspect of changes in Irish households and has an environmental focus. Both chapters are motivated by the construction boom, urban sprawl and the potential lack of enforcement of regulation regarding green spaces and outdoor play facilities in new developments. One of the particularities of the Irish property boom was that with rising property prices any piece of land, no matter how small became worth developing. This resulted in a reduction in green spaces and an increasing number of dwellings built in homeowners' gardens. As a consequence proximity to green spaces became a luxury as did access to outdoor play spaces for children.

Chapter 4 highlights the value homeowners put on living close to green spaces and parks. There is increasing evidence that people value access to green spaces, particularly in urban areas. However the value of an environmental amenity such as a park or green space is difficult to measure and different techniques exist that try to put a value on an amenity which does not have a price. These methods can be split into two types: stated preference methods (where people are directly asked how much they would be willing to pay for a specific type of amenity) and revealed preference methods (where other market prices are used to infer the value of the environmental amenity). Chapter 4 is co-authored with Dr. Seán Lyons, Dr. David Duffy and Professor Richard S.J. Tol at the Economic and Social Research Institute in Dublin. It uses hedonic house price regressions, a type of revealed preference method, to estimate the premium put on living close to green spaces and parks by homeowners. This is the same methodology as the one used in Chapter 3, however in this case as many dwelling related and neighbourhood related characteristics as possible are included in the regression. Using data on house price sales in Dublin presents a unique opportunity to look at the value of parks and green spaces in an urban centre and capital city. Measures for the premium paid by homeowners for living in proximity of a park but also relating to the level of greenness of the neighbourhood were estimated. The results show that homeowners value living within walking distance of a park but value living very close to green spaces more.

Chapter 5 returns to the motivation of Chapter 4; outdoor spaces are valued by homeowners, but what are their benefits? Ireland's economic growth over the Celtic Tiger years had the effect of increasing household incomes but at the same time put Ireland on the same health path as other industrialised countries with more unhealthy diets and growing levels of adult and childhood obesity. This chapter looks at one of the potential benefits of outdoor spaces, namely health outcomes. Exercise and physical activity have been proven to have a positive effect on health and on bodyweight. In children, outdoor play is a way of promoting physical activity, however the level of outdoor play by children is likely to be related to the neighbourhood they live in. Chapter 5 first looks at the effect of perceptions of neighbourhood characteristics as well as the availability of facilities on outdoor activity in children. In a second stage, outdoor activity as well as genetic, diet and lifestyle variables are used to determine overweight and obesity outcomes in children. Neighbourhood perceptions of mothers have been used to determine outdoor play in previous studies however the dataset used here, the Longitudinal Survey of Children in Ireland, offers the unique opportunity of also including children's neighbourhood perceptions in the analysis. This study

aims to determine if it is just the availability of outdoor resources or also the maternal and child perceptions of the neighbourhood that affect outdoor play and if play relates to bodyweight outcomes in children. The results show that maternal perceptions of neighbourhood quality are a highly significant determinant of outdoor play alongside the availability of facilities. Children's perceptions of neighbourhood quality do not have an association with their outdoor play, however, non-infrastructural characteristics such as the presence of friends and "good play" spaces matter more. Outdoor play itself has a direct impact on children's health outcomes. Along with genetics, lifestyle factors and income, outdoor play positively influences a child's BMI. These results hold when within area effects are controlled for.

This thesis concludes with a final section outlining the broad results from each of the chapters described above.

Part I

Taxation in Ireland

Chapter 2

Tax Evasion in Ireland before and during the Celtic Tiger

This chapter was presented at the Trinity College Economics Department PhD Seminar, Dublin (2009), the Irish Society of New Economists 6th Annual Conference, Limerick (2009) and the Spring Meeting of Young Economists Annual Conference, Groningen, the Netherlands (2011). Special thanks is owed to the Central Statistics Office for supplying the data, Dr. John Cullinan for his input and to Professor Patrick Honohan for his comments on earlier drafts of the paper.

2.1 Introduction and Motivation

Rising unemployment, falls in consumer spending and investment as well as a significant reduction in the number of housing transactions have resulted in significant tax shortfalls in Ireland in 2009. Amid growing financial and economic uncertainty a new interest in the identification and location of tax revenue has arisen. Since the beginning of 2009, the Irish Times reported that "one third of self-assessed [households] under declared their income" and that the revenue commissioners had an increasing interest "in the cash trade from tradesmen to barristers". Locating those who are purposefully avoiding the tax net is a politically popular way of increasing tax revenues without an outcry from the general public and was one of the measures mentioned by the Minister for Finance during the 2009 Budget.

Tax evasion is a serious problem. Not only does it result in tax shortfalls at a national level but it can also have redistributive effects as households that do pay tax end up having to pay more to offset the shortfall. If large scale tax evasion occurs it can end up having a negative impact on the provision of public services. Consequently the detection of those most likely to under-report their incomes is crucial. It can however be costly. Indeed, tax auditing is resource and time consuming and accurate targeting of those most likely to conceal is beneficial.

A category of households usually considered to have the most scope to under-declare their incomes are self-employed. This statement is not meant to imply that the self-employed households are less honest than the rest of the population. Rather, as they self-declare their incomes, they have more *scope* for concealment than employees. In Ireland, the national tax system is such that employees pay their taxes at source and hence it is difficult to under-report that income unless they collude with their employer. On the other hand, the self-employed self-declare their earnings to the revenue authorities which gives them more scope to not disclose their full income.

In Ireland, the detection of income under-reporting is done by the Revenue Commissioners under the audit and investigation policy unit. In 1998, they screened over 55,000 income tax returns for discrepancies and audited approximately 4,400. Audits are conducted randomly. Two thirds of these random audits produced no yield and the average return was £1,000. However, under the targeted audits three quarters of all audits produced a yield and these were on average significantly higher at £18,700 in 1998.¹ Total audit activities yielded £160 million in 1998 (this is including business

¹This is the latest year of data published by the Tax Strategy Group.

under-reporting, personal income tax, property tax, and benefits fraud).

The self-employed tend to be the focus of audit activities. However, as stated by the Department of Finance (2000:23): “The construction industry is the single biggest *sectoral focus* of Revenue’s counter evasion programmes”. Indeed, cash based industries also allow scope for non-compliance and under-reporting as earnings in these sectors are more difficult to track. Consequently the present study will also estimate under-reporting for households working in the construction sector.

Most of the previous research on tax evasion in Ireland dates back to the 1980s and uses macroeconomic techniques. This study aims to contribute to research in the area of income under-reporting in Ireland by using a micro-economic, survey based method to identify those most likely to under-report their incomes. A number of factors can change the size of the black economy in a country, in particular the level of economic growth in the country, the size of the cash based economy in the country as well as the size of the tax burden. During the period under examination, Ireland went from a situation of poor economic performance (in the 1980s) to extremely high growth (Celtic Tiger years). In parallel, the income tax burden fell from 1997 onwards and Ireland in 2007 “had the lowest effective tax rate on labour in Western Europe” (Christensen, 2010:2). This study will not allow us to identify which of these effects impacted on under-reporting. It will however first attempt to estimate the level of under-reporting by the self-employed in Ireland over those years. Second, it will aim to identify whether there has been a trend in the change in the level of under-reporting between the late 1980s and 2005 by attempting to highlight who is most involved in shadow economy activities, by age, occupation and geographical location.

The following section will present an overview of the literature in the area. In particular, the aspect of the black economy analysed in this study will be discussed. A review of past studies using various methods in countries worldwide, and in Ireland in particular, will be presented along with their results. Section 2.3 will detail the model and methodology. Section 2.4 will discuss the datasets used in this study and the different variables used in the analysis. Section 2.5 will present and discuss the results of the analysis and Section 2.6 will conclude.

2.2 Literature Review

The Irish economy has changed drastically in the last three decades. Not only has the country recorded an unprecedented level of growth but the structure of the economy has also significantly changed. Ireland moved from being an agricultural and manufacturing economy in the 1980s to being an economy spurred by construction, the financial sector as well as a large pharmaceutical sector in the 2000s. In this section the literature surrounding the estimation of the black economy in general and tax evasion in particular are outlined. First, the definition of the black economy chosen for the purpose of this analysis is underlined. Then, the different techniques available to estimate the black economy are presented and in Sections 2.2.3 and 2.2.4 various estimates for different countries around the world and for Ireland specifically are highlighted.

2.2.1 Definitions of the black economy

There are numerous definitions of the “black economy”, just as there are various names used to describe that sector of the economy. A few examples are the underground economy, the shadow economy, the unobserved economy, the hidden economy, the informal economy or the parallel economy. The OECD (2010:13) states that “the definition of the informal economy is not settled in the economics literature”. All these expressions have subtle differences in meaning but tend to be used to describe the same thing. The general definition of the black economy as defined by the OECD is as follows: “economic activity which is not fully compliant with laws and regulations (or, in some cases, it can refer to economic activity which is itself illegal such as smuggling or selling illegal drugs).”

The component of the black economy investigated in this paper is that involving tax evasion, in particular that of households and the self-employed. The exact area of illegal economic activity analysed in this paper is defined as: the under-reporting of income by self-employed households in the economy. As stated by Fagan (1993:7): “Black economy activities are likely to have their main impact on estimated self-employment income and to a lesser extent corporate profits and undeclared wage income”. The self-employed can conceal income from the tax authorities when the probability of getting caught is low. There exists a substantial literature on the measurement of the black economy both in Ireland and in the rest of the world. Nevertheless divergence remains on the different estimation procedures used. The next section will present the different types of estimation procedures available to measure the size of

the black economy and its sub-component tax evasion as well as their advantages and limitations.

2.2.2 Techniques to measure the black economy

Illegal activities are difficult to measure as by their nature the individuals that are active in these areas are purposefully concealing their activities from the authorities. Hence there are no official, measured figures of the black economy in general or tax evasion in particular available to researchers and all figures relating to these areas are estimates. Nevertheless, various techniques have been devised using available data and sources to infer their level in a country without having to resort to direct surveys. These can be broadly categorised into top-down or macro-economic approaches and bottom-up or microeconomic approaches some of which will be presented in this section. Some estimate the overall black economy (including drug smuggling, prostitution etc.), others estimates only refer to overall tax evasion (corporate and individual) and others to the specific category examined in this paper, tax evasion by self-employed households. As the former encompass the latter, all are presented in this section for completeness.

Top-down or Macroeconomic Approaches

Most of the literature on the estimation of the black economy concentrates on macroeconomic relationships. There are three main techniques of this type, which use macroeconomic or aggregate data to infer the size of the black economy in a country. These are outlined below.

The monetary, currency demand, or currency ratio approach

The monetary or currency demand approach is the most popular technique for estimating the size of the black economy in a country. The basic assumption for the technique is that people who are conducting hidden or illegal transactions will primarily use cash. An increase in the size of the black economy can then be seen through an increase in the demand for currency and excess currency holdings. The method was developed by Cagan (1958) who calculated the correlation between currency demand and tax pressure in the US. The method was then formalised by Tanzi (1999). Schneider (2000) re-emphasised the methodology which tends to encompass the broadest definition of the black economy available and does not focus solely on tax evasion.²

²Similar studies have been conducted on Spanish and Canadian data (Gadea and Serrano-Sanz, 2002,

MIMIC/DYMIMIC approach

The MIMIC approach is a multiple indicator multiple cause approach (and a variant of the linear independent structural relationships model) and the DYMIMIC is a dynamic version of the same model. The basic model was developed by Zellner (1970) and extended by Frey and Weck-Hannemann (1984) who used it on a cross-section of 17 OECD countries over the 1960s and 1970s. The basis of this model is that one should look at the multiple determinants of the black economy as well as different indicators of its size. Generally the causes of the black economy included in these types of models are the level of taxation, increased regulation, and tax morality estimates. Indicators are for instance the development of monetary indicators, the labour market, hours worked and the production market. The model is a structural equation approach which takes account of the causes of illegal economic activities and indicators of the size of these activities simultaneously.³

This type of model is considered more thorough than most models because as its name indicates it considers multiple causes of the existence of the black economy. Even though it has been widely used (see Section 2.2.3) it is not without its criticisms. The outputs from the model are relative estimates of the size of the black economy. It only allows us to see the evolution of the size of the black economy and to rank it. It is still necessary to use another technique to get an exogenous estimation of its size and a baseline figure. Due to the nature of the inputs into the model it does however tend to measure a more narrow definition of the black economy, which is closer to tax evasion than the previous method.

National accounts method

The final top-down approach to measuring the size of the black economy is based on discrepancies between income and expenditure using the national accounts. In national accounting, the income measure of GNP should be equal to the expenditure measure of GNP. This technique was used by Fagan (1997) for Ireland. This type of model has a more narrow focus than the previous models discussed. The latter estimated the whole of the black economy and the former looks specifically at tax evasion by business and individuals. Of the macro-economic approaches it is the closest method to estimating tax evasion as opposed to the overall black economy.

Hill and Kabir, 2000).

³Studies of this kind have been conducted on Portugal (Dell'Anno, 2007), Germany (Pickhardt and Sarda Pons, 2006), Spain (Alanon and Gomez-Antonio, 2005), Italy (Dell'Anno, 2003), France and Greece (Dell'Anno et al., 2007) and a range of countries worldwide (Schneider, 2007).

The types of techniques outlined above, which are based on aggregate figures, do not take account of differences in individual behaviour. A growing literature based on micro-econometric analyses of the black economy has developed. The main techniques of this type are presented in this section.

Bottom-up or Microeconomic Approaches

Tax auditing and Surveys

Tax auditing involves conducting tax audits and aggregating results according to certain criteria and is not an estimation approach *per se*. As it is extremely costly and resource and time consuming audits need to be targeted at those most likely to evade.⁴ Micro-economic studies such as the present paper and those below can help in defining the most probable candidates.

Surveys are another costly way of measuring the size of the black economy. This is a type of direct method which provides very detailed individual information on respondents but due to its cost it tends to be only cross-sectional. Surveys asking people about their involvement in the black economy have been conducted in Norway, Italy, Switzerland, Belgium and Hungary. This type of estimation technique although direct, is however rife with problems. The biggest issue is non-response to the survey as well as interviewees lying in their responses. Despite being assured of the anonymity of their responses, respondents (due to the nature of their transactions and hidden labour) are often unwilling to answer truthfully. It is however possible to get some general guidelines on the nature of hidden labour income from these types of surveys which can then guide any further analysis.

Isachsen and Stroem (1985) analyse data collected using a postal survey on 1,000 respondents in Norway. Although they consider their estimates of the size of tax evasion to be imprecise they do provide some general guidelines regarding the nature and characteristics of those engaged in hidden labour activities that are useful to analysts. For instance they find that men are more active in hidden labour activities than women and that participation decreases with age. Participation in these activities first increases with education and then decreases from a working perspective. Finally, they also find that they are equally likely to engage in hidden labour activities in the cities as in the countryside. These hypotheses will be useful in the analysis part of this paper.

⁴See Feinstein (1999) for a discussion of this issue.

Discrepancies between income and expenditure

This indirect approach uses estimates of the relationship between food consumption and income for self-employed and wage earners and computes "true" income levels for the self-employed, consequently estimating the degree of non-compliance among that group. Individual survey data are used to look for systematic differences in behaviour that may be explicable in terms of concealed income. The technique, which uses household expenditure micro-survey data, was developed by Pissarides and Weber (1989). This is the method used in this analysis and it will be examined and explained further in Section 2.3. The reasons for choosing this method are twofold. First, it allows the estimation of a more focused aspect of the black economy than the macroeconomic methods. The micro-economic approach attempts to estimate the scope of under-reporting for self-employed households. All of the methods described above have their limits as they are attempting to put a figure on an area of the economy that cannot be validated numerically. As tax evasion is a sub-component of the black economy it is often indirectly estimated as part of larger illegal economic activities when using macro approaches. The micro approach is favoured in this instance as it is more targeted. Second, it provides estimates of under-reporting by the household characteristic of choice (age, professional group etc.). As the focus of this thesis has a micro perspective and favours household level surveys this specific methodology was favoured. The following section will present the most important papers using a microeconomic approach for countries around the world and then research relating specifically to Ireland.

2.2.3 Estimates of the black economy and tax evasion in the literature

The economic literature on the measurement of the black economy spans many countries over many years. This section will attempt to present the main contributions to the literature and give an overview of the estimates found in each of these analyses. More details will be given on estimates for Ireland in a subsequent section.

Bottom-up estimates for the rest of the world

There are many methodologies used to estimate the size of a black economy as shown in the previous section. The studies applying these methods to different countries are even more numerous. The following presentation of the literature in this area does not aim to be exhaustive but rather to give an overview of the different estimates available using microeconomic techniques.

Studies using micro-economic approaches are less common, mainly due to a lack of proper time series data. As mentioned previously, Isachsen and Stroem (1985) are one of the rare studies that uses a direct survey to estimate the size of the tax gap in Norway. This is a type of microeconomic approach which as noted before suffers from the problem of non-response. They estimate however that hidden labour income represents around 4% to 6% of GDP in Norway. Micro-expenditure approaches are based on the seminal paper by Pissarides and Weber (1989), which will be described in greater detail in Section 2.3. The approach is based on estimates of "typical" food consumption and income by wage earners. Engel curves for food are compared to those of the self-employed in order to discover their "true" level of income. This allows the degree of non-compliance with taxation in an economy to be uncovered. They find that income under-reporting ranges from 28% to 64% in the UK in 1982 depending on the occupational group examined. Lyssiotou, Pashardes and Stengos (2004) use a consumer demand system approach to estimate the size of income under-reporting in the UK in 1993. They find that the level of under-reporting is lower than 10 years earlier (in the 1982 Pissarides and Weber study) and ranges from 9% to 41% depending on the occupational group (white collar workers under-report less than their blue collar counterparts).

Schuetze (2002) uses six years of Family Expenditure Surveys (the Canadian equivalent of the Household Budget Survey) between 1969 and 1992. The interesting aspect of this study is that the estimates are disaggregated across years and demographic characteristics as well as occupation. He finds that there is a significant degree of non-compliance among the self-employed depending on the occupation and age of the head of household as well as the number of household members who are self-employed. He finds that between 1969 and 1992 households which were self-employed concealed on average between 11% and 23% of total household income. He does not find any significant trends depending on education but does find however that occupation is important. The level of concealment is highest among construction and services, occupations which involve cash transactions. Men and younger workers were more likely to under-report their income. These results confirm the ones from the Norwegian survey. His results are similar to those found by Mirus and Smith (1996) who found that Canadian households concealed 12.5% of their incomes.

Besim and Jenkins (2005) use the 1998 Household Consumption Expenditure Survey to estimate the incidence of income under-reporting in North Cyprus. Their control group consists of government employees, on the basis that their opportunity to

evade tax is smaller than other types of workers. They find that the levels of income reporting in North Cyprus are high for self-employed workers (11-14%) and in some cases just as high for other non-government employees (9-14%). They attribute these high levels of tax evasion to the very high tax rates and economic instability in the country.

Studies using the same methodology also exist for a number of other countries. Johansson (2000) examines income under-reporting in Finland and finds that self-employment income is under-reported by on average 27% between 1994 and 1996. Engström and Holmlund (2006) find a level of income under-reporting of 30% among households in Sweden between 1999 and 2004 while Kim et al. (2008) use panel data from Korea and Russia and find that on average 38% and 47% of incomes in these countries are not reported.

Other similar types of micro-economic studies include Feldman and Slemrod (2007) who use a similar approach to the Pissarides and Weber method but use charitable cash contributions as the dependent variable. They assume, as with food expenditure, that self-employed and employee preferences are the same in this respect. They use unaudited tax returns in the United States in 1999 and find that self-employment income must be multiplied by 1.54 to obtain true income. Their figure for farm incomes is even higher, at 3.87.

Lacko (1999) bases her study on electricity consumption and finds that previous studies using aggregate variations in electricity consumption in transition economies to estimate the level of the black economy were not accurate and that any changes in the level of electricity consumption during that period could be explained by structural changes. However she does find that other indices such as residential electricity consumption give satisfactory estimates. Lang et al. (1997) use a national survey of German household budgets and the relationship between household gross income and taxed income to determine the distributional implications and magnitude of tax avoidance. They find a 33% tax gap in Germany in 1983 and that individual tax evasion increases with income.

Finally, the latest study using the Pissarides and Weber method is that by Gorodnichenko et al. (2007, 2009) on tax evasion in Russia before and after Russia's flat rate income tax reform. They use the same expenditure based method to estimate the level of tax evasion of public sector workers versus private sector workers. They use these categories as it is widely believed that corruption in the public sector in Rus-

sia is high and bribes can explain any discrepancies in consumption by public sector workers compared to their private sector counterparts. The originality of their study is that they not only have a panel dataset but it covers a period of tax reform in the country. This allows them to use a difference-in-differences approach and they find that the introduction of a flat rate income tax helped decrease tax evasion in Russia and reduced the consumption-income gap by 9% down to 12%. The following section presents previous estimates of the black economy for Ireland, all of which are based on macro-economic approaches.

2.2.4 Measures for Ireland

The Office of the Revenue Commissioner's of Ireland stated in their report on tax compliance that:

"the gradual growth in the numbers of self-employed (about 10-15% in the 1990's) has meant a growth potential for irregular economy activity. Every country's counter evasion programme detects substantial under-declarations of tax liability particularly in businesses where transactions are largely in cash [...] Experience suggests that by far the greater part of black economy activity is represented by people who are on record but not disclosing their full activities."

Only a few studies estimating the size of tax evasion have been conducted in Ireland and all are macroeconomic studies. Boyle (1982, 1984) uses the monetary method; the basic assumption of the technique is that people who are conducting hidden or illegal transactions will primarily use cash. An increase in the size of the shadow economy can then be seen through an increase in the demand for currency and excess currency holdings. He finds that the level of the black economy in Ireland in between 1975 and 1983 was approximately 6% of GDP. The issues with his analysis concern the time period being reviewed. As Ireland joined the EMS over that period, the change in currency holdings could be explained by another factor altogether (i.e. sterling balances with the public had to be replaced by Irish pounds).

Fagan (1993) provides estimates of the black economy in Ireland using a few different methodologies. Using a technique developed by Frey and Weck (1984) based on an econometric analysis of the relationship between the determinants of the black economy and indicators of its size, he finds that the Irish black economy accounted for

7.25% GNP in 1978. He also provides estimates based on trends in the currency current account ratio. Using this method he shows that the black economy gradually increased from low levels in the 1960s and reached a peak of 15% of GNP in 1983 before falling back to 11% in 1992. He acknowledges that there are limitations to this technique. He also uses currency demand equations and finds a figure of 10% of GNP in 1983 and 8% in 1992. His main conclusions are that the methodologies provide estimates which are uncertain in size. The size of the estimates varies between 5 and 10% depending on the year and technique used. Nevertheless he does conclude that the methodologies do provide an indication of the existence of a black economy in Ireland. In a later paper, Fagan (1997) provides a monetary estimate of the size of the black economy in Ireland between 1964 and 1995. He finds that the size of the black economy in the country and represented approximately 11% of GNP in 1995.

An article in *The Economist* of May 3rd 1997 estimated the size of the Irish black economy (as a percentage of the 1994 official GNP) at 15.3% based on preliminary studies by Schneider. Schneider (2007) uses the DYMIMIC approach to estimate the size and evolution in size of the black economy for 145 countries between 1999 and 2005. His estimates for Ireland are 15.9% of GDP in 1999-2000 and 14.1% of GDP in 2004-2005. These estimates put Ireland mid-range compared to the rest of Europe and falling with economic growth.

Based on the previous literature the hypotheses regarding how the level of tax evasion changed in Ireland between 1987 and 2004 are not entirely clear. In the period under examination, the tax burden eased implying from the literature that the level of tax evasion would fall. High levels of economic growth would also imply a fall in tax evasion however they also resulted in more funds available for compliance and audits which could indicate lower tax evasion opportunities. The present analysis will hopefully shed some light on the possible trends.

2.3 Methodology

This section presents the empirical approach used in the analysis. The technique uses household expenditure micro-survey data and was developed by Pissarides and Weber (1989). It is often described as the "expenditure method". Individual survey data are used to look for systematic differences in behaviour that may be explicable in terms of concealed income. This indirect approach uses estimates of the relationship between food consumption and income for self-employed and wage earners and com-

putes "true" income levels for self-employed households, consequently estimating the degree of non-compliance among the self-employed.

There are various assumptions associated with the Pissarides and Weber methodology. The first is that all respondents report expenditures on food correctly. The second is that wage and salary earners report income correctly while the self-employed do not.⁵ Finally, they assume that the relationship between food expenditure and after-tax income is the same for all families in the survey.

The first assumption would seem to be valid for food. Indeed, as individual food items are small and not costly in nature there would be no reason for survey respondents to lie about their purchases. This would be quite different if looking at electronic equipment for example. It is for this reason that food expenditure is usually used in these types of analyses.⁶ The literature in the area of non-compliance and under-reporting of income suggests that self-employed earners are more likely to under-report their incomes to a greater extent than wage earners. This is due to the fact that they are responsible for their tax returns making it easier to omit income revenue.

It is also assumed that wage earners and the self-employed have the same preferences for food consumption. This third assumption is more problematic. The self-employed will tend to adjust their consumption faster during shocks. While this may be the case, it is also true that the adjustment in the level of consumption of a necessity such as food will be less rapid and important than, for instance, luxury goods. Moreover, food expenditure in restaurants and bars is not included in the computation of the dependent variable which consists purely of food *consumed within the home*.⁷ Household characteristics and wealth (mortgage repayments, home ownership type) are included in the regressions to take account of differences in consumption that are due to a household's stage in the life-cycle.

Finally, it is possible that there is an element of self-selection into self-employment. For instance workers who are more risk averse might be less likely to under-report

⁵In the case that wage earners also report incomes incorrectly, the estimated under-compliance for the self-employed would be a lower bound.

⁶The main hypothesis in this model is that there is a systematically lower Engel curve for the self-employed than employed. This hypothesis is tested in the same way regardless of the shape of the curve (flat or curved) since income is controlled for. The choice of which good to study depends on the degree of homogeneity of demand within the sample. If it were extremely heterogeneous, there would be bigger error bands around the estimated Engel curve, and it would be harder to establish whether the hypothesis was true or not. Food is chosen as it is an essential good (and thus there will be no zero consumption cases) and is the good that is most likely to be similar in composition across income groups.

⁷However, the share of food consumed out of the home is included as a control variable in the analysis.

their income but they could also be less likely to choose to be self-employed. It will not be possible to control for this using this methodology but it is a factor worth considering when analysing the results.

The basis of the Pissarides and Weber (1989, henceforth PW)⁸ analysis is an expenditure equation which is defined as follows:

$$\ln X f_i = \alpha Z_i + \beta \ln Y_i^P + \varepsilon_i \quad (2.1)$$

where $\ln X f_i$ is the log of food expenditure for household i , Z is a vector of exogenous variables such as household characteristics, ε_i is a white noise error term, β is the expenditure elasticity of food and Y_i^P is the measure of income that influences household consumption decisions, referred to as "permanent income". Pissarides and Weber suggest that reported income (Y_i^R) will differ from permanent income (Y_i^P) for two reasons: first because of non-compliance by certain groups (here, the self-employed) and second because of transitory shocks. It is consequently necessary to take account of these differences in the analysis. First, they assume that the self-employed are more likely to under-report their income. Hence their true income (Y_i^T) will differ from their reported income (Y_i^R) by a factor k_i . The relationship between true income (Y_i^T) and reported income (Y_i^R) is therefore:

$$Y_i^T = k_i Y_i^R \quad (2.2)$$

where $k = 1$ for wage earners (it is assumed they do not lie at all) and $k > 1$ for the self-employed (who have the potential to lie). Consequently, the larger the k the larger the degree of non-compliance by the self-employed. There is then also a problem with how current income is defined and how it relates to "permanent" income. In the PW analysis permanent income is not used in the usual economic theory sense (i.e. smoothed over the full life cycle). Rather, it is a measure of income smoothed to take account of unexpected income shocks and the level of income that defines consumption. Pissarides and Weber define the relationship between true income (Y_i^T) and permanent income (Y_i^P) as:

$$Y_i^T = p_i Y_i^P \quad (2.3)$$

where p is a random variable measuring the effect of aggregate events on current income. This allows them to take account of transitory shocks affecting income. The shocks affecting household incomes are assumed to have the same mean across em-

⁸ A more extensive presentation of the model is available in the Appendix (Section 2.9). The full details of the equation derivations are detailed in Pissarides and Weber (1989).

ployment groups but have a different variance depending on whether households are self-employed or not. It is expected that the variance of p will be larger for the self-employed (as they are more sensitive to income shocks).⁹ In a first stage, Y^P is estimated using Y^R and instruments for permanent wealth (such as home ownership, education etc.) and in a second stage the expenditure equation (2.1) is estimated with the inclusion of a self-employment indicator SE_i , giving:

$$\ln X f_i = \alpha Z_i + \beta \ln Y_i^P + \delta SE_i + \varepsilon_i \quad (2.4)$$

It is then possible to estimate upper and lower bounds to compliance using the following formulae:

$$\ln k_u = \frac{\delta}{\beta} + \frac{1}{2}(\sigma_{\lambda SE}^2 - \sigma_{\lambda EE}^2) \quad (\text{upper bound})$$

$$\ln k_l = \frac{\delta}{\beta} - \frac{1}{2}(\sigma_{\lambda SE}^2 - \sigma_{\lambda EE}^2) \quad (\text{lower bound})$$

where $\sigma_{\lambda SE}^2$ is the residual variance from the first stage regression for the self-employed and $\sigma_{\lambda EE}^2$ is the residual variance from the first stage regression for employees.

In summary, the PW method is based on a two stage estimation technique. Stage 1 predicts the relationship between reported income and permanent income and stage 2 predicts the relationship between permanent income and food expenditure. The results of this analysis are then used to estimate upper and lower bounds of income under-reporting.

2.4 Data

The Household Budget Survey is a survey conducted at varying frequencies in Ireland since 1951 by the Irish Central Statistics Office (CSO, 2007, 2009). It was conducted every 7 years from 1973, and since 1995 it has been conducted every five years. Surveys are conducted in the first few months and last few months of the calendar year. The unit of analysis is private households. Each household that participates in the survey completes a daily diary of household expenditure for two weeks. Detailed information on the composition of households and income is also collected. This results in a wealth

⁹As per Pissarides and Weber (1989): "We do not have data on either p or k so assumptions are made on their distributions over households to make estimation tractable. Thus we assume that p and k are log-normal." For employees $k=1$ and the mean of p is the same for each occupational group.

of information composed of 1,200 items of expenditure of which over 300 are food and drink items, as well as close to 500 household characteristic variables.

The objective of such surveys is to get detailed information on the consumption patterns of households. The statistical purpose of the HBS is to construct and update the consumer price index (CPI) weighting, i.e. the representative basket of goods used to determine the CPI level. The Irish HBS studies have been used in a number of published research papers covering topics such as the distributional implications of taxes (Callan et al., 2009b), appliance use (O'Doherty et al., 2008) and household meals (Newman et al. 2003).

This paper uses data from four cross sections of the HBS: the 1987, 1994, 1999 and the 2004 waves.¹⁰ On average about 7,000 households completed the surveys across the years. The unit of analysis is the household spending unit, which is assumed to act as a single decision-maker for decisions about income reporting and expenditures. Some observations were however not used in the analysis and the following section describes the sample restrictions imposed on the survey by the assumptions of the model.

2.4.1 Sample selection

The model detailed in Section 2.3 requires a certain number of assumptions which themselves impose restrictions on the sample used in the analysis. In order to apply the PW method to the data a fairly homogeneous group of households is required, both for the self-employed and for wage earners.

The first main requirement is that the households examined have a member that is in employment (be it self-employed or earning wages). As a consequence, all households where the chief economic supporter (CES)¹¹ was classified as unemployed or retired were excluded from the analysis. Furthermore, chief economic supporters who work part-time were excluded from the analysis as part-time workers are likely to smooth their consumption patterns differently.

The second assumption relates to the household composition. In this study the sample examined consists of households composed of adults with or without children and households consisting of married couples with or without children. Having a rel-

¹⁰These are the waves for which electronic versions of the micro-data are available.

¹¹Person in the household with the highest gross income.

actively homogenous group is important. It was not possible to estimate a measure of under-reporting for single person households, as not enough observations were available. Moreover, these households tend to have different spending patterns than larger household groups. Consequently, single-adult households were excluded from the analysis.¹²

Age restrictions were also employed. The sample was restricted to households where the chief economic supporter is aged between 25 and 64. This, as with the first restriction, eliminates retired households from the sample as well as students. Only households with positive after-tax income were included in the analysis. The top and bottom 1% of households were also cut according to their disposable income and gross income to eliminate income outliers. The HBS is top coded and thus the top earning households' incomes are not included in detail. By eliminating the top 1% of households the top coding is eliminated.

The HBS microdata file provided to researchers is anonymised and does not contain any information on household members' occupations. However, the CSO provided these variables for the 2004 dataset. The data were however not available for the other years. A separate analysis was conducted on construction industry workers using the 2004 data.

Finally, the issue of rural farming households was also examined. Most previous studies of this nature did not include households where the head of the household was employed in a primary occupation. Farm households are likely to have quite different food expenditure patterns than those in other non-primary occupations as consumption of food from own production can be quite high in rural households and this can distort results. Excluding them would restrict the sample for Ireland, particularly in the 1987 wave, and simply eliminating these households was problematic. Interestingly, the Irish HBS also includes questions on consumption of food and other goods *produced at home*. These items are then included in the survey by the CSO who estimates their "market value" depending on the amount consumed. Consequently, farm households were examined separately with the food expenditure variable adjusted to include food consumed from own produce.

After applying these restrictions, the overall sample in 1987 consisted of 3,650

¹²An analysis was run on a sub-sample of single adult households only and for part-time workers but the results were not conclusive. This is likely to be due to the fact that these types of households smooth their consumption differently to large full-time working households or the lack of sufficient observations for a reliable estimation.

households of which 10% were self-employed. The associated figures for 1994 were 3,191 observations and 11% and for 1999, 3,773 observations and 12%. In 2005, the overall sample consisted of 3'779 households and 11% of those were self-employed. The following section defines the different variables used in the analysis and how they were constructed. It also presents the descriptive statistics for these variables.

2.4.2 Specification of variables used

This section presents the different variables included in the analysis and some descriptive statistics. In general, the statistics presented are only for the overall sample (i.e. households with more than one adult, employees and self-employed and including farming households), for ease of presentation.

Food and Income

The key variables in the analysis are household food consumption and household reported income. Food expenditure is recorded daily in a diary over two weeks. It is unlikely that small items of food expenditure will be concealed for tax purposes. Moreover, the person filling in the diary is probably not the person earning the main income in the household. If concealment were to happen, expenditure on more ostentatious consumption such as cars and high-tech electronic appliances would be more likely. Two versions of household food consumption were used in the various specifications. The first is logged yearly household expenditure on food, which is derived from the disaggregated expenditures on food products and beverages by the household to be consumed in the home. The second food related variable is logged yearly food expenditure including own produce (which is food produced and consumed within the home). The CSO estimates the corresponding "expenditure" by using the market price of these products and their quantities. The share of food and drink consumed outside the home (for example in restaurants and cafes) is included as a control variable but is not part of the dependent variable.

The income variable used in the analysis is logged yearly household disposable income.¹³ This includes all household income (including the spouse's and children's

¹³"Gross household income is defined as direct income plus State transfer payments. Disposable income is arrived at after the deduction of income tax and social insurance. The gross figure includes income tax, social insurance and all other deductions made at source by the employer. Additions are made to this gross figure to include the weekly equivalent value of any regular bonuses, commission etc. received during the 12 months preceding the interview" (CSO, 2007: 257). Income was not equivalised as

incomes if present) after income tax and social insurance. All monetary variables are deflated and expressed in constant 2006 euro and expenditures are yearly. It should be noted that the income variable is income as reported in the household survey. It is not income as derived from actual tax returns. As mentioned in footnote 5, it is possible that wage earners will also mis-report their incomes. Should wage earners under-report their income then the estimated under-compliance for the self-employed would be a lower bound. As it is not necessarily the actual income earner that answers the questions on income it is also possible that there will some level of misreporting occurring naturally. Finally, the self-employment dummy takes a value of 1 if a household is described as being self-employed and zero otherwise.

The summary statistics for these variables across years are presented in Table 2.1. It is clear that the share of food in total income has fallen drastically over the last two decades from 22% to 13%. However the *amount spent* on food has increased, as has the share of income spent on food outside the home.¹⁴ The same variables are detailed by employment type (self-employed, employees and farmers) in Tables 2.2 to 2.4. Statistical tests of the significance in the difference between these averages were conducted. Overall the self-employed spend more on food than employees. This also holds for farming households (the difference is statistically significant in all years except 2004). Although the average wage of self-employed households is lower than that of employees the difference is not statistically significant. However, if the food expenditure figures are correctly reported by respondents, they are a first indication that the self-employed might be concealing part of their earnings. However, further analysis is needed to check whether these differences are maintained when other factors and household characteristics are taken into account. It is also reassuring to see that the standard deviation in income for the self-employed is higher than that for employees. This is consistent with the assumptions presented in Section 2.3 that underlie the model.

Household characteristics

A vector of household characteristics that are expected to determine a household's food consumption is also included in the regressions. This includes a variable for family composition (which details the number of people in the household and their relation to each other) as well as variables for the number of children in different age groups (under 5 years of age, between 5 and 13 and 14 and over).

household composition dummies as well as child age dummies were included in the analysis.

¹⁴Food as a share of total expenditure followed similar patterns across the years being examined.

Information on the chief economic supporter (CES) is added to the household level information. The CES is defined as the adult in the house who has the highest level of gross income. An age of CES dummy is included (equal to 1 if the CES is over 45 years old) as well as a gender dummy equal to one if the CES is female. Occupational dummies equal to one if the CES is in a manual occupation or in the lower professional category, are also included in the regression. Dummies relating to the spouse's employment participation (full-time employment, part-time employment, self-employed spouse and other work) and age are also included as control variables. Dummies for the CES's education level (junior certificate level, leaving certificate level and college education) are included in the instruments.

Wealth characteristics

The second set of variables consists of indicators of a household's wealth, and these variables are used in the first stage of the regression. The first relates to the household's home. A variable relating to the number of bedrooms in the residence is included as well as the type of household accommodation (this consists of three categories in the HBS: apartment, detached house or terraced/semi-detached house).¹⁵ Whether the household is renting, owns its own home or is mortgaged is also included as a set of dummy variables, and if it has a mortgage the mortgage amount is included. A dummy equal to one if the household has a savings account and a set of dummies relating to the ownership of different household appliances were included depending on the years examined (washing-machine, spin dryer, microwave, fridge-freezer, central heating, gas, home computer, etc.). In certain years it was not possible to use these as instruments as too high or too low a percentage of the sample owned these facilities, so their inclusion differs according to the year examined.

Time and Location Dummies

Finally, in order to take account of time fixed effects and location fixed effects, time and location dummies are also included in the analysis. There are various regional variables in the HBS which include town size, regional location and urban-rural categories. In the context of this analysis the town size variable was used as this was the

¹⁵Due to insufficient observations for the apartment category it was merged with the semi-detached category. The dummy included in the regression is equal to 1 if the house is a detached house and zero otherwise.

most detailed option available over all years. Month dummies were included to take account of seasonal variations in consumption.

The town size categorical variable consists of 5 categories: rural households, households residing in Dublin, households residing in towns with more than 20,000 inhabitants, households residing in towns with between 3,000 and 20,000 inhabitants, and households residing in towns with less than 3,000 inhabitants. It should also be noted that all variables in the first stage regression were interacted with the self-employment dummy. Interactions of the self-employment dummy with the CES age dummy, Dublin dummy, and manual and low professional dummies were also included in the main regression in order to facilitate the calculation of under-reporting for these categories.

Tables 2.17 to 2.20 in the Appendix detail the summary statistics for all the explanatory variables presented above (see Appendix, Section 2.9, Tables 2.21 to 2.24 for the summary statistics for the first stage variables and instruments). According to the summary statistics, the self-employed in the sample are slightly older and are less likely to have young children or a college education. They are also less likely to have a female CES which is in line with the idea that there are fewer female entrepreneurs than male ones and they also have a lower level of income contribution from a spouse. The self-employed also have larger houses and they are less likely to rent. This is also in line with the assumption in Section 2.3 that the self-employed smooth their consumption over the life-cycle differently from employees. More of their wealth is invested in assets and it is hence essential that these are taken into account when estimating permanent income.

2.5 Results and discussion

2.5.1 Non-parametric estimates

As mentioned in the previous section the summary statistics seem to indicate that on average the self-employed spend more on food than employees and have lower wages. The data are plotted in Figures 2.1 and 2.2. The first set of figures shows the plots and the second the fitted trend lines for the smoothed (Nadaraya-Watson non-parametric) regressions of food expenditure on disposable income for each employment type (SE, EE and farming households) for the four years of data. This again confirms the assumptions: the self-employed seem to consume more food at each level of income

than employees. The farming household variables here are slightly lower than the SE but still higher than employees. This issue will be returned to in the results section. Although these figures confirm the assumptions of the model, a full analysis is still required to see if they hold once household characteristics and differing levels of wealth are taken into account.

2.5.2 Estimates of under-reporting for the self-employed

These sections present the results of the various estimations. First, the full regression results are presented for the different years to highlight the important variables affecting consumption of food. The later sections concentrate on the estimates of under-reporting. For ease of reading, only the coefficients relating to variables required to calculate the level of income under-reporting are presented in that section. The coefficients of interest are β , the expenditure elasticity of food, δ the coefficient on the self-employment dummy and k_l and k_u , the lower and upper bounds of income under-reporting for the group under examination. Equation (2.4) from Section 2.3 is estimated using generalised method of moments (GMM)¹⁶ and the dependent variable is the natural log of household deflated expenditure on food. The independent variables include the log of household disposable yearly income and a self-employment dummy which is 1 if the CES is self-employed and zero otherwise. The instruments used vary by year but consist overall of the CES's education level, house size, accommodation type, and dummies for ownership of appliances.¹⁷

Regression results

Tables 2.5 to 2.8 present the main regression results across all years.¹⁸ All the family composition dummies are significant. The number of people and structure of the family unit all contribute in a significant way to the level of food expenditure in the household. The larger the family size the larger the coefficients across all years. Another household characteristic which is significant across all years is the effect of children of different ages on food expenditure. The coefficients get larger the higher the children's ages and the more children included. Most of the regional location variables are also significant as are most of those relating to the spouse's occupation type. With

¹⁶In the presence of heteroskedasticity, the use of a GMM estimator is advised.

¹⁷Hausman F Tests confirmed disposable income should be treated as endogenous, the validity of the instruments were also confirmed and all the regressions passed tests for over-identifying restrictions (Hansen's J statistic). The full first stage regression results are in the Appendix, Table 2.31.

¹⁸The coefficients for the monthly dummies are not shown to save on space.

regards to age and food expenditure, households headed by a CES over the age of 45 tend to spend more on food than younger ones (however, even though this coefficient is positive it is not significant across years). The coefficients on the spouse's age are all positive and significant. They show that the older the spouse the more is spent on food until the age of 55 where there is a slight drop (except in 2004) which could be an indication of lower food consumption at this age. None of the coefficients on the explanatory variables are unexpected and most of them conform to prior theory. The coefficients of interest in capturing income under-reporting will now be presented.

Under-reporting results, self-employed households

Table 2.9 presents the key regression coefficients and the lower and upper bound estimates for income under-reporting by the self-employed compared to the rest of the population in 1987, 1994, 1999 and 2004. This sample excludes farmers and the dependent variable is food purchased outside the home for consumption within the home. The expenditure elasticity of food is positive and significantly different from zero at the 1% level in all years. For instance in 1999, households spent on average 0.19% extra on food for every 1% increase in income.

The coefficient on the self-employment dummy is also positive and significant in all years (in 2004 it is only significant at the 10% level). Using these coefficients and the residual variances from the first stage equation it is possible to calculate the lower and upper bounds of income under-reporting which are reported in Table 2.9. In 1987, self-employment income needed to, on average be multiplied by between 1.40 to 1.49 to get "true" income. This fell to an average of 1.35 in 1994 and then rose slightly to 1.37 times reported income in 1999. It further fell to an average of 1.31 in 2004. These results are in the same range as the estimates for Canada (Schuetze, 2002) and the 1987 figures are fairly similar to the results for the UK around the same period (Pissarides and Weber, 1989).

Estimates of under-reporting by household characteristic

The regressions were run for sub-samples of the data depending on household characteristics. This allows the calculation of the under-reporting bounds for different household groups (see full regressions in Appendix).

The level of urbanisation of an area has been found to have an effect on the level of under-reporting (Engström and Holmlund (2006), Pahl (1988)). It was not possi-

ble due to data restrictions to test this assumption specifically, however a check was done to see if those living in the capital city under-reported in the same manner as the overall sample. The results for the Dublin sub-sample are presented in Table 2.10. Results for 1999 and 2004 are not significant and consequently the coefficient on δ is replaced by zero for the calculation of bounds. An interesting trend emerges. In the first two years of data, the level of under-reporting by a household based in Dublin is between 31% and 43% while in the latter two years it falls to an average of 5.5%. The fall in under-reporting values for the Dublin sub-sample is much more pronounced than that for the full sample. Although, as the coefficient on the self-employment indicator is not significant for the latter two years this results should be taken with caution.

The trend for households with an older CES was also examined in Table 2.11. Although more detailed age groupings for the chief economic supporter are available in the dataset it was necessary to combine age groups in order to have sufficient observations to estimate the equations for a large enough sub-sample. Hence, an older CES variable was created: consisting of a CES aged between 45 and 64 years of age. Using the older CES sub-sample, under-reporting bounds for this age category were calculated. There does not seem to be a clear chronological trend. Across years, under-reporting bounds do not change much with the average level ranging between 24% and 32% depending on the survey year. There is a slight difference in the average bounds: the first two years of data have average bounds in the low 20s and the second two years of data have average bounds in the early 30s. However, again as the coefficient on the self-employment indicator is not significant for two of the years it is difficult to confidently identify a trend. It is also difficult to pick out a cohort effect as the data is not panel data. However it is possible that the lack of change across years is due to the fact that the households which are under-reporting are getting older.

Finally, two other sub-sample were examined depending on the CES's professional group. This follows Pissarides and Weber who estimated under-reporting for blue-collar and white-collar workers separately. In this case the categories were CES in a manual occupation and CES in a lower professional group and the under-reporting bounds are presented in Tables 2.12 and 2.13. In the case of the both sub-samples, the coefficients on the self-employment dummies are not significant for the 1994 and 2004 survey years. However, if these are replaced by zero and the bounds are calculated a trend emerges. In both cases the average level of under-reporting is much higher in 1987 than in any other year. In 1987, self-employed manual workers on average under-report their income by 66% compared to employed manual workers. For those in the low professional category the average is 69%. In both sub-samples the average

under-reporting drops to a level below 30% in the later years. This may be an indication of a move towards less cash-based transactions for manual and low professional self-employed households in later years and consequently a reduction in the *scope* for under-reporting. A much steeper fall in under-reporting is seen for these two categories than that for the full sample but all follow the same downward trend.

Overall the results seem to indicate that under-reporting of incomes by the self-employed changed over the course of the years covered by the datasets in the analysis. For the full sample the trend is a gradual fall in the levels of under-reporting. The fall in the level of under-reporting for manual and low professional groups was much steeper and no real trend is observed for the older CES sub-sample. We now turn to other occupational groups.

2.5.3 Estimates of under-reporting for other occupational groups

Farming households

The same estimations using the same dependent variable were run for farming households (using employees as the reference category for the first three years of data). The results for each of the years are presented in Table 2.14. The coefficients on the income variable are significant at the 1% level across all years, as are the farming dummy coefficients. Under-reporting varies from 40% in 1987 to 51% in 1994 back down to 35% in 1999. However, as discussed in Section 2.4.2, farming households may get a significant proportion of their food from own produce. As stated in Section 2.3, households in a primary sector occupation tend to be excluded from these types of analyses as their food consumption is difficult to measure. The originality of the Irish household budget surveys is that food produced within the home is measured and valued. Hence it would be logical to include “own produce” food expenditure in the dependent variable. The regressions were re-estimated using the latter and the new under-reporting coefficients are presented in Table 2.15. The level of under-reporting for farming households increases significantly and is much higher than that of self-employed households. The estimates range on average between 60% and 100% depending on the year and bound. These results may seem quite large, however they are similar to those estimated by Pissarides and Weber for manual occupations.¹⁹

¹⁹It is worth noting that farming households may have a greater variance in income than other households types across years as well as within years. Unfortunately, as this is not a panel survey it is not possible to control for this here.

Construction sector households

Access to the non-anonymised version of the 2005 dataset was provided by the CSO. This version of the dataset includes the chief economic supporter's occupation. Consequently it was possible to separately examine specific occupational groups, namely those in construction. The results for this section relate to the level of under-reporting by those employed in the construction industry compared to other employees and are reported in Table 2.16. The under-reporting bounds for this category are 1.10 to 1.12. This is an indication that self-employed workers in a cash based industry such as construction under-reported more than their employed counter-parts in 2004. However, it is possible, as with any of the results of the analysis, that the level of under-reporting is higher than predicted. As stated in the assumptions, if the base sample also under-report, the calculated bounds are an indication of the *additional* level of under-reporting. For instance, if all employees in construction under-report (e.g. due to cash based transactions) the results here would indicate that those who are self-employed under-report 11% *more* on average. This holds for other sub-samples too.

To conclude, households with a CES working in farming or the construction industry have more of an opportunity to conceal and as a consequence the under-reporting bounds for these categories are higher than for the average self-employed, in line with previous literature.

2.6 Conclusions

The Irish recession has highlighted what has been described in national newspapers as a re-emergence of the black economy (Burke, 2009). Income under-reporting is a phenomenon in all countries but the extent of tax evasion has been found to depend on economic growth in a country as well as the level of the tax burden. In particular, self-employed households tend to be targeted by tax audits as they have the most scope to conceal their incomes.

The approach adopted in this paper involved using individual survey data to look for systematic differences in behaviour between the self-employed and employees that may be explicable in terms of concealed income in Ireland. The methodology was developed by Pissarides and Weber (1989) and rests on the assumptions that reporting of food expenditures by both groups is accurate whereas if concealment occurs it relates to reported incomes. Comparisons of food expenditures at different levels of income can then be used to uncover what can be described as "differential reporting".

Four waves of household expenditure surveys were used spanning from 1987 to 2004, a period of high economic growth in Ireland. Under-reporting estimates were calculated for self-employed households across years as well as farming households and households where the CES is employed in the construction sector. Sub-sample based on the characteristics of the household head were also examined to check whether trends in under-reporting were visible according to different characteristics such as age, regional location and professional type.

There is strong evidence of differential reporting with regards to food consumption between the self-employed and employees for a given level of income. The results indicate that on average concealment of income was higher in the earlier years of the surveys and gradually fell. On average self-employed household incomes in Ireland need to be multiplied by between 1.30 and 1.49 depending on the year.²⁰ No trend was found for the older households and manual and low professional groups followed the same downward trend as the main sample, albeit with a steeper decline. Dublin based households also showed reduced under-reporting over the course of the survey years. The estimates for farming sector households were found to increase significantly when food produced within the home was taken into account and average under-reporting bounds ranged between 1.6 and 2.07. The levels of under-reporting are similar in magnitude to recent estimates for North Cyprus and Canada. This is promising as the latter study uses a similar dataset and methodology and the self-employed households in the Canadian dataset had similar characteristics to those in Ireland. The results are also consistent with the conclusions made by Pissarides and Weber who found that manual occupations under-reported significantly more than non-manual occupations in the 1980s in Britain.

It is not surprising that the levels of income concealment found in the present study are higher for manual and low professional households in the 1987 survey and for farming households across all years. Indeed, the former categories (in particular manual) were cash based occupational sectors. The main basis of this analysis is that self-employed individuals have more opportunity to evade tax as they self-declare their earnings. This *opportunity* to evade is further compounded in a sector where transactions are cash-based leading to more opportunities to conceal earnings. With regards the results for the farming households, the situation is fairly similar. Revenues from

²⁰Based on an average income for self-employed households in 1999 of €41,427 and an average under-reporting range for that group of 1.37, this would represent under-reporting of approximately €15,300. This is in the same order of magnitude as the results of the targeted audits conducted by the Revenue Commissioners which yielded on average £18,700 in 1998.

farming are dependent on extremely variable sales of products at volatile prices and consequently difficult to check. Higher non-compliance seems to relate to the level of opportunity for concealment. Although the level of income under-reporting by self-employed households seems to have decreased over the period, income lost from this under-reporting as a proportion of total taxable income increased (see Figure 2.3).²¹ This study cannot attribute a reason to the changes in under-reporting uncovered. Tax evasion can be linked to the level of the tax burden, economic growth as well as to the growth of certain occupational groups (e.g. those based in cash based industries). Ireland changed significantly during the period under examination in all those areas. Although the results of this study also do not allow the exact targeting of those evading tax they do give some guidelines as to those most likely to evade and audits concentrating on the groups identified above would be effective.

²¹Note, this is for the sample analysed in the present study, which excludes single-person households. Consequently, total lost income from under-reporting is likely to be higher.

2.7 Tables

Table 2.1: Food and income across years, full sample

Variable	1987	1994	1999	2004
Deflated disposable yearly income	30013.92	32090.11	43921.31	61763.16
Deflated yearly amount spent on food	6037.95	6202.58	6772.07	7158.81
Share of food expenditure in total income	22.59	21.59	17.49	13.44
Own produce share	0.57	0.23	0.13	0.05
Eating out share	2.51	3.54	3.86	4.17
Observations	3650	3191	3773	3779

Table 2.2: Food and income across years, Self-Employed

Variable	1987	1994	1999	2004
Deflated disposable yearly income	27789.23	29845.63	41427.45	61344.98
	(12892)	(12255)	(16920)	(30756)
Deflated yearly amount spent on food	6483.15	6840.83	7387.83	7379.22
	(2629)	(2776)	(3032)	(3267)
Share of food expenditure in total income	27.34	26.18	20.55	14.34
Own produce share	0.28	0.11	0.02	0.03
Eating out share	3.10	4.00	4.90	4.97
Observations	371	360	467	401

Table 2.3: Food and income across years, Employees

Variable	1987	1994	1999	2004
Deflated disposable yearly income	30400.54 (11342)	32834.54 (11172)	44711.09 (17125)	61826.94 (26474)
Deflated yearly amount spent on food	5838.08 (2357)	5979.02 (2439)	6596.77 (2749)	7012.95 (3067)
Share of food expenditure in total income	21.07	19.90	16.51	13.00
Own produce share	0.14	0.03	0.03	0.02
Eating out share	2.64	3.66	3.85	4.16
Observations	2835	2434	2956	3040

Table 2.4: Food and income across years, Farmers

Variable	1987	1994	1999	2004
Deflated disposable yearly income	29404.18 (15466)	29561.31 (12987)	40578.5 (18485)	61685.64 (30803)
Deflated yearly amount spent on food	6942.19 (3230)	6994.44 (2821)	7430.97 (3092)	8209.26 (3355)
Share of food expenditure in total income	28.27	27.83	21.69	16.29
Own produce share	3.54	1.60	1.09	0.41
Eating out share	1.22	2.40	2.56	3.39
Observations	444	397	350	338

Table 2.5: Results 1987

	(1)	
	Log yearly food consumption	
Log disp. income	0.208***	(0.0325)
Self-employed	0.0774***	(0.0176)
Married cpl 1kd	0.0896***	(0.0274)
Married cpl 2kd	0.185***	(0.0280)
Married cpl 3kd	0.218***	(0.0318)
Married cpl 4+kd	0.241***	(0.0399)
1 adult with kd	0.168**	(0.0751)
Other HH	0.368***	(0.0295)
Share food outside	0.00195	(0.00188)
Lower professional	0.00917	(0.0154)
Manual worker	0.0326**	(0.0145)
Female dummy	-0.0333	(0.0221)
Kids less than 5	0.0513***	(0.0108)
Kids 5-13	0.0798***	(0.00849)
Kids 14-18	0.0947***	(0.00837)
CES over 45	0.0112	(0.0202)
Spouse 25-34	0.0979***	(0.0320)
Spouse 35-44	0.193***	(0.0326)
Spouse 45-54	0.249***	(0.0356)
Spouse over 55	0.219***	(0.0375)
Spouse parttime	0.0179	(0.0273)
Spouse SE	0.00516	(0.0402)
Spouse other occ	0.0455**	(0.0214)
Dublin	-0.0343**	(0.0155)
Town less 20000	-0.0241	(0.0182)
Town 3000-20000	-0.0307*	(0.0176)
Town less than 3000	-0.0400	(0.0278)
Monthly dummies	Yes	Yes
Observations	3206	
R ²	0.406	
Underid test χ^2 (p-value)	622.62	0.000
Weak identification test	43.90	
Hansen's J (p-value)	33.23	0.125

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.6: Results 1994

	(1)	
	Log yearly food consumption	
Log disp.income	0.254***	(0.0360)
Self-employed	0.0764***	(0.0189)
Married cpl 1kd	0.147***	(0.0279)
Married cpl 2kd	0.230***	(0.0278)
Married cpl 3kd	0.245***	(0.0339)
Married cpl 4+kd	0.234***	(0.0461)
1 adult with kd	0.217***	(0.0595)
Other HH	0.323***	(0.0300)
Share food outside	0.00155	(0.00156)
Lower professional	-0.00770	(0.0161)
Manual worker	0.0360**	(0.0159)
Female dummy	0.0626***	(0.0221)
Kids less than 5	0.0598***	(0.0118)
Kids 5-13	0.0900***	(0.00962)
Kids 14-18	0.142***	(0.00945)
CES over 45	0.0166	(0.0209)
Spouse 25-34	0.194***	(0.0395)
Spouse 35-44	0.275***	(0.0393)
Spouse 45-54	0.317***	(0.0404)
Spouse over 55	0.264***	(0.0450)
Spouse parttime	0.0766***	(0.0251)
Spouse SE	0.0373	(0.0372)
Spouse other occ	0.0953***	(0.0216)
Dublin	-0.0439***	(0.0165)
Town less 20000	-0.0590***	(0.0204)
Town 3000-20000	-0.0217	(0.0171)
Town less than 3000	0.0692**	(0.0273)
Monthly dummies	Yes	Yes
Observations	2794	
R^2	0.449	
Underid test χ^2 (<i>p-value</i>)	517.05	0.000
Weak identification test	39.76	
Hansen's J (<i>p-value</i>)	31.15	0.17

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.7: Results 1999

	(1)	
	Log yearly food consumption	
Log disp. income	0.193***	(0.0404)
Self-employed	0.0609***	(0.0230)
Married cpl 1kd	0.164***	(0.0286)
Married cpl 2kd	0.242***	(0.0280)
Married cpl 3kd	0.250***	(0.0346)
Married cpl 4+kd	0.179***	(0.0447)
1 adult with kd	0.170***	(0.0498)
Other HH	0.325***	(0.0297)
Share food outside	0.00466***	(0.00141)
Manual worker	0.0139	(0.0196)
Female dummy	0.0216	(0.0192)
Lower professional	0.000816	(0.0174)
Kids less 5	0.0688***	(0.0121)
Kids 5-13	0.100***	(0.00932)
Kids 14-18	0.140***	(0.00916)
CES over 45	0.0256	(0.0188)
Spouse 25-34	0.143***	(0.0358)
Spouse 35-44	0.223***	(0.0358)
Spouse 45-54	0.267***	(0.0354)
Spouse over 55	0.226***	(0.0391)
Spouse parttime	0.0570***	(0.0195)
Spouse SE	0.0255	(0.0322)
Spouse other occ	0.0820***	(0.0204)
Dublin	-0.0428**	(0.0174)
Town less 20000	-0.0486**	(0.0198)
Town 3000-20000	-0.0239	(0.0185)
Town less than 3000	0.00293	(0.0312)
Monthly dummies	Yes	Yes
Observations	3423	
R^2	0.382	
Underid test χ^2 (p-value)	523.76	0.000
Weak identification test	37.8	
Hansen's J (p-value)	31.15	0.17

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.8: Results 2004

	(1)	
	Log yearly food consumption	
Log disp. income	0.164***	(0.0463)
Self-employed	0.0449*	(0.0252)
Married cpl 1kd	0.134***	(0.0285)
Married cpl 2kd	0.198***	(0.0322)
Married cpl 3kd	0.186***	(0.0412)
Married cpl 4+kd	0.148***	(0.0568)
1 adult with kd	0.0149	(0.0586)
Other HH	0.313***	(0.0301)
Share food outside	0.00510***	(0.00185)
Manual worker	0.0614**	(0.0244)
Female dummy	0.0128	(0.0189)
Lower professional	-0.00431	(0.0197)
Kids less than 5	0.103***	(0.0153)
Kids 5-13	0.124***	(0.0134)
Kids 14-18	0.162***	(0.0109)
CES over 45	0.0199	(0.0220)
Spouse 25-34	0.141***	(0.0430)
Spouse 35-44	0.224***	(0.0411)
Spouse 45-54	0.271***	(0.0398)
Spouse over 55	0.322***	(0.0382)
Spouse parttime	0.00852	(0.0198)
Spouse SE	0.0107	(0.0313)
Spouse other occ	0.0383	(0.0235)
Dublin	-0.0343*	(0.0205)
Town less 20000	-0.0592**	(0.0265)
Town 3000-20000	-0.0220	(0.0183)
Town less than 3000	-0.0484**	(0.0238)
Monthly dummies	Yes	Yes
Observations	3441	
R^2	0.333	
Underid test χ^2 (p-value)	436.18	0.000
Weak identification test	32.36	
Hansen's J (p-value)	47.52	0.101

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.9: Under-reporting bounds SE HH (1987-2004)

Variable	Income (β)	SE (δ)	Lower bound (k_l)	Upper bound (k_u)
1987	0.208*** (0.033)	0.077*** (0.018)	1.40	1.49
1994	0.254*** (0.036)	0.076*** (0.019)	1.30	1.39
1999	0.193*** (0.040)	0.061*** (0.023)	1.31	1.43
2004	0.164*** (0.046)	0.045* (0.025)	1.30	1.33

Table 2.10: Under-reporting bounds SE Dublin based CES (1987-2004)

Variable	Income (β)	SE dummy (δ)	Lower bound (k_l)	Upper bound (k_u)
1987	0.211*** (0.045)	0.056* (0.031)	1.20	1.42
1994	0.440*** (0.069)	0.132*** (0.044)	1.05	1.86
1999	0.312*** (0.077)	0.000 (0.046)	1.00	1.11
2004	0.165** (0.080)	0.000 (0.050)	1.01	1.11

Table 2.11: Under-reporting bounds SE Older CES (1987-2004)

Variable	Income (β)	SE dummy (δ)	Lower bound (k_l)	Upper bound (k_u)
1987	0.250*** (0.063)	0.052* (0.028)	1.10	1.38
1994	0.280*** (0.066)	0.000 (0.031)	1.10	1.11
1999	0.236*** (0.059)	0.066* (0.035)	1.22	1.42
2004	0.143*** (0.065)	0.000 (0.037)	1.00	1.07

Table 2.12: Under-reporting bounds SE Low Professional CES (1987-2004)

Variable	Income (β)	SE dummy (δ)	Lower bound (k_l)	Upper bound (k_u)
1987	0.289*** (0.063)	0.148*** (0.038)	1.41	1.97
1994	0.280*** (0.060)	0.000 (0.041)	1.00	1.40
1999	0.169** (0.075)	0.118*** (0.057)	1.59	2.54
2004	0.132* (0.070)	0.000 (0.044)	1.00	1.18

Table 2.13: Under-reporting bounds SE Manual CES (1987-2004)

Variable	Income (β)	SE dummy (δ)	Lower bound (k_l)	Upper bound (k_u)
1987	0.229*** (0.060)	0.116*** (0.028)	1.56	1.77
1994	0.252*** (0.054)	0.078*** (0.026)	1.27	1.46
1999	0.165*** (0.057)	0.000 (0.070)	1.00	1.25
2004	0.264*** (0.088)	0.000 (0.073)	1.00	1.29

Table 2.14: Under-reporting bounds Farm HH (1987-2004)

Variable	Income (β)	Farm dummy (δ)	Lower bound (k_l)	Upper bound (k_u)
1987	0.215*** (0.028)	0.072*** (0.019)	1.29	1.50
1994	0.237*** (0.030)	0.098*** (0.020)	1.43	1.59
1999	0.202*** (0.030)	0.059*** (0.021)	1.22	1.47

Table 2.15: Under-reporting bounds Farm HH inc. own produce (1987-2004)

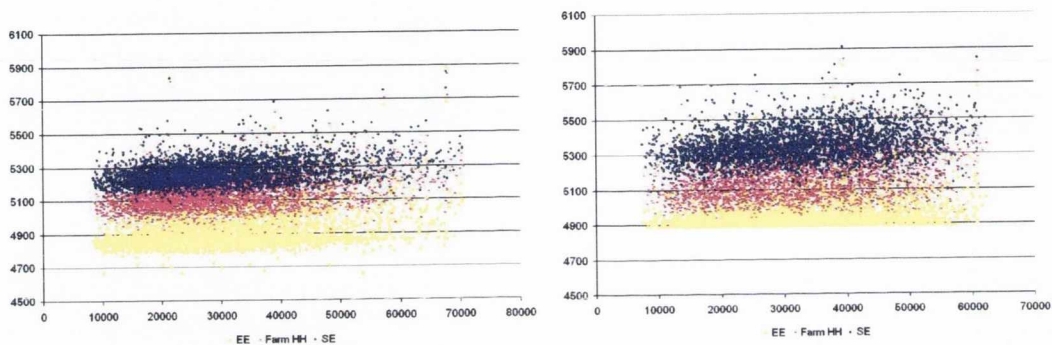
Variable	Income (β)	Farm dummy (δ)	Lower bound (k_l)	Upper bound (k_u)
1987	0.221*** (0.028)	0.161*** (0.020)	1.91	2.24
1994	0.244*** (0.030)	0.136*** (0.020)	1.70	1.90
1999	0.211*** (0.030)	0.104*** (0.022)	1.49	1.79

Table 2.16: Under-reporting bounds, construction industry (2004)

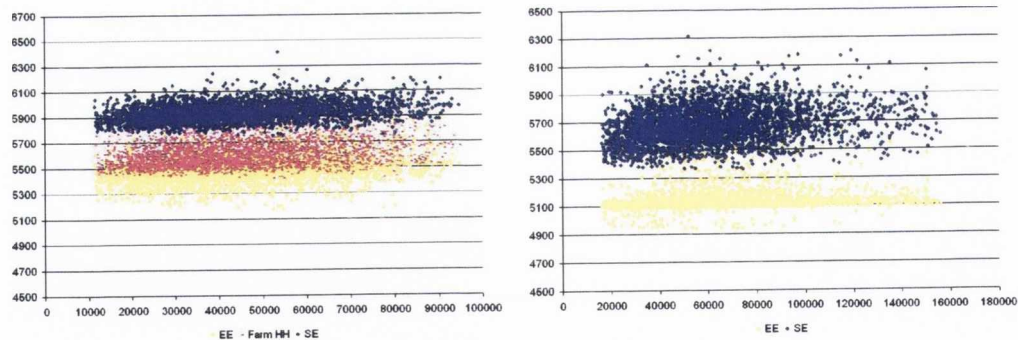
Variable	Income (β)	Coefficient (δ)	Lower bound (k_l)	Upper bound (k_u)
SE	0.378*** (0.099)	0.000 (0.074)	1.10	1.12

2.8 Figures

Figure 2.1: Non-parametric SE, EE, Farm plots

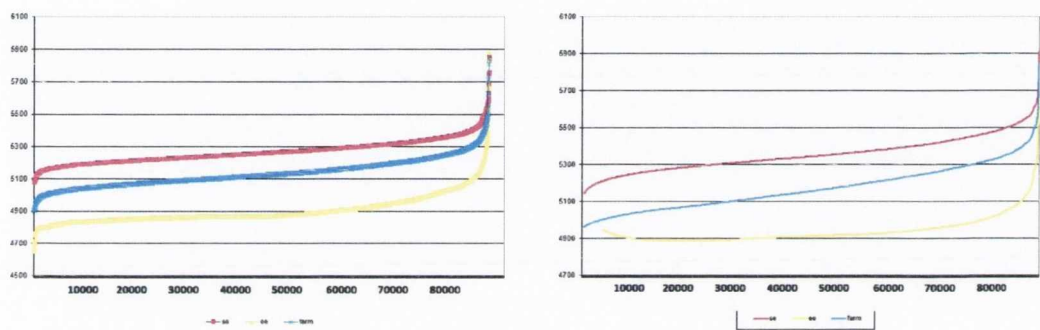


(a) 1987 and 1994

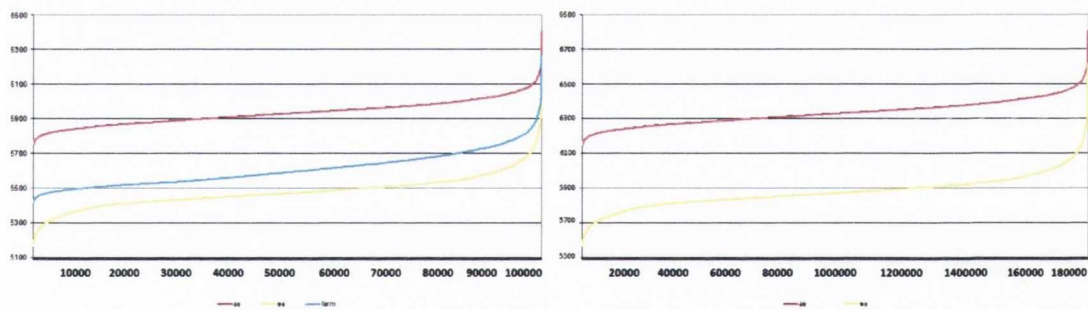


(b) 1999 and 2004

Figure 2.2: Non-parametric SE, EE, Farm smoothed

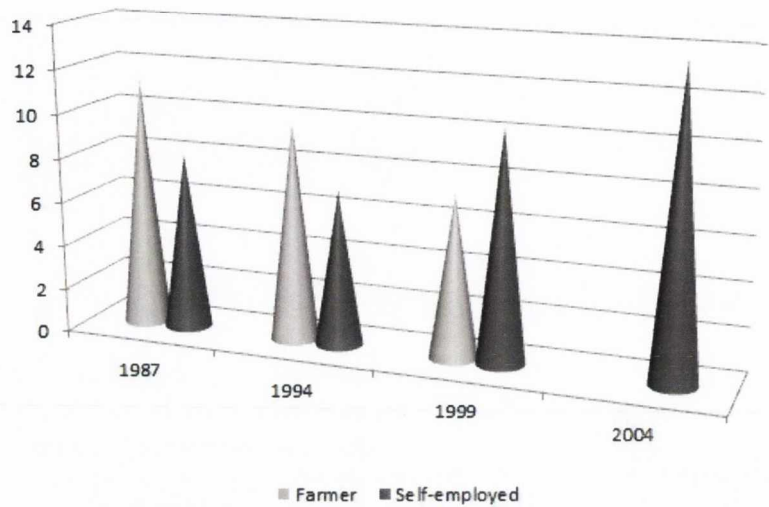


(a) 1987 and 1994



(b) 1999 and 2004

Figure 2.3: Lost income due to tax evasion as a proportion of total taxable income



2.9 Appendix

2.9.1 Formula derivations

The following section presents the derivation of the lower and upper bounds of compliance as detailed in Pissarides and Weber (1989).

Y_i^R is income as reported by household i
 Y_i^T is household i 's true income
 Y_i^P is household i 's permanent income
 Xf_i is the expenditure on food of household i

Subscripts SE and EE refer to the self-employed and employees respectively. The relationship between true income (Y_i^T) and reported income (Y_i^R) is:

$$Y_i^T = k_i Y_i^R \quad (2.5)$$

where k is a random variable and $k = 1$ for wage earners (it is assumed they don't lie) and $k > 1$ for self-employed (who have the potential to lie). Consequently, the larger the k the larger the degree of non-compliance by the self-employed.

The expenditure equation is:

$$\ln Xf_i = \alpha Z_i + \beta \ln Y_i^P + \varepsilon_i \quad (2.6)$$

where $\ln Xf_i$ is the log of food expenditure for household i , Z is a vector of exogenous variables such as household characteristics, ε_i is the error term, β is the expenditure elasticity of food and Y_i^P is the measure of income that influences consumption decisions, referred to as "permanent income". It is assumed that households correctly report their food expenditure.

It is also assumed that for a given level of permanent income, self-employed income will be subject to more variation than that of employees. This will affect the measures of under-reporting. PW assume that permanent and measured income are linked as follows:

$$Y_i^T = p Y_i^P \quad (2.7)$$

where p is a random variable. The expected value of p in each year will depend on aggregate events in that year. For instance, in a financially beneficial year p will on

average be larger than one. PW assume that the mean of p is the same for all households regardless of their work sector but that the variance of p may be different for the self-employed and employees. It is expected that the variance of p will be bigger for the self-employed as they are more sensitive to economic shocks and their incomes will adjust to shocks faster. Using 2.5 and 2.7 permanent income is:

$$\ln Y_i^P = \ln Y_i^R - \ln p_i + \ln k_i \quad (2.8)$$

As no information on p_i and k_i is available it is assumed that they are distributed as log-normal random variables. Writing p_i and k_i as deviations from their means:

$$\ln p_i = \mu_p + u_i \quad (2.9)$$

and

$$\ln k_i = \mu_k + v_i \quad (2.10)$$

The random variables u_i and v_i have zero means and constant variances, σ_u^2 and σ_v^2 .

A self-employment indicator SE_i is introduced into the expenditure equation (2.1), giving:

$$\ln X f_i = \alpha Z_i + \beta \ln Y_i^P + \delta SE_i + \varepsilon_i \quad (2.11)$$

where α and β are the same for all households. Concealed income by the self-employed will result in the intercepts for the two groups being different, as well as differences due to the difference in the variance of p . Consequently, PW show that:

$$\delta = \beta[\mu_k + 0.5(\sigma_{uSE}^2 - \sigma_{uEE}^2)] \quad (2.12)$$

What we are aiming to estimate is the average value of k_i . Assuming log normality, the log of k_i is:

$$\ln \bar{k} = \mu_k + \frac{1}{2}\sigma_{vSE}^2 \quad (2.13)$$

Using equations (2.9), (2.10), (2.11) and (2.13) $\bar{k}$ can be written as:

$$\ln \bar{k} = \frac{\delta}{\beta} + \frac{1}{2}(\sigma_{vSE}^2 - \sigma_{uSE}^2 + \sigma_{uEE}^2) \quad (2.14)$$

We do not observe p_i and k_i so no single estimate of the degree of under-reporting can be calculated. However lower and upper bounds can be calculated by substituting in values for δ , β , $\sigma_{\lambda SE}^2$ and $\sigma_{\lambda EE}^2$.

It is then possible to estimate upper and lower bounds to compliance using the fol-

lowing formulas:

$$lnk_u = \frac{\delta}{\beta} + \frac{1}{2}(\sigma_{\lambda SE}^2 - \sigma_{\lambda EE}^2) \quad (2.15)$$

$$lnk_l = \frac{\delta}{\beta} - \frac{1}{2}(\sigma_{\lambda SE}^2 - \sigma_{\lambda EE}^2) \quad (2.16)$$

where $\sigma_{\lambda SE}^2$ is the residual variance from the first stage regression for the self-employed and $\sigma_{\lambda EE}^2$ is the residual variance from the first stage regression for employees.

2.9.2 Summary statistics explanatory variables, by year

Table 2.17: Summary Statistics Explanatory Variables 1987

Variable	Mean	Std. Dev.	Min.	Max.
Number of people in household	4.39	1.71	2	13
Female dummy	0.134	0.341	0	1
Number of kids < 5 years of age	0.536	0.774	0	4
Number of kids 5-13 years of age	1.086	1.242	0	7
Number of kids 14-18 years of age	0.461	0.811	0	6
CES over 45	0.313	0.464	0	1
Dummy if spouse is less than 25 years old	0.108	0.311	0	1
Dummy if spouse is 25-34 years old	0.331	0.471	0	1
Dummy if spouse is 35-44 years old	0.321	0.467	0	1
Dummy if spouse is 45-54 years old	0.166	0.372	0	1
Dummy if spouse is over 55 years old	0.074	0.261	0	1
Rural area	0.375	0.484	0	1
Dublin Metropolitan Area	0.272	0.445	0	1
Town bigger than 20000	0.163	0.369	0	1
Town between 3000 and 20000	0.135	0.342	0	1
Town less than 3000	0.055	0.228	0	1
Urban non farm household	0.625	0.484	0	1
Rural non farm household	0.218	0.413	0	1
Rural farm household	0.157	0.364	0	1
Household surveyed in January	0.048	0.214	0	1
Household surveyed in February	0.079	0.27	0	1
Household surveyed in March	0.145	0.352	0	1
Household surveyed in April	0.09	0.286	0	1
Household surveyed in May	0.085	0.278	0	1
Household surveyed in June	0.089	0.284	0	1
Household surveyed in July	0.079	0.27	0	1
Household surveyed in August	0.062	0.241	0	1
Household surveyed in September	0.08	0.271	0	1
Household surveyed in October	0.084	0.278	0	1
Household surveyed in November	0.09	0.286	0	1
Household surveyed in December	0.07	0.255	0	1
N		3650		

Table 2.18: Summary Statistics Explanatory Variables 1994

Variable	Mean	Std. Dev.	Min.	Max.
Number of people in household	4.092	1.546	2	14
Female dummy	0.188	0.391	0	1
Number of kids < 5 years of age	0.409	0.676	0	4
Number of kids 5-13 years of age	0.939	1.117	0	6
Number of kids 14-18 years of age	0.498	0.818	0	5
CES over 45	0.348	0.477	0	1
Dummy if spouse is less than 25 years old	0.118	0.323	0	1
Dummy if spouse is 25-34 years old	0.254	0.435	0	1
Dummy if spouse is 35-44 years old	0.349	0.477	0	1
Dummy if spouse is 45-54 years old	0.21	0.407	0	1
Dummy if spouse is over 55 years old	0.069	0.254	0	1
Rural area	0.399	0.49	0	1
Dublin Metropolitan Area	0.232	0.422	0	1
Town bigger than 20000	0.116	0.32	0	1
Town between 3000 and 20000	0.187	0.39	0	1
Town less than 3000	0.042	0.201	0	1
Urban non farm household	0.601	0.49	0	1
Rural non farm household	0.259	0.438	0	1
Rural farm household	0.139	0.346	0	1
Household surveyed in January	0.071	0.257	0	1
Household surveyed in February	0.074	0.262	0	1
Household surveyed in March	0.087	0.283	0	1
Household surveyed in April	0.071	0.257	0	1
Household surveyed in May	0.084	0.277	0	1
Household surveyed in June	0.109	0.311	0	1
Household surveyed in July	0.08	0.271	0	1
Household surveyed in August	0.082	0.275	0	1
Household surveyed in September	0.1	0.3	0	1
Household surveyed in October	0.077	0.266	0	1
Household surveyed in November	0.09	0.286	0	1
Household surveyed in December	0.076	0.265	0	1
N		3191		

Table 2.19: Summary Statistics Explanatory Variables 1999

Variable	Mean	Std. Dev.	Min.	Max.
Number of people in household	3.866	1.426	2	12
Female dummy	0.217	0.412	0	1
Number of kids < 5 years of age	0.384	0.676	0	5
Number of kids 5-13 years of age	0.773	0.987	0	5
Number of kids 14-18 years of age	0.457	0.783	0	5
CES over 45	0.39	0.488	0	1
Dummy if spouse is less than 25 years old	0.133	0.34	0	1
Dummy if spouse is 25-34 years old	0.198	0.398	0	1
Dummy if spouse is 35-44 years old	0.339	0.473	0	1
Dummy if spouse is 45-54 years old	0.226	0.418	0	1
Dummy if spouse is over 55 years old	0.104	0.306	0	1
Rural area	0.488	0.5	0	1
Dublin Metropolitan Area	0.215	0.411	0	1
Town bigger than 20000	0.13	0.337	0	1
Town between 3000 and 20000	0.143	0.351	0	1
Town less than 3000	0.024	0.153	0	1
Urban non farm household	0.512	0.5	0	1
Rural non farm household	0.382	0.486	0	1
Rural farm household	0.105	0.307	0	1
Household surveyed in January	0.053	0.224	0	1
Household surveyed in February	0.074	0.262	0	1
Household surveyed in March	0.088	0.283	0	1
Household surveyed in April	0.078	0.269	0	1
Household surveyed in May	0.08	0.271	0	1
Household surveyed in June	0.136	0.343	0	1
Household surveyed in July	0.138	0.344	0	1
Household surveyed in August	0.082	0.275	0	1
Household surveyed in September	0.079	0.27	0	1
Household surveyed in October	0.071	0.256	0	1
Household surveyed in November	0.064	0.246	0	1
Household surveyed in December	0.056	0.23	0	1
N		3773		

Table 2.20: Summary Statistics Explanatory Variables 2004

Variable	Mean	Std. Dev.	Min.	Max.
Number of people in household	3.668	1.356	2	10
Female dummy	0.28	0.449	0	1
Number of kids < 5 years of age	0.352	0.662	0	4
Number of kids 5-13 years of age	0.696	0.950	0	6
Number of kids 14-18 years of age	0.406	0.736	0	5
CES over 45	0.422	0.494	0	1
Dummy if spouse is less than 25 years old	0.14	0.347	0	1
Dummy if spouse is 25-34 years old	0.177	0.382	0	1
Dummy if spouse is 35-44 years old	0.31	0.463	0	1
Dummy if spouse is 45-54 years old	0.241	0.428	0	1
Dummy if spouse is over 55 years old	0.132	0.338	0	1
Rural area	0.379	0.485	0	1
Dublin Metropolitan Area	0.22	0.414	0	1
Town bigger than 20000	0.084	0.277	0	1
Town between 3000 and 20000	0.216	0.412	0	1
Town less than 3000	0.101	0.301	0	1
Urban non farm household	0.621	0.485	0	1
Rural non farm household	0.266	0.442	0	1
Rural farm household	0.113	0.317	0	1
Household surveyed in January	0.061	0.24	0	1
Household surveyed in February	0.077	0.267	0	1
Household surveyed in March	0.079	0.27	0	1
Household surveyed in April	0.082	0.274	0	1
Household surveyed in May	0.084	0.277	0	1
Household surveyed in June	0.077	0.267	0	1
Household surveyed in July	0.075	0.264	0	1
Household surveyed in August	0.083	0.275	0	1
Household surveyed in September	0.07	0.255	0	1
Household surveyed in October	0.086	0.281	0	1
Household surveyed in November	0.138	0.345	0	1
Household surveyed in December	0.088	0.283	0	1
N	3779			

2.9.3 Summary statistics instrumental variables, by year

Table 2.21: Summary Statistics Instrumental Variables 1987

Variable	Mean	Std. Dev.	Min.	Max.
Dummy CES has junior certificate	0.538	0.499	0	1
Dummy CES has leaving certificate	0.295	0.456	0	1
Dummy CES has college education	0.167	0.373	0	1
Dummy spouse has junior certificate	0.537	0.499	0	1
Dummy spouse has leaving certificate	0.354	0.478	0	1
Dummy spouse has college education	0.109	0.312	0	1
Deflated mortgage amount	16573.691	21587.39	0	112966.023
Dummy for renters	0.13	0.337	0	1
Dummy for households with a mortgage	0.569	0.495	0	1
Dummy for households who own their home	0.301	0.459	0	1
Number of bedrooms	3.362	0.826	1	10
Live in a flat	0.022	0.148	0	1
Live in a detached house	0.468	0.499	0	1
Live in a terraced or semi-d house	0.509	0.5	0	1
Household has a savings account	0.446	0.497	0	1
Household owns a washing maching	0.941	0.235	0	1
Household owns a dish washer	0.121	0.326	0	1
Household owns a fridge	0.362	0.481	0	1
Household owns a deep freeze	0.24	0.427	0	1
Household has double glazing	0.168	0.374	0	1
Household has a burglar alarm	0.048	0.214	0	1
Household has a bank account	0.718	0.45	0	1
Household has a spin dryer	0.318	0.466	0	1
Household owns a second home	0.031	0.174	0	1
Household has a phone	0.652	0.476	0	1
Household has gas heating	0.114	0.317	0	1
Household owns a television	0.932	0.252	0	1
Household has central heating	0.700	0.458	0	1
Household owns a microwave	0.094	0.291	0	1
Household owns a computer	0.102	0.303	0	1

Table 2.22: Summary Statistics Instrumental Variables 1994

Variable	Mean	Std. Dev.	Min.	Max.
Dummy CES has junior certificate	0.533	0.499	0	1
Dummy CES has leaving certificate	0.293	0.455	0	1
Dummy CES has college education	0.174	0.379	0	1
Dummy spouse has junior certificate	0.44	0.496	0	1
Dummy spouse has leaving certificate	0.343	0.475	0	1
Dummy spouse has college education	0.104	0.305	0	1
Deflated mortgage amount	23798.808	28260.591	0	186178.594
Dummy for renters	0.101	0.302	0	1
Dummy for households with a mortgage	0.575	0.494	0	1
Dummy for households who own their home	0.323	0.468	0	1
Number of bedrooms	3.399	0.822	1	8
Live in a flat	0.026	0.158	0	1
Live in a detached house	0.526	0.499	0	1
Live in a terraced or semi-d house	0.448	0.497	0	1
Household has a savings account	0.25	0.433	0	1
Household owns a washing machine	0.977	0.15	0	1
Household owns a dish washer	0.264	0.441	0	1
Household owns a fridge	0.613	0.487	0	1
Household owns a deep freeze	0.307	0.461	0	1
Household has double glazing	0.405	0.491	0	1
Household has a burglar alarm	0.138	0.345	0	1
Household has a bank account	0.685	0.465	0	1
Household has a spin dryer	0.356	0.479	0	1
Household owns a second home	0.021	0.142	0	1
Household has a phone	0.843	0.364	0	1
Household has gas heating	0.165	0.372	0	1
Household owns a television	0.976	0.154	0	1
Household has central heating	0.846	0.361	0	1
Household owns a microwave	0.625	0.484	0	1
Household owns a computer	0.229	0.421	0	1
Household has access to the internet	0	0	0	0

Table 2.23: Summary Statistics Instrumental Variables 1999

Variable	Mean	Std. Dev.	Min.	Max.
Dummy CES has junior certificate	0.413	0.492	0	1
Dummy CES has leaving certificate	0.305	0.46	0	1
Dummy CES has college education	0.281	0.45	0	1
Dummy spouse has junior certificate	0.341	0.474	0	1
Dummy spouse has leaving certificate	0.346	0.476	0	1
Dummy spouse has college education	0.187	0.39	0	1
Deflated mortgage amount	26914.991	32233.211	0	128495.469
Dummy for renters	0.109	0.312	0	1
Dummy for households with a mortgage	0.53	0.499	0	1
Dummy for households who own their home	0.361	0.48	0	1
Number of bedrooms	3.505	0.908	1	9
Live in a flat	0.022	0.148	0	1
Live in a detached house	0.538	0.499	0	1
Live in a terraced or semi-d house	0.439	0.496	0	1
Household has a savings account	0.336	0.472	0	1
Household owns a washing maching	0.99	0.097	0	1
Household owns a dish washer	0.45	0.498	0	1
Household owns a fridge	0.394	0.489	0	1
Household owns a deep freeze	0.394	0.489	0	1
Household has double glazing	0.622	0.485	0	1
Household has a burglar alarm	0.243	0.429	0	1
Household has a bank account	0.805	0.396	0	1
Household has a spin dryer	0.537	0.499	0	1
Household owns a second home	0.043	0.203	0	1
Household has a phone	0.938	0.241	0	1
Household has gas heating	0.233	0.423	0	1
Household owns a television	0.984	0.124	0	1
Household has central heating	0.912	0.283	0	1
Household owns a microwave	0.822	0.382	0	1
Household owns a computer	0.4	0.49	0	1
Household has access to the internet	1	0	1	1

Table 2.24: Summary Statistics Instrumental Variables 2004

Variable	Mean	Std. Dev.	Min.	Max.
Dummy CES has junior certificate	0.311	0.463	0	1
Dummy CES has leaving certificate	0.301	0.459	0	1
Dummy CES has college education	0.388	0.487	0	1
Dummy spouse has junior certificate	0.255	0.436	0	1
Dummy spouse has leaving certificate	0.314	0.464	0	1
Dummy spouse has college education	0.29	0.454	0	1
Deflated mortgage amount	56709.299	84238.267	0	956521.75
Dummy for renters	0.116	0.32	0	1
Dummy for households with a mortgage	0.523	0.5	0	1
Dummy for households who own their home	0.361	0.48	0	1
Number of bedrooms	3.631	0.942	1	10
Live in a flat	0.016	0.126	0	1
Live in a detached house	0.574	0.495	0	1
Live in a terraced or semi-d house	0.41	0.492	0	1
Household has a savings account	0.349	0.477	0	1
Household owns a washing machine	0.995	0.069	0	1
Household owns a dish washer	0.697	0.46	0	1
Household owns a fridge	0.655	0.475	0	1
Household owns a deep freeze	0.458	0.498	0	1
Household has double glazing	0.833	0.373	0	1
Household has a burglar alarm	0.355	0.479	0	1
Household has a bank account	0.895	0.306	0	1
Household has a spin dryer	0.757	0.429	0	1
Household owns a second home	0.061	0.24	0	1
Household has a phone	0.898	0.302	0	1
Household has gas heating	0.3	0.458	0	1
Household owns a television	0.993	0.086	0	1
Household has central heating	0.961	0.195	0	1
Household owns a microwave	0.917	0.275	0	1
Household owns a computer	0.758	0.428	0	1
Household has access to the internet	0.597	0.491	0	1

2.9.4 Farming regression results

Table 2.25: Results 1987, Farmers

	(1)	
	Log yearly food consumption	
Log disp. income	0.215***	(0.0282)
Farmer	0.0718***	(0.0187)
Married cpl 1kd	0.101***	(0.0276)
Married cpl 2kd	0.171***	(0.0281)
Married cpl 3kd	0.203***	(0.0316)
Married cpl 4+kd	0.233***	(0.0384)
1 adult with kd	0.140*	(0.0725)
Other HH	0.354***	(0.0286)
Share food outside	0.00118	(0.00196)
Female dummy	-0.0130	(0.0213)
Kids less than 5	0.0535***	(0.0106)
Kids 5-13	0.0848***	(0.00792)
Kids 14-18	0.0974***	(0.00813)
CES over 45	0.0279	(0.0190)
Spouse 25-34	0.111***	(0.0307)
Spouse 35-44	0.185***	(0.0314)
Spouse 45-54	0.250***	(0.0337)
Spouse over 55	0.220***	(0.0351)
Spouse parttime	0.0252	(0.0271)
Spouse SE	-0.0367	(0.0598)
Spouse other occ	0.0637***	(0.0213)
Dublin	-0.0400**	(0.0158)
Town less 20000	-0.0338*	(0.0191)
Town 3000-20000	-0.0339*	(0.0180)
Town less than 3000	-0.0638**	(0.0309)
Monthly dummies	Yes	Yes
Observations	3279	
R^2	0.422	
Underid test χ^2 (p -value)	782.95	0.000
Weak identification test	64.19	
Hansen's J (p -value)	28.47	0.16

Robust standard errors in parentheses.

No self-employed households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.26: Results 1994, Farmers

	(1)	
	Log yearly food consumption	
Log disp. income	0.237***	(0.0295)
Farmer	0.0984***	(0.0196)
Married cpl 1kd	0.150***	(0.0277)
Married cpl 2kd	0.213***	(0.0275)
Married cpl 3kd	0.233***	(0.0329)
Married cpl 4+kd	0.213***	(0.0443)
1 adult with kd	0.213***	(0.0558)
Other HH	0.333***	(0.0289)
Share food outside	0.00219	(0.00156)
Female dummy	0.0461**	(0.0208)
Kids less than 5	0.0639***	(0.0119)
Kids 5-13	0.0945***	(0.00904)
Kids 14-18	0.134***	(0.00916)
CES over 45	0.0220	(0.0201)
Spouse 25-34	0.164***	(0.0375)
Spouse 35-44	0.250***	(0.0367)
Spouse 45-54	0.277***	(0.0379)
Spouse over 55	0.267***	(0.0410)
Spouse parttime	0.0779***	(0.0253)
Spouse SE	0.0502	(0.0389)
Spouse other occ	0.0984***	(0.0207)
Dublin	-0.0568***	(0.0171)
Town less 20000	-0.0648***	(0.0215)
Town 3000-20000	-0.0314*	(0.0177)
Town less than 3000	0.0570**	(0.0289)
Monthly dummies	Yes	Yes
<i>Observations</i>	2831	
<i>R</i> ²	0.453	
<i>Underid test</i> χ^2 (<i>p-value</i>)	635	0.000
<i>Weak identification test</i>	56.19	
<i>Hansen's J</i> (<i>p-value</i>)	34.28	0.05

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.27: Results 1999, Farmers

	(1)	
	Log yearly food consumption	
Log disp. income	0.202***	(0.0299)
Farmer	0.0593***	(0.0213)
Married cpl 1kd	0.133***	(0.0284)
Married cpl 2kd	0.240***	(0.0275)
Married cpl 3kd	0.245***	(0.0345)
Married cpl 4+kd	0.204***	(0.0431)
1 adult with kd	0.156***	(0.0505)
Other HH	0.326***	(0.0275)
Share food outside	0.00548***	(0.00166)
Female dummy	0.0184	(0.0178)
Kids less than 5	0.0688***	(0.0119)
Kids 5-13	0.101***	(0.00943)
Kids 14-18	0.138***	(0.00908)
CES over 45	0.0503***	(0.0182)
Spouse 25-34	0.152***	(0.0339)
Spouse 35-44	0.209***	(0.0329)
Spouse 45-54	0.230***	(0.0327)
Spouse over 55	0.188***	(0.0356)
Spouse parttime	0.0525***	(0.0186)
Spouse SE	0.0448	(0.0411)
Spouse other occ	0.0923***	(0.0187)
Dublin	-0.0453**	(0.0179)
Town less 20000	-0.0551***	(0.0209)
Town 3000-20000	-0.0281	(0.0196)
Town less than 3000	-0.00694	(0.0338)
Monthly dummies	Yes	Yes
Observations	3306	
R^2	0.389	
Underid test χ^2 (p -value)	694.10	0.000
Weak identification test	44.89	
Hansen's J (p -value)	33.32	0.260

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2.9.5 Older CES regression results

Table 2.28: Results 1987, Farmers including own produce

	(1)	
	Food outside	
Log disp. income	0.221***	(0.0284)
Farmer	0.161***	(0.0198)
Married cpl 1kid	0.104***	(0.0280)
Married cpl 2kid	0.170***	(0.0285)
Married cpl 3kid	0.201***	(0.0323)
Married cpl 4+kid	0.229***	(0.0392)
1 adult with kid	0.133*	(0.0723)
Other HH	0.360***	(0.0290)
Share food outside	0.000492	(0.00196)
Female dummy	-0.00732	(0.0214)
Kids less than 5	0.0579***	(0.0108)
Kids 5-13	0.0886***	(0.00808)
Kids 14-18	0.0988***	(0.00819)
CES over 45	0.0254	(0.0193)
Spouse 25-34	0.108***	(0.0310)
Spouse 35-44	0.188***	(0.0317)
Spouse 45-54	0.260***	(0.0338)
Spouse over 55	0.227***	(0.0355)
Spouse parttime	0.0185	(0.0275)
Spouse SE	-0.0317	(0.0587)
Spouse other occ	0.0644***	(0.0215)
Dublin	-0.0620***	(0.0161)
Town less 20000	-0.0540***	(0.0193)
Town 3000-20000	-0.0581***	(0.0181)
Town less than 3000	-0.0889***	(0.0311)
Monthly dummies	Yes	Yes
Observations	3279	
R^2	0.443	
<i>Underid test χ^2 (p-value)</i>	782.95	0.000
<i>Weak identification test</i>	64.19	
<i>Hansen's J (p-value)</i>	32.91	0.16

Robust standard errors in parentheses.

No self-employed households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.29: Results 1994, Farmers including own produce

	(1)	
	Food outside	
Log disp. income	0.244***	(0.0297)
Farmer	0.143***	(0.0203)
Married cpl 1kid	0.151***	(0.0278)
Married cpl 2kid	0.210***	(0.0276)
Married cpl 3kid	0.232***	(0.0330)
Married cpl 4+kid	0.211***	(0.0443)
1 adult with kid	0.214***	(0.0560)
Other HH	0.329***	(0.0291)
Share food outside	0.00214	(0.00158)
Female dummy	0.0483**	(0.0208)
Kids less than 5	0.0648***	(0.0119)
Kids 5-13	0.0965***	(0.00896)
Kids 14-18	0.135***	(0.00910)
CES over 45	0.0236	(0.0201)
Spouse 25-34	0.163***	(0.0378)
Spouse 35-44	0.250***	(0.0370)
Spouse 45-54	0.279***	(0.0381)
Spouse over 55	0.276***	(0.0414)
Spouse parttime	0.0756***	(0.0253)
Spouse SE	0.0451	(0.0389)
Spouse other occ	0.101***	(0.0208)
Dublin	-0.0600***	(0.0171)
Town less 20000	-0.0682***	(0.0216)
Town 3000-20000	-0.0361**	(0.0178)
Town less than 3000	0.0519*	(0.0291)
Monthly dummies	Yes	Yes
Observations	2831	
R^2	0.461	
Underid test χ^2 (<i>p-value</i>)	635.00	0.000
Weak identification test	56.19	
Hansen's J (<i>p-value</i>)	37.28	0.05

Robust standard errors in parentheses.

No self-employed households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.30: Results 1999, Farmers including own produce

	(1)	
	Food outside	
Log disp. income	0.211***	(0.0300)
Farmer	0.104***	(0.0216)
Married cpl 1kid	0.129***	(0.0285)
Married cpl 2kid	0.235***	(0.0275)
Married cpl 3kid	0.242***	(0.0345)
married cpl 4+kid	0.202***	(0.0433)
1 adult with kid	0.156***	(0.0503)
Other HH	0.323***	(0.0276)
Share food outside	0.00532***	(0.00166)
Female dummy	0.0222	(0.0178)
Kids less than 5	0.0693***	(0.0120)
Kids 5-13	0.101***	(0.00947)
Kids 14-18	0.138***	(0.00909)
CES over 45	0.0485***	(0.0184)
Spouse 25-34	0.152***	(0.0339)
Spouse 35-44	0.210***	(0.0329)
Spouse 45-54	0.232***	(0.0328)
Spouse over 55	0.192***	(0.0359)
Spouse parttime	0.0545***	(0.0187)
Spouse SE	0.0428	(0.0419)
Spouse other occ	0.0959***	(0.0188)
Dublin	-0.0512***	(0.0180)
Town less 20000	-0.0603***	(0.0209)
Town 3000-20000	-0.0335*	(0.0196)
Town less than 3000	-0.0122	(0.0338)
Monthly dummies	Yes	Yes
Observations	3306	
R^2	0.393	
<i>Underid test</i> χ^2 (<i>p-value</i>)	694.00	0.000
<i>Weak identification test</i>	44.89	
<i>Hansen's J</i> (<i>p-value</i>)	34.17	0.233

Robust standard errors in parentheses.

No self-employed households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.31: First stage regressions - main sample

	1987	1994	1999	2004
SE	-0.0669	0.0974	-0.1084	0.1057
Lower professional	0.0162	0.0134	0.0152	0.0141
Manual worker	-0.0331**	-0.0375**	-0.0338**	-0.058**
Female dummy	-0.0983***	0.0184	-0.0605	-0.1536***
Kids less 5	0.0031	0.0095	-0.1026**	-0.0906***
Kids 5 to 13	0.017**	0.0216*	-0.007	-0.0241*
Kids 14 to 18	0.0097	0.008	-0.0111	-0.0352***
Older CES	-0.0141	0.0088	-0.0112	-0.0308***
Spouse part-time	-0.1587***	0.0187	0.0335**	0.0099
Spouse SE	-0.1167**	0.0225	-0.1345**	-0.0903***
Spouse other occ	-0.2537***	0.0465	-0.0924***	-0.1238***
Dublin	0.0813***	0.0157	-0.1621**	0.0351
Town less 20000	0.0105	0.0192	-0.2426***	-0.2579***
Town 3000-20000	0.009	0.0195	0.0203	0.1147***
Town less than 3000	-0.003	0.0191	0.0564***	0.0239
Savings	0.1098***	0.0245	0.017	0.0073
SE-Savings	0.003	0.0429	-0.0493	0.0384
Savings	0.1098***	0.0109	0.0892**	0.0118
SE-Bedrooms	-0.0037	0.0277	0.0108	0.0219
Bedrooms	0.0427***	0.0082	0.0423**	-0.0105
SE-Detached	0.0145	0.0429	0.0089	0.0078
Detached	-0.0026	0.016	-0.0297	0.0371
SE-deepfreeze	-0.0387	0.0464	0.0483	0.0216
Deepfreeze	0.0631***	0.0139	0.0394**	0.0536***
SE-Second home	-0.2817***	0.0923	0.0821**	0.0151
Second home	0.08***	0.0296	0.0939**	0.0371
SE-Medical card	0.039	0.0643	0.0133	0.097***
Medical card	-0.1639***	0.0167	0.028	-0.0159
SE-Cas	-0.0884	0.0795	-0.0105	0.0625***
Cas	-0.0321*	0.0171	0.0291**	0.0084
SE-Central heating	-0.0715	0.0541	0.0152	0.0078
Central heating	0.0562***	0.0134	0.0668	-0.0782
SE-Microwave	0.0676	0.0562	0.0169	-0.0016
Microwave	0.0311	0.0195	0.1026	0.032
Bank account			0.0517**	0.0222
SE-Bankaccount			0.0304**	0.0222
Computer			0.1146***	0.0222
SE-Computer			0.0092	0.0121
Spin dryer			0.0269*	0.0138
SE-Alarm			0.0083	0.0462
Burglar alarm			0.0442***	0.0124
SE-Double glazing			0.0144	0.0368
Double glazing			0.0845***	0.041
SE-Internet			0.0027	0.0137
Internet			0.0066	0.0832***
Constant	9.6643***	0.469	0.1026	0.0751
First stage F test (p value)	52.06	0	40.62	0

Table 2.32: First stage regressions - farmers

	1987	1994	1999
Farmer	-0.3718***	-0.1368	0.1044
Female dummy	-0.0546***	0.0192	-0.0808***
Kids less 5	0.0245***	0.0095	0.0175
Kids 5 to 13	0.0308***	0.0075	0.0408***
Kids 14 to 18	0.0316***	0.0089	0.0318***
Older CES	-0.0277	0.019	0.0043
Spouse part-time	-0.1856***	0.0242	0.0892
Spouse SE	-0.1915***	0.054	-0.0331*
Spouse other occ	-0.3262***	0.0267	0.0243
Dublin	0.0637***	0.0146	0.0579
Town less 20000	0.0086	0.0161	-0.1387***
Town 3000-20000	0.0074	0.0163	-0.3866***
Town less than 3000	-0.0121	0.024	0.0295
Savings	-0.0109	0.0455	0.0147
Farm*Savings	0.1108***	0.0108	0.038***
Farm*Bedrooms	0.0802***	0.026	0.0401**
Bedrooms	0.0288***	0.0081	0.0344**
Farm*deepfreeze	0.0184	0.0525	0.0159
Deepfreeze	0.05***	0.0135	0.0391
Farm*Second home	0.078***	0.013	
Second home	-0.1459***	-0.012	0.0207
Farm*Medical card		0.0466	0.0724**
Medical card		0.0165	
Farm*Gas			
Gas			0.3121***
Farm*Central heating			0.0338**
Central heating		0.0033	0.0618
Farm*Microwave		0.0618***	0.0161
Microwave			0.0514**
Bank account	0.0239	0.0532	0.0181
Farm*Bankaccount	0.1043***	0.0136	0.1376
Computer		-0.1181**	0.0499***
Farm*Computer			0.0843***
Spin dryer	-0.0469	0.0471	0.0778
Farm*Alarm	0.0261**	0.0118	0.0728***
Burglar alarm			0.075***
Constant	9.6762***	0.0454	0.0288**
First stage F test (p-value)	81.56	0	0.0906*
		67.24	-0.0474
		0	9.8734***
		0	0.063
		50.76	0

Table 2.33: First stage regressions - farmers own produce

	1987			1994			1999		
Farmer	-0.3718***	0.0878	-0.1368	0.1044	0.0175	-0.2848*	0.1487		
Female dummy	-0.0546***	0.0192	-0.0808***	0.0175	-0.0685***	0.0161	0.1487		
Kids less 5	0.0245***	0.0095	0.0408***	0.012	0.0124	0.011	0.091		
Kids 5 to 13	0.0308***	0.0075	0.0318***	0.0093	-0.011	0.0095	0.0095		
Kids 14 to 18	0.0316***	0.0089	0.0043	0.0092	-0.0028	0.0087	0.0087		
Older CES	-0.0277	0.019	-0.0311*	0.0188	0.026	0.0169	0.0169		
Spouse part-time	-0.1856***	0.0242	-0.1793***	0.0243	-0.1989***	0.0188	0.0188		
Spouse SE	-0.1915***	0.054	-0.2394***	0.0579	-0.1387***	0.0428	0.0428		
Spouse other occ	-0.3262***	0.0267	-0.3698***	0.0295	-0.3866***	0.0238	0.0238		
Dublin	0.0637***	0.0146	0.0826***	0.0147	0.038***	0.019	0.019		
Town less 20000	0.0086	0.0161	0.0349*	0.0182	0.0401**	0.0189	0.0189		
Town 3000-20000	0.0074	0.0163	0.0102	0.0153	0.0344**	0.0167	0.0167		
Town less than 3000	-0.0121	0.024	-0.0543**	0.0277	-0.0159	0.0391	0.0391		
Farm Savings	-0.0109	0.0455	0.0037	0.0455	0.0466	0.046	0.046		
Savings	0.1008***	0.0108	0.0849***	0.0128	0.0882**	0.0118	0.0118		
Farm Bedrooms	0.0802***	0.026	0.0641**	0.0284	-0.0371	0.0501	0.0501		
Bedrooms	0.0288***	0.0081	0.0367***	0.0087	-0.1495*	0.0765	0.0765		
Farm Deepfreeze	0.0184	0.0525	0.0743*	0.0452	0.0164	0.0461	0.0461		
Deepfreeze	0.05***	0.0135	0.0157	0.013	0.0684	0.0655	0.0655		
Farm Second home					0.0724***	0.0246	0.0246		
Second home	-0.1459***	0.0165	-0.1759***	0.0165	0.3121***	0.0618	0.0618		
Medical card					0.0338**	0.0161	0.0161		
Farm Cas									
Cas									
Farm Central heating	0.0969*	0.051	0.0033	0.0603					
Central heating	0.0496***	0.0135	0.0618***	0.0168					
Farm Microwave					0.1057**	0.0525	0.0525		
Microwave					0.0499***	0.0159	0.0159		
Bank account	0.1043***	0.0136	0.0894***	0.0131	0.0843***	0.0149	0.0149		
Farm Bankaccount	0.0239	0.0532	-0.1181**	0.0544	0.0778	0.0622	0.0622		
Computer					0.0728***	0.012	0.012		
Farm Computer					-0.0474	0.0895	0.0895		
Spin dryer	0.0261**	0.0118	0.0381***	0.012					
Burglar alarm					0.075***	0.0139	0.0139		
Constant					9.8734***	0.063	0.063		
First stage F test (p-value)	81.56	0	67.24	0	50.76	0	0		

Table 2.34: Results 1987 Older CES subsample

	(1)	
	Log yearly food consumption	
Log disp. income	0.250***	(0.0632)
Self-employed	0.0516*	(0.0278)
Married cpl 1kd	0.140***	(0.0475)
Married cpl 2kd	0.213***	(0.0524)
Married cpl 3kd	0.291***	(0.0564)
Married cpl 4+kd	0.280***	(0.0616)
1 adult with kd	0.228	(0.142)
Other hhds	0.377***	(0.0517)
Share food outside	0.00197	(0.00409)
Lower professional	0.0365	(0.0242)
Manual worker	0.0350	(0.0259)
Female dummy	0.0397	(0.0427)
Kids less than 5	0.00774	(0.0359)
Kids 5-13	0.0692***	(0.0112)
Kids 14-18	0.0721***	(0.0106)
Spouse 25-34	0.0351	(0.156)
Spouse 35-44	0.294***	(0.0615)
Spouse 45-54	0.309***	(0.0601)
Spouse over 55	0.275***	(0.0622)
Spouse parttime	0.0619	(0.0496)
Spouse SE	0.0194	(0.0678)
Spouse other occ	0.0801**	(0.0380)
Town less 20000	0.0228	(0.0249)
Town 3000-20000	-0.00297	(0.0261)
Town less than 3000	0.00499	(0.0489)
Observations	919	
R^2	0.444	

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.35: Results 1994 Older CES subsample

	(1)	
	Log yearly food consumption	
Log disp. income	0.280***	(0.0658)
Self-employed	0.0485	(0.0309)
Married cpl 1kd	0.224***	(0.0470)
Married cpl 2kd	0.322***	(0.0483)
Married cpl 3kd	0.346***	(0.0552)
Married cpl 4+kd	0.325***	(0.0750)
1 adult with kd	0.419***	(0.141)
Other hhds	0.374***	(0.0453)
Share food outside	0.00452	(0.00283)
Lower professional	-0.0207	(0.0275)
Manual worker	0.0295	(0.0286)
Female dummy	0.0895*	(0.0467)
Kids less than 5	0.0652**	(0.0331)
Kids 5-13	0.0569***	(0.0150)
Kids 14-18	0.123***	(0.0125)
Spouse 25-34	0.248*	(0.132)
Spouse 35-44	0.377***	(0.0699)
Spouse 45-54	0.371***	(0.0641)
Spouse over 55	0.328***	(0.0677)
Spouse parttime	0.0489	(0.0472)
Spouse SE	0.0161	(0.0711)
Spouse other occ	0.101**	(0.0417)
Town less 20000	-0.0612*	(0.0352)
Town 3000-20000	-0.0130	(0.0302)
Town less than 3000	0.0796*	(0.0482)
Observations	893	
R^2	0.471	

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.36: Results 1999 Older CES subsample

	(1)	
	Log yearly food consumption	
Log disp. income	0.236***	(0.0586)
Self-employed	0.0661*	(0.0352)
Married cpl 1kd	0.195***	(0.0452)
Married cpl 2kd	0.251***	(0.0466)
Married cpl 3kd	0.242***	(0.0575)
Married cpl 4+kd	0.198***	(0.0716)
1 adult with kd	0.291***	(0.0985)
Other hhds	0.291***	(0.0446)
Share food outside	0.00361	(0.00237)
Manual worker	0.0304	(0.0333)
Female dummy	0.0468	(0.0382)
Lower professional	-0.0328	(0.0282)
Kids less than 5	0.0613**	(0.0311)
Kids 5-13	0.0835***	(0.0131)
Kids 14-18	0.128***	(0.0116)
Spouse 25-34	0.229**	(0.106)
Spouse 35-44	0.217***	(0.0661)
Spouse 45-54	0.235***	(0.0592)
Spouse over 55	0.168***	(0.0629)
Spouse parttime	0.0876***	(0.0337)
Spouse SE	0.141***	(0.0494)
Spouse other occ	0.115***	(0.0319)
Town less 20000	-0.0716**	(0.0346)
Town 3000-20000	-0.0448*	(0.0268)
Town less than 3000	0.0322	(0.0364)
Observations	1273	
R^2	0.404	

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.37: Results 2004 Older CES subsample

	(1)	
	Log yearly food consumption	
Log disp. income	0.143**	(0.0647)
Self-employed	0.0367	(0.0365)
Married cpl 1kd	0.157***	(0.0442)
Married cpl 2kd	0.177***	(0.0512)
Married cpl 3kd	0.260***	(0.0594)
Married cpl 4+kd	0.125	(0.0817)
1 adult with kd	0.203**	(0.102)
Other hhds	0.321***	(0.0429)
Share food outside	0.00300	(0.00313)
Manual worker	0.0594*	(0.0354)
Female dummy	0.0349	(0.0289)
Lower professional	-0.0382	(0.0299)
Kids less than 5	0.0817*	(0.0443)
Kids 5-13	0.115***	(0.0188)
Kids 14-18	0.145***	(0.0136)
Spouse 25-34	0.337	(0.212)
Spouse 35-44	0.293***	(0.0686)
Spouse 45-54	0.291***	(0.0571)
Spouse over 55	0.327***	(0.0576)
Spouse parttime	-0.00265	(0.0277)
Spouse SE	0.0141	(0.0421)
Spouse other occ	0.0249	(0.0345)
Dublin	-0.0391	(0.0309)
Town less 20000	-0.0585	(0.0383)
Town 3000-20000	-0.000119	(0.0292)
Town less than 3000	-0.0622*	(0.0360)
Observations	1369	
R ²	0.344	

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2.9.6 Manual CES regression results

Table 2.38: Results 1987 Manual subsample

	(1)	
	Log yearly food consumption	
Log disp. income	0.229***	(0.0595)
Self-employed	0.116***	(0.0280)
Married cpl 1kd	0.0377	(0.0422)
Married cpl 2kd	0.143***	(0.0423)
Married cpl 3kd	0.212***	(0.0485)
Married cpl 4+kd	0.253***	(0.0617)
1 adult with kd	0.0890	(0.127)
Other HH	0.371***	(0.0447)
Share food outside	0.00531	(0.00336)
Female dummy	-0.00434	(0.0454)
Kids less than 5	0.0277	(0.0170)
Kids 5-13	0.0577***	(0.0130)
Kids 14-18	0.0865***	(0.0134)
CES over 45	-0.0146	(0.0267)
Spouse 25-34	0.0923*	(0.0506)
Spouse 35-44	0.178***	(0.0512)
Spouse 45-54	0.178***	(0.0554)
Spouse over 55	0.220***	(0.0619)
Spouse parttime	0.00765	(0.0444)
Spouse SE	-0.00235	(0.0748)
Spouse other occ	0.0788**	(0.0322)
Town less 20000	0.00421	(0.0237)
Town 3000-20000	0.0344	(0.0224)
Town less than 3000	-0.0352	(0.0350)
Observations	1172	
R^2	0.431	

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.39: Results 1994 Manual subsample

	(1)	
	Log yearly food consumption	
Log disp. income	0.252***	(0.0540)
Self-employed	0.0780***	(0.0263)
Married cpl 1kd	0.131***	(0.0471)
Married cpl 2kd	0.239***	(0.0441)
Married cpl 3kd	0.255***	(0.0518)
Married cpl 4+kd	0.194***	(0.0717)
1 adult with kd	0.0825	(0.0928)
Other HH	0.319***	(0.0473)
Share food outside	0.00115	(0.00259)
Female dummy	-0.00866	(0.0411)
Kids less than 5	0.0619***	(0.0172)
Kids 5-13	0.0842***	(0.0139)
Kids 14-18	0.130***	(0.0143)
CES over 45	-0.00785	(0.0293)
Spouse 25-34	0.0733	(0.0585)
Spouse 35-44	0.169***	(0.0584)
Spouse 45-54	0.232***	(0.0600)
Spouse over 55	0.160**	(0.0715)
Spouse parttime	0.0271	(0.0365)
Spouse SE	0.0842	(0.0694)
Spouse other occ	0.0893***	(0.0324)
Town less 20000	-0.00581	(0.0300)
Town 3000-20000	-0.0351	(0.0253)
Town less than 3000	0.0690*	(0.0382)
Observations	994	
R^2	0.468	

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.40: Results 1999 Manual subsample

	(1)	
	Log yearly food consumption	
Log disp. income	0.165***	(0.0569)
Self-employed	-0.0141	(0.0694)
Married cpl 1kd	0.220***	(0.0495)
Married cpl 2kd	0.309***	(0.0477)
Married cpl 3kd	0.300***	(0.0562)
Married cpl 4+kd	0.242***	(0.0698)
1 adult with kd	0.242***	(0.0913)
Other HH	0.451***	(0.0481)
Share food outside	0.00383	(0.00250)
Female dummy	0.0204	(0.0412)
Kids less than 5	0.0915***	(0.0193)
Kids 5-13	0.105***	(0.0147)
Kids 14-18	0.133***	(0.0140)
CES over 45	0.0646**	(0.0292)
Spouse 25-34	0.189***	(0.0568)
Spouse 35-44	0.269***	(0.0575)
Spouse 45-54	0.272***	(0.0596)
Spouse over 55	0.291***	(0.0600)
Spouse parttime	0.0426	(0.0295)
Spouse SE	0.00637	(0.0811)
Spouse other occ	0.0546*	(0.0283)
Town less 20000	-0.0375	(0.0277)
Town 3000-20000	-0.0169	(0.0309)
Town less than 3000	0.0810*	(0.0464)
Observations	1223	
R^2	0.421	

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.41: Results 2004 Manual subsample

	(1)	
	Log yearly food consumption	
Log disp. income	0.264***	(0.0877)
Self-employed	-0.0240	(0.0734)
Married cpl 1kd	0.142***	(0.0470)
Married cpl 2kd	0.156***	(0.0533)
Married cpl 3kd	0.177**	(0.0697)
Married cpl 4+kd	0.138	(0.0939)
1 adult with kd	-0.0498	(0.117)
Other HH	0.256***	(0.0578)
Share food outside	0.00809**	(0.00321)
Female dummy	0.0228	(0.0442)
Kids less than 5	0.0664**	(0.0265)
Kids 5-13	0.101***	(0.0224)
Kids 14-18	0.166***	(0.0182)
CES over 45	0.0156	(0.0374)
Spouse 25-34	0.0971	(0.0787)
Spouse 35-44	0.181**	(0.0770)
Spouse 45-54	0.216***	(0.0763)
Spouse over 55	0.307***	(0.0736)
Spouse parttime	0.0295	(0.0295)
Spouse SE	0.0779	(0.0605)
Spouse other occ	0.1000***	(0.0358)
Dublin	-0.0911***	(0.0331)
Town less 20000	-0.107**	(0.0419)
Town 3000-20000	-0.0568*	(0.0310)
Town less than 3000	-0.0493	(0.0351)
Observations	1014	
R^2	0.374	

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2.9.7 Dublin CES regression results

Table 2.42: Results 1987 Dublin subsample

	(1)	
	Log yearly food consumption	
Log disp. income	0.211***	(0.0446)
Self-employed	0.0556*	(0.0308)
Married cpl 1kd	0.130***	(0.0412)
Married cpl 2kd	0.156***	(0.0438)
Married cpl 3kd	0.177***	(0.0503)
Married cpl 4+kd	0.197***	(0.0697)
1 adult with kd	0.174	(0.121)
Other HH	0.359***	(0.0409)
Share food outside	0.00134	(0.00248)
Lower professional	-0.00349	(0.0217)
Manual worker	-0.0122	(0.0229)
Female dummy	-0.0424	(0.0329)
Kids less than 5	0.0410**	(0.0176)
Kids 5-13	0.103***	(0.0147)
Kids 14-18	0.0989***	(0.0148)
CES over 45	0.0310	(0.0304)
Spouse 25-34	0.116***	(0.0431)
Spouse 35-44	0.196***	(0.0436)
Spouse 45-54	0.235***	(0.0471)
Spouse over 55	0.193***	(0.0528)
Spouse parttime	0.0919**	(0.0393)
Spouse SE	0.0134	(0.0658)
Spouse other occ	0.0632**	(0.0304)
Observations	985	
R^2	0.504	

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.43: Results 1994 Dublin subsample

	(1)	
	Log yearly food consumption	
Log disp. income	0.440***	(0.0685)
Self-employed	0.132***	(0.0437)
Married cpl 1kd	0.189***	(0.0519)
Married cpl 2kd	0.214***	(0.0556)
Married cpl 3kd	0.226***	(0.0738)
Married cpl 4+kd	0.180*	(0.101)
1 adult with kd	0.365***	(0.0986)
Other HH	0.276***	(0.0577)
Share food outside	-0.00148	(0.00265)
Lower professional	-0.0251	(0.0285)
Manual worker	0.0869***	(0.0289)
Female dummy	0.117***	(0.0394)
Kids less than 5	0.0622***	(0.0241)
Kids 5-13	0.0971***	(0.0203)
Kids 14-18	0.168***	(0.0193)
CES over 45	0.0468	(0.0421)
Spouse 25-34	0.248***	(0.0725)
Spouse 35-44	0.281***	(0.0733)
Spouse 45-54	0.286***	(0.0744)
Spouse over 55	0.246***	(0.0790)
Spouse parttime	0.103**	(0.0471)
Spouse SE	0.0852	(0.0714)
Spouse other occ	0.136***	(0.0415)
Observations	738	
R^2	0.426	

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.44: Results 1999 Dublin subsample

	(1)	
	Log yearly food consumption	
Log disp. income	0.312***	(0.0774)
Self-employed	0.0525	(0.0461)
Married cpl 1kd	0.167***	(0.0618)
Married cpl 2kd	0.332***	(0.0549)
Married cpl 3kd	0.297***	(0.0746)
Married cpl 4+kd	0.268***	(0.0897)
1 adult with kd	0.203**	(0.0871)
Other HH	0.351***	(0.0589)
Share food outside	0.00220	(0.00313)
Manual worker	0.0296	(0.0386)
Female dummy	0.0757**	(0.0366)
Lower professional	-0.0232	(0.0317)
Kids less than 5	0.0747***	(0.0245)
Kids 5-13	0.0840***	(0.0191)
Kids 14-18	0.135***	(0.0203)
CES over 45	0.0336	(0.0377)
Spouse 25-34	0.106	(0.0696)
Spouse 35-44	0.204***	(0.0672)
Spouse 45-54	0.190***	(0.0680)
Spouse over 55	0.106	(0.0713)
Spouse parttime	0.0794**	(0.0405)
Spouse SE	0.0654	(0.0653)
Spouse other occ	0.148***	(0.0428)
Observations	809	
R ²	0.367	

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.45: Results 2004 Dublin subsample

	(1)	
	Log yearly food consumption	
Log disp. income	0.165**	(0.0804)
Self-employed	0.0897	(0.0493)
Married cpl 1kd	0.0876	(0.0572)
Married cpl 2kd	0.185***	(0.0684)
Married cpl 3kd	0.131	(0.0899)
Married cpl 4+kd	0.0220	(0.122)
1 adult with kd	0.107	(0.0995)
Other HH	0.331***	(0.0583)
Share food outside	0.00133	(0.00450)
Manual worker	0.00704	(0.0499)
Female dummy	0.0144	(0.0378)
Lower professional	-0.0222	(0.0369)
Kids less than 5	0.103***	(0.0314)
Kids 5-13	0.182***	(0.0273)
Kids 14-18	0.130***	(0.0219)
CES over 45	0.0208	(0.0444)
Spouse 25-34	0.167**	(0.0741)
Spouse 35-44	0.210***	(0.0738)
Spouse 45-54	0.262***	(0.0701)
Spouse over 55	0.313***	(0.0700)
Spouse parttime	0.0772*	(0.0416)
Spouse SE	0.0995	(0.0638)
Spouse other occ	-0.00293	(0.0518)
Observations	830	
R^2	0.327	

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2.9.8 Low Professional CES regression results

Table 2.46: Results 1987 Low professional subsample

	(1)	
	Log yearly food consumption	
Log disp. income	0.289***	(0.0630)
Self-employed	0.148***	(0.0378)
Married cpl 1kd	0.0932**	(0.0441)
Married cpl 2kd	0.138***	(0.0445)
Married cpl 3kd	0.127**	(0.0523)
Married cpl 4+kd	0.124*	(0.0645)
1 adult with kd	0.0359	(0.110)
Other HH	0.260***	(0.0476)
Share food outside	-0.00222	(0.00302)
Female dummy	-0.0428	(0.0320)
Kids less than 5	0.0942***	(0.0198)
Kids 5-13	0.113***	(0.0151)
Kids 14-18	0.155***	(0.0158)
CES over 45	0.00552	(0.0387)
Spouse 25-34	0.0478	(0.0539)
Spouse 35-44	0.117**	(0.0574)
Spouse 45-54	0.262***	(0.0607)
Spouse over 55	0.253***	(0.0635)
Spouse parttime	0.0585	(0.0481)
Spouse SE	-0.0743	(0.0584)
Spouse other occ	0.0310	(0.0409)
Town less 20000	-0.00236	(0.0268)
Town 3000-20000	-0.0445	(0.0274)
Town less than 3000	0.000630	(0.0499)
Observations	1020	
R^2	0.434	

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.47: Results 1994 Low professional subsample

	(1)	
	Log yearly food consumption	
Log disp. income	0.280***	(0.0605)
Self-employed	0.0604	(0.0413)
Married cpl 1kd	0.0880*	(0.0495)
Married cpl 2kd	0.190***	(0.0470)
Married cpl 3kd	0.185***	(0.0603)
Married cpl 4+kd	0.197**	(0.0775)
1 adult with kd	0.142	(0.0891)
Other HH	0.299***	(0.0509)
Share food outside	0.000149	(0.00318)
Female dummy	0.0980***	(0.0318)
Kids less than 5	0.0461**	(0.0223)
Kids 5-13	0.114***	(0.0177)
Kids 14-18	0.158***	(0.0171)
CES over 45	0.0430	(0.0372)
Spouse 25-34	0.192***	(0.0624)
Spouse 35-44	0.238***	(0.0605)
Spouse 45-54	0.255***	(0.0615)
Spouse over 55	0.212***	(0.0716)
Spouse parttime	0.164***	(0.0434)
Spouse SE	0.0861	(0.0547)
Spouse other occ	0.139***	(0.0362)
Town less 20000	-0.0670*	(0.0368)
Town 3000-20000	0.0123	(0.0297)
Town less than 3000	0.0276	(0.0502)
Observations	910	
R^2	0.474	

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.48: Results 1999 Low professional subsample

	(1)	
	Log yearly food consumption	
Log disp. income	0.169**	(0.0751)
Self-employed	0.118**	(0.0566)
Married cpl 1kd	0.166***	(0.0496)
Married cpl 2kd	0.211***	(0.0470)
Married cpl 3kd	0.256***	(0.0587)
Married cpl 4+kd	0.154*	(0.0805)
1 adult with kd	0.180**	(0.0799)
Other HH	0.293***	(0.0502)
Share food outside	0.00650**	(0.00309)
Female dummy	0.0380	(0.0267)
Kids less than 5	0.0623***	(0.0213)
Kids 5-13	0.106***	(0.0161)
Kids 14-18	0.131***	(0.0181)
CES over 45	0.00674	(0.0313)
Spouse 25-34	0.128**	(0.0595)
Spouse 35-44	0.198***	(0.0608)
Spouse 45-54	0.238***	(0.0583)
Spouse over 55	0.161**	(0.0626)
Spouse parttime	0.0578*	(0.0322)
Spouse SE	0.0143	(0.0602)
Spouse other occ	0.106***	(0.0387)
Town less 20000	-0.0467	(0.0345)
Town 3000-20000	-0.000619	(0.0332)
Town less than 3000	-0.0524	(0.0554)
Observations	1048	
R^2	0.367	

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.49: Results 2004 Low professional subsample

	(1)	
	Log yearly food consumption	
Log disp. income	0.132*	(0.0695)
Self-employed	0.0592	(0.0435)
Married cpl 1kd	0.0961*	(0.0520)
Married cpl 2kd	0.154***	(0.0553)
Married cpl 3kd	0.117*	(0.0708)
Married cpl 4+kd	0.0242	(0.0939)
1 adult with kd	0.0441	(0.0855)
Other HH	0.315***	(0.0499)
Share food outside	0.00965***	(0.00369)
Female dummy	0.00102	(0.0281)
Kids less than 5	0.139***	(0.0269)
Kids 5-13	0.149***	(0.0229)
Kids 14-18	0.143***	(0.0178)
CES over 45	0.0105	(0.0352)
Spouse 25-34	0.116	(0.0703)
Spouse 35-44	0.234***	(0.0624)
Spouse 45-54	0.301***	(0.0567)
Spouse over 55	0.344***	(0.0589)
Spouse parttime	-0.0463	(0.0373)
Spouse SE	0.0513	(0.0536)
Spouse other occ	-0.0715	(0.0493)
Dublin	-0.0390	(0.0342)
Town less 20000	-0.0197	(0.0449)
Town 3000-20000	-0.0213	(0.0308)
Town less than 3000	-0.0385	(0.0462)
Observations	1241	
R^2	0.318	

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2.9.9 Construction CES regression results

Table 2.50: Results 2004, Construction subsample

	(1)	
	Log yearly food consumption	
Log disp. income	0.378***	(0.0995)
Self-employed	0.0973	(0.0736)
Married cpl 1kid	0.209***	(0.0596)
Married cpl 2kid	0.278***	(0.0699)
Married cpl 3kid	0.243***	(0.0855)
Married cpl 4+kid	0.261**	(0.116)
1 adult with kid	-0.127	(0.162)
Other HH	0.304***	(0.0670)
Share food outside	0.00551*	(0.00314)
Manual worker	0.0846	(0.0716)
Female dummy	0.195***	(0.0681)
Kids less than 5	0.0516	(0.0316)
Kids 5-13	0.0864***	(0.0251)
Kids 14-18	0.175***	(0.0259)
CES over 45	0.0628	(0.0485)
Spouse 25-34	0.101	(0.0941)
Spouse 35-44	0.167*	(0.0935)
Spouse 45-54	0.121	(0.0981)
Spouse over 55	0.287***	(0.0977)
Spouse parttime	-0.0172	(0.0415)
Spouse SE	0.0204	(0.0555)
Spouse other occ	0.147***	(0.0411)
Dublin	-0.0475	(0.0441)
Town less 20000	-0.0435	(0.0540)
Town 3000-20000	0.0626*	(0.0355)
Town less than 3000	-0.0599	(0.0458)
Observations	557	
R ²	0.366	

Robust standard errors in parentheses.

No farming households included in this regression.

Omitted categories: married couple no children, higher professional, spouse under 25, spouse full time, town more 20000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Chapter 3

Designing a Property Tax without Property Values: Analysis in the case of Ireland

This chapter is developed from a joint working paper with Dr. Seán Lyons and Prof. Richard S.J. Tol. It was presented at the Economic and Social Research Institute Policy Seminar "The property tax: why, how, who pays" Dublin (2010), the Trinity College Economics Department PhD Seminar, Dublin (2010) and the Royal Economic Society PhD Job Market Conference, London (2011).

3.1 Introduction

The collapse in house prices in Ireland over the past three years has resulted in a much thinner property market with owners waiting for house prices to increase before selling and buyers waiting for prices to drop further before buying. The thin property market has had a direct effect on the tax yield from property in Ireland, as the main source of revenue in this area is a transactions based tax: stamp duty.

The Commission on Taxation Report of 2009 recommended the implementation of a property tax. In the first instance, the Commission suggests, house values under a property tax scheme should be self-assessed. The report also recommends that a national register of house prices should be established. The latter suggestion was recently taken up by the government but it will still require time and a lot of resources to establish a full dataset of dwelling prices.

In addition to self assessment and in advance of developing a national house price register, we suggest that hedonic methods can assign most residences to the correct tax bands. Indeed, hedonic methods are widely used as part of the valuation process in systems where property values are assigned by a valuation agency (McCluskey et al., 2007). However, very simple hedonic methods, despite assigning to a correct tax band, can still result in large inequities when it comes to the actual valuation figures. In this paper we evaluate how closely one can predict house prices by using simple hedonic regression analysis, using readily available and easily verifiable variables. Inevitably when using a very simple valuation methodology (and in the case of this study, the most basic form possible) there will be estimation errors. The purpose of this study is to evaluate the potential inequities which could emerge if such a methodology were to be used. Who loses out from a valuation method of this kind? How much information on a home would be required in order to accurately predict into which tax *band* this home falls? Could a system of this type be used to identify potential mis-reporting by home-owners through targeted audits?

Hedonic analysis decomposes the price of a good or service (a house in this case) into its components. With information on the structural characteristics of a dwelling, its precise location and neighbourhood characteristics, it is possible to closely predict the price of a home. The trade-off in this case is between more information (which is administratively costly) and more accurate predictions, versus less information but more inaccurate predictions. In addition how the overall cost and tax burden is distributed is equally important.

We study the errors in the predictions depending on the different explanatory variables used (area, rooms, age, dwelling type), the number of variables used and who they affect the most (by age groups, income groups, region). That is, we do not aim for the “perfect” regression model. Rather, we use a regression model that is deliberately simple: the kind that may be used by tax authorities because of lack of data. We then study the patterns in the forecast errors of the simple regression, and test whether these errors correlate with distributional variables. We look at the absolute value of the errors in euro as well as the “assessment sales ratio” a measure developed by the International Association of Assessing Officers (IAAO), which indicates whether the valuation system produces errors which are progressive or regressive. Further aspects of the equitability of the different valuations are examined using a dispersion coefficient. It should be noted that the deliberate omission of key variables in the regression will result in mis-specified model and the estimated parameters are likely to be biased. When looking at the errors produced by such a method what will be key will be the examination of the types of inequities which emerge.

Property taxes existed previously in Ireland but due to the high number of exemptions only a very small proportion of households were taxed. Research on the design of a new property tax in the country has been published. Callan et al. (2009a) look at how a property tax in Ireland could be designed in order to take account of ability to pay. They highlight the issues regarding which tax base is used (all land, owner-occupied, all residential) and how to set tax rates. They focus on the design of the tax and use the Economic and Social Research Institute of Ireland’s tax benefit model to estimate the distributional impact on household incomes for a given overall tax revenue target. The model also allows the examination of the effect of exemptions for poorer households. They find that even if the poorest households in the Republic were to be exempt it would still be possible to raise close to 1 billion euro a year. The provision of an income exemption limit, marginal relief, as well as income-related reliefs would safeguard those on lower incomes. The present study looks at how the property valuations would be estimated and the consequences of different valuation methods on tax equity.

The remainder of this chapter is set out as follows. Section 3.2 presents some background statistics on the housing sector during the Celtic Tiger and the current fiscal system relating to property in Ireland. Section 3.3 describes the Commission for Taxation recommendations in detail as well as the features of the previously existing property tax in Ireland. Section 3.4 reviews the literature in the area of property tax design, property assessment as well as errors in assessment and studies on property tax equity

issues. Section 3.5 presents the different methodologies used to predict house prices and evaluate the accuracy of the predictions that are used in this paper. Section 3.6 provides an overview of the data used in the analysis while Section 3.7 presents the results. Finally, Section 3.8 provides a discussion and conclusions.

3.2 The Irish housing market and fiscal system

The Celtic Tiger years in Ireland were characterised by rapid and sustained income growth. Increasing disposable incomes and a culture of home ownership resulted in massive increases in property prices as investors bought multiple properties, home owners sought to upgrade and first time buyers tried to get onto the property ladder. As shown in Figure 3.1, average national house prices increased from €65,000 to €323,000 between 1990 and 2007. But, as is also clear from Figure 3.2, this rapid growth came to a sudden end that year.

In Ireland, property features in the taxation system in a number of ways. Capital gains tax is paid on the gains made from the disposal of assets including property, although principal private residences are excluded from this tax as it is aimed at investors. Capital acquisitions tax applies to property that is inherited or gifted. Property also features in relation to income tax through mortgage interest relief. Home owners receive tax relief in relation to their mortgages for a set number of years. In 2009 the government introduced a tax on second homes which was a flat rate tax of 200 euro on all non primary residences. The most important property related tax in Ireland over the last ten years and the main source of tax revenue from residential property during the boom years has been stamp duty.

Stamp duty is a transactions tax and is paid to the government at the sale of a property. The rules regarding the incidence of stamp duty did change over the course of the Celtic Tiger years but overall the level of duty to be paid was based on the value of the home and exemptions were made for first time buyers and purchases of second hand properties. Stamp duty receipts were very high during the boom years as the number of sales grew and prices rose. As is shown in Figure 3.3, property as a percentage of total stamp duty grew steadily over the period and stamp duties from residential property resulted in 1.3 billion euro in tax revenue to the Exchequer in 2006.

One downside of using stamp duty as the main source of tax revenue for property is that receipts from it tend to be volatile and pro-cyclical. When the current property

market slump took hold in 2007 and the number of transactions fell dramatically, this had an immediate effect on tax revenues from property.¹ The OECD (2010) stated that “tax revenues became too dependent on construction and housing transaction-related receipts as the market boomed”. In one year alone (2006-2007) stamp duties from residential property fell by 300 million euro. The now relatively stagnant property market has resulted in a huge shortfall in tax revenue for the government which, combined with the reductions in tax revenue due to the recession, is problematic. Consequently, the government has decided to move away from stamp duty in favour of a more stable property tax.²

A property tax is a tax on owners of property and can be levied as a flat rate (i.e. all owners pay the same amount for each property) or as an *ad valorem* tax (i.e. based on the value of the property). Receipts from such taxes tend to be more stable than stamp duty as they are based on the value of the housing stock and not on the flow of transactions. They may also decrease speculation and vacancy rates as people do not want to pay taxes on empty properties. A property tax can help lower transactions costs when purchasing property (as it would replace stamp duty). This would in turn make the housing market more liquid, so that fewer people would live in the “wrong” house (e.g., young families who cannot afford the transactions costs of moving to larger houses, or empty nesters who stay in large family homes). Furthermore, property taxes should give rise to less deadweight loss and be less damaging to the competitiveness of Irish exports than income taxes; and they should be less damaging to Ireland as a destination for foreign direct investment than profit taxes. In sum, if properly implemented, property taxes are seen as an efficient and stable source of revenue. The implementation of a property tax in Ireland was one of the subjects of a comprehensive review by the Commission on Taxation in 2009, which is discussed in the following section.

3.3 Commission on Taxation recommendations and the old property tax

In 2009, the Commission on Taxation (CoT) issued a comprehensive report on possible changes to the structure of the Irish fiscal system. The relevant sections relating to the current study will be presented in Section 3.3.1 while Section 3.3.2 details the charac-

¹While the collapse in stamp duty did affect tax revenue from property, the main impact of the house price collapse on tax revenues occurred through the collapse in activity in building and construction and its impact through direct and indirect tax receipts.

²The latest announcements have indicated that the Government is now looking to introduce a flat rate tax as of January 2012. <http://www.irishtimes.com/newspaper/breaking/2011/0531/breaking52.html>.

teristics of the previously existing property tax in Ireland.

3.3.1 Commission on Taxation

The Commission on Taxation was established in 2008 to “review the structure, efficiency and appropriateness of the Irish taxation system”.³ The final report was published in 2009 and Part 6 deals with the taxation of property. The six main recommendations made by the Commission are presented in Table 3.1 and are discussed in further detail below along with the main comments on implementation issues and design.

“We consider that an annual property tax (APT) should be implemented at the earliest possible date, taking account of the very significant administrative challenge for the Revenue Commissioners who will have to develop an assessment, collection and accounting system” (CoT, 2009:157).

The Commission acknowledges the implementation difficulties of introducing a property tax. The main difficulties usually cited with regards the implementation of this sort of tax are administration procedures, monitoring, setting up an appeals system, deciding rates, enforcing penalties and, evidently, assessment (Szalai and Tassonyi, 2004). Although introducing an ad valorem property tax for inclusion in the next Budget (2012) is looking less likely, the government has indicated that it has not yet been examined in any detail⁴ and recent media reports suggest that the government is planning on a flat rate “household charge” to be announced in the next Budget.⁵ The complexity of introducing a tax of this kind in an equitable manner is high and if a property tax based on house size or value is to be put into operation it will require time. The main issue and the one examined in this paper will be the assessment of house values.

The Commission on Taxation also broaches this subject.

“Options for the future taxation of property in Ireland are constrained by the absence of an up-to-date valuation database on which an annual property tax could be based. An up-to-date valuation database will also provide a database against which self-assessed property tax returns can be checked as part of the monitoring of our proposed annual property tax” (CoT, 2009:160).

³<http://www.commissionontaxation.ie/>

⁴“Fall in values reduces scope for property tax - Lenihan”, *The Irish Times*, 19th July 2010.

⁵“Minister defends new 100 euro charge ahead of full property tax”, *The Irish Times*, 27th July 2011.

The Minister for Housing has since the publication of the CoT report announced the establishment of an expert group to look at the matter of establishing a national register of house prices.⁶ To our knowledge the group has not yet produced any publications.

A flat rate tax (where all owners pay the same fee) has been the recent focus of media attention. Unfortunately, this type of tax would be the most inequitable as owners of very expensive homes would be charged the same as the owners of the lowest priced homes. Consequently an ad valorem tax is an option which should be favoured. This system requires an estimate of *relative* house values. The Commission suggests self-assessment as a solution to the lack of property register.

"We consider self-assessment is an appropriate method of assessment for the tax subject to appropriate monitoring and audit mechanisms. It is greatly simplified by the banding of valuations. [...] We acknowledge that direct assessment would provide greater certainty for taxpayers. However, in our view, it is not capable of being implemented within an appropriate time-frame due to the lack of an existing valuation database for residential properties" (CoT, 2009:165).

Although self-assessment may seem an attractive option, due to a lack of national house price database, there are a number of issues with the method. First, had a self-assessing property tax been implemented during the Celtic Tiger years, when house transactions and sales were numerous, home owners would have had a lot more information about the possible value of their homes and the government would have had a lot more transactions to compare prices with when monitoring self-assessed values. Unfortunately, this policy would now be introduced at a time when the market is extremely thin. Volumes are so low that the ESRI/PTSB house price index was initially reduced to a quarterly index (instead of monthly) because of insufficient observations on transactions and was then cancelled entirely.⁷ Consequently, not only will it be difficult for home owners to self-assess due to a lack of prices to compare their property with, but with prices still falling, the task is made even more difficult. Moreover, as is discussed in Section 3.4, self-assessment only tends to work if proper enforcement and penalties are put into place. This will take time and will require significant resources to become effective, which may be problematic in the current tight fiscal climate.

"As with any self-assessment tax but particularly so in the case of a new tax such as the APT it is vital that there is a clear understanding from the

⁶"Government group looks at house price index", *AIB Global Treasury fxcentre*, 10th March 2010.

⁷It has now been replaced by the CSO Residential Property Price Index.

outset by taxpayers that the tax will be subject to appropriate audit and compliance actions monitoring" (CoT, 2009:169).

As all properties registering for the tax will be newly registered, monitoring at the national scale will be a very extensive enterprise and could leave the door open to under-valuations and cheating on the part of home-owners. Consequently, in this paper we investigate whether or not an additional method of valuation is feasible.

"Each householder who is liable to pay the tax and who gets a professional assessment of the value of his or her property should get a tax credit of up to €75 in the first year to compensate for costs incurred" (CoT, 2009:165).

Although the use of professional valuers is a good idea and more likely to result in more accurate valuations, the thin property market issue still applies. Furthermore, determining house prices in the current market really depends on finding a buyer and to which year the current price is being compared to. It will be difficult to set a baseline "house value" without more accurate guidelines, even for valuers.

In this paper, we use property characteristics to examine how accurately one can estimate house prices using a limited set of house characteristics and county dummies. We are particularly interested in whether using the property characteristics would result in estimates of house values that are considered equitable. For example, "We examined market value and floor space as possible bases for the tax. We consider that using floor space alone as a base for the tax would offend the principle of equity. A residential property owner who lives in or owns a 1'200 square foot house in an affluent area should pay more tax than one who lives in a similar sized house in a less affluent area" (CoT, 2009:165). Taking into account regional differences could mitigate this problem. We will underline the importance of regional characteristics in the Section 3.6.

"We also considered using a combination of floor space and market value as the tax base. The introduction of a new annual property tax will be a significant challenge for the Government, who will have to convince taxpayers that it is an appropriate policy approach. For this reason, the rules that govern the assessment and calculation of the tax should be as simple as possible so that taxpayers can calculate their tax with reference to as few parameters as is feasible. We consider that the application of floor space in addition to the market value would add undue complexity for householders liable to pay the tax" (CoT, 2009:165).

Although the addition of parameters inevitably complicates a tax assessment, the calculation of the tax incidence tends to fall on the government side. Consequently, a professional assessment as recommended above could also include the number of rooms in the home and its square footage without any extra undue complexity. The reporting of additional parameters should not be difficult; however the administration of a tax based on a more complex system may be. The purpose of this analysis will also be to investigate the trade-off between the cost of additional variables and the resulting benefits in terms of equity and the reduction in the error in estimation.

“There should be a fixed valuation date, set at a date in advance of the commencement of the tax, so that all house owners self-assess at the same time” (CoT, 2009:165).

One advantage when starting a property tax scheme from scratch is that all valuations have a common date. The issue usually faced by other countries is that valuations do not have a common date. The main issue then is to determine how often the valuations need to be updated. The Commission does not recommend a rate for the property tax and states that it is “a matter for Government” (CoT, 2009:159).

Although the CoT does not recommend a rate for the tax it does give an example of possible valuation bands.

“The annual property tax should be calculated by reference to valuation bands within which a property owner would value his or her house. Our rationale is that there should be certainty about the tax base. Therefore, the classification of a property for the tax should be easily determined” (CoT, 2009:161).

These bands increase in €150'000 steps and are presented in Table 3.2. We will try and determine whether using property characteristics can allow the estimation of house values to within the correct bands.

The issue of who should pay the tax is also addressed by the Commission.

“The tax should apply to all residential housing units including holiday homes, second homes and houses that are let or available for letting. It should also apply to vacant houses [...] and to bed and breakfast and guest-house accommodation” (CoT, 2009:164).

The dataset used in our analysis is presented in detail in Section 3.6. We concentrate our analysis on owner-occupiers as it allows us to analyse the impact of the tax based on the owners' household characteristics, however robustness checks were done by including all households (owner occupiers and renters) in the analysis.

The Commission stresses the issue of taking into consideration the ability to pay of the most vulnerable. "Exemptions for local authorities and other social housing providers, care facilities, nursing homes, boarding school accommodation, charities" (CoT, 2009:164). Owners who paid stamp duty during the previous seven years "would be exempted from paying the annual property tax for a seven year period from the year they paid stamp duty" (CoT, 2009:168). This would be to avoid double taxation of those who not only bought at the height of the boom and paid stamp duty but who are also those most likely to currently be in negative equity.⁸ "The annual property tax should be applied to all housing stock, including vacant units" (CoT, 2009:168). Property taxes are said to discourage speculation as they make holding on to vacant properties more expensive.

Finally, the main point stressed by the Commission is the issue of equity.

"We considered a flat charge per property but rejected it as inequitable. We consider that owners of more valuable properties should pay more tax than those who own less valuable properties. The banded valuation system we are proposing provides for a proportionate system of taxation [...]. We also considered but rejected a progressive tax as it would be more difficult to administer than a proportionate tax using valuation bands" (CoT, 2009:167) and "In making these proposals we are aware of two issues which we consider are important aspects of an APT on residential property: [...] the provision of a scheme to mitigate the impact of the recurrent property tax on low-income house owners, including those with relatively low income and large more valuable houses" (CoT, 2009:161).

The property tax that existed in Ireland in the late 1980s and early 1990s was perceived by most to be inequitable. It is crucial, especially in a time of recession that those with more valuable properties pay more tax and those with less are given additional support. However, it will also be essential to take account of those in society who by circumstance have low incomes despite owning large properties. Making allowances

⁸It is worth noting however, that the suggestion that people who paid stamp duty in the previous seven years should have property tax exemption would only lead to a very limited degree of compensation for those people as the level of property tax would be likely to be small in comparison to the stamp duty paid.

for these cases will enhance the equity of the tax and make it more likely to be accepted by the public.⁹

The research conducted by the Commission on Taxation into the implementation of a property tax in Ireland is extensive and thorough. The main issue as underlined by the Commission is the lack of a property database in Ireland. Unfortunately, self-assessment as recommended by the Commission in response to the lack of data is more difficult in a thin property market. The objective of the present paper is to examine the types of inequities that may or may not arise when using house characteristics such as floor space or number of rooms to value a dwelling. Estimating the scale of these inequities and who they would affect is important prior to a tax being implemented.

3.3.2 Historical property tax in Ireland

Property taxes are not new to Ireland. Indeed a Residential Property Tax (RPT) existed in Ireland between 1983 and 1997.¹⁰ "The tax was charged at the rate of 1.5% on the excess of the market value of all relevant residential properties of a person over a market value exemption limit and was payable provided the income of the household exceeded an income exemption limit."¹¹ It was an annual tax, chargeable on the market value of residential property, owned and occupied on the 5th April of each year which was the valuation date.

The RPT was a national tax but due to the number of exemptions related to it, only a very small number of home-owners actually paid it. Indeed, exemptions existed for owners of homes under a certain value (£101'000) and whose income was below a threshold (£30'100). Owners provided an assessment of their home value and if the Revenue Commissioners considered it to be too low it could re-value the property. The tax was very unpopular with the general public for its perceived lack of equity. It was considered a "Dublin tax as the capital accounted for almost two thirds of the revenue collected".¹² As a consequence it generated very little revenue and according to the Commission (CoT, 2009:157) it resulted in only 17 million euro in receipts in 1996 and

⁹Two other points of interest are: "The tax should be applied on the gross value of the property without regard to borrowing to fund its purchase" (CoT, 2009:164) and "After an appropriate introductory period, all of the revenues from an annual property tax should be used for local government financing" (CoT, 2009:168).

¹⁰Commission on Taxation (CoT, 2009:156) provides an overview of the previous residential property taxation systems in Ireland. The main one of note is the RPT.

¹¹<http://www.revenue.ie/en/tax/rpt/index.html>

¹²"Property Investor", *Irish Times*, June 17th, 2010.

was paid by only 2% of all households.

Making sure that these mistakes are not repeated will be crucial in the implementation of a new property tax in Ireland. Indeed the CoT states that “An important lesson to be learned from the residential property tax system that operated between 1983 and 1997 is that an overly narrow tax base led to an insignificant revenue flow for the Exchequer, high administrative costs and a perception of inequity” (CoT, 2009:157). The following section reviews the academic literature regarding various aspects of property tax design and assessment as well as the case of Northern Ireland.

3.4 Literature

This section reviews the literature in the area of property tax design, property value assessment and property tax equity. First the concepts of horizontal and vertical equity as well as regressivity and progressivity are defined. Then, existing studies on various aspects of property taxes are presented, first regarding the efficiency of self-assessment, then the assessment of property taxes in thin markets and implementation issues. Section 3.4.4 describes the case of Northern Ireland which has recently had to revalue its property database.

3.4.1 Equity and Fairness in Taxation

The issues of equity and fairness in taxation are central to most of the academic literature in the area. While they will not be discussed in depth in this paper it is worth underlining the main characteristics a tax can have and what features are used to assess it. The three aspects of a tax that tend to be used to evaluate it are: efficiency, equity and administrative simplicity.

Inevitably, the introduction of a new tax at a national level will incur additional administrative costs for the body or bodies in charge of collecting it and administering it. In the case of Ireland and in the case of the tax as proposed by the CoT the bulk of these administrative costs would be in compliance and monitoring of tax assessments. Costs would also need to be taken into account depending on the appeals system. If dwelling characteristics were used to determine house value, additional information and estimations would be required, resulting in more costs. The trade-off is then between additional information and hence cost, versus greater accuracy and fairness.

Tax assessment methods can lead to inequity in several ways depending on how and whom a tax targets. According to Allen and Dare (2002) a tax is vertically inequitable if there is systematic variation in the assessed value compared to the market value across property types and property value bands. Using the methodology described in Section 3.5, vertical inequity would occur if there was a systematic difference in the assessment sales ratios for properties that have low versus high market values. Horizontal inequity occurs when properties of similar market value are assessed differently. Inevitably some variation in assessment will occur depending on who values the property and a small amount of horizontal inequity is to be expected, however horizontal inequity can also be due to valuer bias. For instance, some valuers may systematically place a higher value on certain property attributes compared to others. It can also arise when neighbourhood characteristics are not known to the valuer.

A tax can also be deemed regressive or progressive depending on how it affects households. A regressive tax is a tax where the rate decreases as the amount subject to the tax increases, i.e. relative to their total income the poor suffer more than the rich. A progressive tax is a tax where the rate increases as the taxable amount increases; personal income taxes are often progressive. The rich pay a higher percentage of income in tax than the poor. With regard to vertical inequity, if lower value house prices are always over-valued compared to high valued houses then the valuation method is regressive. If lower value house prices are always assessed at a lower rate than their market rate compared to high value houses then the valuation method is progressive. All of these possible impacts should be assessed when implementing a new tax and we will be looking at whether valuations based on dwelling characteristics satisfy the equity principles.

3.4.2 Self-assessment and the tax system

Plassmann and Tideman (2008) provide an accurate description of the common issue related to self-assessment: "A person who is subject to self-assessment can be expected to provide a truthful report of the (monetary equivalent of the) value that he attaches to a good that he owns if and only if truthful reporting maximises his utility". Consequently, when using self-assessment for tax purposes, any utility maximising agent under-estimates the value of the good that is taxed. Potential solutions to this problem date back centuries.

Bird (1984) presents examples of countries that use a property tax system with self-assessment that gives the government the right to purchase at declared value. The earliest documented example he gives is of New Zealand in the late 19th century. If home owners deemed their valuation excessive they could value it themselves at a lower rate on condition that the government could purchase the property at that rate plus a "fair margin". In parallel the land owner could force the government to purchase the property at the higher official value. Bird also quotes the example of Columbia in the 1950s and 1960s. However due to the economic situation at the time there was no credible threat from the government as it could not afford to purchase the properties. This might be the case of Ireland at the moment, especially in the aftermath of NAMA purchases. Niou and Tan (1994) discuss a similar method proposed in China in 1905. They examine this "truth-revelation mechanism" and find that with two forms of penalties (compulsory purchase and taxes) there is still a small probability that some land owners will under-report. This probability is reduced with important random audits but this entails an additional cost to the government. Their model does not discover a mechanism that induces correct self-assessment at no cost. In cases where the threat of purchase on both sides is credible, self-assessment can work.

This is shown by Gstach (2009) who uses a theoretical model to prove that a Nash equilibrium among tax payers exists in a game where the property tax is based on the owner's self-assessment but only in the case where the credibility of the purchase threat is real and the government purchases a randomly selected number of properties based on the distribution of declared property values. This leads to all homeowners declaring their true property values and no purchases are required on the government's side. The equilibrium situation however rests on the credible threat of repurchase and the repurchasing price. Wolinsky (2000) offers a "softer" version of the scheme whereby an option would be given to home owners to have their assessment value defined through a centralised process. Homeowners wanting to determine their own house price could still do so but this would constitute an obligation to sell. In reality and in Ireland this is unlikely to occur and the use of self-assessment will hence in all likelihood lead to dwelling values being under-declared. The problems relating to self-assessment are exacerbated when the property market is thin as it makes it difficult for home owners to correctly assess their house value when they do not have many prices to compare it to.

3.4.3 Thin markets and property tax assessment

Most of the literature on errors in property tax assessment is based on studies of markets with existing property taxes. Consequently the questions asked by authors of these studies are whether the current tax regime is equitable and if not, how to best remedy the situation. The following papers, mostly from the United States, detail some of the methodologies used to assess the horizontal and vertical inequity of existing property tax regimes.

McMillen and Weber (2008) investigate the horizontal and vertical inequities in property taxation in Chicago. They use property tax assessment ratios (i.e. a property's assessed value divided by its market value) to determine when these inequities appear. Assessment ratios can vary between and within price segments of the market. Horizontal inequity occurs when there is a lack of information on property and neighbourhood factors available to agents (buyers, sellers and valuers). Vertical inequity occurs when higher priced properties appreciate more quickly compared to repeat assessment times. The authors in this case investigate the effect of a thin market and low sales in certain areas on relative assessment ratios. They test the notion "that assessment ratios will be not only less uniform but also higher in areas where prices are relatively unpredictable. Assessments of land and improvements in inactive markets may vary significantly for nearly identical properties in the same price strata and may well be biased systematically." The authors use data on sale prices in Chicago in 2004 and 2005 as well as assessment data from 2003 to conduct their analysis. They check to see if assessments are more uniform in areas with more sales. They look at thin markets (at the top and bottom ends of the scale) and find that there is regressivity: as prices increase the assessment ratios fall. They also find that the assessment ratio declines over time. The authors conclude that in areas with more sales, assessment ratios are more likely to cluster in the centre of the distribution, i.e. having more sales makes it easier to produce uniform assessments. Thick markets improve horizontal equity.

Goolsby (1997) analyses house prices as assessed by valuers to determine whether systematic errors occur in valuation. In the United States assessors are legally required to assess a property at a level equal to the market value but when property prices are increasing this will tend to result in a ratio of smaller than one. If the assessment ratios were the same across the board there would not be a problem however they are generally not uniform. Goolsby uses data from Washington State in 1993 on owner-occupied single family houses. He regresses the assessment ratio on variables that are thought to cause a consistent bias in the estimates. He finds that valuers tend to under-assess

high value houses and older properties and over-assess very large houses across all counties. He concludes that using valuers' estimates of house values should be taken with caution and in the best case should be adjusted to take account of consistent assessor bias.

Allen and Dare (2002) also look at assessment error, in particular its impact on equity when using an *ad valorem* property tax. When dealing with this type of tax the issue of correctly assessing the relative house value is crucial. The authors concentrate on the issue of horizontal inequity using sample data from Florida. Florida uses computer aided mass appraisal models to estimate a property's value for tax purposes each year. Similarly to Goolsby (1997), they find that older and larger homes as well as properties in neighbourhoods with a higher percentage of minority residents present the highest level of assessment error. They do not find much evidence of vertical inequity but do find indications that the middle quartiles of house values systematically differ from the bottom and top quartile, similarly to the results in Goolsby (1997). Allen et al. (2009) revisit the issue of vertical inequity due to property taxation for owners of hotels in Florida between 1999 and 2004. The issue of valuing hotels is more complex than homes as the value of the business must be separated from the value of the property. They find that lower value properties are systematically over assessed and that this is not the case for high value properties. This suggests that owners of lower value hotels should contest their valuations and get re-assessed.

Spahr and Sunderman (1998) also look at property tax inequities but concentrate on agricultural property taxation. In Wyoming, agricultural property tax is based on the land's productive value. They use land sales in Wyoming from 1989 to 1995 and data on the characteristics of the land to create a predictive hedonic model. They then estimate assessed value ratios and calculate the coefficients of dispersion and price related differentials. They compare their results to the standards used for residential property and find that there is excessive vertical, horizontal and market inequity in the taxing of agricultural land, and that farms and ranches under the existing system are being under-assessed by nearly 50%. Correct valuation could result in between 8 and 14 million extra dollars in revenue a year.

Chapman et al. (2009) focus on a land value tax and look at the effect of levying taxes on land that is mis-valued. As the use of a plot of land determines its economic value it can be difficult to determine its price and estimates can be prone to error. They find that mis-valuing the land essentially results in applying two tax rates: one to the true value of land and a different one to the capital. Land value taxation has neutral-

ity advantages but these are eliminated if the value of the land is not assessed correctly.

Smith (2000) tests for vertical inequity in Indiana for sales data from 1993 to 1998 using the assessed value to market value ratio, and tests whether the assessment process in Indiana should be overhauled. Indeed, Indiana does not use the assessed market value of a property as the basis for the tax but rather the "true tax value" which uses rules, definitions and photographs as well as depreciation tables to determine value. The system was deemed unconstitutional in 1996 because it ignored hard data on values. The full ruling was subsequently overturned but amendments were made to regulations requiring the valuation to be equitable. Smith employs the different methodologies available in the literature (most of which are explained in more detail in Section 3.5) to test Indiana's property tax. He finds that in nearly all cases a progressive assessment system is observed. He concludes that focusing just on inequity in general is limiting and that it should be related to other aspects of properties (such as location, age and type of dwelling).

Bell and Kirschner (2009) look at the effects that public pressure has had on the level of property tax in the US. They look at two ways of calculating property tax burdens that are used by States and different jurisdictions in the US, namely the tax per capita and the tax as a percentage of housing value. They outline the different methodologies and tax bases used by different US bodies for calculating burdens. Some calculate the effective tax rates by using the assessment ratio times the nominal tax rate, some use the net tax liability, some present burdens by property type or by jurisdiction. They conclude that not using microdata for the calculations of these burdens hinders any type of equity related research at a within State or within jurisdiction level.

Szalai and Tassonyi (2004) look at the issues involved in implementing a value based property tax in Hungary using the United States as an example. Taxes on residential properties, which were imposed at a municipal level were terminated in the early 1990s mostly by local politicians looking for public support. In light of EU accession, Hungary's very mixed fiscal structure needed to be overhauled and an *ad valorem* property tax coupled with economic growth and rising house prices would result in local governments having the necessary revenue to invest in the community. Hungary has much more defined local municipalities compared to Ireland and they already have fiscal responsibilities which would allow property taxes to be implemented more easily. The authors present the different administrative issues that would need to be taken into account in order to implement the new tax, such as the rates, the assessment method, the appeals procedure, and how to determine who should be exempt.

Existing property taxes in the US tend to result in some form of inequity. It is possible to determine which factors cause the inequity and consequently try and find a solution to the tax design but most of the issues which arise relate to the assessment of the property's value. The following section presents the case of Northern Ireland which recently had to update its property value database.

3.4.4 The case of Northern Ireland

In 2007, the domestic housing sector in Northern Ireland was subject to a complete revaluation of all properties for property tax purposes.¹³ The last valuation of the domestic housing sector had occurred in 1976 and in the meantime the property market had changed drastically. Previously unwealthy areas had recorded very high growth rates and the rate system had not managed to keep up with price increases in wealthy neighbourhoods. The resulting system was unfair and considered to "disadvantage the less well off" and "the system was not progressive" (Parry, 2006:3). Householders were not paying rates in line with the more recent value of their home. After public consultations and numerous reviews, a new system was introduced in 2007. Discussion had centred around whether bands should be used for valuation (as used in Great Britain) or whether to introduce discrete individual capital assessments. Government commissioned research found that the introduction of a "banded system was likely to be a regressive step. Analysis on the likely impact of an individual capital value system shows it to be more positive" (Parry, 2006:4). The new system was based on individual assessments conducted by the Valuation and Lands Agency which aimed to value 700,000 homes with a common valuation date. Discounts and special procedures were put in place for those facing a significant price hike due to the new system, those in full-time education, and disabled persons. The next revaluation is due to take place in 2012.

The system used in Northern Ireland is one that is quite common in the United States and is based on the use of a type of Automated Valuation Model (AVM). The IAAO (2003) defines AVM as a: "mathematically based computer software program that produces an estimate of market value based on market analysis of location, market conditions, and real estate characteristics from information that was previously and separately collected. The distinguishing feature of an AVM is that it is a market appraisal produced through mathematical modeling. Credibility of an AVM is de-

¹³ Referred to as "rates".

pendent on the data used and the skills of the modeler producing the AVM".¹⁴ The system is based on property market sales between 2002 and 2005 which were recorded in a specially designed Sales Inspection Database, the data from which was then used as the basis for the AVM (Gloudemans and Montgomery, 2008). The most important property characteristics affecting sales price were determined across 25 "sub-markets". Categories of dwellings (such as apartment, mixed properties) were examined separately. The final analysis produced capital values for homes based on their potential sales price on the 1st January 2005. These will then be updated again in 2012. The house price inflation followed by deflation that occurred between 2005 and 2007 was ignored for the most recent valuation.

The AVM used in Northern Ireland included significantly more variables than the present study as the valuation agency already had information about the house characteristics (McCluskey et al. 2007). The variables used were: habitable space, ancillary space, outbuildings size, primary classification (public or private build), sub-classification (detached, semi-detached, terraced), grade of construction, age band, number of storeys, heating, sewerage provision, water provision, power provision, external repair, garage, site positive features, site negative features, neighbourhood, location (rural, urban, suburban) and access type. As rates were pre-existing, the basic property database could be used for the analysis, and the addition of more information was not as costly. However the Government committed the funds to conduct a thorough revaluation of the full property stock so as to have a solid system and database on which to base any further revaluations. In Ireland, as currently no database associating geographical information system data on the location of dwellings (and then matched to their characteristics) exists, producing a database of this kind would be very expensive. In the present study we concentrate on a small number of variables, however it is clear that the more information on a dwelling the more accurate the valuation and the fairer the system.¹⁵ The following section presents the different assessment methods used in this study.

3.5 Methodology

This section reviews the different methodologies used to assess the valuation method for a property tax. First, the use of hedonic analysis in the estimation of property

¹⁴The system in Northern Ireland is called CAMA (Computer Assisted Mass Appraisal).

¹⁵In the US, some states estimate each property's taxable value from the self-assessed values of the surrounding comparable properties, which also reduces the information cost.

prices is discussed. Then, the various errors in house price assessments are presented, followed by the techniques used by the International Association of Assessing Officers (IAAO), which produces assessment standards for ratio studies and property tax policy.

3.5.1 Predicting house prices

The hedonic technique is based mainly on work by Griliches (1961) and Rosen (1974) and originated in the development of value indices for manufactured products that combined measures of quantity and quality. The seminal paper by Griliches (1961) derived a hedonic price index for motorcars. The technique centres on consumers' choices regarding composite goods. The assumption is that goods are valued for their utility-bearing attributes and that these attributes are internalised into the price of the good. A house has several attributes, including attributes of the house itself, such as the number of rooms, bathrooms and the availability of car park spaces, as well as attributes of its environs, such as crime rates, quality of schools, and air pollution. All of these attributes make different contributions to the price of the house. If you have a large enough sample of housing market transactions, it is possible to use econometrics to separate out the implicit price of the attributes. This is done using a hedonic house price model. The basic technique involves regressing the property price on the set of variables measuring quality, while controlling for unobserved time and area effects. The regression coefficients are then interpreted as the marginal implicit prices of the quality components.

The hedonic price function takes the following form:

$$Price = f(S, N, E) + \varepsilon \quad (3.1)$$

where the price (or logged price of the house) is a function of the house's structural or physical characteristics (number of bedrooms, size in square metres etc.), neighbourhood or location characteristics (such as location in the city, access to transport routes etc.) and environmental characteristics (such as proximity to green spaces, to coast, quality of ambient air); ε is an error term. The house price is thus a function of all of the attributes relating to the house and the resulting coefficients are the marginal implicit prices of the attributes.

The attributes can also be used to produce estimates of the house value. We regress house price on one or more readily observable and verifiable house attributes and use

the results to estimate a house price. How closely this estimate relates to the “correct” house price will determine the accuracy of the estimation method - and thus horizontal equity if this valuation is used as the basis for taxation. The variables included in the analysis will be based on those used in Northern Ireland and in previous hedonic analyses, such as area in square metres, number of rooms and bedrooms, dwelling type (detached, semi-detached, terraced) and geographical indicators. It will then be possible to determine how much of house value is determined by a single criterion. It is worth reemphasising at this point, that the goal of the analysis is to identify if and which inequities emerge using this basic form of regression. By deliberately making the regression simple, the estimates are likely to be biased. Just how biased and to the detriment of which types of households is what is of interest. Consequently, if we were to use hedonic valuation for the precise estimation of house prices, we would be more likely to use a detailed model such as the one presented in Chapter 4. We are not advocating the use of such simple models but rather seeing of the consequence of using such methods. It is possible that the “bulk” of a house price can be determined using just one dwelling characteristic (e.g. size) but it maybe that three or four variables are necessary in order to reach a suitably precise house price estimate. The following section details how we define “suitably precise”.

3.5.2 Evaluating the house price assessments

We look at the errors in the predictions depending on the different explanatory variables used (area, rooms, age, dwelling type), the number of variables used and who they affect the most (by age groups, income groups, areas). We consider both the absolute value of the error in euro and the percentage of the house value. The first type of error we identify is the difference between the estimated price and the true house price in absolute value. This error is calculated as follows:

$$ABV = EstimatedValue - TrueValue \quad (3.2)$$

That is, if the ABV is positive (negative), the dwelling’s value is over-estimated (under-estimated). This is a negative (positive) outcome for the home owner as they end up paying more (less) tax than they should. Inevitably, some ABVs will be positive and some will be negative. There is an implicit cross-subsidy from those being over-valued to those being under-valued. The ABV error will give an indication as to the absolute amount in euro that is being mis-valued but does not allow us to evaluate the equity of the valuation system. For that, we need to check whether there is a *systematic* mis-valuation of certain kinds of houses.

3.5.3 The assessment to sales price ratio

The assessment to sales price ratio (A/S) measures the degree of accuracy of a valuation. The A/S ratio is the ratio of a property's assessed value to its sales price. It is used extensively in the US to assess the equity of property taxes and is one of the guideline ratios recommended by the International Association of Assessing Officers (IAAO, 2010). It is calculated as follows:

$$\frac{A}{S} = \frac{\text{AssessedValue}}{\text{MarketValue}} \quad (3.3)$$

In the case of this study, we consider the properties to be assessed at 100% of their market value so the assessed value will be equal to the appraisal value (i.e. the estimate of house price). Most studies of property tax equity are based on data from existing property taxes. These studies use the assessed or estimated values of properties as per government guidelines and compare them to the market value of homes when these are then sold. In this paper we compare the assessed value of the home, as estimated through hedonic regression, to the market value, as estimated by the interviewer. If a property is correctly valued then the A/S ratio will be one. A tax is regressive if the A/S ratio declines with increasing property values and is progressive when the A/S ratios increase with higher property values. A/S ratios tend to be evaluated using the median A/S ratio for the full group or specific sub-groups. According to the IAAO (2010:13): "the median is the generally preferred measure of central tendency for evaluating overall appraisal level" as it is less affected by ratio outliers.¹⁶

From the A/S ratios it is possible to calculate measures of dispersion. Measures of variability can be calculated for different groups or sub-groups of data. The main one used in studies is the coefficient of dispersion (COD) which Spahr and Sunderman (1998) define as: "a non parametric statistic that measures the average deviation of a group of A/S ratios, taken around the median and expressed as a percentage of that measure". The COD taken around the median can be calculated using the following equation:

$$COD_{MED} = \frac{100}{A/S_{MED}} \left[\left(\sum |A_i/S_i - A/S_{MED}| \right) / n \right] \quad (3.4)$$

where COD_{MED} is the coefficient of dispersion based on the median value, A_i/S_i are the individual ratios, n is the number of ratios, and A/S_{MED} is the median ratio for the group being examined. It measures the average percentage deviation of the ratios from

¹⁶It is also possible to use the mean ratio to compare appraisals.

the median ratio. The interpretation of the COD does not depend on the assumption that the ratios are normally distributed which makes it the favoured method of evaluation. According to the IAAO (2010) "in general more than half the ratios fall within one COD of the median".¹⁷ A lower coefficient of dispersion means more consistent valuation estimates and equitable treatment.

The International Association of Assessing Officers also offers guidelines as to the acceptable level of these figures. The median A/S ratio should be between 0.9 and 1.1 (Spahr and Sunderman, 1998). An acceptable coefficient of dispersion (COD) for residential dwellings in densely populated, large cities and new properties is between 5.0 and 10.0. For large to medium sized cities the COD should be between 5.0 and 15.0 and for rural areas with older properties it can be between 5.0 and 20.0. Figures above these boundaries indicate a low degree of assessment uniformity and the presence of horizontal inequity. These assessments would be considered more variable than acceptable. The following section presents the survey data and the variables used in this analysis as well as summary statistics and checks on the reliability of the house price variables.

3.6 Data

Section 3.6.1 presents the survey and variables used in the analysis. It also includes details on any possible sample restrictions applied to the data. Section 3.6.2 discusses the "house value" variables in more detail.

3.6.1 The survey, sample and variables

The dataset used in this analysis is the Irish National Survey of Housing Quality (NSHQ) which was conducted in 2002 by the Economic and Social Research Institute on behalf of the Department of the Environment, Heritage and Local Government. It obtained information from a "representative sample of 40,000 householders on the characteristics and problems of the dwelling" (Watson and Williams, 2003). Previous national housing surveys had been conducted in 1981 and 1991 however these had been smaller surveys collected at a local authority level and then grouped. The 2002 survey is the latest one of its type to be conducted in Ireland.

¹⁷In fact, if the uncertainty about the assessed value is normally distributed, as assumed in our regression model, 67% of the observations fall in that range.

The survey contains a number of different variables useful for this analysis. The main variable of interest is house value which will be discussed in further detail in Section 3.6.2. Variables relating to the properties' characteristics are used in the hedonic regressions. These include the type of dwelling, which contains 5 categories: detached house/bungalow, semi-detached house/bungalow, terraced house, flat-apartment and flat apartment in a converted house.¹⁸ As too few observations were available, the final two categories were grouped into a single "apartment" category. Two thirds of the sample are detached houses followed by smaller fractions of semi-detached houses and terraced houses. This is not unexpected as the survey is a national survey including rural areas of Ireland where the majority of dwellings would be one-off detached houses. There are very few apartments in our sample for two reasons: first, apartment living in Ireland is a fairly recent phenomenon which is restricted to the cities. Second we only kept owner-occupiers in this sample and most apartments bought at the beginning of that decade were for rental purposes. Although this was a nationally representative survey when the survey was conducted, it is worth cautioning that this survey may not be entirely representative of the current housing stock in Ireland. In particular, a large proportion of the Irish apartment stock was built during the Celtic Tiger years and this type of dwelling is under-represented in the sample. We cannot therefore make any inferences on the potential inequalities arising from valuing apartments in this way.¹⁹

Information on the number of rooms and bedrooms in the house is available. Also stated is the age of the dwelling, which is grouped into 5 categories: pre-1940, 1941 to 1970, 1971 to 1980, 1981 to 1996 and post-1996. Respondents were also asked to state, if known, the size of their dwelling in square feet or square metres. Approximately 30% of respondents declared they knew the area size of their home and gave a figure. Consequently our sample size for the regressions using a square metre variable is much smaller. Despite the lower number of available observations, the declared areas are fairly normally distributed (Figure 3.4).²⁰ The average size in square metres is 150m².

¹⁸Caravans/mobile homes were excluded from the analysis.

¹⁹We are not advocating the use of this particular survey for the valuation of the housing stock in Ireland. We have used this particular survey for this analysis due to its availability and its national aspect but acknowledge the caveats associated with using a survey this old.

²⁰The subsample containing values for the area in square metres is representative of the larger sample. Both samples have similar proportions of dwelling types and rooms however the sample with an estimate of area size have a larger proportion of newer dwellings (post 1996) than the larger sample. This can be taken as an indication that households who made more recent purchases are more knowledgeable about the size of their home, which is to be expected. Due to this slight difference between the samples, the analysis uses both area size and number of rooms throughout.

Geographical dummies are crucial variables as these take into account any area effects that influence house prices. A number of these variables are available in the NSHQ. It states which region (Border, Dublin, Mid-East, Midlands, Midwest, Southeast, Southwest, West), location (Dublin, urban or rural) and the size of the location the household is situated in (village, town, city, open country). These variables are useful when comparing the incidence of the tax by geographical area. However, more detailed geographical location data are needed for inclusion in the hedonic models to take account of geographic house price differences. There are 34 local authorities in Ireland including 29 county councils and 5 city councils. The NSHQ data list which local authority the dwelling is located in and these data are included in the hedonic regressions and are interacted with the variables of interest.

The full sample of the NSHQ comprises over 40,000 observations. We imposed some restrictions on the data. First, for the main analysis we focus on owner-occupiers as these are the households that will directly pay the tax. Rented dwellings will also be subject to the tax and this may be passed on to renters in whole or in part. Due to Ireland's very high ownership rates, this only reduced the sample by 5,000 observations. In order to take account of coding errors relating to the price of the house (e.g. some dwellings were stated as having a 1 value) we picked a lower end cut-off point for house prices. This cut-off was 40,000 for homes outside of Dublin and 90,000 for homes in Dublin. These cut-offs are based on the minimum value in 2002 (the year of the survey) for dwellings in and outside Dublin from the ESRI/PTSB house price index micro-data. This restricted the sample to only "realistic" house prices. In order to take account of the effect of outliers, the top and bottom quarter of a percent of observations according to house value were also excluded. Due to a lack of sufficient observations we did not include owners of caravans or mobile homes in the sample. Finally, houses purchased from a local authority were also excluded as they are likely to be exempt from tax. All monetary values were converted from Irish pounds into euro. The final sample comprises 25,016 observations and 7,023 of these have a value for area size, in square metres.

The summary statistics for the variables included in the analysis are available in Table 3.3. It also includes household characteristics for the analysis of groups and sub-groups. Variables such as the occupational class of the household (manual, professional, skilled, unskilled), the age of the chief economic supporter, the principal economic status of the chief economic supporter (at work, unemployed, home duties, student, retired, ill) or which income quintile the household falls into are also available. As indicated in Table 3.3, the number of rooms in the dwelling ranges from 2 to

30 while the number of bedrooms ranges from 1 to 15. This variable seems to have a very wide range however, consistent with expectations, 50% of all dwellings have 3 bedrooms or less and 6 rooms or less.

3.6.2 The house value

The NSHQ provides three variables relating to house value. The first is the original purchase price of the home. This variable is not used in the analysis as it would result in homes being valued at different points in time. As we require a consistent estimate of house value, we considered the two other variables. The first was an estimate of house value made by the owner. The owner was asked: "What price do you think you would be able to get for the accommodation if you were to sell it now?". The second estimate was made by the interviewer. Survey interviewers tend to have a good knowledge of the area and the local prices and in 2002 the market was far from thin, making estimation easier and more likely to be accurate. The interviewer was asked the "Estimated value of the accommodation if it were to be sold" (Watson and Williams, 2003).

We subtracted the owner's estimate from the interviewer's estimate. We would hope to see fairly similar figures and the corresponding difference would then centre around zero. The differences between the two estimates for the full sample are plotted in the top panel of Figure 3.5. Approximately 83% of all values are within €20,000 of the estimate. As is expected, the graph is skewed to the left, indicating that the home owners tend to value their homes at a higher value than the interviewers do.²¹

The final mean house value is €118,000 euro with a standard deviation of €73,000. These figures vary quite a bit by region (Table 3.4) with Dublin having the highest mean house value (nearly double the mean house value at €198,000) however they are well distributed within regions (Figures 3.6 and 3.7).²² The following section presents the results and analysis.

²¹As there are a few extremely high outliers (e.g. a difference of for instance 4.6 million euro) which could have been due to coding mistakes (or very optimistic owners!) we cut the top and bottom quarter percent of values and the final sample differences are presented in panel (b) of Figure 3.5.

²²For comparison purposes, data from PTSB which provides house values by county was aggregated to the regions as set out in the NHSQ. The average house price by region in the NHSQ was lower than the data from PSTB. We cannot rule out that these differences do not relate to a downward bias on the interviewers' valuations or an underrepresentation of the most expensive houses in the NSHQ.

3.7 Results

This section presents the different segments of the analysis. First, the importance of geographical dummies when using hedonic analysis is highlighted. Then the absolute errors in estimation using different dwelling characteristics are presented followed by the A/S ratio analysis. Section 3.7.4 looks at who would be favourably and adversely impacted by different estimation methods followed by tests for vertical inequity and Section 3.7.5 looks at what would happen if property tax bands were to be used.

3.7.1 The importance of geographic dummies in hedonic regressions

The first step in the analysis is to estimate house values using dwelling characteristics. The top panel of Table 3.5 shows the results of a regression, where house value is regressed on the dwelling size in square metres. The results are as expected; as dwelling size increases, the house value increases. The adjusted R^2 for this regression is very low, at 12%. If we estimate house values using these regression results and then plot the ABV error we get the top panel of Figure 3.8. This shows, in increasing order, by how many euro the model mis-predicts the house price. Anything below the x-axis represents an under-valuation of house price by the model and anything over the x-axis represents an over-estimation. There are some very high value outliers on either end of the spectrum and the median error is €18'300. This is not unexpected as we would anticipate that regional differences would have a significant effect on house prices. Consequently the regression is re-estimated using interactions with local authority dummies and the results are presented in the bottom panel of Table 3.5. The adjusted R^2 for this regression is already much higher, at 53%. Just one variable, the size in square metres of the house when coupled with geographical dummies accounts for 53% of the variation in house prices.²³ It is immediately clear that the ABV errors have reduced and the graph in the bottom panel of Figure 3.8 is leaner. The median error falls to €7,300. Consequently, it is important to include geographical effects in the analysis. As it is possible that local authorities would administer the property tax and this is the most detailed location variable we have available, we use local authority dummies for the remainder of the analysis.²⁴ The following section takes a closer look at the different ABV errors depending on the dwelling characteristics used to estimate

²³Table 3.16 presents the full regression results including the interaction terms. The significant coefficients are of the expected sign. As an example, houses in Dublin city council command a premium of €98,000 compared to Carlow. This rises to €119,000 in Dun Laoghaire but is €26,000 less in Mayo. The effect of square metres is €218 in Dublin city council and rises to €367 in Fingal but is as low as €12 in Mayo.

²⁴An interaction of town size and county produced similar results.

price.

3.7.2 Absolute Value Errors

The least administrative costly taxes are those which require the least amount of information to collect (and are as such uncomplicated). Consequently, a property tax which could equitably be implemented using just one dwelling characteristic would be very efficient. However, the predictions based on such a simple model are very likely to result in biases as the model suffers from omitted variable bias. The previous section showed the absolute value errors using just one variable, size in square metres. This section compares the different dwelling characteristics, both separately and combined.

We estimated house value using a series of single explanatory variables, namely, number of rooms, number of bedrooms, year built (pre-1940, 1941-1970, 1971-1980, 1981-1996, post-1996) and dwelling type (detached, semi-detached, terraced house or bungalow and apartment). The summary statistics for the ABV errors for these regressions are presented in Table 3.6. The median errors are very low with number of rooms being the lowest, followed by size in square metres, number of bedrooms, dwelling type and year built. In contrast the standard deviations are much higher at approximately €50,000 to €60,000, while the outliers (minimum and maximum) are fairly similar regardless of the variable used. The results when using size in square metres and number of bedrooms are quite similar, an indication that rooms are a fairly good proxy for area. The number of rooms also has a smaller standard deviation.

Table 3.7 presents the same statistics but relating to multivariate regressions, where combinations of dwelling characteristics have been used to estimate house value. Including all variables in the regression improves the median error and standard deviation however it is clear that it is the inclusion of size related variables (number of rooms or square metres) that has the largest effect. The inclusion of five variables instead of one reduces the median error to €3,900.

Figure 3.8 shows the level of error (in euro) for each of the houses in the sample using two different regressions. The first (top panel) is the error from a regression using square metres but does not include any geographical dummies, and the second (bottom panel) does. The errors are numerically ordered. It is clear from the graph that the level of error varies from -600,000 to +300,000 in both cases. However, the number of dwellings with an error in the range of 100,000 reduces when geographical dummies

are introduced. Ideally we would want as large a number as possible to be very close to zero. Figure 3.9 shows the same results but for all the explanatory variables on the same graph. In this case we want the curve to stay as close to the zero concentric line as possible. The bottom panel shows that a large part of the errors remain very close to this zero line regardless of the explanatory variable used. There are however differences at the extremes but overall the number of room tends to stay the closest to the zero circle. If these results were presented individually, as in Figure 3.8, the bottom part of the graph would have been larger, i.e. dwellings are being under-valued at a higher level than those being over-valued. We would expect this to be a good result for homeowners, however, it depends on the proportion of the error in relation to total house price. Indeed, under-valuation could only occur on very high priced homes and over-valuations on lower priced homes, which would be inequitable. Consequently, we need to look at the errors in the context of house value, this is the subject of the following section.

3.7.3 A/S ratios and assessment errors

An estimation error of 50,000 euro is not very large for a home worth 1 million euro but is quite important in the case of a home worth 250,000 euro. We therefore look at relative errors. Table 3.8 presents the summary statistics for the assessment ratios for the full sample depending on which explanatory variable or variables are used to estimate house price. The mean and median A/S ratios are all larger than one, indicating that on average properties are overvalued. This is confirmed by the skewness values which are all positive signifying that there is more over-assessment than under-assessment. Indicators of size and type result in median A/S ratios that are smaller than 1.1, however year built, as a single explanatory variable, is beyond the recommended A/S boundary (IAAO). The addition of all variables into the model significantly reduces the median A/S ratio and brings it very close to 1. These results indicate in the first instance that part of the sample is estimated to within a quite close range. However, it is important to look at the dispersion of the estimates. The CODs for estimates using size in square metres only, number of rooms only and all explanatory variables are 28, 27 and 27 respectively. This is well above the range of 5 to 15 considered acceptable by the IAAO. This result indicates that the assessments are too variable. Figure 3.10 shows the different AS ratios plotted against house value. It is clear that the outliers at both ends of the scale are causing the variability in the AS ratios regardless of the model used. Although some are correctly assessed, due to the large variability in the assessment for the outliers there is considerable horizontal inequity. This was to be

expected due to the very simplified nature of the estimation model.

Lin (2010) estimates the number of incorrectly assessed properties using the A/S ratio and defines incorrectly assessed properties as those with A/S ratios either one standard deviation above or below the mean A/S ratio. According to this criterion, the area-only model correctly values 93% of properties. An equal proportion is correctly estimated when using all five dwelling characteristics, and 72% when using the number of rooms only (inclusive of regions in each case). This confirms the result above. Although a large number of properties are “correctly” assessed, those that are not have very large deviations. The next stage in the analysis is to investigate whether these assessment differences relate to a specific group of properties or owners. For the rest of the analysis we will only present results for three estimations: size in square metres only, number of rooms only, and the set of all 5 variables. We include number of rooms even though it relates quite well to size in square metres as this allows us to use the much larger sample as a check of the consistency of the results.

3.7.4 Horizontal inequity across groups

If the assessment model correctly estimates house prices, we should not detect any difference in assessment ratios among different types of properties or across owner groups. As indicated above, the three estimations produce median results that are within an acceptable range, however the dispersion of the results is large. By breaking down the analysis into sub-groups we identify which properties are causing the bulk of the errors. We can also identify which household groups are more likely to be mis-assessed. Consequently, we break down the A/S ratios by region, by income, by age of the chief economic supporter, by employment class and occupational status and rural versus urban dwellings. The results are presented in Tables 3.9 to 3.14. Figure 3.11 presents the median A/S ratios with their confidence intervals by the household characteristic of interest. In an ideal world we would want the bars to centre on 1. The inequities of the results are clearly visible - through the household professional group and income quintile results. The error is staggered in favour of the richest. The main groups affected are also clearly visible on the graph: the under 30s, rural households and students and unemployed are all disadvantaged. However it is worth noting that the results for the students and unemployed groups must be taken with caution due to the lack of observations available.

The regional breakdown indicates that the A/S ratios using 5 variables are all be-

low 1.1 (both for the mean and the median figures) across all regions. Using just one dwelling characteristic increases the ratios, however the median ratios remain below 1.1 across all regions. A result worth noting in the regional breakdown is that the median A/S ratios for Dublin (1.102 and 1.094) are the largest compared to all other regions when using just one explanatory variable (rooms or square metres). However it is much lower when all five dwelling characteristics are included. This is not unexpected as price variability within Dublin is much higher than anywhere else and the geographical dummies only split the city into four regions. Consequently, additional information on dwellings *within* Dublin is required and the age or type of dwelling may well, in the case of Dublin, reflect neighbourhood differences.

Table 3.10 shows that the dwellings of people with high incomes are under-assessed. The median assessment ratio gets smaller the higher the income. This would suggest that any tax imposed would be regressive. The results according to employment class also confirm this result, with high professionals' homes being under-valued (Table 3.12). Estimation results for the unskilled are not within acceptable ranges and the same applies to unemployed owners, and those in the "Home duties" and "Other" categories (Table 3.13). These results however could be due to small samples in these categories. Assessment ratios are too high for the youngest and oldest owners when using just square metres as an explanatory variable (Table 3.11). However they are all within acceptable ranges when using the full set of explanatory variables. Finally, there is only a slight difference between assessment ratios depending on whether the dwelling is located in a rural or urban area. Consequently, the value of properties of younger owners, the unskilled and unemployed tend to be over-assessed while high income and high professional owners win out. The initial indications seem to be that this type of valuation method would be regressive.

3.7.5 Predicting the tax band and difficulties in self-assessment

The Commission on Taxation suggested the use of tax bands for self-valuation. This change would allow some relaxation of the need to precisely assess property.²⁵ The bands suggested by the Commission are presented in Table 3.2 and increase in increments of €150,000. The proportion of dwellings correctly assessed using the different models is presented in Table 3.15. Table 3.15 also shows whether the dwelling was

²⁵With assessment ratios of 1.1 the "acceptable" band is within 10% of the house value (i.e. within 40,000 euro of a 400,000 euro dwelling) which even at much higher average house prices is well below the CoT bands (which increase in increments of 150,000 euro). Consequently, the proportion of homes "correctly" attributed will be higher using this criterion.

incorrectly put into a higher or a lower band than it should have. The proportion of households that are correctly assessed is on average above 78%. The use of all five available explanatory variables results in the lowest mistake rate (16%). It is also worth noting that of those incorrectly assessed, regardless of the explanatory variable used, a higher proportion end up in a lower band than in their correct band.

Therefore, a system based on hedonic regressions could be used to check fairly accurately whether self-assessed households have assessed their property into the right tax band. With such a simple estimation method however, the bands would still need to be quite large as discussed above. However, considering the results in the previous sections, it would still be necessary to put in place checks to more accurately audit very highly priced properties. At the same time, the regressivity of the system could be reduced by introducing exemptions for lower value properties and low income groups.

A tax system using property characteristics as a supplement to self-assessment might also help households who are unsure about the price of their property to more accurately estimate it. Figure 3.12 shows that the error in self-assessment as a proportion of the total property value decreases as household income increases. This suggests that lower income households are more likely to self-assess incorrectly. A property tax based on self-assessment therefore makes regressive errors too, and requires auxiliary policies just like a hedonic-based tax would.

3.8 Summary and Conclusions

This paper examines the implications of using hedonic regressions - a basic type of automated valuation system - to estimate the values of properties in Ireland, for the purpose of introducing an ad valorem property tax in the Republic. Flat rate property taxes are deemed unfair as lower income households pay the same amount as higher income households. Consequently, taxes based on the value of the property are considered more acceptable and a tax on the ownership of property is less volatile than the existing stamp duty. The implementation of a property tax does however require that a price valuation of the property exists.

We find that a system based on a handful of property characteristics (as well as location dummies) can produce estimates which are within the acceptable bounds as defined by the International Association of Assessing Officers. These estimates are produced using county dummies. Size characteristics (such as area in square metres and

number of rooms) produce more accurate estimates compared to, for instance, the age of the house. However, the more variables included, the more accurate the prediction. The precision of the geographical dummies is also important. Moreover, the estimates produced here are biased due to the omission of key explanatory variables from the regressions. It is therefore inevitable that inequities will emerge, as the results have shown. If a system of tax bands, as proposed by the Commission on Taxation (2009) were to be used, the hedonic regressions allocate approximately 80% of the sample properties into the correct band. However, we found that a valuation system of this type is regressive, resulting in lower value homes being over-valued and higher value homes being under-valued. At the extremes, using just one property characteristic does not produce equitable results. The results of this study have a number of caveats. First, the data available for the analysis was limited and is quite old. Second, no "true" house price value was included in the survey and we were forced to use an interviewer based estimate. The interviewer values were lower than average house prices in the region. This would mean that the under-valuation of lower value homes may not be as severe as expected but the under-valuation of higher value homes may be more severe. The analysis nevertheless highlights the types of inequities that will emerge when using this type of basic valuation.

Two important issues arise from this exercise. The first is the issue of timing and the second is the issue of funding. With regard to the former, if an ad valorem property tax were to be introduced in Ireland in the near future, a simple valuation system is the only feasible option. Self-assessment has been offered as a potential solution. However in a thin recessionary property market, it is difficult for households to estimate the value of their property. Without adequate checks and penalties, there are incentives to under-report house values. Hedonic house prices, or self-assessment with hedonic validation, are alternative solutions but as has been shown they result in inequities. More detailed hedonic regression models which do not suffer from the biases seen here would result in more equitable valuation but would require more data. A very simple system of this type could in parallel be used as a basis for collecting information about dwelling characteristics and help build a national geo-coded database of self-assessed price and house features. It could also be envisaged that if a banded system is used, a model based on house characteristics combined with geographical dummies could be used to help homeowners identify their own tax band. However, as was demonstrated in Northern Ireland, a system based on specific (rather than banded) assessments would be fairer.

Moreover it is worth noting that the design of a property tax has two dimensions:

the valuation of the property and the setting of the tax rate. The introduction of a property tax in Ireland in the current economic climate will inevitably be linked to specific revenue targets. These can be reached through various combinations of the valuations and the tax rates. Lower valuations combined with a higher tax rate versus a high valuation and a lower tax rate despite achieving similar revenues will yield different levels of deadweight loss for society. The balance of these should be taken into consideration in conjunction with the inequities involved in the valuation.

Regardless of the potential accuracy of such a basic estimation, a system of this type should not be seen as an adequate replacement for a properly administered automated valuation model. The ultimate objective with a tax of this type should be to create a more equitable and accurate valuation system as soon as possible. The creation of a national house price database is already underway. However, with existing house price volatility it will be ineffective unless combined with the gathering of information on the house characteristics. All homes included in the database will also have to be geo-coded so as to facilitate any future valuations and so that exact spatial sub-markets can be identified. The range of variables used in this analysis is constrained by the variables available in the dataset. Collecting information on the total size of the property site including gardens and any land would be more useful than solely using the size of the constructed property. Special consideration would also need to be given regarding the valuation of apartments. If funds are going to be put towards creating a database, this should be done properly and the experience of neighbours should also be considered. As was found in Northern Ireland, "the need for good data is imperative and the data must be in place before successful modelling is possible" (Gloudemans and Montgomery, 2008: 30). As with any type of tax, putting in place measures to protect vulnerable groups is essential.

Finally, using a consistent sample period for valuations is important. The Irish property market has fluctuated enormously in the last few years and house prices have been on a downward course since the beginning of the recession. According to the CSO Residential Property Price Index (July 2011)²⁶, property prices in Ireland have fallen by 43% since the height of 2007.²⁷ Clearly there would be distortions if some houses were valued using data from the boom period and others from the recession. If a national valuation of properties is to take place, it will be crucial that all properties are valued over the same period. What this chapter has shown is that inequities emerge when using too simple a hedonic model for valuations of this type. clearly, other variables

²⁶<http://www.cso.ie/releasespublications/documents/prices/current/rppi.pdf>

²⁷They are now back to 2002 levels.

relating to neighbourhoods and housing quality influence house prices. A more detailed model based on the same technique will be used in the next chapter to estimate the premium home owners are willing to pay to live close to green spaces.

3.9 Tables

Table 3.1: Recommendations on the Taxation of Property by the Commission on Taxation

Number	Recommendation
6.1	The provision of an up-to-date valuation base for all property and land in Ireland should be addressed as a priority issue.
6.2	Provide for an annual property tax on all residential housing units with the broad exceptions of local authority and social housing units and some other limited exceptions set out in section 4.2 of Part 6.
6.3	Stamp duty for purchasers of principal private residences should be zero-rated.
6.4	Stamp duty should continue to apply to investor purchasers of residential housing units. The rate should be competitive having regard to the transaction tax rates and thresholds that apply across the EU.
6.5	The windfall gains increases in land values due to rezoning decisions should be subject to an additional capital gains tax charge.
6.6	A recurrent property tax on land zoned for development should be introduced.

Source: Commission on Taxation (2009:154).

Table 3.2: Valuation bands as proposed by the Commission on Taxation

Group	Valuation band
A	0-150000
B	150001-300000
C	300001-450000
D	450001-600000
E	600001-750000
F	750001-1000000
G	1000001-1500000
H	1500001 and higher

Table 3.3: Summary Statistics for the main variables in the analysis

Variable	Obs	Mean	Std. Dev.	Min.	Max.
Interviewer Value	25016	118078	73178	33077	763937
Owner Value	25016	122528	76087	31502	930310
Number of rooms	25016	7.04	1.84	2	30
Number of bedrooms	25016	3.68	0.99	1	15
Year built	25016	2.60	1.24	1	5
Dwelling type	25016	1.55	0.75	1	4
Local authority	25016	16.81	9.93	1	34
Region	25016	4.30	2.35	1	8
Occupational class	25016	3.86	1.98	1	7
Rural	25016	0.50	0.50	0	1
Age of CES ^a	25016	2.91	0.81	1	4
Employment status	25016	2.25	1.74	1	7
Square metres	7023	147.26	66.90	2.00	928.57

^a Chief Economic Supporter

Table 3.4: House values by region, full sample

Variable	Obs	Mean	Std. Dev.	Min.	Max.
Border	4031	96253	48086	33865	590673
Dublin	3930	198456	102448	70881	763937
Mideast	2015	141695	67360	39378	669429
Midland	3068	93315	45644	35440	551295
Midwest	2731	92671	47067	33078	393782
Southeast	3895	98679	50721	34114	708808
Southwest	2423	115636	58219	35440	630051
West	2923	101431	56207	33078	630051

Table 3.5: Regression results: house value regressed on square metres, with (bottom panel) and without (top panel) geographical dummies and interactions

House Value	Coefficient	Std. Err
Square metres	457.38***	14.317
Constant	70024.51***	2315.762

HouseValue	Coefficient	Std. Err.
Square metres	580.20***	99
Dublin city council	93802.45***	16771
Dublin south	-745.07	18742
Dublin Fingal	49571.09***	17329
Dun Laoghaire/Rathdown	119598.6***	17184
Kildare	21687.85	19459
Kilkenny	-4191.7	19695
Laois	3883.7	18716
Longford	-16938.8	19303
Louth	48653.31***	17989
Meath	39773.84**	18040
Offaly	-5346.65	18003
Westmeath	8807.76	19764
Wexford	5951.04	18938
Wicklow	-8200.98	20452
Clare	22940.94	18321
Cork city	16198.2	17674
Cork county	29187.22	17824
Kerry	26347.85	18536
Limerick city	19079.13	18249
Limerick county	40626.15**	17884
Tipperary north	-16306.4	17730
Tipperary south	-31983	23533
Waterford city	55392.86***	17927
Waterford county	-57643.69**	23226
Galway city	25276.53	20240
Galway county	-37438.72*	19975
Leitrim	14462.84	17510
Mayo	-26377.7	18761
Roscommon	22634.56	17301
Sligo	24993.97	17466
Cavan	2121.43	19131
Donegal	21123.51	17607
Monaghan	6419.79	20422
Constant	36387.13**	15320

*** p<0.01, ** p<0.05, * p<0.1, Obs: 7023, R²: 0.17 and 0.53

Interactions of square metres and geographical dummies included in regression but not shown.

Table 3.6: ABV Error statistics depending on which individual explanatory variable is used

Variable	Obs	Std. Dev.	Median.	Min.	Max.
Square metres	7023	58319	7322.373	-490157	633198
Dwelling type	25016	56582	9585.908	-606166	244145
Number of rooms	25016	49083	6100.632	-542236	369864
Number of bedrooms	25016	53876	8299.977	-575954	415567
Year built	25016	59508	12353.17	-594469	228144

Geographical dummies were included in these regressions.

Table 3.7: ABV Error statistics depending on which combinations of explanatory variable are used

Variable	Obs	Std. Dev.	Median.	Min.	Max.
All variables	7023	48031	3888.27	-449697	333510
SqM, Brm, Yb, Dt	7023	50483	3950.39	-448209	336543
SqM and Dt	25016	55299	5711.44	-489675	502641
Rm, Brm, Yb, Dt	25016	45887	4526.06	-531562	377553
Rm and Dt	25016	47241	4993.05	-541901	359081
Brm, Rm and Dt	25016	47162	5026.39	-542436	354930

Note: "SqM" is square metres, "Brm" is number of bedrooms, "Yb" is year built, "Dt" is dwelling type and "Rm" is number of rooms.

Geographical dummies were included in these regressions.

Table 3.8: A/S ratio summary statistics for entire sample depending on explanatory variables used in the hedonic regression

Variable	Mean	Std. Dev	Median	Min	Max	Skewness	Kurtosis
SqM	1.131	0.427	1.066	0.119	8.385	2.72	26.42
Dt	1.166	0.448	1.098	0.145	4.875	1.03	5.12
Rm	1.114	0.372	1.066	0.129	4.306	1.03	5.58
Yb	1.176	0.449	1.141	0.138	4.353	0.75	4.07
Brm	1.142	0.410	1.09	0.145	4.196	0.92	4.74
All	1.084	0.340	1.036	-0.049	5.235	1.53	9.99
Rm, Brm, Yb, Dt	1.100	0.356	1.048	0.159	4.823	1.21	6.87
SqM, Brm, Yb, Dt	1.095	0.360	1.037	-0.466	5.038	1.60	10.12
Rm, Brm, Dt	1.107	0.367	1.053	-0.022	4.817	1.20	6.62
Rm, Dt	1.107	0.367	1.053	0.12	4.875	1.20	6.65
SqM, Dt	1.117	0.403	1.053	-1.129	5.820	2.12	16.12

Note: "SqM" is square metres, "Brm" is number of bedrooms, "Yb" is year built, "Dt" is dwelling type and "Rm" is number of rooms.

Table 3.9: Assessment ratios by region, using just square metres (top), just number of rooms (middle) and using all variables (bottom)

Region	Obs	Mean	Std. Dev.	Median.	Min.	Max.
Border	1241	1.135	0.430	1.059	0.17	5.80
Dublin	1225	1.138	0.428	1.102	0.24	5.32
Mideast	604	1.111	0.399	1.061	0.19	3.85
Midland	884	1.116	0.376	1.069	0.28	4.92
Midwest	784	1.167	0.528	1.074	0.32	8.38
Southeast	835	1.119	0.403	1.040	0.22	3.78
Southwest	677	1.141	0.446	1.051	0.29	4.22
West	773	1.114	0.387	1.065	0.12	3.50
Border	4031	1.113	0.376	1.059	0.19	3.96
Dublin	3930	1.103	0.347	1.094	0.20	4.24
Mideast	2015	1.107	0.373	1.050	0.24	3.42
Midland	3068	1.114	0.371	1.052	0.18	3.33
Midwest	2731	1.125	0.388	1.071	0.24	3.20
Southeast	3895	1.114	0.375	1.064	0.15	3.90
Southwest	2423	1.120	0.382	1.069	0.13	3.21
West	2923	1.119	0.371	1.094	0.17	4.31
Border	1241	1.085	0.343	1.044	0.08	3.21
Dublin	1225	1.075	0.342	1.031	-0.05	5.23
Mideast	604	1.077	0.331	1.029	0.43	3.08
Midland	884	1.089	0.329	1.037	0.28	2.86
Midwest	784	1.093	0.348	1.050	0.32	3.79
Southeast	835	1.071	0.316	1.021	0.34	2.82
Southwest	677	1.099	0.367	1.043	0.38	3.30
West	773	1.084	0.340	1.029	0.29	3.27

Table 3.10: Assessment ratios by income quintile, using just square metres (top), just number of rooms (middle) and using all variables (bottom)

Quintile	Obs	Mean	Std. Dev.	Median.	Min.	Max.
Quintile 1	534	1.288	0.477	1.230	0.32	5.54
Quintile 2	1282	1.209	0.456	1.151	0.22	8.38
Quintile 3	1517	1.152	0.426	1.088	0.12	5.80
Quintile 4	1790	1.106	0.401	1.038	0.17	4.92
Quintile 5	1900	1.041	0.393	0.979	0.19	5.32
Quintile 1	2971	1.200	0.392	1.164	0.20	3.90
Quintile 2	5380	1.159	0.374	1.115	0.24	3.96
Quintile 3	6026	1.135	0.374	1.084	0.17	3.21
Quintile 4	5870	1.088	0.360	1.038	0.17	3.50
Quintile 5	4769	1.014	0.343	0.967	0.13	4.31
Quintile 1	534	1.113	0.349	1.090	0.33	2.74
Quintile 2	1282	1.132	0.356	1.085	0.37	3.79
Quintile 3	1517	1.112	0.347	1.064	0.28	3.27
Quintile 4	1790	1.078	0.328	1.028	-0.05	3.30
Quintile 5	1900	1.025	0.322	0.980	0.08	5.23

Table 3.11: Assessment ratios by age of reference person, using just square metres (top), just number of rooms (middle) and using all variables (bottom)

Age	Obs	Mean	Std. Dev.	Median.	Min.	Max.
Under 30	324	1.201	0.498	1.125	0.45	5.80
30-45	2280	1.098	0.391	1.046	0.19	5.32
46-64	3133	1.125	0.427	1.058	0.12	8.38
65+	1286	1.186	0.462	1.112	0.17	5.54
Under 30	994	1.078	0.345	1.018	0.26	3.39
30-45	6449	1.069	0.350	1.020	0.13	4.31
46-64	11417	1.116	0.366	1.070	0.15	3.38
65+	6156	1.162	0.402	1.112	0.17	3.96
Under 30	324	1.085	0.333	1.046	-0.05	3.21
30-45	2280	1.074	0.328	1.025	0.29	5.23
46-64	3133	1.091	0.339	1.046	0.32	3.79
65+	1286	1.082	0.362	1.035	0.08	2.92

Table 3.12: Assessment ratios by employment status of reference person, using just square metres (top), just number of rooms (middle) and using all variables (bottom)

Employment	Obs	Mean	Std. Dev.	Median.	Min.	Max.
High Professional	1293	1.021	0.393	0.961	0.19	5.32
Low Professional	1543	1.072	0.387	1.012	0.17	4.39
Other non	1228	1.148	0.411	1.076	0.12	4.37
Skilled	1113	1.161	0.381	1.114	0.22	4.11
Semi-skilled	571	1.194	0.461	1.128	0.36	5.54
Unskilled	357	1.302	0.427	1.308	0.39	4.48
Unknown	918	1.221	0.525	1.136	0.35	8.38
High Professional	2963	1.006	0.345	0.956	0.13	4.24
Low Professional	4635	1.048	0.362	0.991	0.15	4.31
Other non	4610	1.124	0.374	1.070	0.26	3.96
Skilled	4178	1.138	0.358	1.098	0.24	3.20
Semi-skilled	2498	1.145	0.357	1.105	0.26	3.33
Unskilled	1731	1.204	0.368	1.164	0.31	3.42
Unknown	4401	1.168	0.393	1.117	0.25	3.38
High Professional	1293	1.028	0.340	0.978	0.28	5.23
Low Professional	1543	1.056	0.334	1.013	0.08	2.98
Other non	1228	1.095	0.334	1.042	0.32	3.27
Skilled	1113	1.100	0.322	1.058	0.43	2.48
Semi-skilled	571	1.098	0.317	1.069	-0.05	2.77
Unskilled	357	1.158	0.348	1.124	0.41	2.78
Unknown	918	1.135	0.370	1.085	0.33	3.79

Table 3.13: Assessment ratios by employment class of reference person, using just square metres (top), just number of rooms (middle) and using all variables (bottom)

Employment	Obs	Mean	Std. Dev.	Median.	Min.	Max.
At work	4924	1.110	0.409	1.048	0.12	8.38
Unemployed	163	1.341	0.591	1.212	0.43	5.45
Home duties	537	1.170	0.424	1.119	0.37	3.33
Student	28	1.328	0.928	1.068	0.78	5.80
Retired	1328	1.162	0.446	1.083	0.30	5.54
Other	43	1.188	0.319	1.170	0.54	1.87
At work	15417	1.093	0.363	1.046	0.13	4.31
Unemployed	602	1.207	0.377	1.154	0.28	3.24
Home duties	2792	1.140	0.387	1.103	0.24	3.96
Student	92	1.076	0.282	1.064	0.53	1.76
Retired	5877	1.142	0.383	1.101	0.20	3.90
Other	236	1.217	0.401	1.126	0.37	2.88
At work	4924	1.084	0.335	1.033	-0.05	5.23
Unemployed	163	1.143	0.401	1.080	0.38	3.30
Home duties	537	1.082	0.331	1.044	0.44	2.82
Student	28	1.142	0.451	1.099	0.79	3.21
Retired	1328	1.073	0.347	1.026	0.32	2.92
Other	43	1.156	0.354	1.164	0.57	2.07

Table 3.14: Assessment ratios by rural versus urban, using just square metres (top), just number of rooms (middle) and using all variables (bottom)

Rural/Urban	Obs	Mean	Std. Dev.	Median.	Min.	Max.
Urban	3288	1.104	0.437	1.043	0.19	8.38
Rural	3735	1.155	0.417	1.088	0.12	5.54
Urban	12556	1.078	0.331	1.049	0.15	4.24
Rural	12460	1.150	0.406	1.096	0.13	4.31
Urban	3288	1.044	0.311	1.003	-0.05	5.23
Rural	3735	1.118	0.360	1.067	0.08	3.30

Table 3.15: Proportion of households being put in the incorrect estimation band by explanatory variable used

Variable	Obs	Inc. band %	% Lower	% Higher
SqM	7023	22.53	13.58	8.94
Dt	25016	20.03	11.86	8.17
Rm	25016	17.15	10.39	6.76
Yb	25016	21.40	12.77	8.63
Brm	25016	19.05	11.99	7.06
All 5 variables	7023	19.95	10.85	9.10
Rm, Brm, Yb, Dt	25016	16.18	9.32	6.86
SqM, Brm, Yb, Dt	7023	20.65	11.28	9.37
Rm, Brm, Dt	25016	16.63	9.71	6.92
Rm and Dt	25016	16.67	10.02	6.65
SqM and Dt	7023	22.40	12.07	10.32

Note: "SqM" is square metres, "Brm" is number of bedrooms, "Yb" is year built, "Dt" is dwelling type and "Rm" is number of rooms.

3.10 Figures

Figure 3.1: Average national house prices (1975-2008, nominal terms) in Ireland for new and second hand houses. Source: Department of the Environment, Heritage and Local Government.

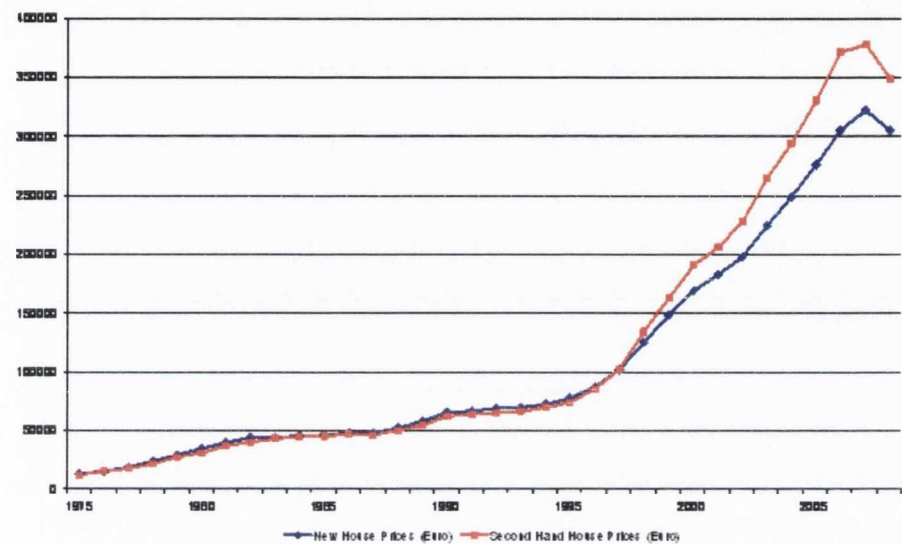


Figure 3.2: Quarterly house price growth (1996-2010) in Ireland for national and Dublin prices. Source: ESRI/PTSB House price index, Base Q3 2003

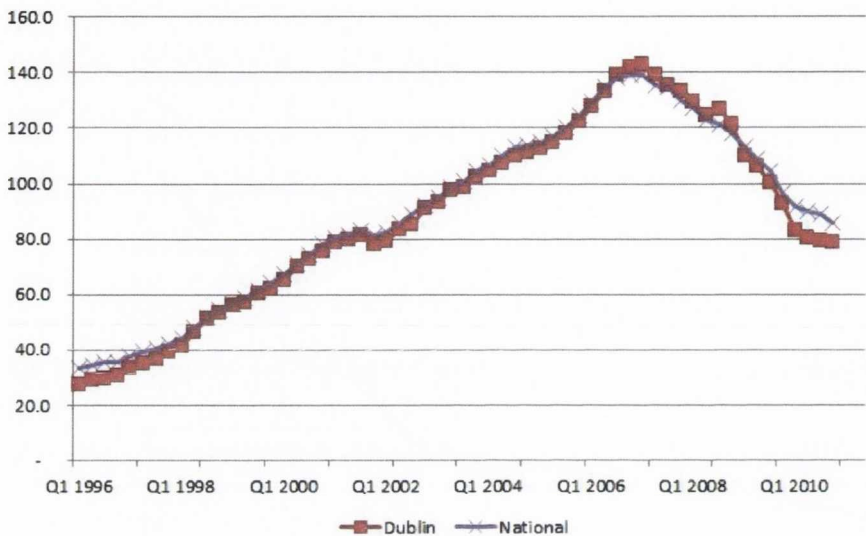


Figure 3.3: Net receipts of stamp duties (2002-2007) in Ireland (€millions). Source: CSO Construction and Housing in Ireland Report, 2008 Edition. Based on figures in Table 5.4.

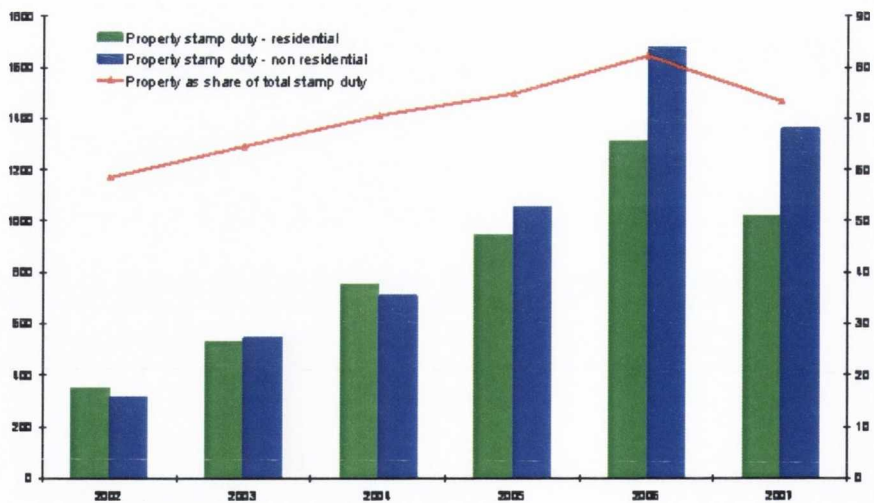


Figure 3.4: Distribution of dwelling area size (in square metres) for respondents who knew the size of their home with a normal distribution plot.

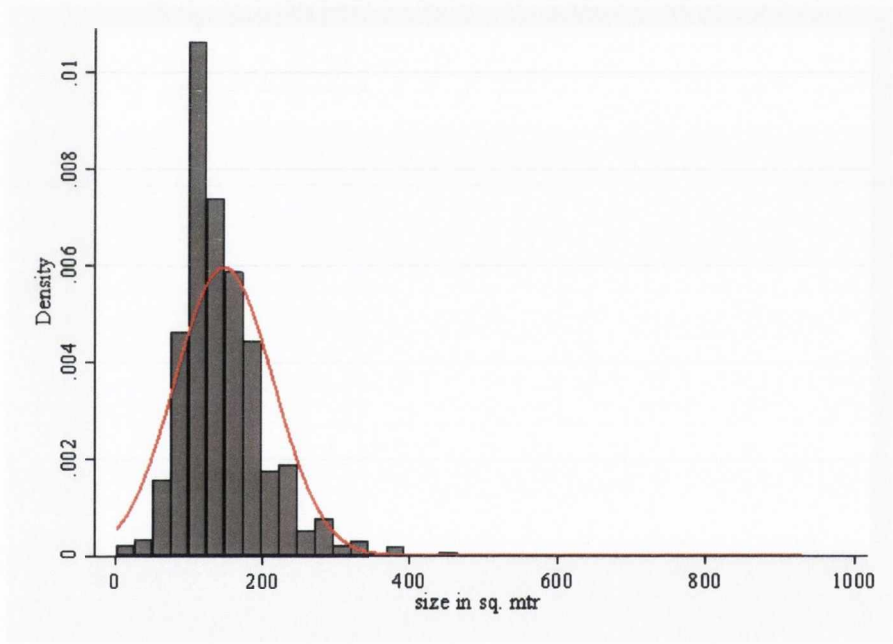


Figure 3.5: Plot of interviewer estimate minus owner estimate of house value. The top panel is the original sample. The bottom panel is the reduced sample.

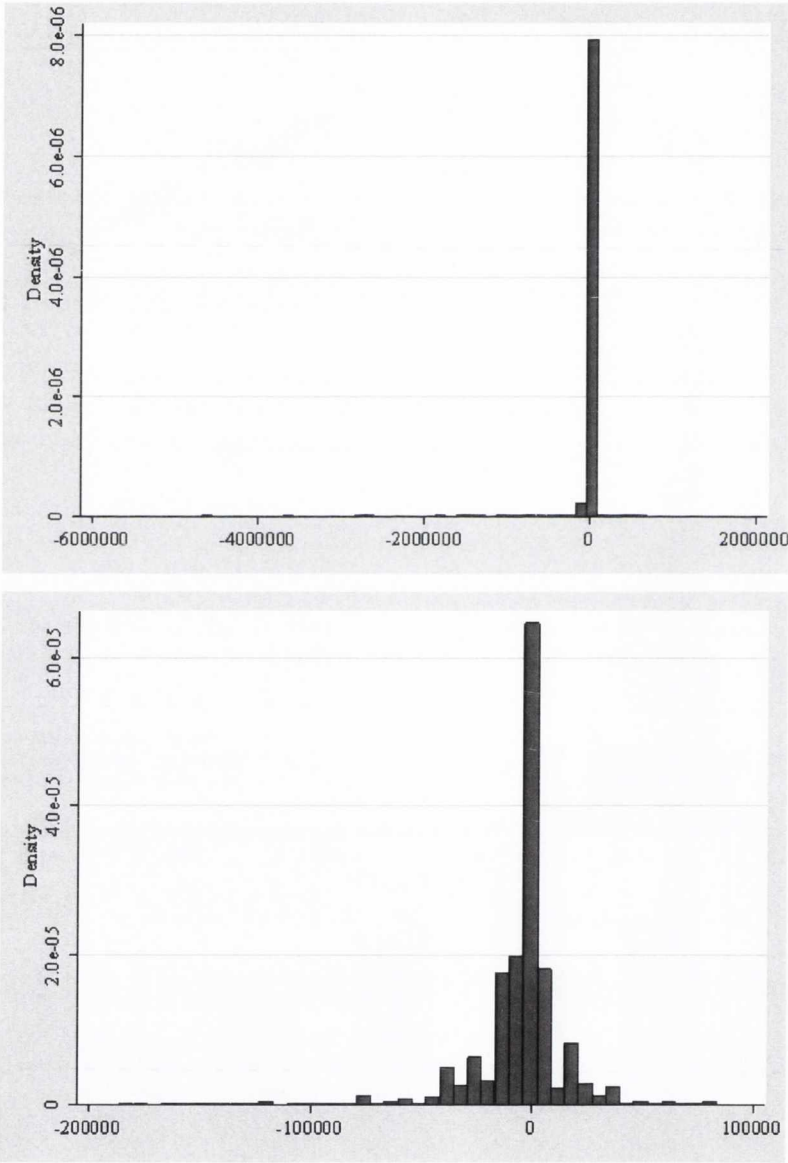


Figure 3.6: Histogram plots of house value by region (from left to right, top to bottom): Dublin, Border, Mideast, Midlands.

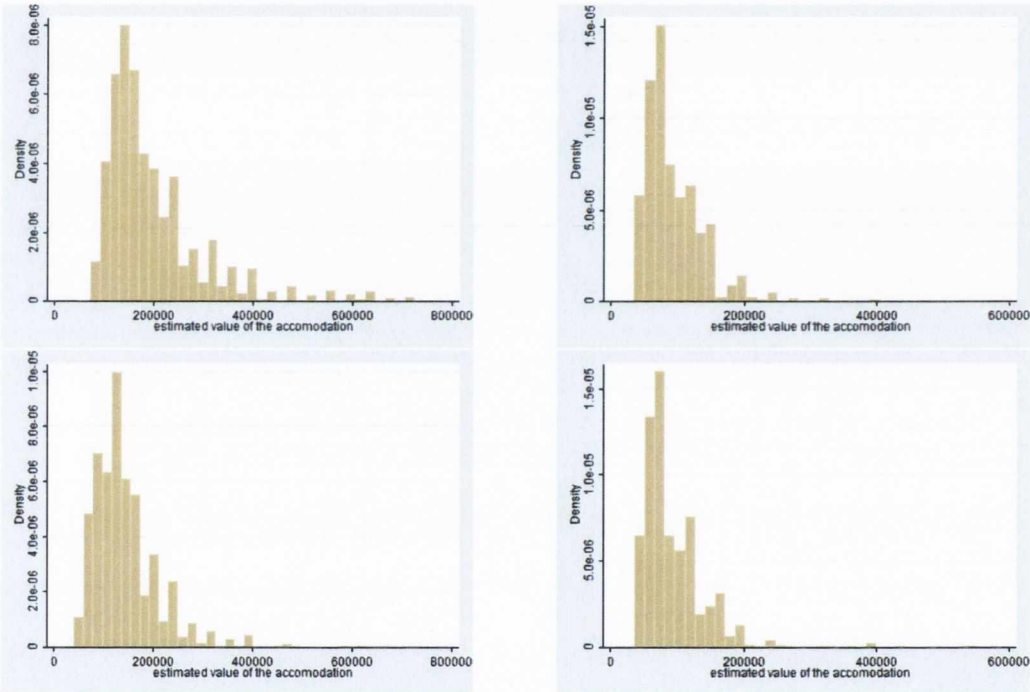


Figure 3.7: Histogram plots of house value by region (from left to right, top to bottom): Midwest, Southeast, Southwest, West.

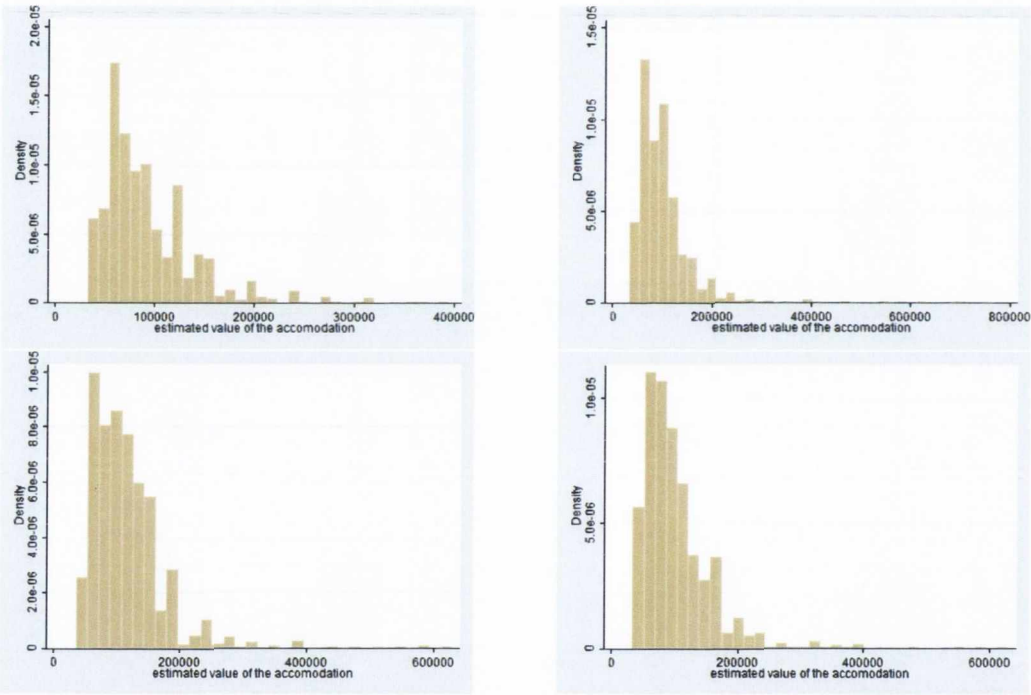


Figure 3.8: ABV errors from a regression using square metre dummies without geographical dummies (top panel) and with geographical dummies (bottom panel).

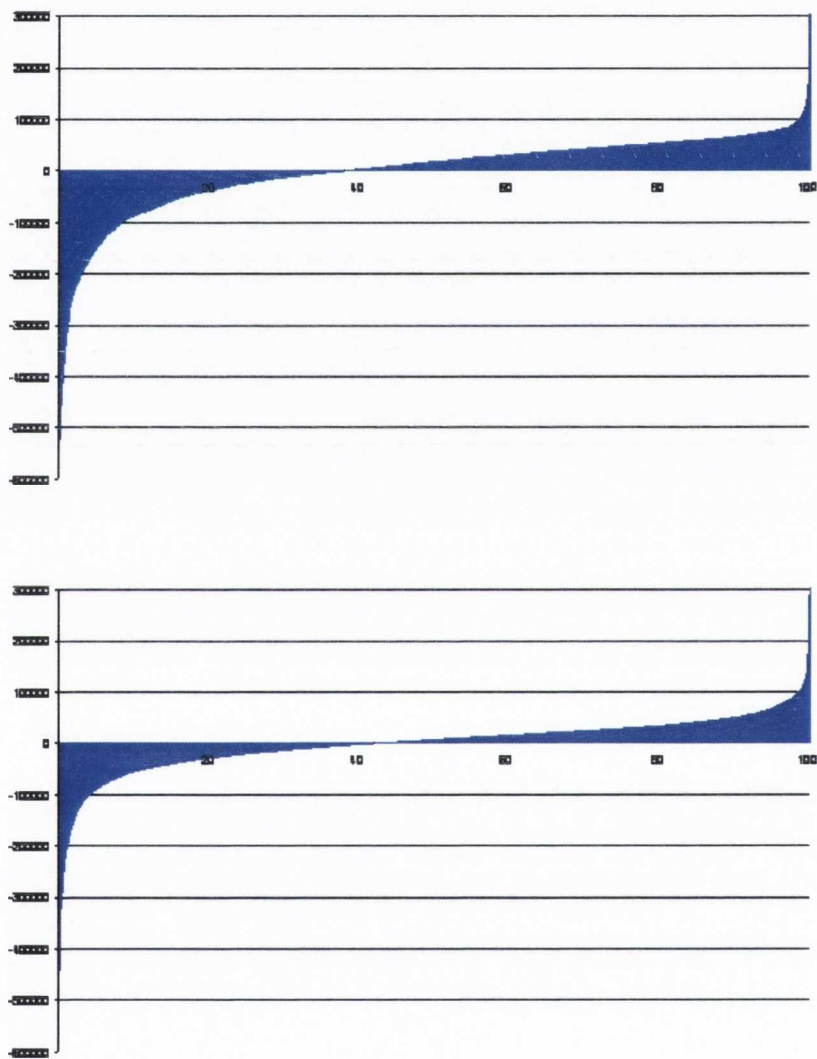


Figure 3.9: ABV errors from regressions using dwelling type, number of rooms, number of bedrooms and year built as individual explanatory variables with local authority dummy variables.

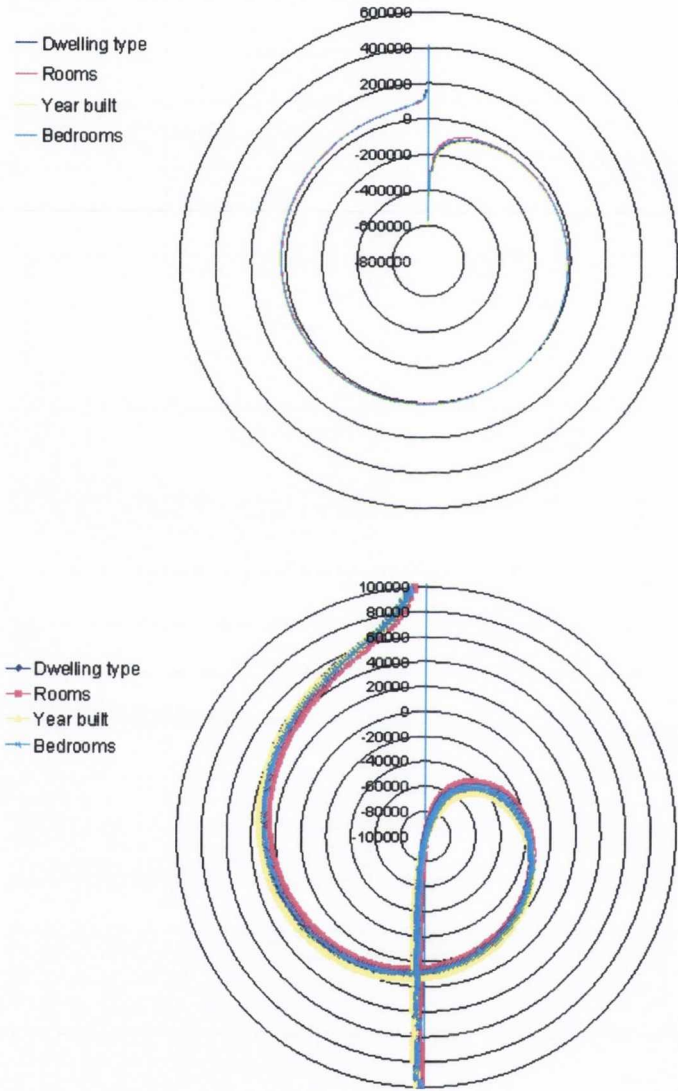


Figure 3.10: Assessment ratios against property price calculated using just size in square metres (top), just number of rooms (middle) and all five explanatory variables (bottom).

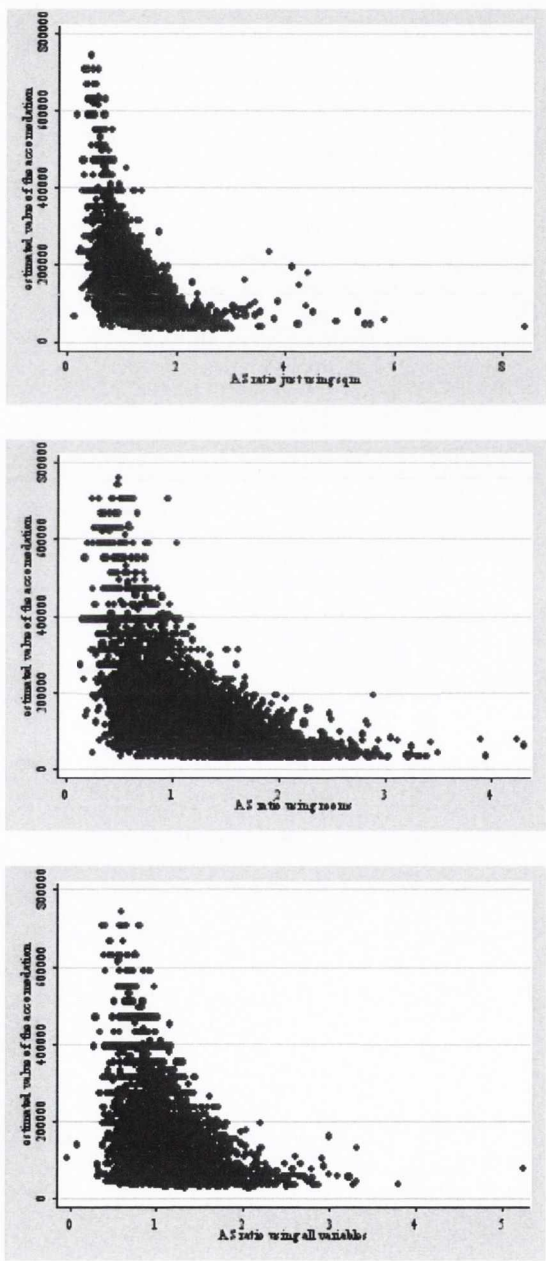


Figure 3.11: Summary of A/S ratios by household characteristics.

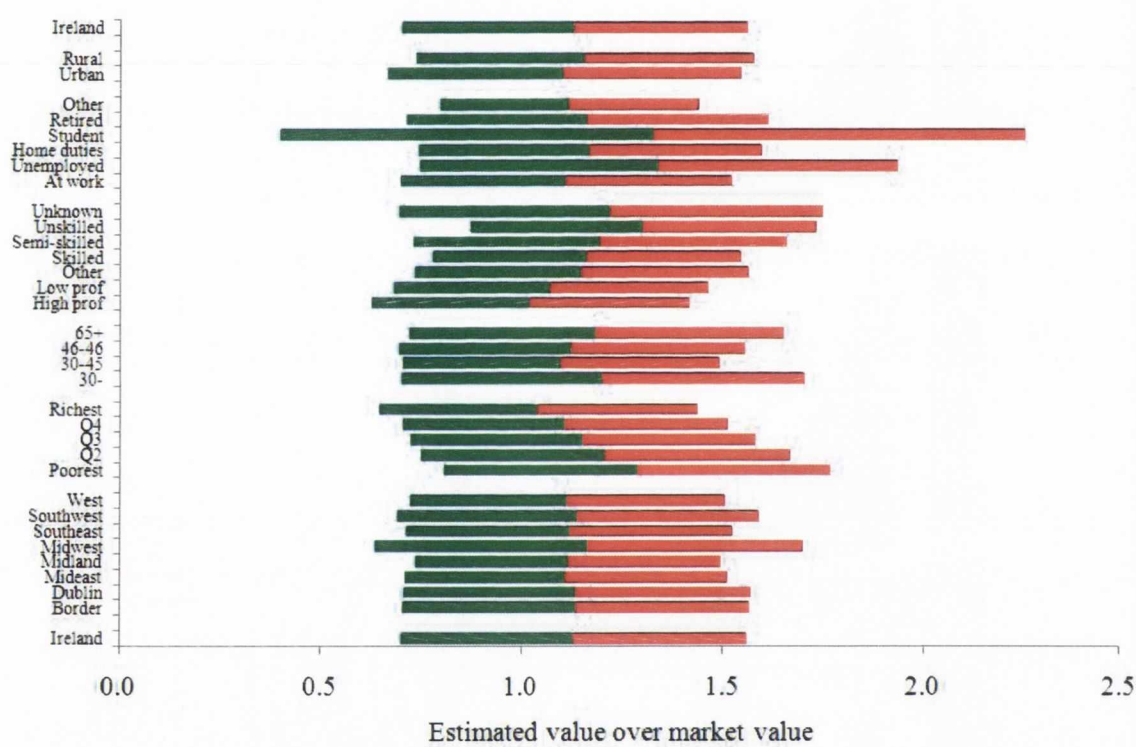
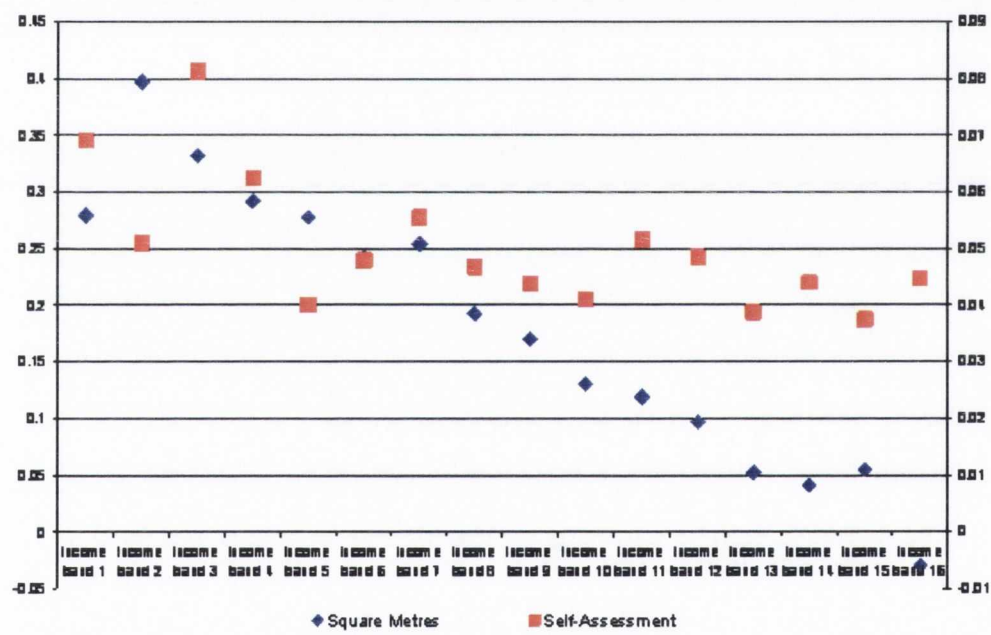


Figure 3.12: Mean assessment error (for self-assessment and for hedonic house price model results using just square metres) as a proportion of the property value, by income band.



3.11 Appendix

Table 3.16: Full regression results

House Value	Coefficient	Std. Err.	Constant	
Square metres	580.2***	98.71	Interactions	36387.13
Location dummies				15319.64
Dublin city council	93802.45***	16771.06	Dublin city council	218.8292**
Dublin south	-745.0656	18741.8	Dublin south	109.1378
Dublin Fingal	49571.09***	17329.48	Dublin Fingal	604.3988***
Dun Laoghaire/Rathdown	119598.6***	17183.85	Dun Laoghaire/Rathdown	367.3266***
Kildare	21687.85	19458.8	Kildare	228.2088**
Kilkenny	-4191.703	19695.41	Kilkenny	99.556
Laois	3883.696	18716.07	Laois	122.6063
Longford	-16938.75	19302.73	Longford	-81.9312
Louth	48653.31***	17989.41	Louth	122.5231
Meath	39773.84**	18040.29	Meath	-132.6515
Offaly	-5346.651	18003.31	Offaly	115.8565
Westmeath	8807.755	19764	Westmeath	-93.2265
Wexford	5951.035	18937.67	Wexford	123.2555
Wicklow	-8200.981	20452.03	Wicklow	-216.0703**
Clare	22940.94	18321.43	Clare	110.232
Cork city	16198.2	17674.18	Cork city	-24.2581
Cork county	29187.22	17823.96	Cork county	113.7302
Kerry	26347.85	18535.91	Kerry	-45.8232
Limerick city	19079.13	18249.07	Limerick city	-67.3766
Limerick county	40626.15**	17884.07	Limerick county	126.6113
Tipperary north	-16306.38	17730.43	Tipperary north	-133.6084
Tipperary south	-31983.04	23533.07	Tipperary south	120.3989
Waterford city	55392.86***	17927.34	Waterford city	638.6155***
Waterford county	-57643.69**	23225.83	Waterford county	-275.1704**
Galway city	25276.53	20239.97	Galway city	115.6622
Galway county	-37438.72*	19974.52	Galway county	77.7382
Leitrim	14462.84	17510.28	Leitrim	-135.1805
Mayo	-26377.72	18761.23	Mayo	112.5526
Roscommon	22634.56	17300.68	Roscommon	-280.5543**
Sligo	24993.97	17466.47	Sligo	-122.3376
Cavan	2121.43	19131.34	Cavan	-386.0359***
Donegal	21123.51	17607.16	Donegal	-115.2546
Monaghan	6419.794	20422.38	Monaghan	97.6427

*** p<0.01, ** p<0.05, * p<0.1, Obs: 7023, Adj R²: 0.53.

Omitted category: Carlow.

Part II

Environmental amenities in Ireland

Chapter 4

A Hedonic Analysis of the Value of Parks and Green Spaces in the Dublin Area

This chapter is developed from a joint working paper with Dr. Seán Lyons, Dr. David Duffy and Prof. Richard S.J. Tol. Special thanks is owed to Sherry FitzGerald for allowing access to the data. This chapter was presented at the Agricultural Economics Society of Ireland Annual Conference (2010), Dublin, the 1st Annual Irish Environmental Economics Network Conference (2010), Galway, and the Environmental Protection Agency, Environmental Policy Seminar (2010), Dublin. This is the second paper that is part of a research series on hedonic analyses conducted at the Economic and Social Research Institute, Dublin, Ireland. Three working papers using the Sherry FitzGerald data have been produced to date. The first focuses on transport links and the rail network in Dublin. Mayor, K., S. Lyons, D. Duffy and R.S.J. Tol, Forthcoming. "A Hedonic Analysis of the Value of Rail Transport in the Greater Dublin Area". *Journal of Transport Economics and Policy*. The third paper looks at the value of cultural heritage and is under review at *Environment and Planning A*. Moro, M., Mayor, K., S. Lyons, and R.S.J. Tol, 2011. "Does the housing market reflect cultural heritage? A case study of Greater Dublin". *Stirling Economics Discussion Paper*, 2011-07.

4.1 Introduction

In rapidly expanding urban areas with increasing land prices, construction of housing and commercial property tend to become very lucrative and green spaces can be crowded out. Construction of new housing in Dublin was one of the driving forces of the economy in the last decade and the construction of new builds maintained a very high rate throughout the Celtic Tiger years, with the national average reaching 18 new builds per 1,000 inhabitants in 2007 compared to a European average of 5.3 per 1,000 inhabitants.¹ While the construction of housing itself is not necessarily a problem, the management of large scale construction projects without the necessary or recommended level of master-planning that is required to maintain coherent urban spaces can be problematic. Due to the speed and scale of construction projects in the Irish capital, there are now areas of Dublin which have inadequate school, transport or retail infrastructure. The absence of such amenities has an effect on the house prices in these areas. Much of public policy, understandably, concentrates on amenities of public infrastructure such as education and transport. The benefits of the provision of green spaces and parks in residential areas have however only recently been emphasised by policymakers.

According to Dublin City council there are approximately 3,500 acres of park land and open spaces within Dublin City.² The types of green spaces in Dublin are wide ranging. Dublin also boasts the largest enclosed urban public park in Europe, Phoenix Park. Proximity to green spaces is valued by homeowners for the recreational and aesthetic value of the space. It has also been shown that access to green spaces can positively affect people's health. Hartig and Fransson (2009) looked at the effect of access to leisure homes in Sweden on the probability of early retirement for health reasons. They found that men with access to green spaces were less likely to have health problems later in life. They also refer in their paper to the "grey:green ratio" which is the ratio of green spaces to urban built areas. They state that people might be:

"able to live in and enjoy high urban residential densities if outdoor spaces were designed to better afford visual and physical access to natural areas, even if those areas were small in size" (2009: 93).

Hence, access to green spaces provides recreational, aesthetic and physical benefits, however as access to most urban parks and green spaces is free it can be difficult to establish the monetary value attached to them. In this regard, we need to use non-market

¹CSO (2008), *Construction and Housing in Ireland* 2008 Edition, Table 11.1, p.64, Total House Completions in Ireland, Ireland and Euroconstruct countries.

²<http://www.dublincity.ie/RECREATIONANDCULTURE/DUBLINCITYPARKS/Pages/parks.aspx>

valuation techniques, such as the hedonic house price model which will be further explained later.

In this chapter we ask whether private homeowners value living close to green spaces and parks and if they differentiate between types of outdoor spaces. In order to do this we built a unique GIS dataset comprising the location and characteristics of houses purchased between 2001 and 2006 in the Dublin Region, the Republic of Ireland's capital city, and the location and type of green spaces in proximity of each home. Although studies on the value of parks have been conducted before (as detailed in the following literature section), the focus of this chapter is on a major urban centre and a capital city, which to our knowledge is quite unique.³ The remainder of this chapter is set out as follows. The next section presents some of the most relevant papers in the area. Section 4.3 presents the model used in this paper: the hedonic house price model. Section 4.4 describes the data used in the study. Section 4.5 presents the results for two different aspects of green spaces. Finally, Section 4.6 provides a discussion and conclusions.

4.2 Literature

The value of green spaces has become quite a topical research question and quite a few hedonic studies have been conducted in a variety of countries. One of the earliest papers on the value of parks is by Weicher and Zerbst (1973). They refer to work by Milton Friedman (1962) who identified the problem of valuing city parks and how to charge those benefiting from the park but who do not access it: passers-by enjoying the scenery. Friedman identified the externality to be valued, the aesthetic value of the park, and Weicher and Zerbst made an attempt at valuing it for five parks in Columbus, Ohio. They included three variables relating to parks: whether or not the house was close to and facing a park, whether it was close to and backing onto a park and if the house was close to the park and facing an area with heavy recreational use. They found that the highest premium was paid for houses facing a park and backing onto parks (23% more). However, houses close to recreational areas sold for less. They attribute this to a loss of privacy and security. Since this paper, various other studies of the value of green spaces have been conducted in the United States and elsewhere. These have focussed on different types of green spaces, ranging from large spaces such

³Since writing this paper, a working paper on the city of London has been issued and is discussed in the Literature section. Smith, D. 2010. "Valuing housing and green spaces: Understanding local amenities, the built environment and house prices in London". *GLA Economics*, Working Paper 42, Greater London Authority: London.

as forests and parks to smaller areas such as the landscaping in front of specific houses.

On a larger level, Cho et al. (2008) use data of sales of over 9,000 houses in Knoxville, Tennessee in 2001 to value urban forests. They find that evergreens and more natural looking forests are valued in more rural areas whereas areas that are more urban place a higher value on having tree variety (mixed forest species). Tyrvaainen and Miettinen (2000) look at the same type of urban forests but in Finland in 1996. They use four different variables relating to green spaces: distance to the nearest wooded recreational area, the direct distance to the nearest forested area, the relative amount of forested areas in the housing district, and the view from the dwelling window. They find that a one kilometre increase in the distance to the nearest forested area leads to an average 5.9% decrease in the market price of the dwelling. Views onto forests increase prices by 4.9%.

On a smaller scale, Voicu and Been (2008) focus on the effect of community gardens on property prices in New York city between 1974 and 2003. They use a difference-in-difference specification of a hedonic regression model on residential properties and find significant positive effects in the poorest neighbourhoods. They also find that high quality gardens have the highest impact indicating that the quality of the space is also important.

Finally, Des Rosiers et al. (2002) look at the green areas directly related to the house and the property's outdoor landscaping. They use data on 760 houses in Quebec sold between 1993 and 2000. This is a very detailed analysis which looks at the percentage tree cover on the property and in the neighbourhood as well as landscaping attributes (e.g. landscaped patio and curbs, rock plants etc.). They find that the positive effect of trees on property prices is higher when there is a larger proportion of retired people living in the area. They also find that hedges and landscaped walls add 4% to the price of a property through their visual impact. Anderson and Cordell (1988) provide even more detail and find that houses in Georgia with on average five trees in the front yard sold for 4% more than houses without, and that trees of medium to large sizes were valued more.

Studies on the value of other types of green spaces are numerous and use similar methodologies. Bolitzer and Netusil (2000) use dummies for homes located within different distance thresholds of green spaces while Bender (1997) uses a similar variable: distance to green areas. Hobden et al. (2004) concentrate on park sizes and whether or not they include trails. Leigh and Coffin (2005) look at the effect of an announcement

of a change of regulation on the price of houses close to brownfields in Atlanta. Hand et al. (2008) concentrate on the effect of forests on house prices and wage markets and find that living very close to a forest increases house prices in Arizona and New Mexico but also that having access to forests in nearby areas results in increased house values and lower wages. Luttik (2000) uses separate variables for being within the vicinity of a park and having a view of an open space in a study in the Netherlands while Espey and Owusu-Edusai (2001) look at park size and available amenities in South Carolina. Crompton (2005) provides a comprehensive literature review of studies conducted in the US and McConnell and Walls (2005) survey the literature since the 1970s.

Three similar papers to this study are those by Mansfield et al. (2005), Geoghegan (2002) and Wachter and Wong (2008). Mansfield et al. (2005) use the density of green areas within three circular buffers around the houses in their sample. They also use the distance to the edge of a parcel of greenness defined as an area of vegetation. Finally they include dummies for whether or not a house is close to the perimeter of a green parcel. We use similar variables in this analysis. Mansfield et al. (2005) use a dataset consisting of 11,500 house sales in North Carolina between 1996 and 1998. They find that the closer a house is to green areas the higher the value of that property and that the overall level of "greenness" in an area also adds value to a property. They conclude that parcel greenness can substitute for proximity to a forest. Geoghegan (2002) conducts a similar study on residential land values in Maryland and uses similar distance and density variables. An original feature of this paper is that her data allow her to split her open space variable into developable open space and permanent open space. She finds that people value permanent open space much more than green space that might be developed in the future. This has important policy implications at a zoning level. Finally, Wachter and Wong (2008) also investigate the effect of green spaces on house prices, in particular curbside treeplanting in Philadelphia. They use GIS data to look at the potential effects of two types of planting schemes (group and individual) and find that the premium of being situated close to a tree parcel varies between 7% and 11% depending on the distance to the parcel and the type of scheme, but also find that the price difference can be linked to the creation of social capital.

Hence studies have shown that green spaces are valued for their aesthetic value, amenity value, health benefits and visual impacts. Secure and well-kept spaces are favoured, however being located too close to a very popular park can be a negative externality (for noise, nuisance and security reasons). Parks with bigger, more mature trees are favoured, as are mixed species forests. Homeowners will pay a premium to live in a "green" neighbourhood and will pay even more to live somewhere that is not

likely to be developed in the future. Most of these results pertain to small cities in the United States. Here we will attempt to see if they are also valid for a European capital city undergoing major urban change: Dublin.

To our knowledge the only other study of a capital city is that by Smith (2010). This study builds on a previous GLA study published in 2003 which did not use Geographic Information System data for the locational variables. Consequently, the 2010 study uses a more detailed spatial scale. They define two types of spaces : "formal green space" and "general open space" which are the equivalent of "parks" and "green spaces" in the present study. Applying a similar hedonic methodology he finds that each hectare of formal green space within 1km of the dwelling attracts a premium of approximately 0.08% and a regional park within 600m between 1.9 and 2.9%. The buffers differ from the ones used in this study but our results are of a similar range. These figures tend to be larger than those of studies in other urban centres. For London this may be an indication that an extra premium exists when households are located in a dense capital city. This may also be the case for the more densely populated areas of Dublin. The next section will describe the model used in more detail.

4.3 Model

Due to the non-market nature of parks and open spaces, it is difficult to measure their benefits to society. Different non-market valuation techniques have been used to try and estimate the value of open spaces. Contingent valuation is a stated preference technique where respondents are directly asked how much they would be willing-to-pay for a specific amenity. It has the advantage of taking a range of benefits (recreational, aesthetic, etc.) into account, but is subject to bias as the responses are entirely dependent on the respondent's decision and not based on any market transactions. The travel cost method has also been used to estimate the value of certain major park amenities. The use of this method is however only possible if access to the park is costed. In that case the value of the park is estimated depending on the number of visitors, cost of access and distance travelled by those visiting. On the other hand, hedonic analyses use unrelated market-based transactions as indicators of a person's preferences. The hedonic technique is based mainly on work by Griliches (1961) and Rosen (1974) and originated in the development of value indices for manufactured products that combined measures of quantity and quality. A hedonic analysis decomposes the price of a good into the prices of the separate characteristics of that good and centres on consumers' choices regarding composite goods.

The most common hedonic analysis is the hedonic house price model which was used in Chapter 3. A house can be decomposed into its constituents, for instance, number of rooms, bathrooms, size of the garden and car park spaces. All of these attributes make different contributions to the price of the house. On top of these house characteristics, neighbourhood characteristics also contribute to house prices. If you have a large enough bundle of houses, it is possible to econometrically separate out the implicit price of the attributes.

The hedonic price function takes the following form:

$$Price = f(S, N, E) + \varepsilon \quad (4.1)$$

where the price (or logged price) of the house is a function of the house's structural characteristics (number of bedrooms, size in square metres etc.), neighbourhood characteristics (such as location in the city, access to transport routes etc.) and environmental characteristics (such as proximity to green spaces, to the coast, quality of ambient air). ε is the error term. The house price is thus a function of all of the attributes relating to the house and the resulting coefficients are the marginal implicit prices of the attributes. This analysis uses ordinary least squares and a semi-log specification which is the most common specification in these types of studies. Consequently the results will represent the percentage change in the price of the house for a one unit change in the explanatory variable.

Our basic model looks as follows:

$$\ln(H) = \alpha X + \beta Z + \delta Y + \gamma Q + \eta T + \lambda E + \kappa Greenspace + \phi Parks \quad (4.2)$$

where $\ln(H)$ is the log sale price of the property, X is a vector of property related characteristics, Z are a series of location fixed effects which take account of unobserved neighbourhood features, Y and Q are a series of year and quarter of sale dummies to take account of any time or seasonality trends. T are a series of transport related variables and E are a series of environmental variables. Finally we include park and green space variables. We estimate two types of models. The first uses the density of green spaces and parks and the second looks at the effect of specific named parks on house prices. The following section details the house price data used in the analysis and the GIS based environmental variables included.

4.4 Data

4.4.1 Housing data

The dataset used in this analysis is composed of a spatially referenced house price and characteristics dataset, which is related to information on the location of green spaces and parks using Geographical Information Systems (GIS) software. The house price dataset was provided by Sherry FitzGerald, Ireland's largest property advisory group. The dataset consists of a representative sample of house sales facilitated by Sherry FitzGerald in the Dublin area between January 2001 and December 2006. This amounted to just over 9,700 dwellings. The complete addresses were used, along with the An Post Geodirectory⁴, to geo-code the data. Not all addresses in the original database were amenable to geo-coding. Our valid sample size after geo-coding was 6,956, covering most of the Dublin area (see Figure 4.1) and a wide range of house prices. This is not only a large sample but also very detailed and location specific. A comparison of the dataset with other sources of housing market data indicates that our sample has an average price for houses that is higher than other sources. However, this reflects the fact that the majority of transactions within our dataset take place in South Dublin, the part of the city that is generally more expensive than other areas. Indeed, Sherry FitzGerald focuses on the top end of the housing market.

The available structural variables are the floor space measured in square metres, the number of bedrooms, the presence or not of a utility room, of parking and of a garden, whether the heating system is gas heating or not and the condition of the house as assessed by the real estate agent (excellent, fair, poor, very poor). The type of dwelling is also included (apartment, detached house, semi-detached house terraced house and cottage) as well as the period in which the house was built (pre-1900, 1900-1950, 1950-1975, 1975-2000, post-2000).

4.4.2 Control variables

We also incorporated GIS data from a number of sources for the environmental, transport and location controls. The environmental variables include the distance (in metres) to the nearest beach and to the coastline. These data were provided by the Environmental Protection Agency (EPA). Transport variables include three types of rail transport: proximity to train stations, commuter rail stations and light rail stations as

⁴The definitive database of buildings in the Republic of Ireland.

well as distance to tracks.

We allow for unobserved heterogeneity in area characteristics through the use of locality dummy variables, and we include quarterly dummy variables (from Q1 2001 to Q4 2006) to control for house price inflation. Potential seasonality in house pricing is also accounted for using a dummy for each calendar month. The locality dummies represent neighbourhoods, and each is made up of one or more electoral divisions (EDs) sharing a common area name. There are more than 284 electoral divisions within the Dublin Region, with an average of 24 houses within each ED in our sample. 90 locality dummies representing neighbourhoods at a lower disaggregate spatial level were built.⁵ Each of these areas is made up of one or more EDs sharing a common area name, which brings the average number of houses per area to 78. The data on ED boundaries come from the national mapping agency, the Ordnance Survey Ireland. We considered using individual electoral division variables as locality controls, but we have too few observations to allow the use of such small area dummies (there are over 200 electoral divisions in our sample area).

4.4.3 Green space data

Green space data was extracted from the CORINE 2000 project courtesy of the EPA and combined with European Environmental Agency data on green urban areas within urban zones.⁶ This allows us to identify all types of green spaces in Dublin (small tree areas, neighbourhood greens, neighbourhood parks and larger city parks). All of these areas are visible in Figure 4.2. We were then able to locate what we call “open access named parks”. We identified 22 of these in Dublin. This was a somewhat subjective selection based on the authors’ knowledge of the largest parks, the most popular tourist parks and those with extra amenities (sports facilities, art installations, memorials). This was the basis of the separation between two of our main “greenness” variables: green spaces and park space (any of these 22 parks).

The next step was to create buffers zones⁷ around each of the houses and calculate

⁵Various levels of aggregation of the ED dummies were tested as were the inclusion of much broader area dummies (based on postcodes or general region within the city). However, it was found that when using these areas the impact of additional explanatory variables was large. However, when including the locality dummy variables the results were stable to the introduction of other explanatory variables.

⁶EEA (2005), *Results of the extraction of Green Urban Areas (GUA) from satellite data within Urban Morphological Zones (UMZ)*, <http://dataservice.eea.europa.eu/dataservice/metadetails.asp?id=912>.

⁷The buffer zones delineate an area around each one of the houses within which the density of green spaces is then calculated.

the density of the two types of green space within those buffer zones. We also created variables for the Euclidean distance to the nearest park boundary and dummies for each of the parks which equal 1 if that park is the closest named park to that property. A specific Phoenix Park dummy was included in all the models, it is equal to 1 if Phoenix park is the closest park to a property and zero otherwise.

The following section presents the results of the analyses. The first section presents the overall neighbourhood and house characteristics. The second section focuses on the density of green spaces and parks and the final section underlines the results and issues relating to individual public access parks.

4.5 Results

We look at the effect of green spaces from two perspectives: first, the density of green spaces and parks in the vicinity of houses; and second the effect of distance to specific named parks. We start with a short description of the results relating to other variables included in the analysis.

4.5.1 Transport, neighbourhood and house characteristics

A hedonic analysis must include a number of different control variables in order to be able to single out the effect of a park. We include in the regressions other variables that might affect house prices. This section presents the results of the different estimations. First, we include a number of area and time dummies to take account of seasonal and neighbourhood effects on price. These included quarter dummies and monthly dummies for the time of sale as well as locality dummies to take account of intrinsic neighbourhood characteristics.

Second, we include some transport related variables to take account of proximity to public transport access. Distances to train stations, tram stations and rapid rail transit stations as well as train tracks are included. The main transport and environmental variables are reported in Table 4.1. We find that living within 2km of a light rail station adds between 9% and 11% to house value (depending on which tram line it is). The older rapid rail transit stations add approximately 3% to house values in the vicinity.⁸

⁸Mayor, K., S. Lyons, D. Duffy and R.S.J. Tol, Forthcoming. "A Hedonic Analysis of the Value of Rail Transport in the Greater Dublin Area". *Journal of Transport Economics and Policy* provides a more in depth analysis of the transport related variables.

Environmental and recreational amenities such as distance to beaches and coasts are also included. We find that living very close to a beach confers a negative externality while living in the vicinity adds to house value. The coefficients on coast are all positive and significant. The positive effect decreases the further the purchased house is located from the coast.

Finally, the house characteristics themselves are also included in the analysis. The results of the coefficients on house attributes are in line with expectations. Floor space, the number of bedrooms, the presence of a utility room, a parking, a garden, a gas heating system are all positive and significant.⁹ Fair, poor and very poor conditions are negatively associated with house price, (with respect to excellent condition); detached houses command a higher price with respect to semi-detached, while the other types of dwelling command a lower premium. Houses built prior to 2000 command a lower price than houses built after the year 2000, with the exception of very old dwellings. As expected, the bigger the house the higher its value. People place a higher value on detached houses with a garden, a parking space, that are in excellent condition and that are either very new or very old (pre-1900). We now turn our attention to the value of green spaces and parks in Dublin.

4.5.2 Density of green spaces

As mentioned above we defined two types of green spaces. The first, which we refer to as “parks” includes all large green spaces in Dublin that would be referred to as an open access park. These include Dublin’s largest park, Phoenix Park and a variety of other local parks. The data also allowed us to identify green areas in Dublin. These were areas with trees and grass that were not necessarily parks but could be used as an outdoor play space. These “greens” are fairly common in housing estates in Ireland. They do not necessarily have any extra amenities (playgrounds, picnic tables) but are visible pieces of nature in the city. Consequently, we analysed the effect of green spaces and separately of parks on house prices in Dublin.

As the green spaces can be quite small and dotted around an area, the distance to the nearest green space may not be a useful indicator. Consequently we placed geo-

⁹By way of comparison, in Chapter 3, an extra bedroom adds €15,554.29 to the price of a home (anywhere in Ireland). In Chapter 4 an extra room adds approximately 3% to the price of a home. On the basis that the average home in the sample in Chapter 4 is €350,000 this represents an average premium of approximately €10,500. These results are approximately in the same range, although it should be noted that the models are not completely comparable.

graphical buffers around each house in our sample and measured the percentage of each buffer that consisted of green spaces or parks. Table 4.2 presents different buffer sizes and their corresponding measure in square metres. For instance, with a 100 metre buffer, a 1% increase of green space within that buffer refers to approximately an extra 313 square metres of green space.

Table 4.3 presents the regression results relating to the different green space models. Model 1 presents the results using a 200 metre and a 2000 metre buffer for the green space and park density variables. A Phoenix park dummy is also included. A 10% increase in green spaces within 200 metres of a house results in a 9.0% increase in house prices. A 10% increase in green spaces in areas between 200 metres and 2 kilometres of a house results in a 7.6% increase in house prices. People seem to be valuing proximity and possibly the visual impact of green spaces more than their larger neighbourhood being green. Nevertheless living in a green neighbourhood is still highly valued by property owners.

Similar variables for parks were included in the regression in Model 1 with 200m and 2000m buffers. However, there was no significant difference between the two buffers and in Model 2 only the 2km buffer was used. A 10% increase in park area within 2km of a house results in a 6.6% increase in the house price. People do not place a significantly different value on parks depending on whether they are very close to their house or within walking distance. We also checked to see if combining the two types of green spaces would affect the results. The results are presented in Model 3. We find that a 10% increase in green spaces or park area within 200 metres of a house increases the house price by 9.0% and by 7.0% if within 2000 metres, both of which are in the same range as the estimates above.

The Phoenix Park is by far the biggest park in Dublin and results for houses in any of the areas surrounding the park could be affected. A Phoenix park dummy was included and is 1 if Phoenix Park is the closest named park to the house and zero otherwise. The dummy variable is positive and weakly significant and its inclusion in the model does not affect the results for the parks and green spaces variables.

It should also be noted that the results are dependent on the type of locational dummies used. The impact of different area dummies on the coefficients for the variables of interest (parks and green spaces) are presented in Table 4.4. The coefficients remain of the same sign and approximately of the same order of magnitude regardless of the locational dummies used. However, as the locality dummies get "finer" the coefficient

reduces and as the locality dummies get “larger” the coefficient increases. Our results are in the middle of the range.

These premium results are consistent in size with the Mansfield et al. (2005) study. However, what the present analysis has shown is that people value different types of green spaces differently depending on their distance from the house. Homeowners will pay a premium to live in an area with more green spaces and place even more value on having these spaces within a very close vicinity to their house. Being able to access parks is also important but property owners do not necessarily require the parks to be on their doorstep. The following section presents the results for the regressions estimating the impact of specific named parks.

4.5.3 Distance to named public access parks in Dublin

Given the likely importance of a park’s quality for its value, it would be interesting to identify the value of individual parks. Dublin has a number of parks of different sizes, quality and offering different amenities located around the city. These are of different sizes and can sometimes be located quite close to each other. We first included the logged distance to each of the 22 parks we had identified. The results are presented in Table 4.6 in the Appendix (Section 4.9). The first model includes all 22 parks and the next few models include just some of the parks. As Dublin is a compact city with a large number of parks, we suspected we may have a multicollinearity problem. When alternate parks were eliminated the coefficients, signs and significance of the other parks changed a lot. As all the parks were spatially quite close to each other it was not possible to identify the specific effects of each park and hence it was not possible to include them all in the analysis.

Consequently we decided to concentrate our efforts on a selection of parks. We then picked five parks: Phoenix Park as it is the biggest park in Dublin, St. Anne’s Park, a popular walking park in North Dublin, Marlay Park in the South, Ringsend in the East and Tymon in the West. We acknowledge that there is no objective reason for picking these parks and not others but we decided to pick the biggest ones in each section of the city.

We used a dummy equal to 1 if the park was within 2500 metres of the house in any direction and zero otherwise. Table 4.5 presents the results including the dummies with the 2500 metre buffer. Again they remain fairly stable regardless of how many

of the dummies are included. Only two variables are statistically significant. Being within walking distance of Phoenix Park confers a positive externality and increases house prices and being within walking distance of Tymon park a negative externality. The negative coefficient on the Tymon dummy may be due to congestion issues in and around the park at weekends as it is used for a variety of sports events. However, it is more likely to be due to its close proximity to Dublin's busiest motorway, the M50.

4.6 Discussion and Conclusions

We use a hedonic house price model to estimate the impact of green spaces on house prices in Ireland. We find that green areas have a positive externality and result in higher house prices in the neighbourhood. People distinguish between green space in the close vicinity of their house and within walking distance in their neighbourhood. The premium associated with having a greener neighbourhood is between 7% and 9% depending on the proximity buffer used. We also find that people value being close to park amenities but do not value close proximity more highly. This implies that the green spaces provide an aesthetic and visual externality whereas the parks are preferred for their recreational value.

We also included a number of transport and structural variables to ensure the model was taking account of as many possible factors affecting the house price. We find that living close to public transport links is considered a positive amenity. However being located close to train tracks is a negative externality due to noise and visual intrusion. Environmental variables such as proximity to coast and beaches also command a premium.

Although it was not possible to measure the exact value of specific city parks, the results of the density regressions do indicate to city planners and developers the value of maintaining green spaces. Home buyers value the advantage of being close to a park but what the results in this paper emphasise is that home buyers place even more importance on having a certain level of "greenness" outside their front door. As advocated by Hartig and Fransson (2009) high density urban living areas, such as some recently developed in Dublin would be considered more pleasant to live in if a certain level of green space was maintained in the area. However, purely providing outdoor spaces and green spaces may not be a final policy solution in itself as their positive benefits on health and well-being depend on their use and not just on their availability. The potential impacts of access to green spaces and the determinants of use of outdoor play spaces for children in particular is the subject of the next chapter.

4.7 Tables

Table 4.1: Transport, Environmental and Structural Variables

	Coefficient (Std.Error)	
Within 1500m of a train station	0.01	(0.01)
Within 2000m of a commuter rail station	0.03**	(0.01)
Within 1000m of a train track	-0.01	(0.01)
Within 250m from a beach	-0.27***	(0.06)
Between 250m to 500m from a beach	0.13***	(0.05)
Between 500m to 1km from a beach	0.07***	(0.02)
Between 1000m to 1500m from a beach	0.03**	(0.02)
Within 250m from the coast	0.22***	(0.02)
Between 250m to 500m from the coast	0.16***	(0.02)
Between 500m to 1km from the coast	0.12***	(0.01)
Between 1000m to 1500m from the coast	0.06***	(0.01)
Within 2000m of the south city light rail line, zone 2	0.11***	(0.01)
Within 2000m of the south city light rail line, zone 3	0.09***	(0.02)
Within 2000m of the north city light rail line,	0.02	(0.03)
Ln(Floor Area)	0.64***	(0.01)
Number of bedrooms	0.03***	(0.00)
Existence of a utility room	0.04***	(0.01)
Gas heating	-0.02***	(0.01)
Garden	0.03***	(0.01)
Parking	0.01**	(0.01)
Good condition	-0.03***	(0.01)
Fair condition	-0.08***	(0.01)
Poor condition	-0.09***	(0.01)
Very poor condition	-0.15***	(0.03)
Unknown condition	-0.02	(0.03)
Apartment	-0.02	(0.02)
Detached house	0.16***	(0.01)
Terraced house	-0.07***	(0.01)
Cottage	-0.08***	(0.03)
Built pre-1900	0.16***	(0.01)
Built pre-1950	0.09***	(0.01)
Built pre-1975	-0.03***	(0.01)
Built pre-2000	-0.04***	(0.01)
Quarter dummies	Yes	Yes
Monthly dummies	Yes	Yes
Locality dummies	Yes	Yes

* significant at 10 per cent level, ** significant at 5 per cent level,

*** significant at 1 per cent level; robust standard error in parentheses.

Omitted categories: Excellent condition, Semi-detached, Built post-2000.

Table 4.2: Buffer zones and their related area size

Buffer Zone	Total area in m^2	1% increase in area in m^2
100m	31374	313
200m	125580	1255
500m	785189	7851
1000m	3141175	31411
2000m	13000000	125655

Table 4.3: Results - park and green space density regressions

Variables	Model (1)	Model (2)	Model (3)
Phoenix Park dummy (within 2000m)	0.02*	0.02*	0.02*
	-0.01	-0.01	-0.01
% of park space within 200m of house	0.90***	0.90***	
	-0.1	-0.1	
% of green space between 200m and 2000m of house	0.76***	0.76***	
	-0.1	-0.1	
% of park space within 200m of house	0.66***		
	-0.08		
% of park space between 200m and 2000m of house	0.67***		
	-0.08		
% of park space within 2000m of house		0.67***	
		-0.08	
% of park and green space within 200m of house			0.89***
			-0.07
% of park and green space within 2000m of house			0.75***
			-0.07
Constant	9.06***	9.06***	9.06***
	-0.06	-0.06	-0.06
Transport dummies	Included	Included	Included
Beach-Coast dummies	Included	Included	Included
Housing dummies	Included	Included	Included
Quarter dummies	Included	Included	Included
Monthly dummies	Included	Included	Included
Locality dummies	Included	Included	Included
Observations	6956	6956	6956
R-squared	0.89	0.89	0.89

* significant at 10 per cent level, ** significant at 5 per cent level,

*** significant at 1 per cent level; robust standard error on second line.

Table 4.4: Sensitivity of Model 1 results to locational dummies used

Variables	ED	Locality	Location	Region
Model 1	288 dummies	90 dummies	70 dummies	17 dummies
Phoenix	0.0316*** (0.0097)	0.0216*** (0.0099)	0.0252*** (0.0092)	0.0262*** (0.0097)
Green spaces 200m	0.9515*** (0.1452)	0.8911*** (0.0868)	0.7926*** (0.0772)	0.4573*** (0.0558)
Green spaces 2000m	0.8860*** (0.1486)	0.7486*** (0.0855)	0.6932*** (0.0768)	0.2974*** (0.0553)

* significant at 10 per cent level, ** significant at 5 per cent level,

*** significant at 1 per cent level; robust standard error in parentheses.

Table 4.5: Results - Dummy park is within 2500m of property

Variables	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)
Ringsend park within 2500m	0.01 -0.02	0.01 -0.02				
Phoenix park within 2500m	0.14*** -0.02		0.15*** -0.02			
Marlay park within 2500m	-0.01 -0.02			-0.01 -0.02		
Tymon park within 2500m	-0.05*** -0.02				-0.06*** -0.02	
St. Anne's park within 2500m	0.02 -0.02					0.02 -0.02
Constant	9.13*** -0.06	9.13*** -0.06	9.13*** -0.06	9.13*** -0.06	9.13*** -0.06	9.13*** -0.06
Transport dummies	Included	Included	Included	Included	Included	Included
Beach-Coast dummies	Included	Included	Included	Included	Included	Included
Housing dummies	Included	Included	Included	Included	Included	Included
Quarter dummies	Included	Included	Included	Included	Included	Included
Monthly dummies	Included	Included	Included	Included	Included	Included
Locality dummies	Included	Included	Included	Included	Included	Included
Observations	6956	6956	6956	6956	6956	6956
R-squared	0.88	0.88	0.88	0.88	0.88	0.88

* significant at 10 per cent level, ** significant at 5 per cent level,

*** significant at 1 per cent level, robust standard error on second line.

4.8 Figures

Figure 4.1: Map of Dublin with location of sample houses used in the analysis (red dots). Source: Map by authors. Data from Sherry FitzGerald.

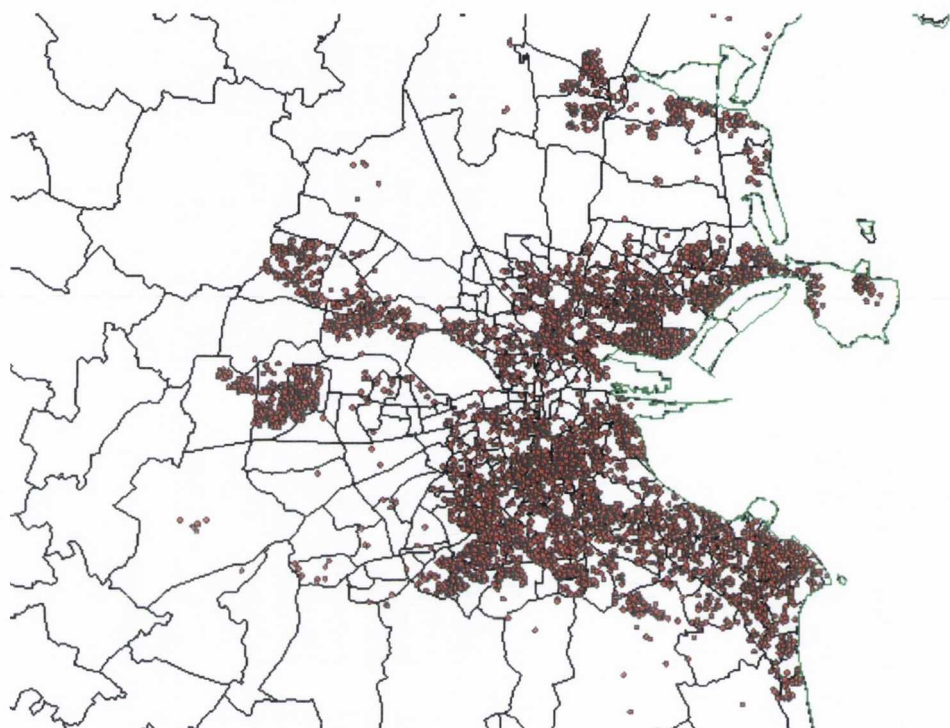
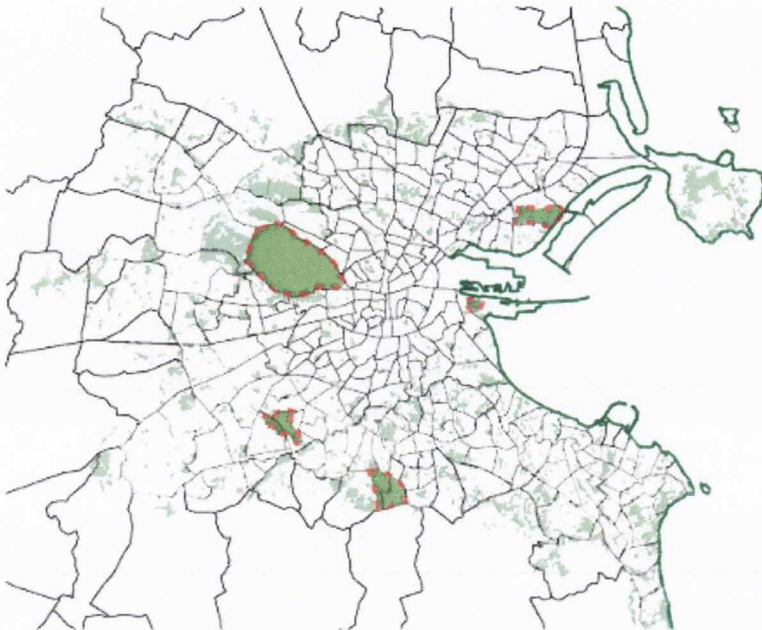


Figure 4.2: Map of Dublin with green spaces used in analysis (green areas)



4.9 Appendix

Table 4.6: Results - Logged distance to individual parks

Variables	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
Poppintree	0.45** -0.21		0.74*** -0.2	0.29 -0.21	0.36* -0.19
Ellenfield	-2.79** -1.41		-4.72*** -1.36	-1.78 -1.43	-2.28* -1.34
St. Anne's	-0.03* -0.02		-0.01 -0.02	-0.05*** -0.02	-0.04** -0.02
Prospect	-0.18*** -0.04		-0.23*** -0.03	-0.17*** -0.04	-0.10*** -0.03
Phoenix	-0.63*** -0.09		-0.62*** -0.09	-0.65*** -0.09	-0.32*** -0.03
Griffith	-0.05*** -0.02	-0.09*** -0.02		-0.06*** -0.02	-0.06*** -0.02
Ringsend	0.21*** -0.04	0.22*** -0.04		0.25*** -0.04	0.22*** -0.04
S.Moore	-0.12*** -0.04	-0.13*** -0.04		-0.23*** -0.04	-0.12*** -0.04
E.Ceannt	0.10*** -0.03	0.09*** -0.03		0.10*** -0.03	0.12*** -0.03
Kenilsworth	-0.10*** -0.03	-0.09*** -0.03		-0.18*** -0.02	-0.10*** -0.03
Palmerstown	-0.12*** -0.02	-0.12*** -0.02	-0.13*** -0.02		-0.12*** -0.02
Tymon	0.24*** -0.04	0.24*** -0.04	0.22*** -0.03		0.33*** -0.03
Bushy	0.11** -0.05	0.12** -0.05	0.10** -0.05		-0.10*** -0.02
Deer	-0.19*** -0.02	-0.19*** -0.02	-0.19*** -0.02		-0.16*** -0.02
Dodder-Knocklyon	-0.10*** -0.02	-0.09*** -0.02	-0.09*** -0.02		-0.12*** -0.02
Riversdale	0.05** -0.02	0.05** -0.02	0.05** -0.02	-0.12*** -0.02	
Longmeadows	0.46*** -0.08	-0.04 -0.04	0.48*** -0.08	0.60*** -0.08	
Cardiffsbridge	-0.24*** -0.06	-0.41*** -0.05	-0.28*** -0.06	-0.23*** -0.07	
Tolka	0.28*** -0.06	0.21*** -0.05	0.31*** -0.06	0.26*** -0.06	
Terenure	-0.27*** -0.05	-0.28*** -0.05	-0.27*** -0.05	-0.20*** -0.02	
St. Ed	0.11*** -0.01	0.11*** -0.01	0.11*** -0.02	0.13*** -0.01	0.13*** -0.01
Lorcan O'Toole	0.04* -0.02	0.07*** -0.02	0.05** -0.02	0.05** -0.02	

* significant at 10 per cent level, ** significant at 5 per cent level,

*** significant at 1 per cent level; robust standard error on second line.

Table 4.7: Results Locality Dummies

	(1)	
	lprice	
Arran Quay	0.422***	(0.0412)
Ashtown	0.138***	(0.0324)
Ayrfield	0.140***	(0.0310)
Balbriggan	-0.126***	(0.0263)
Baldoyle	0.200***	(0.0260)
Balgriffen	0.850***	(0.0257)
Ballinteer	0.204***	(0.0232)
Ballyboden	0.237***	(0.0176)
Ballybough	0.289***	(0.0357)
Ballybrack	0.131***	(0.0463)
Ballygall	0.348***	(0.0266)
Ballymun	0.214***	(0.0309)
Beaumont	0.302***	(0.0142)
Blackrock	0.530***	(0.0224)
Blanchardstown	-0.0150	(0.0111)
Bohernabreena	0.189***	(0.0177)
Botanic	0.462***	(0.0328)
Cabinteely	0.316***	(0.0282)
Cabra	0.281***	(0.0279)
Castleknock	0.192***	(0.0221)
Chapelizod	0.0685	(0.0532)
Cherry Orchard	0.0816*	(0.0436)
Churchtown	0.306***	(0.0298)
Clondalkin	-0.0695*	(0.0419)
Clonskeagh	0.408***	(0.0309)
Clontarf East	0.537***	(0.0232)
Clontarf West	0.359***	(0.0191)
Crunlin	0.417***	(0.0357)
Dalkey	0.536***	(0.0306)
Donabate	0.0684	(0.0484)
Drumcondra	0.394***	(0.0245)
Dun Laoghaire	0.429***	(0.0226)
Dundrum	0.345***	(0.0260)
Edenmore	0.110**	(0.0483)
Edmonstown	0.171***	(0.0274)
Finglas	0.0290	(0.0852)
Firhouse	0.273***	(0.0114)
Foxrock	0.478***	(0.0283)
Glencullen	0.299***	(0.0299)
Grace Park	0.293***	(0.0282)
Grange	0.161***	(0.0215)
Harmonstown	0.214***	(0.0315)
Holmpatrick	0.578***	(0.0303)
Howth	0.375***	(0.0335)
Inchicore	0.243***	(0.0773)
Inns Quay	0.383***	(0.0418)
Killiney	0.240***	(0.0360)
Kilmainham	0.242***	(0.0533)
Observations	6956	

Robust standard errors in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Omitted location category: Lucan.

Table 4.8: Results Locality Dummies

	(1)	
	lprice	
Kilmore	0.157***	(0.0331)
Kilgallaghan	0.583***	(0.0224)
Kimmage	0.559***	(0.0305)
Kinsaley	0.547***	(0.139)
Lusk	0.0287	(0.0300)
Malahide	0.342***	(0.0276)
Mansion House	0.765***	(0.108)
Merchants Quay	0.663***	(0.0327)
Mountjoy	0.223***	(0.0359)
North city	0.632***	(0.0851)
North Dock	0.228***	(0.0321)
Palmerstown	0.0940**	(0.0467)
Pembroke East	0.749***	(0.0286)
Pembroke West	0.857***	(0.0404)
Portmarnock	0.218***	(0.0322)
Priorswood	0.0786	(0.0614)
Raheny	0.276***	(0.0304)
Rathfarnam	0.447***	(0.0179)
Rathmines East	0.701***	(0.0260)
Rathmines West	0.695***	(0.0245)
Rotunda	0.358***	(0.0417)
Royal Exchange	0.493***	(0.0273)
Rush	-0.141***	(0.0451)
Shankill	0.165***	(0.0431)
Skerries	-0.127	(0.187)
South Dock	0.708***	(0.0313)
St Kevin's	0.741***	(0.0420)
Stillorgan	0.586***	(0.0319)
Sutton	0.294***	(0.0265)
Swords	0.140***	(0.0182)
Tallaght	0.139***	(0.0503)
Templeogue	0.466***	(0.0220)
Terenure	0.569***	(0.0214)
The Ward	0.00506	(0.0391)
Tibradden	0.625***	(0.127)
Turnapin	0.178***	(0.0322)
Ushers	0.503***	(0.0398)
Walkinstown	0.405***	(0.0443)
Whitehall	0.360***	(0.0203)
Wood Quay	0.750***	(0.0334)
Green and park dummies	Included	
Transport dummies	Included	
Beach-Coast dummies	Included	
Housing dummies	Included	
Quarter dummies	Included	
Monthly dummies	Included	
Constant	9.063***	(0.160)
Observations	6956	

Robust standard errors in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Omitted location category: Lucan.

Chapter 5

Lazy children or helicopter mothers? Evidence from Ireland on outdoor play, overweight in children and perceived neighbourhood characteristics

This chapter was presented at the Trinity College Dublin Development Working Group Seminar (2011). Special thanks is owed to the Growing Up in Ireland team at the ESRI, Prof. James Williams and the Office of the Minister for Children and Youth Affairs (OMCYA) for granting me access to the data. The views in this study are my own and in no way reflect the views or opinions of the OMCYA.

5.1 Introduction

The intense construction and soaring property prices witnessed in Ireland during the Celtic Tiger boom years resulted in competition for land use with land being prioritised for development, particularly in urban areas. Urban densities also increased and for the first time in Irish history, apartment living was not restricted to inner city lower income households. Green spaces were put under intense pressure due to competing demands for land use and combined with lax planning regulations, the provision of outdoor spaces for play and recreation for children and families were often neglected. This was particularly intensified for lower income households who due to rising housing costs could not afford the luxury of outdoor space. These changes may have also led to environmental inequity reaching the lower middle-income groups. Access to outdoor spaces became a luxury as opposed to a necessity. With the recession and the collapse in the construction industry, have come renewed calls for the better provision of public and outdoor spaces in Ireland by citizens, parents and architects.

Outdoor spaces, green spaces and playgrounds have been valued numerous times in the economics literature in the past for their amenity values. As shown in Chapter 4, they have been found to have an effect on house prices but also on neighbourhood safety as well as individual happiness. In recent years attention has been paid to the health benefits of outdoor spaces with evidence suggesting that access to outdoor spaces can have an effect on physical and mental health. Studies have found that even just the sight of green spaces can have an effect on mental wellbeing (Petty et al. 2005). As more and more families in Ireland and in developed countries start to live in urban centres the focus of such a literature has gradually been moving towards the health effects of outdoor spaces on *children*.

Ireland, as with most other western countries, has seen a growth in the number of overweight and obese children. Overweight and obesity has been identified by the World Health Organisation (WHO) as “one of the greatest public health challenges of the 21st century”. The WHO projections estimate that “one in three of the world’s adult population is overweight and almost one in ten is obese”.¹ WHO data indicate the estimated overweight and obesity prevalence of males in Ireland aged 15 and over in 2010 is 54% and for women is 44%.² A more worrying trend is the increase in the number of overweight and obese children. WHO research has indicated that children who are overweight or obese are at an increased risk of obesity in adulthood and will

¹<http://www.who.int/mediacentre/factsheets/fs311/en/index.html>.

²Source: WHO Global Infobase, <https://apps.who.int/infobase/?id=1>.

also be prone to the diseases associated with overweight and obesity such as cardiovascular disease, diabetes, and cancer. All these diseases require substantial health care and can result in premature death. Consequently the issue of overweight and obesity in children should be, and is, a main government policy issue as it will drive up health care costs in the long-run.

Despite these two issues being at the centre of research, the link between children's health and access to outdoor spaces and recreational facilities has been less closely examined. In particular, the role of parents (especially mothers) in this environment as well as *children's* views on available outdoor amenities has rarely been examined in this context. The availability of outdoor amenities does not necessarily imply their use and perceptions by both mothers and children of neighbourhood safety and play facilities can also affect their choices. Mothers who pay extremely close attention to their children have been described as "helicopter mothers" - over-cautious mothers who view their neighbourhood as too dangerous for outdoor play may not allow their children to venture too far from their sight. At the same time, the child's view of the neighbourhood could also impact on their willingness to venture outside, as will their alternative forms of recreation. Indeed, the availability of other forms of entertainment such as computer games and television are natural substitutes for children when it comes to play activities. Consequently, the determinants of children's outdoor play may be linked not only to their mothers' perceptions of their neighbourhood and their own perceptions of their neighbourhood, but also to the available supply of play spaces and other play alternatives.

This chapter looks at the link between overweight and obesity and the take up of play activities in outdoor spaces by children whilst taking account of maternal and child neighbourhood perceptions and the availability of alternate forms of play. The remainder of this chapter is set out as follows. Section 5.2 first presents some background literature on physical activity and green space availability and then reviews the literature on neighbourhood perceptions and health outcomes in Ireland. Section 5.3 presents the methodologies used in the analysis. Section 5.4 provides an overview of the dataset, the various questionnaires and some summary statistics while Section 5.5 presents the results. Finally, Section 5.6 provides a discussion and conclusions.

5.2 Literature

This section provides an overview of some of the literature relating to the present analysis. In a first part, recent papers investigating the links between physical activity and green space accessibility are presented followed by papers on neighbourhood perceptions. Finally, some of the most important research on obesity in Ireland is described in Section 5.2.3.

5.2.1 Physical activity, neighbourhood and green space availability

The literature on the links between access to green spaces and physical activity is broad and has been the subject of many studies. The following papers are studies that most closely relate to the present analysis or have similar methodologies.

De Vries et al. (2007) look at the links between a child's exercise and the characteristics of his environment in the Netherlands. The study uses a set of surveys conducted in six different cities on children between the ages of 6 and 11 who filled out an activity diary over the course of a week. The authors find that the physical activity of the surveyed children is positively correlated with activity friendly neighbourhoods as well as residential density and the availability of parallel parking spaces. Traffic and waste have a negative effect on physical activity. No correlations were found between the level of physical activity and the availability of recreation facilities, with the exception of sports fields. Care must be taken in interpreting their results as they may have suffered from self-selection bias. Indeed, only half of the children surveyed filled out the activity diary.

Huang et al. (2010) use data from a survey of children in Taiwan to identify the environmental variables that influence their physical activity levels. They focus on the differences between children living in rural and urban areas. The research question in this article is quite similar to the present study in that the authors look at how the perceived environments affect the level of physical activity. However the methodologies differ. First, this study has two sets of perceived environment factors (mothers' and children's) and Huang et al. only use one. Second, their activity level was measured as metabolic equivalents whereas this study looks at *use* of outdoor spaces. They find that the only significant factors determining physical activity are gender and the *accessibility* of the area. They do find a significant difference between urban and rural areas; however, taking account of parental income and education negates this difference. The final results indicate that accessibility is just as important as availability.

Mitchell and Popham (2007) use GIS data on green spaces, UK Census data on self-reported health and distinguish between rural urban and suburban areas. They use a simple linear regression model to test the effect of green spaces on poor health outcomes in adults. They find that a higher percentage of green space in an area results in better health outcomes. The positive relationship does not however exist in higher income rural areas and depends on the level of income deprivation in the area.

Black et al. (2010) look at how factors such as the availability of food and exercise facilities as well as average income levels affect obesity at the neighbourhood level in New York. They focus on the availability of different types of food outlets (supermarkets, restaurants, fast food outlets etc). They find that even after adjusting for individual level characteristics, the average risk of obesity differs between neighbourhoods. The presence of large supermarkets reduces obesity. They however do not take account of social and physical factors relating to the neighbourhood. They conclude that neighbourhood facilities and services are also determinants of obesity and should be taken into account when formulating policy in addition to individual factors.

Coombes et al. (2010) examine how access to green spaces and their use affect obesity in Bristol. They use GIS data on green space characteristics. Specifically, they look at the types of green spaces accessible to respondents and whether the parks are formal, informal, natural, for sports or geared towards young people. They also include measures of road density, land use and census measures of the neighbourhood. They conduct separate logistic regressions with three dependent variables: visiting a green space, engaging in physical activity and being overweight. They find a negative correlation between the distance to the nearest green space and the level of physical activity. Those living further away are also more likely to be obese. Their results suggest that accessibility may be related to use and consequently could be used as a tool to reduce obesity levels.

Finally, Li et al. (2005) concentrate on a different population group - older adults - and one type of physical activity - walking. They use a multi-level model and GIS measures of the environment in Portland in 2001 and assess their effects on neighbourhood walking. They evaluate between and within neighbourhood variability in walking by using the multi-level model. They find variation in walking due to differences between neighbourhoods of 28%. Walking is positively influenced by the availability of recreational facilities and safety. They conclude that the neighbourhood effect is important, as is safety. The provision of facilities and environmental influences can be a deciding

factor in older people's physical activity.

The links between physical activity and access to outdoor spaces have been shown to be important. Previous research has concentrated on the differences between urban and rural areas as well as physical activity in adults. All have found that accessibility affects use and is not the same as availability. However, the effect of neighbourhood *perceptions* is also important and has been the subject of separate research.

5.2.2 Neighbourhood perceptions

The effect of a person's subjective perception of a neighbourhood on his or her interaction with it has become a topical subject of research in the past 3 or 4 years. This is linked to a renewed focus on management of urban centres at a smaller, neighbourhood scale. The next two papers incorporate the neighbourhood perceptions of parents into their analysis.

The analysis by Bringolf-Isler et al. (2010) looks at parental perceptions of the environment on children's outdoor play. The authors combine GIS data with survey data on primary school children and adolescent health conducted in three cities in Switzerland. They also have access to questionnaires on parents' subjective views on the environment. They find that the perceived danger of traffic by parents results in primary school children spending less time playing outdoors. On the other hand, boys, foreign nationals and the presence of younger siblings in the family increased outdoor play times.

Elo et al. (2009) show that scales relating to perceptions of crime and safety are consistent and reliable within neighbourhoods. The authors conclude that "when data on objective neighbourhood characteristics are unavailable, the inclusion of questions about residents' perceptions of neighbourhood conditions in surveys of inner-city residents provides a useful alternative to characterize neighbourhood conditions" (2009:1299). Previous studies had found that interviewees often use different standards when making an evaluation of their neighbourhood. In this study the authors focus on low-income households living in the inner city of Philadelphia. They concentrate on three types of neighbourhood issues, "crime and safety, physical disorder and social disorder". They find that older women, educated women, foreign born women and less disadvantaged women are less likely to be concerned about neighbourhood problems. They conclude that despite some differences, perceptions have good neighbourhood-level reliabilities. These types of neighbourhood level percep-

tions are used in the analysis in this chapter. Finally we look at some relevant literature on obesity in Ireland.

5.2.3 Obesity in Ireland

Lissau et al. (2004) look at overweight in adolescents (13 to 15 years of age) in 15 countries using cross sections of school surveys in 1997 and 1998. All the surveys used were conducted by the WHO and consequently their methods and protocols are similar and result in the surveys being comparable. Ireland (826 observations) was included in the European sample but 39% of BMI observations were missing. The authors state that this was mostly due to the recorded weights and heights being in metric measurements and students leaving the answer blank. They use the data to look at 85th and 95th centiles and find that “Ireland, Finland and Greece had significantly increased prevalence [of obesity] at or above the 85th centile (for both sexes)” compared to the study reference standard. The study reference standard in this analysis was a reference BMI curve constructed from all the data but “smoothed across month of age within year, resulting in references appropriate for each sex by month of age to use in comparisons.” They acknowledge there are issues of self-reporting by teenagers but state that as most “obese adolescents remain obese as adults; this age group is a very important group to reach through preventative programs”.

Not much work has been done in relating the physical environment to health outcomes in an Irish context. The main paper of note is very recent and looks at access to food outlets and its impact on health and diet in adults. Layte et al. (2011) find that socio-economic status at an individual and household level affects an individual’s diet quality but that physical distance to the nearest supermarket is also significant. The variable construction used in their analysis is similar to the methodology employed in this paper and is discussed further in Section 5.4.2.

5.3 Model

This section presents the different methods used to analyse the data in this paper. The first technique used is principle component analysis and the second is a form of multi-level analysis.

5.3.1 Principle Component Analysis

With surveys as extensive as the one used in this paper, some of the difficulty can lie in identifying which variables are relevant and which sets of variables are actually measuring a similar concept. Principle component analysis (PCA) can be used in this case to identify groups of variables which are highly correlated and to reduce multiple measures into one or a few principle components. This allows the number of closely related variables to be reduced and a "summary" index to be produced. PCA is similar to factor analysis in that both look for components.³

Vyas and Kumaranayake (2006) describe PCA as a "multivariate technique used to reduce the number of variables in a dataset into a smaller number of dimensions. PCA creates uncorrelated indices or components where a component is a linear weighted combination of the initial variables. The weights for each principal component are given by the eigenvectors of the correlation matrix." (2006:460). The problem when dealing with a multi-dimensional concept is that indicators for the dimensions will tend to be strongly correlated. A summary indicator or a few summary indicators are required.

A first step with PCA is to calculate Cronbach's Alpha in order to determine which variables to group. Only those closely related will be grouped (and the data reduced) and Cronbach's Alpha is consequently a measure of internal consistency. A high value of Alpha is taken as an indication that the different variables are measuring a similar underlying concept. Cronbach's Alpha is measured as follows:

$$\alpha = \frac{N \cdot \bar{c}}{\bar{v} + (N - 1) \cdot \bar{c}} \quad (5.1)$$

where N is equal to the number of variables, $\bar{c}$ is the average inter-variable covariance among the variables and $\bar{v}$ equals the average variance (Nunnally and Bernstein (1994)). The higher the coefficient the higher the level of internal consistency and it should be over 0.75 to expect reasonable internal consistency.

Once the grouped statements have been identified it is necessary to check the dimensionality of the grouping. It is still possible that the variables that have been grouped are multi-dimensional. Picking the correct number of components requires an additional test. A screeplot can be used to determine the number of dimensions

³With PCA the original matrix analyses the total variance whereas factor analysis looks at the common variance.

and how many variables should be constructed. A screeplot graphs the eigenvalue against the number of components. The final number of components is determined by the number of components where the eigenvalues are larger than one. This is also confirmed by the number of components in the screeplot before the point where the graph forms a “kink”. As the line gets flatter each following component is accounting for smaller and smaller amounts of the total variance. If a screeplot kinks at component number 3, then two components account for a disproportionately large amount of the combined variance.

Once the number of components is chosen the principal component (PC) scores can be created. Each PC is a weighted average of the underlying indicators. The new Principal Component can be used as a variable in a multiple regression analysis. It is then important that the effects of the component are interpretable and consequently the underlying variables should be set up in such a way that they can then be interpreted as one in the final regression. The following section presents the regression models used in the analysis.

5.3.2 Statistical analysis

Various models are used in this analysis. In a first stage, the determinants of outdoor play and use of parks are investigated. The model in this instance is a logit model of the form:

$$Pr(Y) = \alpha + \beta\chi + \epsilon \quad (5.2)$$

where Y is a binary variable representing whether or not the child plays outdoors, goes to parks or lists his or her favorite activity as playing outside. This is determined by a vector of predictor variables χ which relate to the mothers’ views of the neighbourhood, the child’s views of the neighbourhood, the available facilities, characteristics of the child and socio-economic variables. ϵ is the residual.

Principal component analysis is applied to the mother’s views of neighbourhood safety and the resulting Principal Components are included in the analysis. The model is then of the form:

$$Pr(Y) = \alpha + \delta PC + \beta\chi + \epsilon \quad (5.3)$$

The next stage of the analysis relates to health outcomes. First a simple logit model showing the effect of genetic factors on different measures of BMI is estimated. This is then extended by adding in lifestyle factors as well as exercise, behavioural and lo-

cational dimensions. Finally, the two analyses are linked through a two-stage analysis where the predicted values from Equation 5.3 are included in the health related regressions.

Another dimension of the analysis must be discussed at this point. It is likely that a child's BMI will be correlated with the BMIs of other children and adults in the school or electoral division (ED) they are located in due to unobserved area-specific influences. In the case of this survey, the effect is more likely to relate to the school effect as sampling for the survey was done at this level. In a case such as this it is common to use multi-level analysis given that measurements taken within schools are likely to be correlated. This "area" effect could affect the results and so to control for it a multi-level analysis is conducted where variables are clustered to take into account within-school correlations.

When using health related variables such as BMI this *within-area effect* can be controlled for using a different approach. This approach is applied to regressions relating to both school and ED level effects. As in Layte et al. (2011) the mean deviation of the child's BMI from the mean BMI in the school (or ED) is regressed on the mean deviation of the explanatory variables. This estimation is shown in equation 5.4 below.

$$Pr(\widehat{BMI}) = Pr(BMI_{ij} - \widehat{BMI}_j) = \alpha + \beta(\chi_{ij} - \widehat{\chi}_j) + \epsilon_{ij} \quad (5.4)$$

where BMI_{ij} is the BMI of person i in school or ED j , $\widehat{BMI}_j$ is the mean BMI within that school or ED, χ_{ij} is a vector of explanatory variables and $\widehat{\chi}_j$ is the mean value of the explanatory variables. Finally, ϵ_{ij} is the residual. This methodology differences out effects within the school or ED. The measured impact of the explanatory variables on BMI then relates to within cluster effects.

5.4 Data

The following section presents the data used in this paper. It first gives a general description of the survey and questionnaires used. It also presents any restrictions imposed on the sample. It then describes the variables of interest in turn and provides some summary statistics and descriptive statistics for the final sample.

5.4.1 Survey and sample restrictions

The dataset used in this paper is the National Longitudinal Study of Children usually referred to as the “Growing up in Ireland” (GUI) survey. GUI is a national government funded study of children. Its stated purpose is to understand “all aspects of children and their development” (GUI, 2010). The study is funded through the Office of the Minister for Children and Youth Affairs in association with the Department of Social and Family Affairs and the Central Statistics Office. The study was carried out by researchers led by the Economic and Social Research Institute (ESRI) and Trinity College Dublin (TCD). The specific section of the survey used in this paper is the 9-year old cohort of the GUI. The survey is nationally representative and includes 8,570 nine-year olds, their families, teachers, principals and regular caregivers. Its stated aim is to “identify the key factors that most help or hinder children’s development” and is “particularly focused on charting the development of children over time and examining their progress and wellbeing at critical periods from birth to adulthood” (GUI, 2010).

The first wave of the survey was conducted in 2008 and a total of 15 questionnaires were administered regarding each child. These are listed in the Appendix and consist of a range of questionnaires ranging from the child, to the caregivers to the child’s teacher and headmaster. The scope of the questionnaires ranges from education, health history, socio-economics, parenting style, likes and dislikes as well as physical characteristics. The complex system of questionnaires is summarised in Figure 5.1 and in the Appendix.

Certain restrictions were applied to the data for the purpose of this analysis. In order to take account of possible differences in parenting style, only children whose mother was the main caregiver were kept for the analysis. Households where the father was listed as the main care-giver were dropped.⁴ Any children whose parent stated that a chronic illness hampered their daily activities, or had a physical or learning disability were also dropped from the sample as this might affect their likelihood of playing outside regardless of neighbourhood characteristics. Finally, any observations which reported missing values for the main variables used in the analysis were also dropped. The final sample was 6,054 children consisting of 52% girls. The following

⁴It is standard in the perceptions literature to use mothers’ perceptions only (as they tend to be the ones to respond to surveys about children) in order to compare like with like (see Kimbro et al (2010), Strife and Downey (2009), Elo et al. (2009)). Only 90 of over 6,000 households had a father as a main caregiver, so this was not a large restriction on the sample. However, the analysis was run with the fathers in the sample to check for any differences but the results remained the same.

sections take a closer look at the specific sections of the questionnaires of interest for this analysis, describe how additional variables were constructed and provide some summary statistics.

5.4.2 Summary statistics

Neighbourhood perceptions

One of the primary interests of this study is the effect of perceptions on the use of neighbourhood facilities. One of the original features of this paper is that perceptions of the neighbourhood were collected both from the parent and from the child. The mother was asked to state how common she believed different neighbourhood characteristics were and gave one of four answers: very common, fairly common, not very common and not at all common. These questions regarded the physical appearance of the neighbourhood (the presence of rubbish and litter and the condition of homes and gardens) as well as unsocial activity (vandalism and public drug-taking or drunkenness). She was also asked to agree or disagree about statements regarding safety in the neighbourhood, in particular whether it was safe to walk alone after dark or safe for children to play outside. These variables also had four possible answers: strongly agree, agree, disagree and strongly disagree. Summary statistics for these variables are listed in Table 5.1. Mothers' neighbourhood perceptions are varied. The presence of litter and rubbish is the most common neighbourhood problem quoted by mothers, with 31% saying it was fairly or very common. The least common problem was that homes and gardens were in a bad condition (7%). Other issues such as public drunkenness and vandalism were also fairly common (12%). Mothers were also asked about facilities for play in the area. 59% said there were safe parks, playgrounds and play spaces in the area, with 60% saying they thought there were recreational facilities appropriate for a 9 year old. A very high proportion, 92%, agreed or strongly agreed that it was safe for children to play outside during the day.

Child perceptions of their neighbourhood environment were also collected, which is unique in this type of study. The answers to these questions were binary yes/no responses instead of a scale of agreement. Each child was asked about play facilities, for instance if there were good places to play near their house (84% responded yes), if there were green areas for them to play near their house (80% responded yes) and if there were playgrounds near where they lived (55% responded no). They were also asked questions about their perception of the neighbourhood. 22% said there were no places for children to play safely around their house but 96% said they felt safe living

in the area and a very high proportion also said they had friends in the area to play with. They were also asked about the physical appearance of their neighbourhood (e.g. are the streets dirty, is there a lot of graffiti), and questions about traffic in the neighbourhood, which is a factor that has been previously found to affect the amount of outdoor play by children. Approximately a fifth of all surveyed children found that there was too much traffic in the area, that the streets were dirty and that there was graffiti near where they live. The summary statistics for these variables are listed in Table 5.2. Next we look at the health related variables used in the analysis.

Health variables

Data on the survey child's height in centimetres and weight in kilograms were collected by the survey interviewer and these variables were used to construct the health variables used in the analysis. The first of these is the child's Body Mass Index (BMI) which is calculated as follows:

$$BMI = \frac{mass(kg)}{(height(m))^2} \tag{5.5}$$

The mean weight, height and BMI of children in the sample are detailed in Table 5.3 and pictured in Figure 5.2. It is clear that boys are on average slightly taller and lighter and that the range in weights for the girls is wider. The average BMI for girls is 17.93 whereas the average BMI for boys is 17.67. This difference is also highlighted in Figure 5.2. The histogram for girls is a lot more spread out than the one for boys.

There are however certain issues with using BMI as a variable in an analysis such as this. Indeed, it is computed in the same fashion regardless of age or gender. Cole et al. (2000) state that no common standard regarding the definition and monitoring of child obesity exists internationally and that BMI measures for children should be gender and age specific. They use over 10,000 observations on the BMI of children from 6 different countries to create cut-off points for BMI for obesity and overweight in children. This is done by creating centile curves defined to pass through specified BMI cut-off points at the age of 18. They can then extrapolate back through childhood and teenage years (see Figure 5.3 in the Appendix). These cut-off points will be used in this analysis and are summarised in Table 5.4. Using these cut-offs, binary variables indicating if a child is overweight or not and obese or not were created. The summary statistics for these variables are listed in Tables 5.5 and 5.6. Table 5.6 indicates that 5.42% of survey children are considered obese under the Cole measure and 18.27% are considered overweight. However, what the table highlights is the big difference in

obesity and overweight prevalence between boys and girls. Indeed, only 16% of boys are overweight compared to nearly 21% of girls. A larger proportion of girls are also obese compared to boys (4.7% for boys and 6.1% for girls). Table 5.5 uses a different scale to measure overweight and obesity and is based on comparisons between individual children and their peers. The 85th percentile and 95th percentile of weight are listed by gender. The top 5% of girls have a BMI between 23.59 and 24.32 while the top 5% of boys have a BMI between 22.83 and 23.72. According to the Cole et al. study, the obesity cut-off for boys is 23.39 and the cut-off for girls is 23.46. In the study sample, the top 5% of girls have a BMI between 23.59 and 24.32 while the top 5% of boys have a BMI between 22.83 and 23.72. These figures indicate that only part of the boys in the top 5% have a weight above the cut-off point whilst all girls in the top 5% are considered obese.

Health variables are included in the analysis in three ways. First, BMI is included as a linear variable. It is also included as an ordered variable equal to zero if the child is of a normal weight, 1 if the child is overweight and 2 if the child is obese. Finally, the BMI variable is also included as a difference from the area mean, as detailed in Section 5.3.2. Parental health is also included to take into consideration any genetic factors. Indicators of parental health are also included in the analysis to take into account any household related lifestyle impacts that may account for the child's weight. As shown in Table 5.7 a high proportion of children's parents are considered overweight and obese. This is consistent with current health data in Ireland which indicate that a growing number of adults in the Republic are considered obese by medical standards. The next section presents the physical activity variables of interest in this study.

Child's activity

Both child and parent were asked a number of questions regarding the child's outdoor play activities and other favorite activities. The child's mother was asked if the child played games outdoors in the last week or went to the park in the last week. Both of these variables are used as the dependent variables in the analysis. The child was also asked if his or her favorite activity was to play games outside. This was also included as a dependent variable in one stage of the analysis. Other activity related variables were included as explanatory variables in the analysis. Dummies for if the child's stated favorite activity is watching TV or playing computer games were included in the regressions. The presence of a computer and TV in the child's bedroom were also included as potential "inactivity" indicators and if sport was their favorite activity this

was included as an indication of “sporty” children. The amount of times a child exercised a week was also taken into account. The statistics relating to these variables are presented in Table 5.8.

The data show that half of the children played games outside in the last week but just a fifth went to the park. The presence of a computer or a television in their bedroom are included in the analysis as potential “inactivity” explanatory variables. Nearly a third of children list watching television as their favorite activity while 40% have a TV in their bedroom. 20% of children are described by their teacher as being categorised as a “quiet child”. Half of the children exercise every day for 20 minutes but a quarter do so for 60 minutes. A very low percentage, 1%, do no exercise at all.

Socio-economic household variables

Finally, in order to take account of possible differences relating to households specifically, a range of household level variables are also included in the analysis. The type of area the household lives in is taken into account using a NUTS8⁵ region code. Despite more detailed variables being included in the questionnaire (notably the exact GIS co-ordinates of the household’s home) for security reasons they have not been included in the datafile made available to researchers. The present analysis would obviously be enhanced in the future with more accurate locational variables. A quarter of households surveyed live in the Dublin area with the rest fairly equally spread across the country (see Table 5.9).

The socio-economic characteristics of the household were also collected. Those that were found to be significant and were included in the regressions were the age of the main care-giver, marital status, work situation, annual household income and whether or not the child was born in Ireland (see Table 5.8). Other variables such as parental education, the existence and nature of childcare, paternal age, occupancy type and accommodation type were also tested but not found to be significant. 41% of mothers are stay at home mothers, 84% are married and living with their spouse while less than 2% are divorced. The next section presents the various results of the analysis.

⁵Nomenclature of Territorial Units for Statistics.

5.5 Results

This section presents the results of the various stages of analysis. It starts with the analysis of the determinants of outdoor play then details the construction and results of the principle component analysis. Section 5.5.3 presents some basic regressions on BMI and genetics and then the results of the two-stage regression are outlined.

5.5.1 The determinants of play

In a first stage, the determinants of outdoor play in children are investigated. Three separate analyses were conducted using three different dependent variables. All are binary variables indicating whether or not the child has played outdoors, gone to the park or lists playing games outside as their top two favorite activities. The probability of a child playing outdoors is then hypothesised to be linked to maternal perceptions of the neighbourhood quality, child perceptions of neighbourhood quality, availability of facilities, location and socio-economic characteristics. The results are presented in Table 5.10. The first clear result is that nearly none of the explanatory variables have an effect on the probability of a child stating “playing outside” as his or her top two favorite activities. This is likely to relate to the large range of activity choices available to the child and the fact that stating a favorite activity is very subjective. There are however some interesting results from the playing outside and going to the park regressions.

Compared to children living in open country locations, those in cities or on the periphery of cities are much less likely to play outside. On the other hand, the location variables are not significant for the park location, indicating that the use of parks is not geographically related. If a park exists in the vicinity of a child’s home they are as likely to use it in Dublin as they are in Kerry. However, outdoor play is location specific and this is a first indication that neighbourhood characteristics matter. Children whose mother’s work and are older are also less likely to play outside or go to the park whereas those from households with higher incomes, boys, and Irish born children are more likely to play outside, in line with some of the results of the Bringolf-Isler et al. (2010) study. These results are a first indication that children living in nicer (possible wealthier) areas are more likely to play outside as are children with younger mothers.

Only three of the children’s perceptions of the area are significant. Children who have friends in the area and where there are “good places to play” will be more likely

to do so. The availability of green spaces in the area has the reverse effect, probably due to it being a suitable alternative. The availability of playgrounds is important for those going to the park, who tend to not be Irish born. An interesting result is the significant negative coefficient on the “quiet child” variable indicating that the child’s personality type as well as environment can have an association with physical play activities. Interestingly the variables relating to mothers’ perception are all counter intuitive. The more litter and rubbish the more likely children are to be let play outside. The safer it is perceived to be the less likely they are let out. The less play spaces, the more likely they will be let out. The reason behind these counter intuitive results could relate to the fact that the variables of mothers’ perceptions are interlinked. This will be addressed in the following section with the principle component analysis. But first the effect of other forms of activity on outdoor play are investigated.

Table 5.11 presents more extended results including alternative activities. The results for quiet children, mother’s age and working moms are still the same. Location variables are still included but not shown. Children who state that playing computer games or watching TV are their favorite activities are less likely to partake in outdoor play.⁶ This is also true of those who exercise less than 4 times a week. There is a downward gradient on the exercise coefficients. The less exercise children do the less likely they are to have active playtime. This confirms the “quiet child” result: regardless of available facilities, active children will be more likely to have active playtime too. The same effect applies to sport. The use of a computer is associated with a lower likelihood of a child playing in a park.

Finally, area characteristics were included, see Table 5.12. Small Area Population Statistics (SAPS) are electoral division statistics on the population make-up of an area. Various SAPS were included but none were found to be significant apart from the proportion of retired households in an area. This coefficient is positive and significant: i.e. the higher the proportion of retired household in an area, the higher the likelihood of a child having outdoor play. This is likely related to the perceived quality and safety of the neighbourhood by mothers, which is at the centre of the principle component analysis in the following section.

⁶There is however scope for reverse causality here as those who partake in outdoor play activities are also less likely to favour TV or computer games.

5.5.2 Neighbourhood perception index

Number of principle components

Mothers in the survey were asked to give their impressions of various neighbourhood characteristics on a four point scale ranging from very common to not at all common. All of the statements, though different, relate to similar concepts: the neighbourhood's visual aspect, its safety etc. Including all of these variables simultaneously in the analysis could lead to multicollinearity problems. Consequently, the variables can be combined to create an index of visual amenities or safety amenities. This is done using principle component analysis as described in Section 5.3.1. However a first step that is required to determine which groups of variables should be combined to create a "neighbourhood perception" index. This is done by calculating Cronbach's Alpha. Eight variables were first included (litter, housing/gardens in bad condition, vandalism, public drunkenness, safety, safe play spaces, safe parks and playgrounds, appropriate recreational facilities). The alpha test scale when all eight variables were included was 0.7012. This was not high enough to create a principle component.⁷ Cronbach's alpha on four variables (litter, housing/gardens in bad condition, vandalism and public drunkenness) however was 0.82, which was suitable for grouping, see Table 5.18 in the Appendix. In order to determine the number of components that needed to be created, a screeplot was produced, see Figure 5.4. The "kink" in the graph clearly occurs before the second component and only one of the components has an eigenvalue larger than one indicating that one component can be created. The newly created PC index is essentially a visual quality of the neighbourhood index. The higher the index, the nicer the perceived quality of the neighbourhood.

Cronbach's alpha was also calculated using the various questions on neighbourhood quality from the child's perspective but the highest score achieved was 0.4 so no underlying index was supported. The next section presents the results of the outdoor play regressions including the principle component.

Principle component analysis - Play and neighbourhood perceptions

The same regressions were run but the mother's perception of the visual quality of the neighbourhood were replaced by the index. These are presented in Table 5.13. The PC variable is positive and significant at the 1% level. The more mothers view the neigh-

⁷ A cut-off point of 0.7 is sometimes used in the social sciences. On this basis a principle component was created using all eight variables and included in the regressions. The results are broadly similar to the analysis with a stricter alpha cut-off and so are not reported here.

bourhood positively the more likely they are to let children play outside. The use of the principle component has solved the previous multicollinearity problem. The inclusion of the PC variable has also resulted in the other neighbourhood variable results not being in line with expectations. Mothers who are of the view that the area is lacking in play spaces are less likely to let their children play outside. The visual aspect of the neighbourhood is not important for mothers letting their children play in parks; instead, safety after dark and the existence of play spaces is valued. On the other hand, children's perceptions of neighbourhood quality are still mostly insignificant. There are two exceptions. The existence of green outdoor play spaces (i.e. parks) reduces the probability of a child playing outside. This may seem counterintuitive but actually is an indication that children will choose park play over outdoor play if they have the option. Second, for those using parks as a play space, the most important factor for the child is the existence of playgrounds in the park. Children do not seem to value the presence of physical play facilities unless they are going to a park. In this case safety is also a concern to them (significant at the 10% level). The other explanatory variables remain largely the same as when discussed in Section 5.5.1: the child's personality still matters as do alternative activities.

What this principle component analysis has highlighted is that there is a distinction in outdoor play activities. When dealing with a specific location such as a park, facilities matter to mothers and children. When dealing with the area outside their dwelling, neighbourhood quality matters to mothers and a child's personality type is more likely to affect a child's probability of playing outside. Just as the results in Chapter 4 indicated, when dealing with the very close proximity of a house, visual impacts tend to dominate, whereas the existence of facilities tends to be a stronger factor for activities within walking distance of a dwelling. The next section turns to the health aspects of the analysis.

5.5.3 Children's health

The next stage of the analysis looks at health variables.⁸ For each stage of the analysis four types of BMI regression are used: an ordered logit (coded 0 when the child's BMI

⁸The main BMI regressions in this analysis do not include dietary factors as diet is not the main objective of this analysis and has been extensively examined elsewhere. However, the types of food eaten by the children were included in the regressions to test if they had an effect on other explanatory variables. The results did not change. For interest one such regression is included in Table 5.25 in the Appendix. Whether or not the child sits down to eat dinner does not have an effect on BMI. One interesting point to note is that children who consume diet drinks and skimmed milk are more likely to be overweight and obese. This is likely to be a causation effect; the children with weight problems are consuming diet products to aid their weight.

is within the average range, 1 if the child is considered overweight, and 2 if obese), a linear BMI OLS regression and two OLS regressions using area differenced BMI to take account of within area effects at a school and ED level. Location dummies are included at each stage for the first two regression types but not for the differenced regressions. First the effect of genetics related variables and parental lifestyle variables on the child's BMI are presented. Then, additional activity variables are introduced and in a final stage the health related analysis is linked to the outdoor play analysis of Section 5.5.2.

Health and genetics

In a first stage, the analysis includes only parental weight characteristics and some parental lifestyle variables. These results are presented in Table 5.14. The household's location does not seem to have a particular effect on a child's BMI and practically none of the location explanatory variables are significant. However, as expected all parental BMI variables are significant across all regressions. Mothers' weight has the most significant impact on child overweight. Lifestyle factors such as smoking and drinking during pregnancy also have a significant association with the child's BMI and are likely to be a proxy for socio-economic indicators. Women who breastfed are less likely to have overweight and obese children. Boys are less likely to be overweight, confirming the sample statistics presented in Section 5.4.2. The household's location within Ireland does not seem to matter much when looking at the impact on overweight and obesity. However, when examining overweight at a more micro level, i.e. when school level and ED level effects are taken into account the coefficients are overall smaller than in the BMI OLS regression. This is a first indication that school and ED effects do have an impact on childhood BMI. Indeed, as discussed in Section 5.3.2, it is possible that children within the same school or local area will end up having similar health issues (or lack thereof). This can be related to having a similar diet, other area-specific influences or peer effects in schools. Next activity and behavioural variables are added to the analysis.

Children's BMI

Table 5.15 presents the more extensive BMI regressions. As before, genetic, parental lifestyle and location variables are included but not shown. There do not seem to be any direct income effects on weight (the top and bottom income quintiles are not sig-

nificant)⁹ but some wealth related variables are significant. Children living in areas where a high proportion of households own their own home are less likely to be overweight and obese as are children whose household dwelling is a house. The children of working mothers and unmarried mothers were also more likely to be overweight. For instance, children whose mother is married have a BMI 0.546 points lower than children whose mother is unmarried, all other variables equal. A mother working increases the BMI index by 0.203. Activity variables also have an impact on the probability of a child being overweight or obese. The main significant variable across all regressions is whether or not the child has a television in his or her room. Children with televisions in their room are much more likely to be overweight or obese than those who don't. This effect is however not significant for the presence of a computer in their room. As expected children who exercise more are less likely to have health problems. Finally, children reported by their mothers to have played outside are significantly less likely to have weight problems. This is an indication of a link between outdoor play and health. However, it is possible that the causal effect runs in the opposite direction: children with weight problems are less likely to play outside. To take this into account a two-stage model is presented in the next section.

Play and health

The variable indicating outdoor play was included in the BMI regressions directly and showed a negative and significant effect; children who play outdoors have lower BMIs. However, this causality may also run in the opposite direction and in order to confirm the effect of outdoor play on health outcomes, a two-stage model where the probability of playing outdoors is instrumented is required. The Durbin score and Wu Hausman F tests for all regressions indicates that the "outdoor play" variable is endogenous and the use of a two-stage least squares (2SLS) model over an OLS model is justified. The instrument used in the 2SLS is a behavioural indicator variable. The surveyed children were evaluated and a number of behavioural traits were used to compile scores relating to their behavioural types. One of these variables indicates whether the child tends to be shy or quiet and this is used as the instrumental variable in the analysis. Children who are shy may be less likely to participate in outdoor games but there should not be a direct causal link between shy children and health. The validity of the instrument is confirmed for all regressions (the Cragg-Donald Wald F statistic for weak identification is over 16 in all cases). The results of the two-stage regression are presented in Table

⁹Other variants on income were included in the analysis, such a logged income, income deciles, and all income but were not found to be significant. Consequently, these results are not reported here.

5.16. The coefficient on the “playing outdoors” variable is negative and significant at the 1% level. In all three models (BMI, ED-differenced BMI and school-differenced BMI) children who play outdoor have lower BMIs. Conditional on playing outside, a child’s BMI is on average 4 to 5.5 points lower. These results indicate that there is a direct causal effect of outdoor play on BMI.

Sub-sample checks

Additional regressions were run to test for the effects of gender and location on play and health. All play regressions were split into rural and urban samples to check if the determinants of outdoor play depended on a rural/urban location divide. These are listed in the Appendix, Table 5.21. The results across urban and rural households are broadly similar with just a few interesting differences to note. Urban mothers seem to attach more importance to issues such as safety and the availability of play spaces than rural mothers. Interestingly, so do the urban children. Indeed, in the results for the full sample, the child’s perception of neighbourhood safety was not significant, however, for the urban sub-sample it is highly significant. Mother’s age and if the child is Irish born are also significant in the urban regression and not the rural one. Evidently, outdoor play and safety are more important issues in an urban setting than a rural one. On the other hand, rural mothers are more conscious of safety with regards to play in parks as are rural children. Neighbourhood safety is a more important issue in an urban setting whereas safety of park facilities is more important for rural households.

The main sample was also split according to gender on the basis that outdoor play determinants may differ depending on whether the child is a boy or a girl. These are available in Tables 5.22 and 5.23 in the Appendix. The determinants of outdoor play differ in some instances if the child is a boy or a girl. Boys are more interested in the existence of “good play spaces” whereas girls are more likely to play outside if there are playgrounds. Girls also attach more importance to having friends in the neighbourhood with which to play with. The effect of playing sports is stronger for boys. Boys who state that playing computers is their favorite activity are less likely to play outside whereas girls are more influenced by watching TV.

Endogeneity issues

There are a number of potential endogeneity issues in this analysis. First, children who have health problems are less likely to play outside, so there may be reverse causation. This endogeneity issue is dealt with by using the two stage least squares analysis and the results appear to be robust. Second, it is also possible that parents who are healthy,

“breed” healthy children. Parental lifestyle and health indicators are included in the regressions to control for this potential issue and are all very significant. Moreover, parents who want their children to be able to play outside might move to areas that will allow them to do so. This issue is more difficult to address given the available data. However, there are some reasons to think it is not a major problem in these models. First, the scope for households to exhibit this effect is limited by income (every parent would like to pick the nicest possible area for their children to live in). Variations of income variables were included in the analysis but these were not significant. Second, if it is the case that parents who want their children to play outside pick “outdoor play” friendly areas then it is likely that the “availability of facilities” variables would be significant, and this is not the case. Regardless of income, location or facilities, the perceptions of the neighbourhood affect parents choices of whether to let their children play outdoors. If there were a significant endogeneity issue, those three aspects would be expected to be significant. As a consequence it is necessary to apply a caveat to the parks results, for which facilities do matter. It is possible that in relation to parks, parents make an explicit neighbourhood choice decision.

Finally, another possible issue with the analysis is the possibility of reverse causation in the perceptions and play regressions. It is possible that mothers who let their children play outside may in so doing change their perception of the neighbourhood. This is difficult to correct with the current dataset. However, it may be feasible to explore this issue using future waves of the GUI dataset. Indeed, once more data is available it may be possible to specifically examine mothers who previously did not let their children play outside and now do. However, this is not currently feasible.

5.6 Discussion and Conclusions

This chapter examines the determinants of outdoor play for children in Ireland. Two areas of interest were examined in this case: the effect of parental neighbourhood perceptions and the effects of the child’s neighbourhood perceptions on their likelihood of engaging in outdoor play. The results show that when all maternal perceptions are included in the analysis, no clear effect can be discerned. However, when they are grouped into a “visual perception index” using principle component analysis, the influence of maternal perceptions becomes clearer. As expected, mothers place importance on the visual aspect of the neighbourhood as well as its safety when it comes to allowing outdoor play for their children. “Helicopter mothers” are less likely to let their children play outdoors. Interestingly, the existence of facilities is only of importance when relating to a specific activity such as going to the park. Similarl to the conclusions of Chapter 4, it seems that visual and safety aspects dominate play close to

the home whereas facilities and the existence of playgrounds dominate play in parks.

The unique dataset allowed inclusion of the children's perceptions in the analysis. This allowed the comparison of neighbourhood perceptions between children and mothers. Visual aspects of the neighbourhood do not matter as much to children, who place more importance on the existence of good play spaces and friends to play with. Similar to their mothers, they are more likely to play in a park if there are specific facilities there (such as playgrounds). These results have important policy implications. Indeed, getting children to play outdoors and consequently promoting physical activity through play has been the focus of policy from a childhood obesity perspective. Children are becoming more sedentary and are watching television more and exercising less. However, what the results indicate are that the uptake of outdoor play has a range of influencing factors. General neighbourhood quality and safety seems more likely to have an effect than building specific facilities. Interestingly, in this instance, income does not seem to matter but there are some gender effects, with boys being more likely to participate in outdoor activities. It is worth noting however that the possibility of reverse causation with regards to the link between neighbourhood perceptions and outdoor play has not been excluded.

The next stage of the analysis looked at the determinants of children's overweight and obesity. Childhood obesity is a growing problem in Ireland as in other western countries and policy is now focused on ways of tackling it. Variables affecting children's BMIs were examined. As expected, genetic factors such as whether the child's mother or father has weight problems had a very big influence on the child's weight. However, other lifestyle factors, such as having a television in the bedroom or playing sports were also significant. A partial link to income inequality feeds through the health aspect of the analysis: socio-economic characteristics also matter as do wealth factors (a neighbourhood effect is evident). Children who played outside were also less likely to have weight problems. Results from modelling with an instrumental variable suggests that this association is causal. This causation is also important from a policy perspective.

Indeed, Ireland is for the first time facing a childhood obesity problem and policy and planning have focused up to now on promoting healthier lifestyles and providing more facilities. Whilst promoting better diets is undeniably important, in urban centres purely providing play facilities will not in itself result in more outdoor play by children and more emphasis on neighbourhood quality could have a more far reaching effect. What this analysis has shown is that it is not purely the existence of facilities that

results in their uptake, less obvious factors such as maternal confidence in neighbourhood safety as well as neighbourhood quality also matter. The positive links between children's outdoor play and health have been shown in this study and the promotion of outdoor play as a health benefit could be beneficial going forward.

5.7 Tables

Table 5.1: Descriptive statistics for neighbourhood perceptions - mother

Scale	Range of scores	Mean Score	SD
Rubbish and litter	1-4	2.86	0.91
Homes and gardens in bad condition	1-4	3.31	0.66
Vandalism	1-4	3.28	0.75
People being drunk/taking drugs	1-4	3.38	0.78
Safe to walk alone after dark	1-4	2.21	0.81
Safe for children to play outside during day	1-4	1.72	0.66
Safe parks, playgrounds and play spaces	1-4	2.39	0.98
Recreational facilities appropriate to a 9 year old	1-2	1.39	0.49

Table 5.2: Descriptive statistics for neighbourhood perceptions - child

Scale	Range of scores	Mean Score	SD
Good places to play near house	0-1	0.84	0.37
Green area to play	0-1	0.81	0.40
Playground near house	0-1	0.45	0.50
Safe places to play	0-1	0.79	0.41
Do you feel safe	0-1	0.96	0.18
Too much traffic	0-1	0.21	0.41
Streets dirty	0-1	0.18	0.39
A lot of graffiti	0-1	0.20	0.40
Friends to play with	0-1	0.81	0.39

Table 5.3: Summary statistics of health variables

Gender	Weight, kg	Height, cm	BMI
Full sample	33.68 (7.13)	137.12 (6.23)	17.81 (2.96)
Boys	33.65 (6.86)	137.60 (6.23)	17.67 (2.77)
Girls	33.72 (7.36)	136.68 (6.20)	17.94 (3.12)

Table 5.4: International cut-off points for BMI for overweight and obesity by sex

Age (years)	BMI 25kg/m2		BMI 30kg/m2	
	Boys	Girls	Boys	Girls
9	19.1	19.07	22.77	22.81
9.5	19.46	19.45	23.39	23.46
10	19.84	19.86	24	24.11

Defined to pass through BMI of 25 and 30kg/m2 at age 18. *Source:* Cole et al. (2000: 1242).

Table 5.5: BMI at or above overweight and obesity cut-offs by sex

Gender	≥ 85th centile	CI	≥ 95th centile	CI
Full sample %	20.82	[20.65-20.92]	23.64	[23.32-23.94]
Boys %	20.31	[20.10-20.55]	23.25	[22.83-23.72]
Girls %	21.22	[20.93-21.43]	24.00	[23.59-24.32]

Table 5.6: BMI at or above overweight and obesity by sex

Gender	Overweight	Obese
Full sample %	18.27	5.42
Boys %	15.68	4.68
Girls %	20.66	6.10

Table 5.7: Descriptive statistics for genetics

Variable	Mean	SD	Min	Max
Mother overweight	0.31	0.46	0	1
Mother obese	0.17	0.37	0	1
Father overweight	0.43	0.50	0	1
Father obese	0.21	0.41	0	1
Birthweight	3.53	0.57	0.57	6.04
Breastfed	0.52	0.50	0	1
Mother smoked	0.21	0.41	0	1
Mother drank	0.39	0.49	0	1

Table 5.8: Descriptive statistics for child's activity variables

Variable	Mean	SD	Min	Max
Played games outside (MM)	0.50	0.50	0	1
Went to the park (MM)	0.21	0.41	0	1
Play games outside (KD-fav)	0.77	1.03	0	3
Play sports outside (KD-fav)	1.42	1.27	0	3
Exercise 20 minutes	3.35	0.81	1	4
Exercise 60 minutes	5.36	2.1	1	8
Quiet child	0.20	0.40	0	1
Computer in room	0.33	0.47	0	1
TV in room	0.40	0.49	0	1
Computer favorite activity	0.17	0.37	0	1
TV favorite activity	0.29	0.45	0	1

Table 5.9: Description of household location

Variable	Frequency
Border	11.13
Dublin	24.83
Mid-East	11.73
Midland	6.13
Mid-West	10.22
South-East	10.09
South-West	16.27
West	9.60

Table 5.10: Determinants of play

Variable	Play outside		Play in park		Play favorite	
Boys	0.329***	-0.0532	0.142**	-0.0647	0.00958	-0.0604
Litter	-0.180***	-0.0647	-0.0791	-0.0795	0.0205	-0.0729
Bad conditions	0.00233	-0.116	-0.168	-0.142	-0.0113	-0.13
Vandalism	0.169	-0.104	0.152	-0.121	-0.16	-0.116
Drug taking	-0.0763	-0.098	0.129	-0.116	0.0644	-0.11
Unsafe after dark	-0.0853	-0.0585	-0.178**	-0.0728	-0.095	-0.0669
Safe to play	-0.232**	-0.101	-0.447***	-0.117	-0.0817	-0.115
Play spaces	0.245***	-0.061	0.281***	-0.0768	0.0543	-0.069
Play facilities	0.0532	-0.0578	0.0784	-0.0718	-0.0416	-0.0656
KD Good play spaces	0.179**	-0.079	0.135	-0.101	0.144	-0.0916
KD green play spaces	-0.147**	-0.0733	-0.011	-0.0908	0.0734	-0.0834
KD playgrounds	0.0791	-0.0574	0.306***	-0.0699	0.00115	-0.0649
KD safe to play	0.0667	-0.073	0.199**	-0.093	-0.177**	-0.0814
KD Feel safe	0.223	-0.154	0.128	-0.187	-0.052	-0.172
KD Traffic	0.0345	-0.0685	0.111	-0.0815	-0.133*	-0.0795
KD dirty	-0.0426	-0.0737	-0.0268	-0.0895	-0.0725	-0.0853
KD graffiti	0.0644	-0.0707	-0.0482	-0.0849	0.142*	-0.0794
KD friends to play with	0.203***	-0.0712	0.0878	-0.0884	0.0223	-0.0806
Quiet child	-0.316***	-0.0665	-0.0413	-0.081	-0.185**	-0.0774
Older mother	-0.0192***	-0.0052	-0.0122**	-0.00611	0.00544	-0.00586
Working mother	-0.0948*	-0.0545	-0.059	-0.0658	-0.031	-0.0622
Log Income	0.0979*	-0.0544	0.014	-0.0648	0.0203	-0.0629
Child Irishborn	0.191**	-0.0861	-0.297***	-0.0982	0.183*	-0.101
Village	-0.324***	-0.0997	-0.0859	-0.124	-0.142	-0.116
Town 1500-2999	-0.374**	-0.162	-0.261	-0.204	0.0973	-0.178
Town 3000-4999	-0.361**	-0.153	0.00836	-0.182	-0.185	-0.182
Town 5000-9999	-0.333***	-0.128	0.00702	-0.152	0.124	-0.141
Town 10000 or more	-0.382***	-0.0844	-0.15	-0.106	0.124	-0.0938
Waterford	0.163	-0.259	0.446	-0.287	0.314	-0.271
Galway	-0.138	-0.258	0.476*	-0.282	0.355	-0.277
Limerick	-0.753***	-0.212	-0.454	-0.292	-0.0921	-0.242
Cork	-0.756***	-0.154	0.102	-0.177	-0.337*	-0.187
Dublin city	-0.652***	-0.0944	0.192*	-0.11	-0.205*	-0.11
Dublin country urban	-0.816***	-0.105	-0.0721	-0.126	-0.0691	-0.12
Dublin country rural	-0.783**	-0.374	0.431	-0.415	-0.0323	-0.409
Constant	-0.588	-0.612	-1.109	-0.727	-1.525**	-0.705

* significant at 10 per cent level, ** significant at 5 per cent level,

*** significant at 1 per cent level; robust standard error in second column.

Omitted category: Open country.

Table 5.11: Determinants of play - including activities

Variable	Play outside	Play in park			Play favorite	
Boys	0.182***	-0.0598	0.0728	-0.0728	0.259***	-0.0691
Litter	-0.191***	-0.066	-0.0753	-0.0797	0.0266	-0.0744
Bad conditions	-0.0199	-0.118	-0.182	-0.141	-0.00932	-0.132
Vandalism	0.180*	-0.107	0.143	-0.122	-0.213*	-0.118
Drug taking	-0.0473	-0.101	0.15	-0.117	0.112	-0.113
Unsafe after dark	-0.0847	-0.0596	-0.176**	-0.0733	-0.087	-0.0686
Safe to play	-0.236**	-0.103	-0.447***	-0.118	-0.0965	-0.117
Play spaces	0.230***	-0.062	0.271***	-0.0773	0.059	-0.0704
Play facilities	0.0224	-0.059	0.0702	-0.0723	-0.0319	-0.067
KD Good play spaces	0.121	-0.081	0.0979	-0.102	0.112	-0.094
KD green play spaces	-0.178**	-0.0751	-0.0171	-0.0919	0.0779	-0.0848
KD playgrounds	0.0721	-0.0585	0.300***	-0.0703	-0.00719	-0.0659
KD safe to play	0.0635	-0.0744	0.193**	-0.0942	-0.199**	-0.0837
KD Feel safe	0.182	-0.159	0.126	-0.188	-0.00553	-0.179
KD Traffic	0.0317	-0.0701	0.111	-0.0822	-0.147*	-0.0806
KD dirty	-0.0279	-0.0755	-0.0189	-0.0909	-0.0686	-0.0873
KD graffiti	0.0603	-0.0718	-0.0417	-0.0858	0.159*	-0.0811
KD friends to play with	0.121*	-0.0729	0.0454	-0.0896	-0.0232	-0.0829
Quiet child	-0.126*	-0.0691	0.0599	-0.0826	-0.224***	-0.0801
Older mother	-0.0231***	-0.00546	-0.0133**	-0.00633	0.00799	-0.00617
Working mother	-0.0996*	-0.0557	-0.0574	-0.0662	-0.0357	-0.0633
Log Income	0.0563	-0.0565	-0.0106	-0.0659	-0.00335	-0.065
Child Irishborn	0.134	-0.0877	-0.343***	-0.0994	0.142	-0.104
Uses a computer	-0.0127	-0.068	-0.199**	-0.084	-0.256***	-0.0807
Computer in room	-0.0776	-0.0669	-0.0714	-0.0805	0.074	-0.0753
TV in room	-0.00037	-0.065	0.00679	-0.0787	-0.053	-0.0747
Computer favorite activity	-0.263***	-0.0766	-0.0578	-0.0925	-0.643***	-0.0935
TV favorite activity	-0.144**	-0.0603	-0.116	-0.0739	-0.567***	-0.0726
Sport favorite activity	0.0806***	-0.0236	0.0283	-0.0283	-0.277***	-0.0284
Exercises 20min never	-0.758**	-0.307	-0.925*	-0.497	0.223	-0.329
Exercises 20min 1-2t/wk	-0.104	-0.0772	-0.275***	-0.0972	-0.1	-0.089
Exercises 20min 3-4t/wk	-0.0535	-0.0678	-0.126	-0.0819	-0.201**	-0.079
Exercises 60min never	-0.448***	-0.16	-0.119	-0.196	-0.1	-0.179
Exercises 60min 1-2t/wk	-0.281***	-0.0793	-0.283***	-0.0997	-0.158*	-0.0924
Exercises 60min 3-4t/wk	-0.192***	-0.0639	-0.105	-0.0764	-0.0586	-0.0732
Does no sport	-1.319***	-0.246	-0.222	-0.254	-0.430*	-0.253
Sport 1-2t/wk	-0.480***	-0.0716	-0.127	-0.0869	-0.059	-0.082
Sport 3-4t/wk	-0.220***	-0.0704	-0.203**	-0.0867	-0.107	-0.0818
Location Variables	Included		Included		Included	

* significant at 10 per cent level, ** significant at 5 per cent level,

*** significant at 1 per cent level; robust standard error in second column.

Omitted categories: Exercises 20min and 60min 5+ times a week, sport 5+ times a week.

Table 5.12: Determinants of play including SAPS

Variable	Play outside		Play in park		Play favorite	
Boys	0.180***	-0.0598	0.0707	-0.0728	0.260***	-0.0692
Litter	-0.182***	-0.0662	-0.0664	-0.0799	0.0234	-0.0745
Bad conditions	-0.0188	-0.119	-0.186	-0.141	-0.00966	-0.133
Vandalism	0.188*	-0.107	0.145	-0.122	-0.216*	-0.118
Drug taking	-0.0414	-0.101	0.151	-0.117	0.11	-0.113
Unsafe after dark	-0.0769	-0.0597	-0.169**	-0.0733	-0.0894	-0.0687
Safe to play	-0.226**	-0.103	-0.432***	-0.118	-0.1	-0.117
Play spaces	0.221***	-0.0621	0.263***	-0.0775	0.0622	-0.0705
Play facilities	0.0223	-0.059	0.0649	-0.0724	-0.0319	-0.0671
KD Good play spaces	0.122	-0.0811	0.103	-0.102	0.111	-0.0941
KD green play spaces	-0.182**	-0.0752	-0.0104	-0.092	0.0793	-0.0848
KD playgrounds	0.0742	-0.0586	0.297***	-0.0705	-0.00791	-0.066
KD safe to play	0.0723	-0.0745	0.202**	-0.0943	-0.202**	-0.0838
KD Feel safe	0.177	-0.16	0.121	-0.189	-0.00335	-0.179
KD Traffic	0.0304	-0.0702	0.114	-0.0823	-0.146*	-0.0805
KD dirty	-0.0271	-0.0755	-0.0155	-0.0909	-0.0688	-0.0874
KD graffiti	0.0667	-0.0719	-0.0373	-0.0858	0.156*	-0.0811
KD friends to play with	0.116	-0.0731	0.0445	-0.0897	-0.0216	-0.083
Quiet child	-0.128*	-0.0691	0.0578	-0.0826	-0.223***	-0.0801
Older mother	-0.0238***	-0.00547	-0.0144**	-0.00635	0.00821	-0.00618
Working mother	-0.100*	-0.0558	-0.0583	-0.0663	-0.0353	-0.0633
Log Income	0.0623	-0.0566	-0.00526	-0.066	-0.0054	-0.0652
Child Irishborn	0.132	-0.0878	-0.349***	-0.0993	0.143	-0.104
SAPS %pop 15-24	-0.0135	-0.00841	0.00948	-0.00974	0.00439	-0.00961
SAPS %pop retired	0.0125*	-0.0069	0.0183**	-0.00795	-0.00421	-0.00789
Location variables	Included		Included		Included	
Activity variables	Included		Included		Included	

* significant at 10 per cent level, ** significant at 5 per cent level,

*** significant at 1 per cent level; robust standard error in second column

Table 5.13: Including principle component

Variable	Play outside		Play in park		Play favorite	
PC component	0.0553***	-0.0186	-0.00711	-0.0225	0.0164	-0.0212
Unsafe after dark	-0.0444	-0.0592	-0.129*	-0.0717	-0.0817	-0.0681
No play space	-0.204***	-0.0617	-0.230***	-0.0768	-0.0574	-0.0701
No play facility	-0.0291	-0.0588	-0.0635	-0.0721	0.0223	-0.0668
KD Good play spaces	0.123	-0.0809	0.108	-0.102	0.125	-0.094
KD green play spaces	-0.193***	-0.0749	-0.0311	-0.0913	0.0725	-0.0847
KD playgrounds	0.0812	-0.0583	0.317***	-0.0696	-0.0191	-0.0657
KD safe to play	0.0692	-0.0741	0.183*	-0.0939	-0.192**	-0.0836
KD Feel safe	0.126	-0.159	0.0802	-0.189	-0.0204	-0.178
KD Traffic	0.0415	-0.07	0.137*	-0.0818	-0.139*	-0.0802
KD dirty	-0.00572	-0.0752	-0.0113	-0.0903	-0.06	-0.087
KD graffiti	0.0851	-0.0718	-0.0336	-0.0852	0.155*	-0.0807
KD friends to play with	0.0942	-0.0728	0.0124	-0.0887	-0.0358	-0.0824
Quiet child	-0.119*	-0.0693	0.0693	-0.0821	-0.224***	-0.08
Older mother	-0.0241***	-0.00545	-0.0141**	-0.00632	0.0093	-0.00612
Working mother	-0.0988*	-0.0557	-0.0637	-0.0661	-0.0386	-0.0632
Log Income	0.0438	-0.0564	-0.00592	-0.0663	-0.00042	-0.0649
Child Irishborn	0.123	-0.0873	-0.357***	-0.099	0.132	-0.103
Computer favorite activity	-0.217***	-0.0754	-0.0324	-0.0908	-0.575***	-0.0914
TV favorite activity	-0.133**	-0.0599	-0.119	-0.0733	-0.537***	-0.0723
Location variables	Included		Included		Included	
Activity variables	Included		Included		Included	
SAPS variables	Included		Included		Included	

* significant at 10 per cent level, ** significant at 5 per cent level,

*** significant at 1 per cent level; robust standard error in second column

Table 5.14: BMI basic regressions

Variable	BMIord	BMI	BMIdiffsch	BMIdiffed
Boys	-0.415***	-0.331***	-0.361***	-0.332***
Mom overweight	0.632***	0.761***	0.632***	0.688***
Mom obese	1.163***	1.635***	1.434***	1.495***
Dad overweight	0.0551	0.157*	0.0812	0.180**
Dad obese	0.523***	0.908***	0.881***	0.843***
Birthweight	0.325***	0.639***	0.597***	0.627***
Breastfed	-0.214***	-0.0644	-0.0747	-0.245***
Smoker	0.592***	0.861***	0.909***	0.933***
Drinker	-0.176***	-0.0756	-0.0971	-0.175**
Constant		15.10***	-0.0744	-0.0783
Village	-0.184	-0.12	-0.246	7.03E-09
Town 1500-2999	-0.317	-0.191	-0.142	-0.0304
Town 3000-4999	-0.327*	-0.227	-0.222	
Town 5000-9999	0.145	-0.282	-0.204	
Town 10000 or more	0.00195	-0.0641	-0.168	
Waterford	-0.0648	0.0402	-0.114	
Galway	-0.276	-0.203	-0.355	
Limerick	-0.127	0.169	-0.286	
Cork	0.0848	0.0493	-0.258	
Dublin city	-0.218**	-0.154	-0.196	
Dublin country urban	-0.112	-0.355***	-0.111	
Dublin country rural	-0.169	-0.151	-0.134	
cut 1	2.615***	0.0907	-0.44	
cut 2	4.396***	-0.225		
		-0.232		

* significant at 10 per cent level, ** significant at 5 per cent level,

*** significant at 1 per cent level; robust standard error in second column.

Omitted category: Open country.

Table 5.15: BMI regressions full

Variable	BMIord		BMI		BMIdiffsch		BMIdiffed	
Boys	-0.385***	-0.0667	-0.301***	-0.0752	-0.338***	-0.0871	-0.311***	-0.0797
SAPS %pop private HH living	-0.00927***	-0.00285	-0.0109***	-0.00331	-0.00291	-0.00282		
Exercises 20min never	0.654*	-0.358	1.048*	-0.586	0.899*	-0.533	0.385	-0.507
Exercises 20min 1-2t/wk	0.171**	-0.0817	0.154	-0.0996	0.145	-0.0984	0.16	-0.107
Exercises 20min 3-4t/wk	0.112	-0.0769	0.147*	-0.0874	0.11	-0.0863	0.0614	-0.0928
Played outside	-0.123*	-0.0644	-0.167**	-0.0733	-0.188**	-0.0739	-0.177**	-0.0775
Computer in room	-0.0701	-0.0792	-0.0138	-0.093	-0.0575	-0.0907	-0.0744	-0.098
TV in room	0.239***	-0.0766	0.258***	-0.0912	0.328***	-0.0902	0.313***	-0.0973
Top Inc Quintile	0.00957	-0.0758	0.0546	-0.0817	0.118	-0.0834	0.119	-0.0886
Bottom Inc Quintile	-0.00901	-0.105	-0.064	-0.13	-2.26E-02	-0.127	-0.0719	-0.135
Workingmom	0.174***	-0.0674	0.203***	-0.0748	0.204***	-0.0754	0.200**	-0.0796
Married mom	-0.313***	-0.101	-0.546***	-0.125	-0.611***	-0.121	-0.591***	-0.131
Live in house	-0.430*	-0.244	-0.839**	-0.338	-0.989***	-0.363	-1.270***	-0.4
Genetics Variables	Included		Included		Included		Included	
Location Variables	Included		Included		Not included		Not included	

* significant at 10 per cent level, ** significant at 5 per cent level,
*** significant at 1 per cent level; robust standard error in second column.
Omitted category: Exercises 20min 5+times week.

Table 5.16: Two stage - including outdoor play

Variable	BMI	Std. Err.	BMIdiffsch	Std. Err.	BMIdiffd	Std. Err.
Played outside	-5.520***	1.831	-3.998***	1.549	-4.039***	1.285
Computer in room	-0.052	0.121	-0.087	0.105	-0.072	0.107
TV in room	0.271**	0.122	0.328***	0.105	0.291***	0.108
Top Inc Quintile	0.187	0.127	0.208*	0.120	0.210*	0.108
Bottom Inc Quintile	-0.045	0.164	-0.121	0.15	-0.139	0.146
Working mom	0.091	0.111	0.130	0.095	0.117	0.095
Married mom	-0.479***	0.160	-0.530***	0.140	-0.490***	0.142
Live in house	-0.659	0.409	-0.728*	0.373	-1.181***	0.373
Sports Variables	Included		Included		Included	
Genetics Variables	Included		Included		Included	
Location Variables	Included		Not included		Not included	
Constant	19.474***	1.153	6.45E-09	0.039	-9.44E-09	0.036
Underid test (χ^2 p-value)	0.000		0.000		0.000	
Weak identification test	17.94		16.38		16.38	

* significant at 10 per cent level, ** significant at 5 per cent level,

*** significant at 1 per cent level; robust standard error in second column

5.8 Figures

Figure 5.1: Summary structure of instruments used in Child Cohort Survey, Wave 1
Source: GUL, 2010

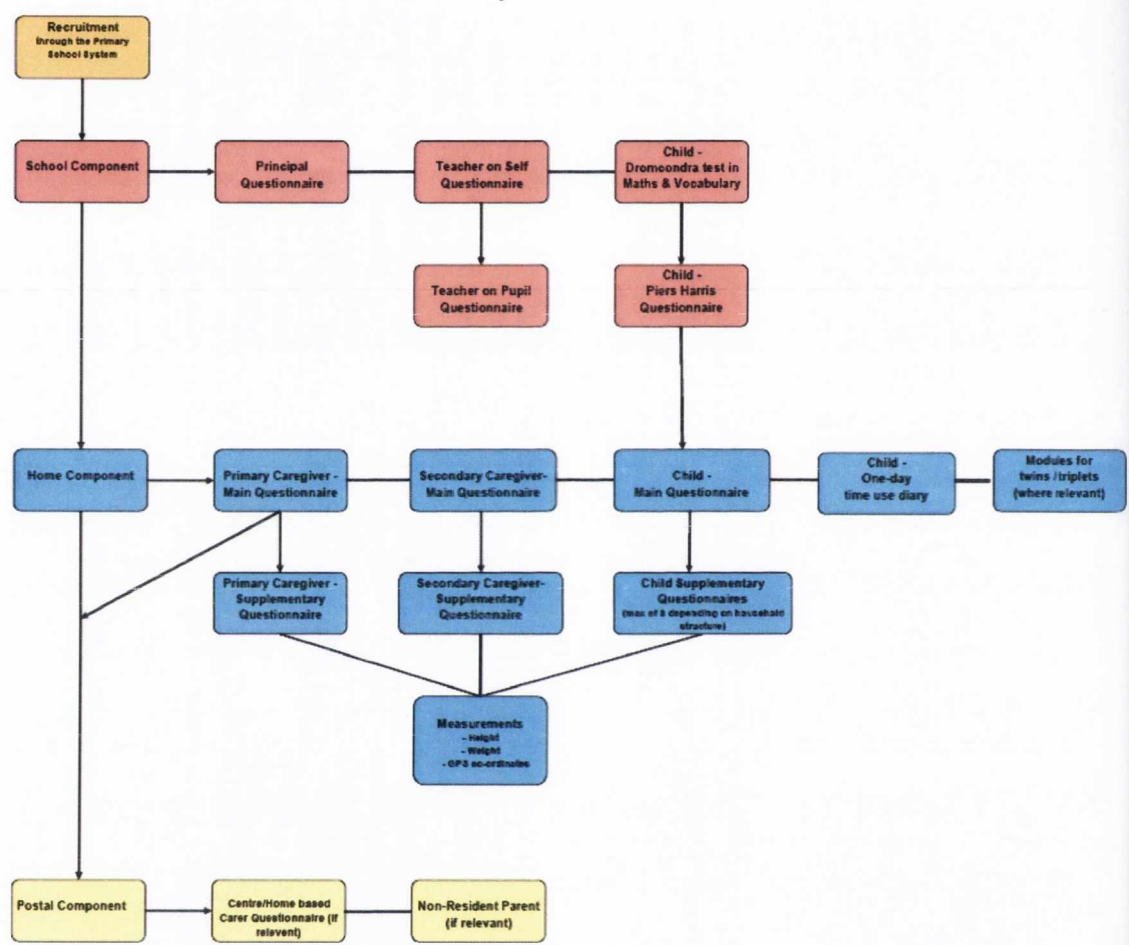


Figure 5.2: Histograms of BMIs for full sample (top), just girls (middle) and just boys (bottom)

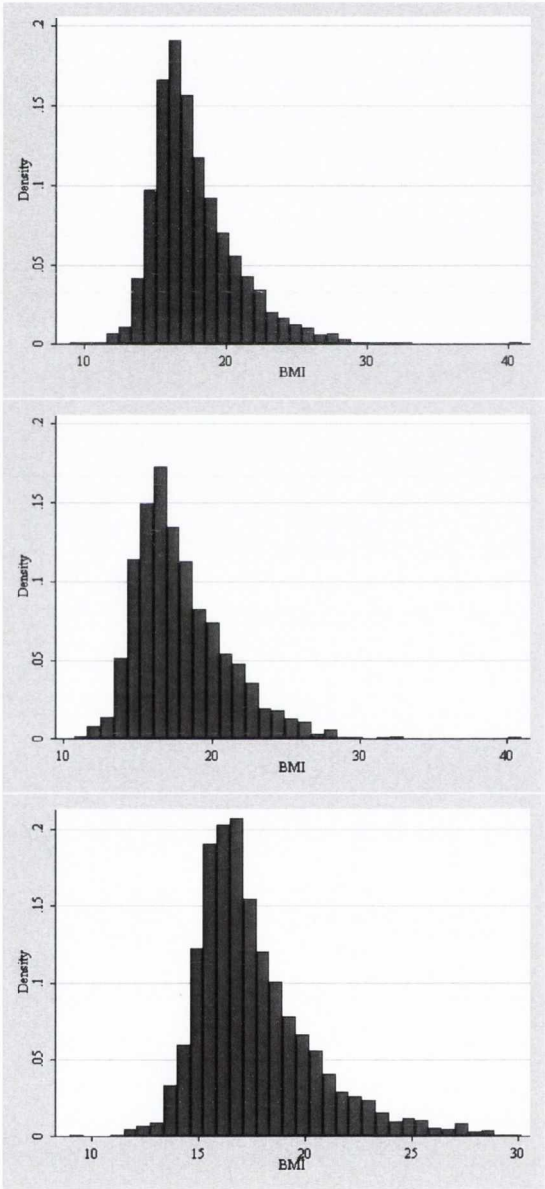
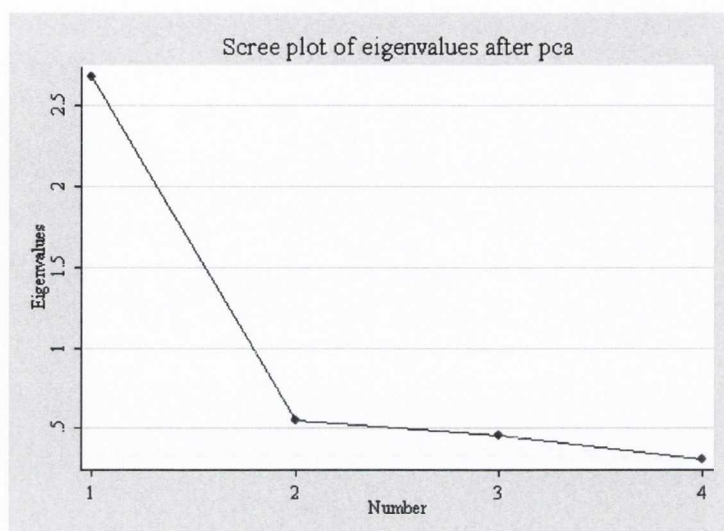


Figure 5.4: Screeplots



5.9 Appendix

Table 5.17: List of questionnaires in the Child Cohort Survey, Wave 1 (GUI, 2010)

Questionnaire	Description
Principal	Recorded general information on school characteristics including size, challenges, ethos etc. and some personal details about the principal.
Teacher on self	Recorded general information about the class(room) characteristics such as size, curriculum, preferred teaching methods and some personal details about the teacher.
Teacher on child	Recorded specific information about the study child such as temperament, academic performance and school preparedness.
Drumcondra Tests	Curriculum-based, standardised tests used to indicate level of ability in vocabulary and maths.
Piers Harris 2	Internationally validated self-concept scale.
Mother/Lone Father	Detailed socio-economic household level questions.
Father/Partner	Individual level behavioural, health and neighbourhood related questionnaire. Also questions on relationship to child.
Child	Individual level behavioural, health and neighbourhood related questionnaire. Also questions on relationship to child.
One-day time use diary	Questionnaire on relationship with parents, friends, school, diet, neighbourhood and home life.
Modules for twins and triplets	Time use diary.
Height and weight	Additional questions on siblings.
Non-resident parent	Additional measurements and weighings.
Carer (home based)	Separate questionnaires for non-resident parents on relationship and behaviour with child.
Carer (centre-based)	Separate questionnaires for carer on relationship and behaviour with child.
	As above.

Table 5.18: Cronbach’s alpha results and eigenvalues

Item	Alpha	Component	Eigenvalue
MMM2a	0.8130	Component 1	2.6748
MMM2b	0.7878	Component 2	0.5529
MMM2c	0.7448	Component 3	0.4594
MMM2d	0.7819	Component 4	0.4980
Test scale	0.8268		

Table 5.19: Descriptive statistics for household level variables

Scale	Mean	SD	Min	Max
Married mother	0.84	0.37	0	1
Mother age	39.91	5.35	22	61
Working mother	0.59	0.49	0	1
Child Irish born	0.89	0.31	0	1
Log income	10.81	0.52	7.60	13.12

Table 5.20: Play analysis Urban/Rural samples

Variable	Urban - Play outside		Rural - play outside		Urban - Play park		Rural - Play park		Urban - Play fav		Rural - Play fav	
Litter	-0.118	-0.1	-0.250***	-0.0887	0.0552	-0.197*	-0.197*	-0.115	-0.0339	-0.112	0.0503	-0.1
Bad conditions	-0.143	-0.159	0.163	-0.186	-0.26	-0.215	-0.235	-0.176	-0.0429	-0.176	0.108	-0.207
Vandalism	0.243*	-0.142	0.113	-0.166	0.0904	0.252	-0.196	-0.255	-0.162	-0.154	-0.174	-0.174
Drug taking	-0.0409	-0.134	-0.0622	-0.16	0.219	0.0688	-0.197	-0.226	-0.154	0.0251	-0.171	-0.171
Unsafe after dark	-0.125	-0.093	-0.0286	-0.0814	-0.106	-0.0766	-0.102	-0.106	-0.247**	-0.106	0.0291	-0.0912
Safe to play	-0.237*	-0.138	-0.169	-0.16	-0.304*	-0.498**	-0.186	-0.159	-0.117	-0.159	-0.0479	-0.181
Play spaces	0.158	-0.0986	0.255**	-0.081	0.112	0.367***	-0.103	-0.103	-0.0817	-0.112	0.0406	-0.0892
Play facilities	0.0246	-0.0918	0.0418	-0.078	-0.0467	0.165*	-0.0989	-0.12	0.0817	-0.103	0.055	-0.107
KD Good play spaces	0.141	-0.134	0.121	-0.102	-0.0639	0.204	-0.137	-0.156	0.0164	-0.141	0.176	-0.119
KD green play spaces	-0.227*	-0.126	-0.153	-0.0942	-0.11	0.0533	-0.122	-0.114	-0.0554	-0.114	0.16	-0.107
KD playgrounds	0.078	-0.0853	0.0575	-0.0817	-0.0433	0.301***	-0.0984	-0.124	0.0138	-0.0971	-0.0654	-0.092
KD safe to play	0.0551	-0.125	0.097	-0.0931	0.186	0.216*	-0.101	-0.121	-0.121	-0.143	-0.212**	-0.105
KD Feel safe	0.344	-0.212	-0.0669	-0.245	-0.0433	0.185*	-0.138	-0.243	-0.0116	-0.106	-0.350***	-0.263
KD Traffic	0.0756	-0.0921	-0.0443	-0.11	-0.0433	0.447	-0.134	-0.126	-0.0627	-0.126	-0.0322	-0.131
KD Feel safe	0.025	-0.109	-0.00844	-0.106	0.185*	-0.0053	-0.136	-0.118	0.226	-0.118	-0.0322	-0.11
KD dirty	-0.0952	-0.109	-0.00722	-0.126	-0.222**	-0.0562	-0.134	-0.124	-0.0116	-0.106	-0.0322	-0.11
KD graffiti	0.0644	-0.0952	0.0571	-0.112	0.0754	0.218*	-0.131	-0.107	0.227*	-0.124	-0.0322	-0.11
KD Friends to play with	0.108	-0.112	0.118	-0.0979	-0.222**	0.0314	-0.123	-0.107	-0.0127	-0.119	-0.0322	-0.11
Quiet child	-0.208*	-0.104	-0.0619	-0.094	0.132	-0.0170*	-0.119	-0.117	-0.215*	-0.119	-0.0322	-0.11
Other mother	-0.0399***	-0.083	-0.0105	-0.074	-0.0170*	-0.0122	-0.0891	-0.093	0.00757	-0.0932	0.00795	-0.0858
Working mother	-0.0691	-0.0833	-0.117	-0.0766	-0.0553	-0.0597	-0.093	-0.093	-0.0932	-0.0932	0.011	-0.0858
Log Income	0.165**	-0.0824	-0.0254	-0.0805	-0.0693	-0.0574	-0.093	-0.093	-0.0932	-0.0932	0.011	-0.0858
Child Irishborn	-0.136	-0.0973	-0.00813	-0.119	-0.453***	-0.297**	-0.138	-0.162	-0.0951	-0.0951	-0.0951	-0.0951
Uses a computer	-0.0203	-0.0973	0.0186	-0.0967	-0.240**	-0.148	-0.122	-0.162	-0.0951	-0.0951	-0.0951	-0.0951
Computer in room	0.0193	-0.0989	-0.102	-0.0904	-0.117	-0.083	-0.112	-0.162	-0.0951	-0.0951	-0.0951	-0.0951
TV in room	-0.0976	-0.1	0.0778	-0.0871	-0.103	-0.114	-0.118	-0.162	-0.0951	-0.0951	-0.0951	-0.0951
Computer favorite activity	-0.134	-0.113	-0.298***	-0.103	-0.0299	-0.182*	-0.103	-0.116	-0.0951	-0.0951	-0.0951	-0.0951
TV favorite activity	-0.175*	-0.0932	-0.11	-0.0797	-0.0299	-0.182*	-0.103	-0.116	-0.0951	-0.0951	-0.0951	-0.0951
Sport favorite activity	0.0719**	-0.0344	0.130**	-0.0308	0.0724*	0.0449	-0.129	-0.116	-0.0951	-0.0951	-0.0951	-0.0951
Exercises 20min never	-0.492	-0.472	-0.963**	-0.41	-0.677	-0.653	-0.77	-0.504	-0.437	-0.437	-0.437	-0.437
Exercises 20min 1-24/wk	-0.135	-0.119	-0.375**	-0.103	-0.147	-0.239*	-0.132	-0.138	-0.138	-0.138	-0.138	-0.138
Exercises 20min 3-44/wk	-0.0989	-0.101	-0.012	-0.0933	-0.0748	-0.175	-0.117	-0.118	-0.118	-0.118	-0.118	-0.118
Exercises 60min never	-0.569**	-0.251	-0.360*	-0.214	-0.0556	-0.181	-0.278	-0.265	-0.265	-0.265	-0.265	-0.265
Exercises 60min 1-24/wk	-0.376***	-0.12	-0.223**	-0.107	-0.343**	-0.137	-0.137	-0.137	-0.137	-0.137	-0.137	-0.137
Exercises 60min 3-44/wk	-0.171*	-0.0949	-0.211**	-0.0872	-0.0329	-0.109	-0.109	-0.109	-0.109	-0.109	-0.109	-0.109
Does no sport	-1.282***	-0.341	-1.381***	-0.349	-0.45	-0.348	-0.372	-0.372	-0.372	-0.372	-0.372	-0.372
Sport 1-24/wk	-0.169	-0.106	-0.480***	-0.0956	-0.202	-0.125	-0.125	-0.125	-0.125	-0.125	-0.125	-0.125
Sport 3-44/wk	-0.106	-0.106	-0.308**	-0.0946	-0.394***	-0.0936	-0.112	-0.112	-0.112	-0.112	-0.112	-0.112
SAPS %pop 15-24	-0.0252*	-0.0107	0.00389	-0.0151	-0.00592	0.0416**	-0.017	-0.017	-0.017	-0.017	-0.017	-0.017
SAPS %pop retired	0.0113	-0.00901	0.012	-0.0113	0.0256**	0.00677	-0.0135	-0.0135	-0.0135	-0.0135	-0.0135	-0.0135

*** significant at 10 per cent level; ** significant at 5 per cent level; * significant at 1 per cent level; robust standard error in second column.

Unitted categories: Exercises 20min and 60min 5+ times a week, sport 5+ times a week.

Table 5.21: PC Analysis Urban/Rural Samples

	Urban sample Play outside	Rural sample Play outside	
PC component	0.0462*	-0.0267	-0.026
Unsafe after dark	-0.0812	-0.0885	-0.0806
No play space	-0.132	-0.0976	-0.0806
No play facility	-0.0198	-0.0916	-0.0779
KD Good play spaces	0.14	-0.134	-0.102
KD green play spaces	-0.249**	-0.125	-0.0941
KD playgrounds	0.1	-0.0849	-0.0818
KD safe to play	0.0343	-0.124	-0.0929
KD Feel safe	0.3	-0.212	-0.245
KD Traffic	0.0907	-0.0916	-0.11
KD dirty	0.0626	-0.108	-0.106
KD graffiti	0.0828	-0.0949	-0.113
KD friends to play with	0.0815	-0.111	-0.0978
Quiet child	-0.193*	-0.105	-0.0938
Older mother	-0.0414***	-0.00829	-0.00736
Working mother	-0.068	-0.0832	-0.0765
Log Income	0.148*	-0.0827	-0.0805
Child Irishborn	0.300**	-0.134	-0.118
Uses a computer	-0.0212	-0.0971	-0.0968
Computer in room	0.0208	-0.099	-0.0903
TV in room	-0.0908	-0.0999	-0.087
Computer favorite activity	-0.124	-0.113	-0.103
TV favorite activity	-0.179*	-0.0931	-0.0796
Sport favorite activity	0.0745**	-0.0344	-0.0308
Exercises 20min never	-0.424	-0.436	-0.411
Exercises 1-2t/wk	-0.136	-0.119	-0.102
Exercises 20min 3-4t/wk	-0.099	-0.101	-0.0932
Exercises 60min never	-0.582**	-0.249	-0.215
Exercises 60min 1-2t/wk	-0.390***	-0.12	-0.107
Exercises 60min 3-4t/wk	-0.173*	-0.0948	-0.0871
Does no sport	-1.270***	-0.34	-0.353
Sport 1-2t/wk	-0.516***	-0.108	-0.0953
Sport 3-4t/wk	-0.172	-0.106	-0.0947
SATS %pop 15-24	-0.0226**	-0.0105	-0.0151
SATS %pop retired	0.0125	-0.00893	-0.0113

* significant at 10 per cent level, ** significant at 5 per cent level, *** significant at 1 per cent level; robust standard error in second column. Omitted categories: Exercises 20min and 60min 5+ times a week, sport 5+ times a week.

Table 5.22: Play analysis - Gender subsamples

	Play outside - boys		Play outside - girls		Play park - boys		Play park - girls		Play fav - boys		Play fav - girls	
Litter	-0.256***	-0.0955	-0.128	-0.0932	-0.0776	-0.113	-0.128	-0.115	-0.00121	-0.109	0.0301	-0.104
Bad conditions	-0.0623	-0.174	-0.355*	-0.164	-0.355*	-0.207	-0.0334	-0.2	0.00885	-0.197	-0.0315	-0.185
Vandalism	0.015	-0.159	0.321**	-0.145	0.117	-0.176	0.184	-0.174	-0.380**	-0.175	-0.111	-0.162
Drug taking	0.0314	-0.152	0.0988	-0.138	0.22	-0.171	0.0742	-0.163	0.122	-0.169	0.0954	-0.156
Unsafe after dark	-0.112	-0.0874	-0.0394	-0.0831	-0.165	-0.104	-0.173*	-0.105	-0.0803	-0.1	-0.0912	-0.0963
Safe to play	-0.222	-0.152	-0.224	-0.142	-0.389**	-0.173	-0.423**	-0.166	-0.232	-0.179	-0.0915	-0.136
Play spaces	0.281***	-0.0912	0.195**	-0.0864	0.244**	-0.111	0.279**	-0.166	0.0989	-0.104	0.0426	-0.0978
Play facilities	-0.0114	-0.0874	0.0241	-0.0813	0.0563	-0.105	0.0805	-0.102	0.0321	-0.101	0.134	-0.0923
KD Good play spaces	0.283**	-0.121	-0.014	-0.11	-0.114	-0.147	0.327**	-0.146	0.0433	-0.143	0.124	-0.127
KD green play spaces	-0.0796	-0.109	-0.284***	-0.104	0.0511	-0.132	0.0878	-0.131	0.0255	-0.126	-0.00139	-0.117
KD play/grounds	0.0409	-0.0843	0.118	-0.0825	0.386***	-0.0997	0.215**	-0.101	-0.0305	-0.096	-0.00139	-0.092
KD safe to play	0.0708	-0.108	0.0561	-0.104	0.00609	-0.139	0.0182	-0.132	-0.17	-0.126	-0.222*	-0.114
KD Feel safe	0.204	-0.217	-0.235	-0.0975	0.122	-0.117	0.289	-0.304	0.242	-0.266	-0.196	-0.253
KD Traffic	0.176*	-0.104	0.0962	-0.107	0.037	-0.129	0.118	-0.118	0.0305	-0.115	-0.314***	-0.114
KD dirty	-0.172	-0.107	0.0904	-0.103	-0.189	-0.123	-0.0242	-0.129	-0.318**	-0.129	0.16	-0.122
KD graffiti	0.0683	-0.102	0.0438	-0.103	-0.0453	-0.126	0.124	-0.121	0.0639	-0.118	0.251**	-0.114
KD friends to play with	0.0105	-0.105	0.226**	-0.104	0.0723	-0.124	0.0253	-0.132	-0.178	-0.12	0.11	-0.119
Quiet child	-0.0601	-0.108	0.170*	-0.0912	0.0723	-0.124	0.122	-0.132	-0.181	-0.123	-0.266**	-0.107
Older mother	-0.0184**	-0.00819	-0.0286***	-0.00751	-0.0103	-0.00915	-0.0196**	-0.00892	0.00115	-0.0922	0.0144*	-0.00849
Working mother	-0.0812	-0.085	0.146*	-0.0779	-0.0533	-0.0944	-0.0476	-0.0948	-0.0666	-0.0971	-7E-05	-0.0908
Log income	-0.0852	-0.085	0.0348	-0.077	0.0478	-0.0964	-0.059	-0.0919	-0.0834	-0.149	0.054	-0.148
Child trashboon	0.129	-0.127	0.125	-0.125	-0.384***	-0.14	-0.326**	-0.142	0.167	-0.149	0.125	-0.113
Uses a computer	0.0689	-0.0979	-0.0805	-0.0976	-0.133	-0.118	-0.276**	-0.124	-0.288**	-0.117	-0.238**	-0.113
Computer in room	-0.0491	-0.106	-0.0789	-0.0929	-0.134	-0.121	-0.0633	-0.115	0.0683	-0.117	0.0712	-0.104
TV in room	-0.0123	-0.104	-0.00836	-0.0859	0.173	-0.123	-0.149	-0.108	-0.0453	-0.12	-0.07	-0.0991
TV favorite activity	-0.288***	-0.0983	-0.224*	-0.125	-0.121	-0.119	0.0527	-0.15	-0.848***	-0.125	-0.402***	-0.145
TV favorite activity	-0.0891	-0.0856	-0.222**	-0.0871	-0.121	-0.102	-0.0447	-0.11	-0.639***	-0.102	-0.522***	-0.106
Sport favorite activity	0.103***	-0.0356	0.0719**	-0.0334	0.0153	-0.0406	-0.0427	-0.0402	-0.311**	-0.0422	-0.256***	-0.0397
Exercises 20min never	-0.227	-0.402	-1.259**	-0.496	0.152	-0.152	-0.282	-0.54	-0.782	-0.655	0.769**	-0.398
Exercises 20min 1-24/wk	0.06	-0.119	-0.222**	-0.103	-0.466***	-0.119	-0.139	-0.131	0.117	-0.132	-0.276**	-0.12
Exercises 20min 3-4/wk	0.138	-0.102	-0.207**	-0.0932	-0.0506	-0.321	-0.199*	-0.115	-0.171	-0.12	-0.258**	-0.107
Exercises 60min never	-0.44*	-0.249	-0.514**	-0.214	-0.391	-0.321	0.0922	-0.246	0.122	-0.261	-0.267	-0.249
Exercises 60min 1-24/wk	-0.275**	-0.122	-0.295***	-0.107	-0.103	-0.148	-0.441***	-0.136	-0.203	-0.144	-0.104	-0.123
Exercises 60min 3-4/wk	-0.193**	-0.0936	-0.189**	-0.089	-0.0804	-0.11	-0.139	-0.108	-0.114	-0.109	0.000318	-0.101
Does no sport	-1.387***	-0.443	-1.203***	-0.302	-0.266	-0.468	-0.18	-0.311	-0.118	-0.405	-0.545	-0.332
Sport 1-24/wk	-0.562***	-0.112	-0.386***	-0.0958	-0.0866	-0.132	-0.15	-0.118	-0.127	-0.131	-0.139	-0.109
Sport 3-44/wk	-0.562***	-0.103	-0.057	-0.0991	-0.209*	-0.125	-0.186	-0.124	-0.193	-0.122	-0.1434	-0.115
SAPS %pop 15-24	-0.0244**	-0.012	-0.00435	-0.0124	-0.0155	-0.0136	0.0334**	-0.0149	0.00955	-0.0133	0.00275	-0.0144
SAPS %pop retired	0.00287	-0.00967	0.0225**	-0.0101	0.0019	-0.011	0.0378**	-0.0118	-0.0202*	-0.0114	0.00963	-0.0112

* significant at 10 per cent level ** significant at 5 per cent level
 *** significant at 1 per cent level robust standard error in second column
 Omitted categories: Exercises 20min and 60min 5+ times a week, sport 5+ times a week.

Table 5.23: PC analysis - gender subsamples

	Play outside - boys		Play outside - girls	
PC component	0.0696***	-0.0269	0.0454*	-0.026
Unsafe after dark	-0.0956	-0.087	0.00346	-0.0823
No play space	-0.268***	-0.0906	-0.167*	-0.0857
No play facility	0.0155	-0.0873	-0.0177	-0.0811
KD Good play spaces	0.270**	-0.121	-0.0205	-0.11
KD green play spaces	-0.0849	-0.109	-0.294***	-0.104
KD playgrounds	0.0608	-0.0842	0.137*	-0.0822
KD safe to play	0.0691	-0.108	0.0567	-0.103
KD Feel safe	2.01E-01	-0.217	0.000889	-0.237
KD Traffic	0.195*	-0.104	-0.0861	-0.0972
KD dirty	-0.17	-0.107	0.128	-0.107
KD graffiti	0.0667	-0.102	0.0692	-0.103
KD friends to play with	-0.00704	-0.105	0.207**	-0.103
Quiet child	-0.0435	-0.108	-0.160*	-0.0914
Older mother	-0.0191**	-0.00817	-0.0295***	-0.00749
Working mother	-0.0202	-0.0811	-0.149*	-0.0777
Log Income	0.0675	-0.0847	0.0113	-0.0771
Child Irishborn	0.118	-0.126	0.128	-0.124
Uses a computer	0.0712	-0.0977	-0.0909	-0.0974
Computer in room	-0.0486	-0.104	-0.0751	-0.093
TV in room	-0.00443	-0.106	0.00249	-0.0854
Computer favorite activity	-0.277***	-0.0985	-0.220*	-0.125
TV favorite activity	-0.0836	-0.0856	-0.228***	-0.0869
Sport favorite activity	0.107***	-0.0356	0.0743**	-0.0324
Exercises 20min never	-0.211	-0.395	-1.241**	-0.495
Exercises 20min 1-2t/wk	0.0494	-0.119	-0.218**	-0.103
Exercises 20min 3-4t/wk	0.126	-0.102	-0.206**	-0.0931
Exercises 60min never	-0.433*	-0.249	-0.531**	-0.214
Exercises 60min 1-2t/wk	-0.277**	-0.122	-0.296***	-0.107
Exercises 60min 3-4t/wk	-0.201**	-0.0935	-0.180**	-0.089
Does no sport	-1.417***	-0.444	-1.171***	-0.303
Sport 1-2t/wk	-0.547***	-0.112	-0.383***	-0.0957
Sport 3-4t/wk	-0.393***	-0.103	-0.0602	-0.0989
SAPS %pop 15-24	-0.0235**	-0.0119	-0.00154	-0.0123
SAPS %pop retired	0.00435	-0.00963	0.0228**	-0.01

* significant at 10 per cent level, ** significant at 5 per cent level,

*** significant at 1 per cent level; robust standard error in second column.

Omitted categories: Exercises 20min and 60min 5+ times a week, sport 5+ times a week.

Table 5.24: BMI analysis - Gender subsamples

	BMI _{old} - boys		BMI _{old} - girls		BMI - boys		BMI - girls	
SAPS %pop private HH living	-0.00523	-0.00429	-0.0124**	-0.00384	-0.00501	-0.00448	-0.0160***	-0.00484
Exercises 20min never	0.849	-0.566	0.549	-0.456	1.034	-0.839	1.123	-0.822
Exercises 20min 1-2h/wk	0.149	-0.13	0.192*	-0.106	0.135	-0.145	0.186	-0.137
Exercises 20min 3-4h/wk	-0.0193	-0.122	0.199*	-0.101	0.116	-0.121	0.179	-0.126
Played outside	-0.266***	-0.0978	-0.0192	-0.0857	-0.339***	-0.0999	-0.0238	-0.107
Computer in room	-0.121	-0.127	-0.0521	-0.105	-0.0182	-0.131	-0.0076	-0.138
TV in room	0.345***	-0.128	0.184*	-0.0977	0.286**	0.133	0.233*	-0.127
Top Inc Quintile	0.0985	-0.116	-0.0552	-0.101	0.154	-0.112	-0.0318	-0.119
Bottom Inc Quintile	0.222*	-0.158	-0.0453	-0.141	0.0136	-0.182	-0.0988	-0.186
Workingmom	-0.103	-0.103	0.147	-0.0913	0.251**	-0.101	0.172	-0.111
Married mom	-0.345**	-0.161	-0.283**	-0.13	-0.601***	-0.188	-0.461***	-0.169
Live in house	-0.648**	-0.326	-0.213	-0.356	-1.134**	-0.461	-0.556	-0.496
	BMI _{isch} - boys		BMI _{isch} - girls		BMI _{old} - boys		BMI _{old} - girls	
SAPS %pop private HH living	-0.00035	-0.00386	-0.00605	-0.00411				
Exercises 20min never	0.425	-0.751	1.341*	-0.742	-0.0451	-0.688	0.719	-0.731
Exercises 20min 1-2h/wk	0.151	-0.145	0.176	-0.134	0.253	-0.156	0.115	-0.148
Exercises 20min 3-4h/wk	0.107	-0.12	0.162	-0.123	0.0613	-0.129	0.0767	-0.133
Played outside	-0.363***	-0.102	-0.0632	-0.107	-0.336***	-0.109	-0.0568	-0.11
Computer in room	-0.0893	-0.122	-0.0835	-0.134	-0.167	-0.134	-0.0487	-0.143
TV in room	0.352***	-0.13	0.316**	-0.126	0.327**	-0.141	0.333**	-0.135
Top Inc Quintile	0.198*	-0.112	0.0457	-0.123	0.213*	-0.12	0.045	-0.13
Bottom Inc Quintile	0.0091	-0.175	-0.042	-0.184	-0.0155	-0.194	-0.104	-0.192
Workingmom	0.222**	-0.103	0.135	-0.11	0.214**	-0.108	0.191	-0.117
Married mom	-0.756***	-0.179	-0.478***	-0.165	-0.750***	-0.197	-0.453**	-0.176
Live in house	-1.025**	-0.469	-0.852	-0.575	-1.089**	-0.551	-1.356**	-0.59

* significant at 10 per cent level, ** significant at 5 per cent level, *** significant at 1 per cent level, robust standard error in second column.

Table 5.25: Diet regression

Variable	BMI _{ord}	BMI	BMI	BMI
Eats at table	-0.0778	-0.129	-0.235	-0.166
Veg	0.0182	-0.0635	-0.0418	-0.072
Coke	-0.0523	-0.0658	-0.0516	-0.0736
Dietcoke	0.167**	-0.0795	0.243**	-0.101
Milk	-0.281***	-0.09	-0.356***	-0.117
Skimmilk	0.345***	-0.102	0.573***	-0.134
% pop owners	-0.00926***	-0.00288	-0.0107***	-0.0033
No exercise 20min	0.642*	-0.366	1.005*	-0.587
1-2 times ex 20min	0.207**	-0.0815	0.177*	-0.0987
3-4 times ex 20min	0.152**	-0.0772	0.182**	-0.0867
Played outside	-0.156**	-0.0646	-0.191***	-0.0728
Computer in room	-0.157**	-0.0791	-0.0821	-0.0909
TV in room	0.247***	-0.0781	0.242***	-0.0912
Top Inc Quintile	-0.00947	-0.0757	0.0188	-0.0809
Bottom Inc Quintile	0.015	-0.106	-0.0412	-0.13
Working mother	0.182***	-0.0676	0.206***	-0.0745
Married mother	-0.312***	-0.101	-0.524***	-0.124
Live in house	-0.402	-0.245	-0.822**	-0.336

* significant at 10 per cent level, ** significant at 5 per cent level,

*** significant at 1 per cent level; robust standard error in second column.

Omitted category: Exercises 20min 5+ times a week.

Table 5.26: First stage regressions

Variable	BMI	Std. Err.	BMI _{diffsch}	Std.Err.	BMI _{diffed}	Std. Err.
Exercises 20 min never	-0.2698***	0.0692	-0.3256***	0.0667	-0.3617***	0.0671
Exercises 20min 1-2wk	-0.1107***	0.017	-0.0983***	0.0169	-0.1154***	0.017
Exercises 20min 3-4wk	-0.0618***	0.0155	-0.0506***	0.015	-0.0513***	0.0153
Computer in room	0.0036	0.0156	0.0069	0.0152	0.0166	0.0153
TV in room	0.001	0.0157	-0.0013	0.0154	-0.0062	0.0157
Top Inc Quintile	0.0241*	0.0152	0.0239*	0.015	0.0239*	0.015
Bottom Inc Quintile	0.0031	0.0211	-0.0274	0.0211	-0.0175	0.0211
Workingmom	-0.0221*	0.0134	-0.0212	0.0132	-0.023*	0.0132
Married mom	0.0133	0.0203	0.0197	0.0199	0.0264	0.02
Live in house	0.0266	0.0522	0.0558	0.0529	0.0167	0.0543
Genetics Variables	Included		Included		Included	
Quietchild	-0.068***	0.0161	-0.0691***	0.0157	-0.0837***	0.0157
Constant	0.5708***	0.0727	0	0.0057	0	0.0053
First stage F test (p-value)	17.95	(0.000)	19.44	(0.000)	28.42	(0.000)
Wu Hausman (p-value)	16.02	(0.000)	8.65	(0.003)	13.10	(0.000)
Durban score (p-value)	16.06	(0.000)	8.67	(0.000)	13.11	(0.000)

* significant at 10 per cent level, ** significant at 5 per cent level,

*** significant at 1 per cent level; robust standard error in second column

Chapter 6

Conclusions

6.1 Thesis conclusions

This thesis was motivated by recent events in the Irish economy. The different chapters provide empirical analyses on different issues related to the state of the Irish economy and Irish society today and in particular some policy issues raised in the aftermath of the boom. This thesis contributes to the existing literature on under-reporting of incomes, property taxation and the environmental and health benefits of outdoor and green spaces. It also raises some policy issues for Ireland going forward.

Chapter 2 examined the potential for under-reporting of incomes by self-employed households in Ireland. The methodology was initially developed in 1989 by Pissarides and Weber and has been employed numerous times to estimate tax evasion in other countries. To date, all published estimates of tax evasion for Ireland have had two characteristics: they are macro-economic analyses and were developed during the 1980s or early 1990s. Chapter 2 provides new estimates of income under-reporting bounds for Ireland from 1987 to 2004. The results suggest that there is strong evidence of differential reporting across all survey years and that under-reporting was high in farming households. The results indicate that on average concealment of income was higher in the earlier years of the surveys and gradually fell. On average self-employed household incomes in Ireland need to be multiplied by between 1.30 and 1.49 depending on the year. Although a downward trend in the level of under-reporting is visible over the period examined, the amount under-reported as a percentage of total taxable income increased. No trend was found for the older households and manual and low professional groups followed the same downward trend as the main sample, albeit with a steeper decline. Differences were uncovered for different professions, with farming households seemingly under-reporting a significant part of their income. Although this analysis does not allow the exact targeting of tax evaders it does provide a platform and some guidelines on which households to target for audits. The recession in Ireland has resulted in a need for additional tax funds and targeting those who are

evading the tax net is one of the least unpopular methods for increasing the government's tax take.

The introduction of new taxes is never a popular policy initiative especially in times of economic difficulties, and consequently any new forms of taxation introduced by the government would at the very least need to be considered fair and equitable by the general public in order to be accepted. This is the motivation of Chapter 3 which investigates the consequences of using a basic type of automated valuation system to estimate the value of properties in Ireland for the purpose of introducing an ad valorem property tax. The Irish government has been discussing introducing a property tax in Ireland to compensate for falls in revenues from the existing stamp duty transactions tax on the sale of homes. The latest announcements have indicated that the Government is favouring the introduction of a flat rate property tax. A flat rate tax would be inequitable but an ad valorem tax requires a price valuation of the property. This chapter investigates whether a system based on a handful of property characteristics (as well as location dummies) can produce estimates which are within the acceptable bounds as defined by the International Association of Assessing Officers and in particular what inequities arise from using such a system. We find that size characteristics produce the most accurate estimates and as expected, the more variables included the more accurate the prediction.

The results indicate however that a valuation system of this type would be regressive. This regressivity could be mitigated by introducing exemptions for lower income groups and having a separate case by case valuation system for extremely high cost dwellings. As discussed in the conclusions (Section 3.8) the paper raises two important policy issues. First, if an ad valorem property tax is to be implemented at short notice, the government will need to use a simple system as no property valuation database exists in Ireland at the moment. Relying on self-assessment in a thin property market will be difficult and could lead to underreporting. Hedonic house prices, or self-assessment with hedonic validation, are alternative solutions. The cost of introducing an ad valorem tax will also need to be taken into account. Using a system of self-assessment with hedonic validation could also help reduce the costs of introducing an automated valuation system in the longer term. Finally, the social implications of using such a simple system and the errors associated with it also need to be weighed up.

House price hedonic regressions of this type have been used in the environmental economics literature for some time to assess the value of environmental amenities. This is the focus of Chapter 4 which examines the premium that households are willing to

pay to live close to green spaces and parks. This chapter contributes to the literature on the valuation of green spaces, in particular in urban centres, and in this case a capital city, Dublin. As opposed to the system used in Chapter 3 where the least amount of information possible was used to evaluate house prices, in this case as much of the house price variation is explained by different house characteristics, neighbourhood characteristics and environmental variables in order to separate out the premium put on proximity to green spaces.

The results indicate that not only do home owners pay a premium to live close to environmental amenities but the premium paid also differs depending on the type of amenity and distance to the amenity. Although not the main focus of the paper, the results show that homeowners will pay a premium in order to live close to transport amenities as well as other environmental amenities such as beaches and coasts. Moreover, homeowners distinguish between green spaces in the close vicinity of their house and within walking distance in their neighbourhood. Living in a green area is valued more highly than living in very close proximity to a park, however living within walking distance of a park is still valued. Depending on the size of the distance buffer used, the premium associated with living in a “green” neighbourhood is between 7% and 9%. As Ireland comes out of a construction boom, these results show that good urban planning, which takes into account environmental amenities, can increase societal well-being. Access to green spaces and outdoor amenities is shown to be valued by homeowners but the potential consequences of access to the outdoors are shown in Chapter 5.

The final paper in this thesis looks at two aspects of access to outdoor and green spaces and their potential links to obesity. In a first stage, the determinants of outdoor play and the use of outdoor spaces and parks by children is examined. The particular focus in this case is how perceptions of the neighbourhood and the outdoor amenities are associated with their use. This analysis contributes to the literature on how neighbourhood perceptions influence outdoor play by children but the originality of this analysis lies in the use of children’s perceptions as well as their mothers’.

The results suggest that mothers are less likely to let their children play outside if in their eyes the neighbourhood quality is poor. This relates to the visual quality of the neighbourhood as well as its level of safety. As in Chapter 4, the visual aspect is not significant when it comes to play in parks, as they are not on the doorstep and with parks the emphasis is placed on facilities (i.e. playgrounds). On the other hand, neighbourhood quality indicators do not matter to children who place more value on

“good play spaces”. In a second stage, the relationship between playing outdoors and children’s weight and health is investigated. Inactivity indicators such as watching television has a significant effect on a child’s BMI. Playing outdoors has a positive effect on overweight and obesity and this causal effect was confirmed using a two stage analysis. Ireland is for the first time facing a childhood obesity problem and policy and planning have focused up to now on promoting healthier lifestyles and providing more facilities. Whilst promoting better diets is undeniably important, in urban centres purely providing play facilities will not in itself result in more outdoor play by children. More emphasis on neighbourhood quality, especially among mothers, could have more far reaching effects.

6.2 Future Research Directions

Some potential avenues for further research emerged during the course of the current studies, and these are now looked at in turn. The results from Chapters 2 and 3 both relate to the Irish fiscal situation and lack of funds. Taxes are by their nature unpopular with the public and even more so in difficult economic times. If new taxes are to be introduced they will need to seem fair or their introduction will add to growing dissatisfaction from the public. Chapter 2 indicates that further research is required looking at the effect of recessions on tax evasion. Anecdotal evidence suggests that the recent economic downturn has resulted in an upsurge in the black economy. Extending the analysis to include a survey conducted *during* the recession would go a long way to confirming the trend that tax evasion is counter-cyclical and its incidence is negatively related to economic growth and taxes. Chapter 3 will need to be further examined sometime next year when a property tax is actually introduced. As the Government has decided to introduce a flat rate tax, the inequities of such a move will have to be investigated and should an ad valorem tax eventually be introduced, the implementation will provide the perfect opportunity to gather further data and allow a more thorough analysis.

Chapters 4 and 5 look at a different aspect of Irish society and deal with more long-term development issues. Chapter 4 would have benefited from the addition of a qualitative aspect to the parks and green spaces areas. At the moment the data available only allow us to look at the density and location of green spaces but should data become available on the quality of parks and the extra infrastructure available, their inclusion would enhance the analysis. Improvements could be made to Chapter 5 once the GIS data on the exact location of the surveyed children and households become

available to researchers. The data exist, but due to the sensitive nature of some of the questionnaires and the dataset having only very recently been made available for research use, the GIS locations are not yet available. These would allow the link between neighbourhood perceptions and true neighbourhood characteristics to be made. As the GUI survey will be repeated in the next few years, it will also be possible to examine long term effects using the panel data.

Bibliography

- Alanon, A. and Gomez-Antonio, M. (2005). Estimating the Size of the Shadow Economy in Spain: A Structural Model with Latent Variables. *Applied Economics*, 37(9):1011–1025.
- Allen, M. T. and Dare, W. H. (2002). Identifying Determinants of Horizontal Property Tax Inequity: Evidence from Florida. *Journal of Real Estate Research*, 24(2):153–.
- Allen, M. T., Dare, W. H., and Riegel, C. (2009). Vertical Inequity in the Tax Assessment of Lodging Properties. *International Journal of Hospitality Management*, In Press, Corrected Proof:–.
- Anderson, L. and Cordell, H. (1988). Influence of Trees on Residential Property Values in Athens, Georgia (U.S.A.): A Survey Based on Actual Sales Prices. *Landscape and Urban Planning*, 15(1-2):153–164.
- Bell, E. (1984). Administrative Inequity and Property Tax Assessment: The Case for the Traditional Approach. *Property Tax Journal*, 3(2):123–131.
- Bell, M. E. and Kirschner, C. (2009). A Reconnaissance of Alternative Measures of Effective Property Tax Rates. *Public Budgeting & Finance*, 29(2):111–136.
- Bender, A. (1997). An Analysis of Perceptions Concerning the Environmental Quality of Housing in Geneva. *Urban Studies*, 34(3):503–513.
- Besim, M. and Jenkins, G. P. (2005). Tax Compliance: When do Employees Behave like the Self-Employed? *Applied Economics*, 37(10):1201–1208.
- Bird, R. (1984). Put up or Shut up: Self-Assessment and Asymmetric Information. *Journal of Policy Analysis and Management*, 3(4):618–620.
- Black, J. L., Macinko, J., Dixon, L. B., Fryer, G. E., and Jr. (2010). Neighborhoods and Obesity in New York City. *Health & Place*, 16(3):489 – 499.
- Bolitzer, B. and Netusil, N. (2000). The Impact of Open Spaces on Property Values in Portland, Oregon. *Journal of Environmental Management*, 59(3):185–193.
- Boyle, G. (1982). A Glimpse at the Non-Accounted Economy. *Central Bank of Ireland Technical Paper*, March.

- Boyle, G. (1984). In Search of Ireland's Black Economy. *Irish Banking Review*, March(1984):32–43.
- Bringolf-Isler, B., Grize, L., Mder, U., Ruch, N., Sennhauser, F. H., and Braun-Fahrlander, C. (2010). Built Environment, Parents' Perception, and Children's Vigorous Outdoor Play. *Preventive Medicine*, 50(5-6):251 – 256.
- Burke, R. (2009). The Black Economy is Back in Business. *Sunday Independent*, Sunday October 18th 2009.
- Callan, T., Keane, C., and Walsh, J. (2009a). What Role for Property Taxes in Ireland? Working Paper Series 322, Economic and Social Research Institute, Dublin, Ireland.
- Callan, T., Lyons, S., Scott, S., Tol, R. S., and Verde, S. (2009b). The Distributional Implications of a Carbon Tax in Ireland. *Energy Policy*, 37(2):407–412.
- Chapman, J. I., Johnston, R. J., and Tyrrell, T. J. (2009). Implications of a Land Value Tax with Error in Assessed Values. *Land Economics*, 85(4):576–586.
- Cho, S., Poudyal, N., and Roberts, R. (2008). Spatial Analysis of the Amenity Value of Green Open Space. *Ecological Economics*, 66:403–416.
- Christensen, J. (2010). How Low Can You Go? Exploring the Evolution of Personal Income Tax Policy and the Welfare State in Ireland. *NordWel*, August.
- Cole, T., Bellizzi, M., Flegal, K., and Dietz, W. (2000). Establishing a Standard Definition for Child Overweight and Obesity Worldwide: International Survey. *British Medical Journal*, 320:1–6.
- Coombes, E., Jones, A., and Hillsdon, M. (2010). The Relationship of Physical Activity and Overweight to Objectively Measured Green Space Accessibility and Use. *Social Science and Medicine*, 70:816–822.
- CoT (2009). Report 2009. Technical report, Commission on Taxation, Dublin.
- Crompton, John, L. (2005). The Impact of Parks on Property Values: Empirical Evidence from the Past Two Decades in the United States. *Managing Leisure*, 10(October):203–218.
- CSO (2007). Household Budget Survey 2004-2005 Final Results. Technical report, Central Statistics Office, Cork.
- CSO (2008). Construction and Housing in Ireland 2008 Edition. Technical report, Central Statistics Office, Dublin.

- CSO (2009). Household Budget Survey - Household Questionnaire. Technical report, Central Statistics Office, Cork.
- De Vries, S., Bakker, I., van Mechelen, W., and Hopman-Rock, M. (2007). Determinants of Activity-Friendly Neighbourhoods for Children: Results from the SPACE Study. *American Journal of Health Promotion*, 21(4):312–316.
- Dell'Anno, R. (2003). Estimating the Shadow Economy in Italy: a Structural Equation Approach. Technical report, Denmark.
- Dell'Anno, R. (2007). The Shadow Economy in Portugal: An Analysis with the MIMIC Approach. *Journal of Applied Economics*, 10(2):253–277.
- Dell'Anno, R., Gomez-Antonio, M., and Pardo, A. (2007). The Shadow Economy in Three Mediterranean Countries: France, Spain and Greece. A MIMIC Approach. *Empirical Economics*, 33(1):197–197.
- Des Rosiers, F., Thriault, M., Kestens, Y., and Villeneuve, P. (2002). Landscaping and House Values: An Empirical Investigation. *Journal of Real Estate Research*, 23(1-2):139–161.
- Elo, I. T., Mykyta, L., Margolis, R., and Culhane, J. F. (2009). Perceptions of Neighborhood Disorder: The Role of Individual and Neighborhood Characteristics. *Social Science Quarterly (Blackwell Publishing Limited)*, 90(5):1298 – 1320.
- Engstrom, P. and Holmlund, B. (2006). Tax Evasion and Self-Employment in a High-Tax Country: Evidence from Sweden. *CESifo Working Paper Category 1: Public Finance*, No. 1736.
- Espey, M. and Owusu-Edusai, K. (2001). Neighborhood Parks and Residential Property Values in Greenville, South Carolina. *Journal of Agricultural and Applied Economics*, 33(3 (December)):487–492.
- Fagan, G. P. (1993). Measuring the Size of Ireland's Black Economy. *Journal of the Statistical and Social Inquiry Society of Ireland*, XXVII(Part I):–.
- Fagan, G. P. (1997). The Black Economy in Ireland. *Irish Banking Review*, Summer:pp 19–30.
- Feinstein, J. S. (1999). Approaches for Estimating Noncompliance: Examples from Federal Taxation in the United States. *The Economic Journal*, 109(456):F360–F369.
- Feldman, N. E. and Slemrod, J. (2007). Estimating Tax Noncompliance with Evidence from Unaudited Tax Returns. *Economic Journal*, 117(518):327–352.

- Frey, B. and Weck-Hannemann, H. (1984). The Hidden Economy as an Unobservable Variable. *European Economic Review*, 26:33–53.
- Friedman, M. (1962). *Capitalism and Freedom*. University of Chicago Press, Chicago, Illinois.
- Gadea, M. D. and Serrano-Sanz, J. M. (2002). The Hidden Economy in Spain - A Monetary Estimation, 1964-1998. *Empirical Economics*, 27(3):499–527.
- Geoghegan, J. (2002). The Value of Open Spaces in Residential Land Use. *Land Use Policy*, 19(1):91–98.
- Gloudemans, R. H. Z. J. and Montgomery, E. J. (2008). Domestic Revaluation in Northern Ireland. *Journal of Property Tax Assessment & Administration*, 5(4):17–31.
- Goolsby, W. C. (1997). Assessment Error in the Valuation of Owner-Occupied Housing. *Journal of Real Estate Research*, 13(1):33–.
- Gorodnichenko, Y., Martinez-Vazquez, J., and Sabirianova Peter, K. (2009). Myth and Reality of Flat Tax Reform: Micro Estimates of Tax Evasion Response and Welfare Effects in Russia. *Journal of Political Economy*, 117(3):504–554.
- Gorodnichenko, Y. and Sabirianova Peter, K. (2007). Public Sector Pay and Corruption: Measuring Bribery from Micro Data. *Journal of Public Economics*, 91(5-6):963–991.
- Griliches, Z. (1961). Hedonic Price Indices for Automobiles. In Griliches, Z., editor, *Price Indices and Quality Change: Studies in New Methods of Measurement*, pages –. Harvard University Press, Cambridge.
- Gstach, D. (2009). A Property Taxation Mechanism with Self-Assessment. *Metroeconomica*, 60(3):400–408.
- GUI (2010). Growing Up in Ireland: Key Findings: 9 Year-Olds.
- Hand, M. S., Thacher, J. A., McCollum, D. W., and Berrens, R. P. (2008). Intra-Regional Amenities, Wages, and Home Prices: The Role of Forests in the Southwest. *Land Economics*, 84(4):635–651.
- Hartig, T. and Fransson, U. (2009). Leisure Home Ownership, Access to Nature, and Health: A Longitudinal Study of Urban Residents in Sweden. *Environment and Planning A*, 41(1):82–96.
- Hill, R. and Kabir, M. (2000). Currency Demand and the Growth of the Underground Economy in Canada, 1991-1995. *Applied Economics*, 32(2):183–192.

- Hobden, D. W., Laughton, G. E., and Morgan, K. E. (2004). Green Space Borders—a Tangible Benefit? Evidence from Four Neighbourhoods in Surrey, British Columbia, 1980-2001. *Land Use Policy*, 21(2):129–138.
- Huang, S.-J., Wen-chi, H., Patricia, A. S., and Jackson, P. W. (2010). Neighborhood Environment and Physical Activity Among Urban and Rural Schoolchildren in Taiwan. *Health & Place*, 16(3):470 – 476.
- IAAO (1978). Improving Real Property Assessment: A Reference Manual. Technical report, International Association of Assessing Officers.
- IAAO (2003). Standard on Automated Valuation Models (AVMs). Technical report, International Association of Assessing Officers, Missouri.
- IAAO (2010). Standard on Ratio Studies. Technical report, International Association of Assessing Officers, Missouri.
- Isachsen, A. J. and Strom, S. (1985). The Size and Growth of the Hidden Economy in Norway. *Review of Income & Wealth*, 31(1):21–38.
- Johansson, E. (2000). An Expenditure-Based Estimation of Self-Employment Income Under-Reporting in Finland. Technical report, Helsinki, Finland.
- Kim, B., Gibson, J., and Chung, C. (2008). Using Panel Data to Exactly Estimate Under-Reporting by the Self-Employed. Technical report, Hamilton, New Zealand.
- Kimbrow, R., Brooks-Gunn, J., and McLanahan, S. (2010). Neighbourhood Context, Poverty, and Urban Childrens Outdoor Play. *Fragile Families Working Paper*, WP10-04-FF.
- Lacko, M. (1999). Electricity Intensity and the Unrecorded Economy in Post-Socialist Countries. In Feige, E. L. and Ott, K., editors, *Underground Economies in Transition*, number 8, pages –. Ashgate, Aldershot.
- Lang, O., Nohrbass, K.-H., and Stahl, K. (1997). On Income Tax Avoidance: the Case of Germany. *Journal of Public Economics*, 66:327–347.
- Layte, R., Harrington, J., Sexton, E., Perry, I and, C. J., and Lyons, S. (2011). Irish Exceptionalism? Local Food Environments and Dietary Quality. *Journal of Epidemiology and Community Health*.
- Leigh, N. and Coffin, S. (2005). Modeling the Relationship among Brownfields, Property Values, and Community Revitalization. *Housing Policy Debate*, 16(2):–.

- Li, F., Fisher, K. J., Brownson, R. C., and Bosworth, M. (2005). Multilevel Modelling of Built Environment Characteristics Related to Neighbourhood Walking Activity in Older Adults. *Journal of Epidemiology and Community Health* (1979-), 59(7):pp. 558–564.
- Lin, T. C. (2010). Property Tax Inequity Resulting from Inaccurate Assessment—The Taiwan Experience. *Land Use Policy*, 27(2):511–517.
- Lissau, I., Overpeck, M. D., Ruan, W. J., Due, P., Holstein, B. E., and Hediger, M. L. (2004). Body Mass Index and Overweight in Adolescents in 13 European Countries, Israel, and the United States. *Archives of Pediatrics and Adolescent Medicine*, 158(1):27–33.
- Luttik, J. (2000). The Value of Trees, Water and Open Space as Reflected by House Prices in the Netherlands. *Landscape and Urban Planning*, 48(3-4):161–167.
- Lyssiotou, P., Pashardes, P., and Stengos, T. (2004). Estimates of the Black Economy based on Consumer Demand Approaches. *Economic Journal*, 114(497):622–640.
- Mansfield, C., Pattanayak, S. K., McDow, W., McDonald, R., and Halpin, P. (2005). Shades of Green: Measuring the Value of Urban Forests in the Housing Market. *Journal of Forest Economics*, 11(3):177–199.
- McCluskey, W. J., Davis, P., and Lay, C. L. (2007). Residential Property Tax Reform in Northern Ireland: Impact Analysis and Spatial Redistribution. *Journal of Property Tax Assessment & Administration*, 4(3):59–70.
- McConnell, V. and Walls, M. (2005). The Value of Open Space: Evidence from Studies of Non-Market Benefits. Technical report, Resources for the Future, Washington.
- McMillen, D. P. and Weber, R. N. (2008). Thin Markets and Property Tax Inequities: A Multinomial Logit Approach. *National Tax Journal*, 61(4):653–671.
- Mirus, R. and R., S. (1996). Self-Employment, Tax Evasion and the Underground Economy: Micro Based Estimates for Canada. Technical Report Working Paper 1002, International Tax Program, Harvard Law School.
- Mitchell, R. and Popham, F. (2007). Evidence Based Public Health Policy and Practice: Greenspace, Urbanity and Health: Relationships in England. *Journal of Epidemiology and Community Health* (1979-), 61(8):pp. 681–683.
- Niou, E. M. and Tan, G. (1994). An Analysis of Dr. Sun Yat-Sen's Self-Assessment Scheme for Land Laxation. *Public Choice*, 78(1):103–114.

- Nunnally, J. and Bernstein, I. (1994). *Psychometric theory*. McGraw Hill, New York, 3 edition.
- OECD (2010). Round Table on Competition Policy and the Informal Economy. In *Global Forum on Competition*, volume BAF/COMP/GF(2009)10.
- Paglin, M. and Fogarty, M. (1972). Equity and the Property Tax: A New Conceptual Focus. *National Tax Journal*, 25:557–565.
- Pahl, R. (1988). The Back Economy in Ireland. *Final Report - European Commission Programme of Research on the Black Economy in Europe*, No.87-931.
- Parry, K. (2006). Rating in Northern Ireland. Standard Notes SN/PC/3648, House of Commons Library, Parliament and Constitution Centre.
- Petty, J., Peacock, J., Sellens, M., and Griffin, M. (2005). The Mental and Physical Outcomes of Green Exercise. *International Journal of Environmental Health Research*, 15:319–337.
- Pickhardt, M. and Sarda Pons, J. (2006). Size and Scope of the Underground Economy in Germany. *Applied Economics*, 38(14):1707–1713.
- Pissarides, C. A. and Weber, G. (1989). An Expenditure-Based Estimate of Britain's Black Economy. *Journal of Public Economics*, 39(1):17–32.
- Plassmann, F. and Tideman, T. (2008). Accurate Valuation in the Absence of Markets. *Public Finance Review*, 36:3(May 2008):334–358.
- Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(Jan/Feb):34–55.
- Schneider, F. (2000). The Growth of the Shadow Economy in the OECD: Some Preliminary Explanations. *Journal of International Affairs*, 53(2):413–.
- Schneider, F. (2007). Shadow Economies and Corruption All Over the World: New Estimates for 145 Countries. *Economics: The Open-Access, Open-Assessment E-Journal*, 1(2007-2009):–.
- Schuetze, H. J. (2002). Profiles of Tax Non-Compliance among the Self-Employed in Canada: 1969 to 1992. *Canadian Public Policy / Analyse de Politiques*, 28(2):219–238.
- Smith, B. C. (2000). Applying Models for Vertical Inequity in the Property Tax to a Non-Market Value State. *Journal of Real Estate Research*, 19(3):321–.

- Smith, D. (2010). Valuing Housing and Green Spaces: Understanding Local Amenities, the Built Environment and House Prices in London. *GLA Economics Working Papers*, September(42).
- Smith, Roger, S. (2002). The Underground Economy: Guidance for Policy Makers?
- Spahr, R. W. and Sunderman, M. A. (1998). Property Tax Inequities on Ranch and Farm Properties. *Land Economics*, 74(3):374–389.
- Strife, S. and Downey, L. (2009). Childhood development and access to nature. *Organization and Environment*, Volume 22(Number 1).
- Szalai, A. and Tassonyi, A. (2004). Value-Based Property Taxation: Options for Hungary. *Environment and Planning C: Government and Policy*, 22(4):495–521.
- Tanzi, V. (1999). Uses and Abuses of Estimates of the Underground Economy. *The Economic Journal*, 109(456):F338–F347.
- Tyrvaainen, L. and Miettinen, A. (2000). Property Prices and Urban Forest Amenities. *Journal of Environmental Economics and Management*, 39(2):205–223.
- Voicu, I. and Been, V. (2008). The Effect of Community Gardens on Neighboring Property Values. *Real Estate Economics*, 36(2):241–283.
- Vyas, S. and Kumaranayake, L. (2006). Constructing Socio-Economic Status Indices: How to Use Principal Components Analysis. *Health Policy and Planning*, 21:6:459–468.
- Wachter, S. M. and Wong, G. (2008). What Is a Tree Worth? Green-City Strategies, Signaling and Housing Prices. *Real Estate Economics*, 36(2):213–239.
- Watson, D. and Williams, J. (2003). Irish National Survey of Housing Quality. Technical report, Economic and Social Research Institute, Dublin.
- Weicher, J. C. and Zerbst, R. H. (1973). The Externalities of Neighborhood Parks: An Empirical Investigation. *Land Economics*, 49(1):99–.
- Wolinsky, A. (2000). A Market Based Approach to Property Tax. Technical report, Northwestern University.
- Zellner, A. (1970). Estimation of Regression Relationships Containing Unobservable Variables. *International Economic Review*, 11:441–454.