# Essays in Applied Microeconomics

## Smart Networks and the Behaviour of Households and Firms

A THESIS SUBMITTED TO THE UNIVERSITY OF DUBLIN, TRINITY COLLEGE
IN APPLICATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
BY

Daire McCoy

Supervised by
Seán Lyons (ESRI)
Eleanor Denny (TCD)

2016

# Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

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.

Daire McCoy

# Summary

This dissertation explores a number of aspects related to the roll-out of network infrastructure. The focus is on the integration of electricity networks and ICT, and the interaction of these technologies with the behaviour of households and firms. This work is empirical in nature and employs simulation and econometric modelling techniques, in order to examine a broad range of questions related to this. There is a particular focus on the spatial distribution of effects, given the heterogeneity that exists in both network infrastructure provision, and household and firm locations across geographies.

Chapter 1 provides the introduction, along with a general background to this research, and then outlines the specific research questions that are explored in this dissertation.

Chapter 2 examines in detail the role in which improved information on electricity usage through ICT can have on household investment decisions related to energy efficiency. This paper uses data from a nationally representative smart-metering trial, in which smart electric meters were placed in Irish dwellings in an effort to examine the potential for improved information and time-of-use pricing to encourage reduction of electricity usage, and time-shifting of demand. This research exploits before and after survey questions related to the stock and flow of energy efficiency technologies within the home to demonstrate a case in which improved information can help reduce consumption, but can also negatively impact the adoption of energy efficient technologies. Binary and count regression models are employed in this analysis. A paper based on this research and co-authored with my supervisor, Dr. Seán Lyons has recently been accepted for publication in the journal *Energy Efficiency*.

Chapter 3 examines the potential of future electric vehicle adoption patterns to concentrate excess demand at certain nodes of electricity distribution networks. This analysis is applied to Ireland and demonstrates also how individual level survey data, from a smart-metering trial, can be combined with spatial census aggregate data in an agent-based model to simulate technology adoption profiles for geographic areas within which diverse ranges of households reside. Results show the overall diffusion level in the population may be determined by the order of adoption. We also find that significant clusters may form in certain geographic areas, even if overall adoption levels remain low.

Ultimately, this work emphasises the uneven spatial nature of technology adoption and the potential impact of this on electricity distribution networks. A paper based on this chapter, co-authored with Dr. Seán Lyons has appeared in the journal *Energy Research and Social Science*, Volume 3 (July 2014).

Chapter 4 examines the impact of broadband infrastructure, electricity networks, other infrastructure including motorways, airports and railways and a range of other factors such as human capital and third level institutes on new business establishments. This chapter is a slight departure in that the emphasis shifts from households to firms; however, there is a degree of symmetry between this work and Chapter 3. While Chapter 3 examined the spatial location of households and the impact this may have on electricity networks, this chapter looks at the spatial configuration of network infrastructure and how this affects the location of firms.

The results show that increased numbers of foreign firms and high-tech firms emerge in areas where broadband is available and they also prefer better broadband. Importantly, human capital and proximity to a third level institution are also important determinants of new firm establishments, including low-tech indigenous firms. Complementarities also exist between human capital and ICT, in that an area's ability to attract new firms with improved ICT depends on its level of human capital. The overall aim of this chapter is to examine how the provision of infrastructure and other factors affect the spatial distribution of economic activity. This is relevant in cases where publicly funded infrastructure projects are initiated in order to disperse economic activity in a more geographically balanced fashion. A paper based on this chapter, co-authored with Dr. Seán Lyons and Dr. Edgar Morgenroth of the ESRI, Dr. Dónal Palcic of the University of Limerick and Ms Leonie Allen of the Commission for Communications Regulation (ComReg), has been submitted to the *Journal of Economic Geography*.

Chapter 5 concludes the dissertation and links the research back to the overall theme of the roll-out of ICT and its integration with electricity networks. Limitations of the research are highlighted and potentially fruitful avenues of future research are discussed.

# Acknowledgements

This research was funded by Science Foundation Ireland (SFI) under grant No. 09/SRC/E1780.

I am indebted to my supervisors Dr. Seán Lyons of the Economic and Social Research Institute (ESRI) and Dr. Eleanor Denny of Trinity College Dublin (TCD) for their guidance and support throughout this dissertation.

I am particularly grateful to Seán, my primary supervisor. His keen insight, breadth of knowledge on an enormous range of topics, patience and good humour have been invaluable to me over the past four years. I count myself extremely lucky to have had him as a mentor and I have learned so much from him.

I am very fortunate to have been based based at the Economic and Social Research Institute for the duration of my dissertation, while also enrolled in Trinity College Dublin. This enabled me to benefit from the knowledge and expertise of colleagues in both institutions.

I would like to thank members of the Economics Department at Trinity College Dublin; in particular the participants at the TCD Micro Working Group for providing excellent comments and feedback on all chapters of this work; and Colette Ding in the Postgraduate Office for being enormously helpful throughout my time at Trinity.

I would also like to thank all staff and colleagues at the ESRI for making my time there such an enjoyable experience. In particular I would like to mention the Energy and Environment Group, and the IT Department - who made the library a lot more fun to work in. Dr. Edgar Morgenroth deserves a special mention for being a fantastic source of both wisdom and practical know-how in the latter stages of my research. Chapter 4 benefited hugely from his knowledge of economic geography and his hoards of wonderful data. The ESRI lunchtime seminar series participants provided useful comments and suggestions on early drafts of Chapters 2 and 4.

Chapter 2 benefitted from comments and discussions with Dr. Laura Malaguzzi Valeri, Dr. Claudia Aravena Novielli, Dr. Geertje Schuitema and Dr. John Curtis. Chapter 3 also benefitted from comments from Dr. Laura Malaguzzi Valeri. Dr. Valeria Di Cosmo provided helpful advice on both the Introduction and Conclusion of this dissertation.

Data used in this work comes from a range of sources. I would like to thank the Central Statistics Office, the Irish Social Science Data Archive, The Department of Jobs, Enterprise and Innovation, enet, Eircom, BT, and ESB Networks.

I owe an enormous debt of gratitude to my parents and family, were it not for their encouragement and support, I would never have been in a position to undertake a PhD in the first place.

Finally, I would like to thank Saoirse for all her love, support and unwavering belief in me.

Daire McCoy
April 2016

# List of Publications Arising from this Thesis

## Journal Publications:

McCoy, D., and Lyons, S. (2014). "Consumer preferences and the influence of networks in electric vehicle diffusion: An agent-based microsimulation in Ireland", *Energy Research and Social Science*, 3, 89-101.

McCoy, D., and Lyons, S. (2016) "Unintended consequences of electricity smart-metering: Trading-off consumption and investment behaviour", *Energy Efficiency*, Forthcoming.

## Journal Articles in Review:

McCoy, D., Lyons, S., Morgenroth E., Palcic, D., Allen, L. "The impact of local infrastructure on new business establishments", submitted to *The Journal of Economic Geography.*

## Conference Papers:

McCoy, D., and Lyons, S. (2014). "Consumer preferences and the influence of networks in electric vehicle diffusion: An agent-based microsimulation in Ireland"

Presented at:

- The Irish Society of New Economists (ISNE) Conference: Maynooth, September 2013
- Young Energy Economists and Engineers Seminar (YEEES): Stockholm, October 2013
- Mannheim Energy Conference (ZEW): Mannheim, May 2014

McCoy, D., and Lyons, S. "Unintended consequences of electricity smart-metering: Trading-off consumption and investment behaviour"

Presented at:

- ESRI/UCC Energy Modelling Workshop: Dublin, May 2014

McCoy, D., Lyons, S., Morgenroth, E., Palcic, D. and Allen L. "The impact of broadband and other infrastructure on new business establishments"

Presented at:

- New Directions for Regional Development Policy in Ireland, Regional Studies Association, Irish Branch Annual Conference: UCC, Cork, September 2015
- Great Transformation: Recasting Regional Policy, Regional Studies Association Winter Conference 2015: London, November 2015. (Presented by Edgar Morgenroth)

# Contents


**1 Introduction** . . . **1**

1.1 Integrating ICT with electricity networks . . . 2

1.2 Thesis motivation and chapter previews . . . 3

1.2.1 Chapter 2: Unintended consequences of electricity smart-metering . . . 3

1.2.2 Chapter 3: Consumer preferences and the influence of networks in electric vehicle diffusion . . . 4

1.2.3 Chapter 4: The impact of broadband and other infrastructure on new business establishments . . . 5

1.2.4 Concluding chapter . . . 7

**2 Unintended consequences of electricity smart-metering** . . . **8**

2.1 Introduction . . . 8

2.2 Related research . . . 9

2.3 Methods . . . 11

2.3.1 Description of data . . . 11

2.3.2 Experimental design . . . 12

2.3.3 Pre-trial stock of energy saving measures . . . 13

2.3.4 In-trial investment in energy efficiency . . . 13

2.3.5 Empirical strategy . . . 14

2.4 Results and discussion . . . 16

2.4.1 Equality of means . . . 16

2.4.2 Regression analysis . . . 17

2.4.3 Discussion . . . 19

2.5 Concluding remarks . . . 21

**Appendices** . . . **21**

2.A Detailed descriptive statistics . . . 22

**3 Consumer preferences and the influence of networks in electric vehicle diffusion** **25**

3.1 Introduction . . . 25

3.2 Agent-based models . . . 26

3.2.1 Agent-based adoption models . . . 27

3.3 Materials and methods . . . 29

3.3.1 Basic principles . . . 29

3.3.2 Entities, state variables and scales . . . 30

3.3.3 Collectives . . . 36

3.3.4 Process overview and scheduling . . . 38

3.3.5 The model . . . 38
3.4 Simulations and results . . . 43
3.4.1 Individual preferences versus peer effects . . . 43
3.4.2 The effect of seeding and network type on adoption level . . . 44
3.4.3 EV adoption for specific neighbourhoods . . . 47
3.5 Concluding remarks . . . 54

**Appendices . . . 55**
3.A Data . . . 56
3.B Results from Ward's clustering algorithm . . . 57
3.C Individual preferences versus peer effects . . . 58
3.D Goodness-of-fit for synthetic agent populations . . . 59
3.E Sensitivity analysis . . . 61
A Threshold distribution . . . 61
B Initial uncertainty . . . 62

**4 The impact of broadband and other infrastructure on new business establishments . . . 63**
4.1 Introduction . . . 63
4.2 Previous literature . . . 65
4.2.1 Broadband infrastructure . . . 65
4.2.2 Agglomeration economies, the labour market and human capital . . . 67
4.2.3 Transport infrastructure and general accessibility . . . 68
4.2.4 Firm location in an Irish context . . . 69
4.3 Empirical approach and econometric considerations . . . 69
4.3.1 Excess zeros . . . 72
4.3.2 Spatial dependence . . . 73
4.3.3 Endogenous explanatory variables . . . 74
4.3.4 Identification strategy . . . 75
4.4 Data . . . 76
4.4.1 Unit of analysis: "Urban Fields" . . . 77
4.4.2 Firm location variables . . . 80
4.4.3 Explanatory variables . . . 83
4.5 Econometric results . . . 91
4.5.1 New business establishments by location of owner . . . 92
4.5.2 New business establishments by skill-level . . . 95
4.5.3 New business establishments by skill-level and location of owner . . . 96
4.5.4 Exploring some interactions . . . 97
4.6 Robustness . . . 99
4.7 Discussion and concluding remarks . . . 102

**Appendices . . . 104**
4.A Incorporating the electricity network . . . 105
A Developing a proxy for network constraints . . . 105
B Electricity results . . . 109
4.B Discussion of spatial dependence in modelling firm counts . . . 110
4.C Detailed firm activity description . . . 112

4.D Descriptive statistics for variables in regressions . . . 114
4.E Robustness . . . 116
4.F Variables and their sources . . . 119

**5 Conclusion . . . 120**

**Bibliography . . . 124**

# List of Figures


2.1 Investments distribution . . . . . . . . . . 15
2.2 Investments boxplot . . . . . . . . . . 17

3.1 Agent environment . . . . . . . . . . 30
3.2 Distribution of preferences for full agent population . . . . . . . . . . 31
3.3 Threshold distributions for $N = 1600$ agents . . . . . . . . . . 35
3.4 Social network diagrams . . . . . . . . . . 37
3.5 Flow-chart of updates for agent$_i$ . . . . . . . . . . 39
3.6 Degree distribution for preferential attachment network of $N = 10$ nodes . . . . . . . 41
3.7 EV adoption levels for alternate seeding strategies on random agent sample . . . . . 46
3.8 Spatial distribution of likely early adopters for Census Small-Areas in Dublin . . . . 48
3.9 Distribution of Income Utility (IU) and Environmental Utility (EU) for different EDs 50
3.10 EV adoption levels for agent samples based on specific areas . . . . . . . . . . 51
3.11 Distribution of cluster size for small-world networks . . . . . . . . . . 53
D1 Actual vs synthetic populations: Social class . . . . . . . . . . 59
D2 Actual vs synthetic populations: Tenure type . . . . . . . . . . 60
E1 The effect of threshold distribution on adoption level . . . . . . . . . . 61
E2 The effect of initial inertia on adoption level . . . . . . . . . . 62

4.1 Map of Urban Fields and Metropolitan Area Networks for South-East of Ireland . . 76
4.2 Map of Urban Fields and Metropolitan Area Networks . . . . . . . . . . 79
4.3 Map of East Coast Urban Fields . . . . . . . . . . 81
4.4 Location of new firms 2002-2011 . . . . . . . . . . 83
4.5 Share of enabled exchanges for DSL over time in Urban Fields by NUTS3 regions . . 86
4.6 International comparison of motorway extension, 1970-2013 . . . . . . . . . . 88
4.7 Location of motorway extensions, Ireland 2002-2011 . . . . . . . . . . 89
4.8 Expected firm counts in the presence of MAN at different levels of educational attainment . . . . . . . . . . 98
4.9 Distribution of educational attainment over time . . . . . . . . . . 100
4.10 Proportion of 2011 population with third-level education and MAN locations . . . . 101
A1 Map of the electricity transmission system . . . . . . . . . . 106
A2 Map of electricity demand categories by Urban Field for the south of Ireland . . . . 108

# List of Tables


2.1 Time-of-use tariffs (cents per kWh excluding VAT) . . . . 12
2.2 Treatment matrix . . . . 13
2.3 Pre-trial stock of energy saving measures . . . . 13
2.4 In-trial adoption of energy saving measures . . . . 14
2.5 Difference in mean in-trial adoption between control and treatment groups . . . . 17
2.6 Effect of treatment on investment in energy efficiency . . . . 18
A1 Pre-trial distribution of socioeconomic characteristics . . . . 22
A2 Pre-trial distribution of house characteristics . . . . 23
A3 Pre-trial distribution of appliance stock and heating type . . . . 24

3.1 Survey answers for Income Utility (IU) . . . . 32
3.2 Survey answers for Environmental Utility (EU) - behavioural . . . . 33
3.3 Survey answers for Environmental Utility (EU) - attitudinal . . . . 34
3.4 Parameter weightings for agent preferences . . . . 42
3.5 Personal preferences vs peer effects . . . . 43
3.6 EV uptake: Preferential-attachment network . . . . 45
3.7 EV uptake: Small-world network . . . . 45
3.8 Neighbourhoods included in simulations . . . . 49
3.9 EV uptake for different neighbourhoods: Preferential-attachment network . . . . 52
3.10 EV uptake for different neighbourhoods: Small-world network . . . . 52
3.11 Clustering results . . . . 53
A1 Census SAPS themes . . . . 56
B1 Clusters generated for Dublin small-areas . . . . 57
C1 Adoption levels for different weighting combinations . . . . 58

4.1 Irish Census geographical areas . . . . 77
4.2 New firms by sector . . . . 82
4.3 New firms by year of entry . . . . 82
4.4 Count of new establishments at Urban Field level 2002-2011, by location of ownership 93
4.5 Count of new establishments at Urban Field level 2002-2011, by skill-level . . . . 95
4.6 Count of new establishments at Urban Field level 2002-2011, by location of ownership and skill-level . . . . 96
4.7 Spearman rank correlations of educational attainment over time . . . . 99
A1 Ratio of residential to business electricity usage . . . . 107
A2 Count of new establishments at Urban Field level 2002-2011. Impact of electricity variables . . . . 109
A3 Count of new establishments at Urban Field level 2002-2011. Impact of electricity variables . . . . 110

C1 Foreign firms: Most frequent NACE 4 digit categories by sector . . . . . . . . . . . . 112
C2 Domestic firms: Most frequent NACE 4 digit categories by sector . . . . . . . . . . . 113
D1 Dependent variables: Firm counts . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 114
D2 Explanatory variables: Main . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 114
D3 Explanatory variables: Agglomeration . . . . . . . . . . . . . . . . . . . . . . . . . . . 115
E1 Comparison of high-tech firm counts from alternative panel model specifications . . 116
E2 Comparison of high-tech firm counts from alternative model specifications . . . . . . 117
E3 Comparison of high-tech firm counts from alternative zero-inflated model specifications 118
F1 Variables and their sources . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 119

# Chapter 1

# Introduction

The electricity network uses power engineering and design concepts dating back over 100 years. Due to mechanical issues, lack of transparency and control, rising populations and increasing adoption of technology, the traditional electricity grid is not fit for the purposes of the 21st century (Gungor et al., 2011). Additionally, moves to de-carbonise the system by integrating increased penetration of intermittent renewable energy sources and distributed generation will place further pressure on transmission and distribution networks. However, in recent years, information and communications technology (ICT) has begun to bring enormous change to the way in which electricity systems are operated, along with many other areas of our economy and society.

Household and firm behaviour is at the heart of this change, both driving it and responding to it. Technology adoption and increased demand for services creates a need for flexibility and control, requiring the integration of all levels of grid infrastructure with ICT. Improved information resulting from this integration is in turn changing the way generators, operators and end-users interact with electricity systems. As both ICT and electricity infrastructure roll-out is uneven, both in time and across geographies, this heterogeneity can also impact the behaviour of economic agents spatially.

Electricity has unique economic properties, in that demand and supply must be matched instantaneously, and providing the optimal level of infrastructure at all points in the network can be a difficult balance. Excess demand on a system can result in increased congestion, increased cost, or blackouts in extreme cases. On the other hand, spare capacity, or excess supply, can be inefficient and result in under-utilisation of network assets. The average costs resulting from both excess demand and supply will generally be socialised across users.

In a submission to the LSE Growth Commission, Newbery (2012) argues that if suboptimal infrastructure investment constrains other investment, it can constrain growth, while excessive investment has no added value. To that extent, examining externalities from behaviour with the potential to create excess demand, or cases in which excessive infrastructure roll-out may not increase productivity, is useful as an economic exercise and may also have important implications for policy.

This thesis explores questions related to the changing nature of network infrastructure and how complementarities between infrastructures interplay with economic activity. Specifically, it examines ICT roll-out, its integration with electricity networks, and how this interacts with the behaviour of households and firms, sometimes in unforseen ways. Factors related to technology adoption, the impact of improved information on behaviour, and the way in which the geographic distribution of economic agents interacts with the spatial configuration of networks are examined.

## 1.1 Integrating ICT with electricity networks

The emerging "smart grid" can broadly be described as the integration of ICT with traditional power systems engineering to produce a more flexible and responsive system. Given the size of and sunk costs relating to current grid assets, it is unlikely that this transition will be revolutionary. It is more likely to evolve through the introduction of distributed control and monitoring alongside the existing grid. In the words of Farhangi (2010), allowing utilities to place a "layer of intelligence over their current and future infrastructure".

Improved ICT should help transmission and distribution system operators (TSOs and DSOs) manage networks better, allowing demand-side-management (DSM) initiatives that improve flexibility, reducing the need for investment in generation. Given the need to balance demand and supply instantaneously, support from distributed storage systems, such as electric vehicle (EV) batteries will be an important element in achieving high penetrations of renewable integration. EVs have the potential to both exacerbate existing peaks in demand, placing pressure on network assets, or relieve pressure on assets through successful aggregation of their storage capacity providing distributed generation.

In the household sector, increasing adoption of appliances may increase demand, while improved efficiency of lighting and heating will help to reduce demand and costs. Smart-electricity meters, smart-devices, and improved heating controls will give households and utilities better information and control over their usage, helping to lower overall and peak demand.

Given these potential benefits, there has been a significant policy push towards widespread roll-out of smart metering in EU member states. Recent EU directives[1] require members to proceed with

[1] See EU Directive 2009/72/EC (European Commission, 2009).

roll-out, covering 80% of consumers in their territory by 2020, provided the cost-benefit analysis (CBA) is positive. Of the twenty member states to have completed CBAs thus far, thirteen have returned positive results, sixteen have decided to proceed with large-scale roll-outs, four have decided not to proceed. Four have neither undertaken a CBA nor have made a decision regarding roll-out (European Commission, 2014).

Ireland is one of the member states with a broadly positive CBA and commitments to roll-out[2]. Currently the Commission for Energy Regulation (CER) in Ireland plan to commence roll-out in 2018.

However sceptical one might be about the likelihood of achieving the 80% target by 2020, we should expect to see increasing integration of ICT with electricity networks in the form of smart meters and smart devices over the coming years.

## 1.2 Thesis motivation and chapter previews

More efficient grid operation may have economic benefits for companies that own generators and grid infrastructure, allowing them to operate their assets in a more cost effective way and potentially passing the benefits on to consumers. Improved information on usage and costs should enable consumers to manage their usage more efficiently and cheaply. However, it is likely that unintended outcomes and negative externalities may also emerge.

The remainder of this chapter focuses on the specific issues this thesis hopes to address, and summarises the following chapters.

### 1.2.1 Chapter 2: Unintended consequences of electricity smart-metering

As discussed above, many countries are implementing electricity smart-metering programmes in order to provide consumers with improved information and feedback on their usage. It is hoped that this will encourage a reduction in consumption, or a move of consumption from peak periods when increasing generation is expensive, to periods when generating electricity is cheaper.

In a review, Faruqui et al. (2010) specifically examine the effect of providing in-house displays (IHD) to consumers in order to encourage reduced consumption. This analysis reviews of a number of North American studies and indicates that in-house displays can induce consumers to reduce

[2] Although under a number of scenarios the results actually returned a negative CBA (CER, 2011).

consumption by 7% on average, and up to 18% in some cases. Relevant to our research, the Commission for Energy Regulation (CER) in Ireland conducted a smart metering trial and found that customers reduced average overall usage by 2.5% and peak usage by 8% (CER, 2011)[3].

Little is known about whether feedback can induce long-term behavioural change. Policies targeting energy reduction have different features; some target efficiency behaviour through the adoption of more energy efficient technologies, others target curtailment behaviour through reducing the usage of existing appliances (Gardner and Stern, 1996; Schuitema and Jakobsson Bergstad, 2012).

Dolan and Galizzi (2015) describes the ripple of behaviour "when a pebble of intervention is thrown in the pond". A limit of many studies is that they capture the immediate, targeted behaviour and not the ripples that may subsequently emerge. We cast some light on this question in Chapter 2.

This chapter examines how smart metering might affect residential energy efficiency investment behaviour. At the time of writing we are unaware of any other research into this question. We use data from a randomised-controlled trial on a final sample of almost 2500 Irish consumers, conducted over a 12 month period to investigate the effect of smart-metering on household investment behaviour. The results show that exposure to time-of-use pricing and information stimuli, while reducing overall and peak usage, can also have the unintended effect of reducing investment in energy efficiency measures within the home.

Our findings indicate that households exposed to treatment were less likely to adopt any energy saving measure (23-28% on average); and being in the treatment group reduced the expected number of energy saving features adopted (15-21% on average). This result highlights the potential for behavioural interventions to have unintended consequences on behaviours other than those specifically targeted. Furthermore it underlines the importance of examining a wider range of outcomes and allowing longer time-scales when evaluating this type of experiment.

### 1.2.2 Chapter 3: Consumer preferences and the influence of networks in electric vehicle diffusion

Electric vehicles (EVs) have been around since the mid-nineteenth century but have only recently been offered as a mass-market alternative to private cars with petrol and diesel engines. They are considered an important element in de-carbonising the transport sector and a number of governments have introduced initiatives to encourage their adoption. While this may have environmental benefits, both in direct emission reductions and in helping achieve $CO_2$ reduction targets in electricity generation (Foley et al., 2013), negative externalities also exist with mass-adoption.

[3]This was through the use of IHDs, other information stimuli and time of use tariffs.

An emerging engineering literature documents the negative effects clustering of electrical load and uncontrolled charging of large numbers of EVs could have on low-voltage distribution networks (Schneider et al., 2008; Shao et al., 2009; Richardson et al., 2010). At the aggregate network level, high penetrations of these vehicles, when charging, may exacerbate existing peaks in electricity demand (Parks et al., 2007). Even if overall adoption levels are very low, this could be concentrated in relatively few areas, due to the heterogeneous spatial distribution of individuals. This effect could be exacerbated if there is spatial dependence in adoption behaviour[4]. While this is an engineering problem, it has economic and social determinants.

To examine the conditions which might give rise to such issues, in Chapter 3 we implement an agent-based, threshold model of innovation diffusion to simulate the adoption of electric vehicles among Irish households. We use detailed survey microdata to develop a nationally representative, heterogeneous agent population. We then calibrate our agent population to reflect the aggregate socioeconomic characteristics of a number of geographic areas of interest. Our data allow us to create agents and assign them socioeconomic characteristics and environmental preferences. Agents are placed within social networks through which the diffusion process propagates.

This model is used to examine a number of questions related to EV diffusion. Firstly we examine how the adoption process is affected by the type of network and by the order of adoption. We find that it is important who adopts first in determining the overall diffusion level in the population. This has implications for policies aiming to promote the adoption of technologies or habits, by targeting so-called "early-adopters" and harnessing social networks. Unless network topology, the conduit through which technology diffuses is better understood, the impact of such policies will be limited.

Applying the analysis to agent populations representative of certain areas in Dublin, we also find that mild local peer-effects can induce significant clustering, in places where there are low financial and attitudinal barriers to adoption. Even if overall adoption in the general population is quite low, this could be highly clustered in certain areas, and peer-effects within those areas could significantly increase cluster size. This could lead to increased costs for electricity network operators and ultimately for consumers, as the average cost of improvements to the network will be socialised.

### 1.2.3 Chapter 4: The impact of broadband and other infrastructure on new business establishments

Successful integration of ICT with electricity networks will require significant provision of communications infrastructure. Chapter 4 is a departure from previous chapters in that the focus moves

[4]This has been shown to occur in the adoption of solar PV panels in California (Bollinger and Gillingham, 2012), High-voltage air conditioning systems in Chicago (Noonan et al., 2013). Also in the adoption of hybrid electric vehicles (Axsen et al., 2009; Axsen and Kurani, 2012).

to the roll-out of ICT infrastructure more generally, and how that effects economic outcomes over time and in different geographies. However, there is a degree of symmetry between this work and Chapter 3. While Chapter 3 examined the geographic location of households and the impact this can have on electricity networks, this chapter looks at the spatial configuration of network infrastructure and the economic impact of this. The impact in question is the effect of infrastructure on new business establishments in an area.

Even as recently as the early 2000s authors were discussing the imminent communications revolution and the substantial benefits it would bring (Parker, 2000). Due to the difficulties in spreading considerable fixed costs among small populations in remote locations, it was considered necessary for policy intervention to expand ICT infrastructure into rural areas where commercial providers had not ventured (Malecki, 2003). Policy-makers were very optimistic about the potential benefits that this could bring (Kandilov and Renkow, 2010).

However, the relationship between broadband and economic growth is nuanced. In macro terms, lowering the cost of transmitting data should improve productivity and raise output, and has been shown to have a positive impact on economic growth, as discussed by Koutroumpis (2009). However, the distributional consequences of this are unclear. If substitution effects dominate, ICT might displace less technically advanced workers and have a negative effect on employment. On the other hand, it may enable higher skill workers to be more productive, complementing employment in these sectors.

We might then expect areas with pre-existing high levels of human capital and knowledge intensive firms better equipped to reap the rewards of ICT roll-out than other areas less endowed with these attributes. Indeed much of the literature finds this to be the case. A number of studies have found that the positive impact of broadband is more pronounced in urban areas and for knowledge-intensive firms (Gillet et al., 2006; Kandilov and Renkow, 2010), and within urban areas the effect can depend on area size and industrial legacy (Mack and Rey, 2014).

Chapter 4 is the first attempt in Ireland to try to bring together maps of broadband infrastructure, electricity networks and other proxies for local infrastructure quality. A unique dataset of infrastructure is created and a model developed which examines how the spatial configuration of this affects the geographic location of new business establishments. This chapter extends previous work in attempting to examine a set of infrastructures that interact with each other in complex ways.

The data for the analysis is a panel of new firms in Ireland for the period 2002 to 2011 covering 192 "Urban Fields" that are defined on the basis of employment densities. The sample period spans the introduction and recent history of broadband in Ireland, and during this period 86% of the current motorway network was constructed. Unusually, our data includes detail on the quality of broadband where broadband is available. As both the roll-out of broadband (in terms of availability and quality) and the construction of the motorway network were not uniform in both time and geography, the data incorporates significant variation over the observed period. The empirical methodology is to model counts of newly established firms using Poisson and Negative-binomial estimators.

The results show that foreign firms and high-tech firms emerge in areas where broadband is available and they also prefer better broadband. Importantly, human capital, measured as the percentage of the population with a third level qualification and proximity to a third level institution are also important determinants of new firm establishments, including low-tech indigenous firms.

We also find that pre-existing levels of human capital appear to be an important indicator of an area's ability to absorb new ICT technologies productively. Previous work such as Akerman et al. (2015) has pointed to a skill complementarity between broadband adoption and skilled labour. Broadband can increase the productivity of skilled graduates, particularly in scientific and technical disciplines, but can act as a substitute for less educated workers, lowering their marginal productivity.

### 1.2.4 Concluding chapter

Chapter 5 presents some concluding remarks and implications for policy. We discuss limitations of this research and avenues for future research.

# Chapter 2

# Unintended consequences of electricity smart-metering

## 2.1 Introduction

Reducing energy consumption and increasing the adoption of energy saving measures and energy efficient appliances are seen as crucial elements in reducing energy demand. A recent EU directive[1], aims to "*remove barriers and overcome market failures that impede efficiency in the supply and use of energy*". The reluctance of users to adopt energy efficient appliances that offer them seemingly positive NPV is known as the "energy efficiency paradox" and has been widely studied[2].

Jaffe and Stavins (1994) provide a discussion of the market barriers to energy efficiency. They denote any barrier that might require a public policy intervention to overcome it as a market failure; these are largely related to imperfect information. They discuss three potential information market failures. First, if improved information is a public good, the market might tend to provide less than the socially optimum level. Second, if information is conveyed by the adopter, and the adopter is not compensated by the market for the positive externality they create by adopting. Third, if the party that possesses the information doesn't benefit from the cost savings, i.e. if they are not the bill payer, a principal-agent problem arises as they have no financial incentive to act on the information.

If a certain proportion of the energy efficiency gap is attributable to an under-provision of information to the relevant parties, it may be possible to remedy this by educating consumers. Allcott

[1] EU Energy Efficiency Directive 2012/27/EU (European Commission, 2012)

[2] See Blumstein et al. (1980), Golove and Eto (1996) and Allcott and Greenstone (2012) for a review

and Mullainathan (2010) cite the growing evidence of the effectiveness of behavioural interventions - such as goal-setting, commitment devices, consumption feedback and social approval, rather than price-based approaches, in changing consumer choices.

A number of years ago in the UK, The Department for Environment, Food and Rural Affairs (DEFRA) commissioned a report entitled “Exploring Catalyst Behaviours” (Brook-Lyndhurst, 2011). The aim of this report was to review the literature on pro-environmental behaviour, in order to determine if policies aimed at encouraging certain types of pro-environmental behaviour “spill-over” into other domains, for instance, if someone is encouraged to recycle, does this make them more likely to reduce electricity consumption?

Reducing energy consumption is an example of “curtailment behaviour”. Replacing household appliances with more energy efficient alternatives is an example of “efficiency behaviour”. This study uses data from a randomised controlled electricity smart-metering trial, based on a nationally representative sample of the Irish population, to provide empirical evidence of an environmental intervention which targets one behaviour, but also induces reduced engagement, or a negative spill-over in another behaviour. The trial targeted (and achieved) a reduction in electricity consumption, but resulted in the perverse side-effect of these households also reducing their investment in energy efficiency during the trial period.

In the following section we discuss other research related to this work, and how our results contribute to this literature. Section 2.3 outlines the experiment, dataset and methodology used, Section 2.4 describes the results and finally in Section 2.5 we explore some potential explanations for our results.

## 2.2 Related research

The provision of feedback in energy efficiency interventions can take a number of forms, summarised by Darby (2001) as direct feedback (e.g. direct displays, smart-meters), indirect feedback (e.g. billing information provided by the utilities) and inadvertent feedback (e.g. indirect observations of energy usage and learning by association). In a review of interventions this author finds that direct feedback has the most promising effect, and can result in energy usage reductions of up to 20%. Generally, direct feedback tends to be most effective when provided in conjunction with other measures, such as pre-pay meters and other information provision.

Abrahamse et al. (2005) provide a comprehensive review of intervention studies aimed at reducing household consumption of energy. They examine both antecedent strategies (e.g. commitment, goal-setting) and consequence strategies (e.g. feedback, rewards) in an analysis of 38 studies evaluating the effect of feedback and improved information on electricity usage. Various forms of information provision such as workshops, mass-media campaigns and tailored home-energy audits were examined. The findings tend to indicate that improved information results in greater knowledge, but this does not necessarily result in behavioural change. Similar findings have also recently been demonstrated in an Irish context (Diffney et al., 2013; Carroll et al., 2014).

Providing feedback specific to the individual is found to be more successful in reducing consumption, and reductions of just over 20% have been achieved in some instances (Midden et al., 1983; Staats et al., 2004). This intervention was found to be even more effective when combined with goal-setting and when provided on a more continuous basis.

A feature of many of the aforementioned interventions seems to be the absence of any evidence of their long-term effectiveness. While a couple of studies have found evidence of long-term effects (Staats et al., 2004; Hirst and Grady, 1983), the vast majority either do not measure, or find no evidence of long-term behavioural change regardless of the type of intervention.

Another feature of much of the above research, is the tendency to examine curtailment behaviour and not efficiency behaviour. While policies targeting efficiency behaviour are generally more expensive for the consumer as they require a greater up-front investment, they also tend to be more acceptable as they involve less effort on an on-going basis (Poortinga et al., 2003). According to Gardner and Stern (1996), efficiency behaviours are also considered to have greater energy-saving potential. For example, replacing an existing energy-intensive boiler with a more-efficient upgrade may reduce energy usage more than simply curtailing usage of the existing boiler. This type of policy intervention also has greater long-term potential as its success does not depend on consumers maintaining a newly developed set of habits over a long period.

There is a body of work within environmental psychology which examines whether pro-environmental behaviour "spills-over" from one domain into another[3]. This research also examines the potential for negative spill-overs. If there are costs associated with increased engagement in one domain, this might reduce engagement in other domains.

This could potentially be due to a "moral licensing" effect and this phenomenon has been observed in many domains of human behaviour, such as political correctness, pro-social behaviour and consumer choice. This theory suggests that individuals who are secure in the knowledge of their past good behaviour, can be more likely to engage in morally bereft actions, freed from the anxiety that normally accompanies these decisions. See Merritt et al. (2010) and Miller and Effron (2010) for a review.

[3]The idea that people strive to be consistent in their beliefs, attitudes and behaviours comes from a range of social-psychological theories rooted in Festinger's theory of cognitive dissonance (Festinger, 1962).

Within the specific domain of pro-environmental behaviour, the empirical evidence for behavioural spill-overs (either positive or negative) is quite limited. A few studies examine this phenomenon (Thogersen and Olander, 2003; Thogersen, 2004; Whitmarsh and O'Neill, 2010). This research finds evidence for positive spill-overs but the magnitude of them is generally quite small. They also find some evidence of negative spill-overs.

Much of the above research on behavioural spill-overs tends to be conducted through lab-based experiments or consumer surveys. Our work is one of the first papers to demonstrate this empirically in a real-world setting. We exploit access to a large smart metering trial[4], and contribute to the literature by examining both curtailment and efficiency behavioural change resulting from a smart-metering intervention.

The results show that feedback and increased information can actually reduce investment in energy efficiency, for a set of consumers who also reduced electricity consumption. The treatment group reduced their overall consumption by 2.5% relative to the control group over a 12 month period, but they were also, on average 23-28% less likely to adopt any energy saving measure during the trial; and being in the treatment group reduced the expected number of energy saving features adopted by 15-21% on average.

As far as we are aware, this trade off between curtailment and efficiency behaviour has not been empirically demonstrated before, however other research has demonstrated perverse side effects between water and electricity usage in energy conservation campaigns (Tiefenbeck et al., 2013).

More generally we demonstrate a case in which a policy targeting one type of behaviour may have unintended consequences on other behaviours within the same domain. This is something that must be considered by policy makers when designing behavioural interventions.

## 2.3 Methods

### 2.3.1 Description of data

We use data from the Irish Commission for Energy Regulation (CER) Smart Metering Customer Behavioural Trial. This is a nationally representative study of households in the Republic of Ireland, containing high frequency energy consumption data along with socioeconomic, attitudinal, behavioural and dwelling data on the participating households. It took place over eighteen months;

[4]CER Electricity Smart Metering Customer Behavioural Trial data.

the benchmark period was from 1st July to 31st December 2009, and the test period was from 1st January to 31st December 2010.

The survey was conducted on Electric Ireland customers, who at that time represented 100% of Irish residential electricity demand. It was designed to quantify the effect of feedback, better information and time-of-use pricing on overall electricity usage and on peak demand, not to examine investments in energy efficiency. However, we can exploit before and after survey questions related to a range of energy efficiency measures adopted within the home both before and during the trial in order to tackle the latter question.

### 2.3.2 Experimental design

Households self-selected into the trial and were then randomly assigned to a control group or various treatment groups[5] Treatments included a range of time-of-use tariffs, as per table 2.1 and information provision. This included providing customers with an electricity usage statement[6] and the use of an in-house display. We will refer to the information treatment groups as Treatment 1, 2 and 3 from this point onwards. Treatment 1 received a bi-monthly bill and energy usage statement; Treatment 2 received a monthly bill and energy usage statement; and Treatment 3 received a bi-monthly bill, energy usage statement and in-house display.

Our starting sample consists of $N = 3488$ households. We drop households who received a financial reward for achieving a reduction target. We also drop households who were on a "weekend" tariff as they did not receive an information feedback stimulus. This leaves us with a sample of $N = 2456$ observations, divided into control and treatment groups as per table 2.2.

Table 2.1: Time-of-use tariffs (cents per kWh excluding VAT)

| Group | Night | Day | Peak |
|---|---|---|---|
| Control | 14.10 | 14.10 | 14.10 |
| Tariff A | 12.00 | 14.00 | 20.00 |
| Tariff B | 11.00 | 13.50 | 26.00 |
| Tariff C | 10.00 | 13.00 | 32.00 |
| Tariff D | 9.00 | 12.50 | 38.00 |

*Source: CER (2011)*

[5]See CER (2011), Di Cosmo et al. (2014) and Carroll et al. (2014) for further information on this trial and related research.

[6]The first page (the bill) was similar to the existing suppliers' bill (with additional lines for time of use tariffs). The second page, the energy usage statement, provided additional detail on usage and tips on energy reduction.

Table 2.2: Treatment matrix

| Group | Control | Treatment 1 | Treatment 2 | Treatment 3 | Total |
|---|---|---|---|---|---|
| Control | 693 | 0 | 0 | 0 | 693 |
| Tariff A | 0 | 199 | 219 | 208 | 626 |
| Tariff B | 0 | 82 | 89 | 67 | 238 |
| Tariff C | 0 | 226 | 220 | 205 | 651 |
| Tariff D | 0 | 81 | 89 | 78 | 248 |
| Total | 693 | 588 | 617 | 558 | 2,456 |

*Source: CER (2011)*

### 2.3.3 Pre-trial stock of energy saving measures

Given our interest in examining how the trial affected participants's investments in energy efficiency, it is important to first examine the pre-trial stock of existing measures installed by households. If systematic differences exist between control and treatment groups, this could lead to post-trial outcomes unrelated to the treatment. From table 2.3 below we can see that these measures are very evenly distributed amongst control and treatment groups.

Table 2.3: Pre-trial stock of energy saving measures

| Proportion of households with: | Control | Treatment 1 | Treatment 2 | Treatment 3 | Mean |
|---|---|---|---|---|---|
| Energy saving lightbulbs | 57% | 57% | 58% | 57% | 57% |
| Double-glazed windows | 91% | 91% | 90% | 92% | 91% |
| Lagging jacket on hot-water tank | 83% | 82% | 85% | 84% | 84% |
| Attic insulation within the last 5 years | 36% | 34% | 35% | 32% | 35% |
| Attic insulation over 5 years ago | 54% | 56% | 55% | 57% | 55% |
| External wall insulation | 56% | 56% | 58% | 58% | 57% |
| Benchmark period (6 months) electricity usage (kWh) | 2048 | 2097 | 2091 | 2074 | 2077 |

*Source: Author's calculations using data from CER (2011)*

Tables A1, A2 and A3 in the Appendix demonstrate there are no systematic differences between control and treatment groups across a whole range of measures that include socioeconomic factors, dwelling characteristics, stock of household appliances and heating type. Di Cosmo et al. (2014) also found no individual or household characteristics to be a significant predictor of being in the control group.

### 2.3.4 In-trial investment in energy efficiency

After the trial was conducted, participants were asked whether they had invested in a range of energy efficiency measures over the previous 12 months. No group received any instructions related to this during the trial. Details are in table 2.4. Unfortunately, we do not know the exact timing of these investments, nor their cost.

Table 2.4: In-trial adoption of energy saving measures

| Energy saving measures installed | Number of Households | Percentage of Households |
|---|---|---|
| Added double glazing to some or all of your windows | 199 | 8% |
| Installed attic or wall insulation | 676 | 28% |
| Replaced appliances with A rated ones | 396 | 16% |
| Fitted a new lagging jacket on your hot water tank | 326 | 13% |
| Fitted other energy saving devices | 206 | 8% |
| Added solar panels | 35 | 1% |
| Added draught-proofing to your doors or windows | 241 | 10% |
| Replaced a central heating boiler with a more efficient one | 164 | 7% |
| Added thermostatic controls to radiators | 181 | 7% |
| None of these | 1170 | 48% |

Note: The total does sum to 100% as some households adopted more than one measure
*Source: Author's calculations using data from CER (2011)*

In total 52% of participants made at least one investment in efficiency, with many adopting a number of measures together. For instance, many of the households who replaced their boilers also added thermostatic controls to their radiators, and lagging jackets to their hot-water tanks.

If the trial had no effect on investment, we should expect no difference between the control group and any treatment group's in-trial investments in efficiency. We test this in a number of different ways.

### 2.3.5 Empirical strategy

First, a series of simple t-tests are employed to check for equality in the mean number of in-trial adoptions between the control group and each of the treatment groups. The hypotheses being tested are:

$$H_0 : \mu_c = \mu_{t_i} \; vs \; H_A : \mu_c \neq \mu_{t_i} \tag{2.1}$$

Where $\mu_c$ is the mean of the control group and $\mu_{t_i}$ is the mean of each treatment group $i$.

Following this, a binary variable is created which represents adoption of any of the efficiency measures listed in table 2.4 in the previous section. As stated above 52% of households adopted at least one measure. Using this we estimate a logistic regression model in order to examine if any treatment altered the probability of being an adopter. Formally, the hypothesis tested is:

$$H_0 : \beta_1 = 0 \; vs \; H_A : \beta_1 \neq 0 \tag{2.2}$$

Where $Prob(Y = 1|x) = p(x)$ and $log\frac{p(x)}{1-p(x)} = \beta_0 + \beta_1 D_i + \epsilon_i$. $D_i$ represents the different treatment groups.

Finally, a count variable was created to measure the number of investments each household made. This was used to examine if treatment changed the expected number of energy saving features adopted. We test a similar hypothesis to ( 2.2) above.

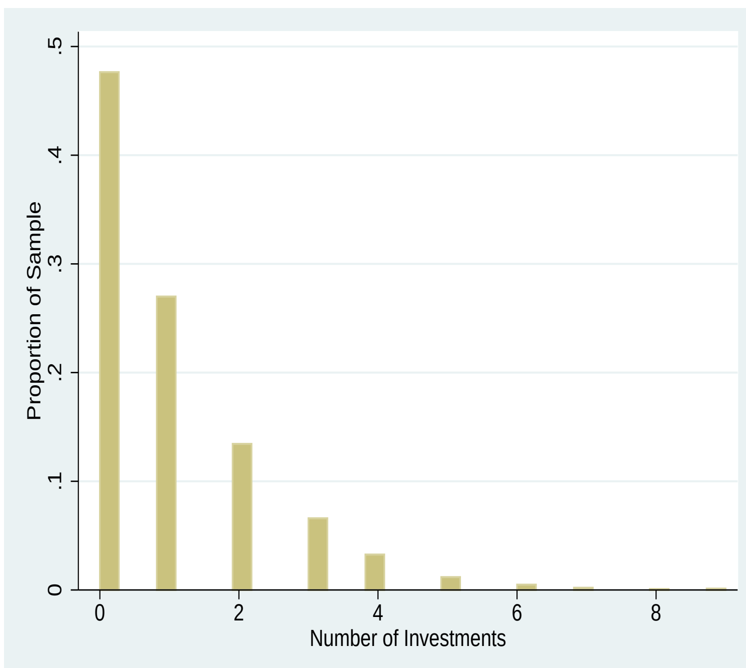


Figure 2.1: Investments distribution

Fig. 2.1 graphically illustrates the distribution of in-trial investments across the entire sample. If the count of household investments in energy efficiency follows a Poisson process, we can represent the probability of observing an event $y_i$ as:

$$Prob(Y = y_i|x_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} \tag{2.3}$$

In this case the expected count of new investments can be represented as follows:

$$E(y_i|x_i) = Var(y_i|x_i) = \mu_i = exp(x_i'\beta) \tag{2.4}$$

The above model assumes equi-dispersion, however, in this case we relax this assumption as $E(y_i|x_i) < Var(y_i|x_i)$. A Negative binomial (NB) model explicitly accounts for unobserved heterogeneity by adding an over-dispersion parameter $\alpha$. $\mu_i$ is replaced by $\mu_i v_i$ and the probability of observing an event can now be expressed as:

$$Prob(Y = y_i|x_i) = \frac{\Gamma(\theta\lambda_i + y_i)}{\Gamma(y_i + 1)\Gamma(\theta\lambda_i)} r^{y_i}(1 - r)^{\theta\lambda_i} \tag{2.5}$$

The expected mean can be expressed as above, while the variance can now be expressed as:

$$Var(y_i|x_i) = \mu_i(1 + \alpha\mu_i) \tag{2.6}$$

## 2.4 Results and discussion

This section presents the results of the tests for equality of means, a logit model examining if treatment altered the probability of adopting any measure, and a negative binomial model testing if treatment influenced the number energy saving measures installed in the home. Following this we will discuss some possible reasons for our results.

### 2.4.1 Equality of means

Exploring the results graphically first, using a box-plot to represent the distribution of in-trial investments for the control and each treatment group, it is immediately obvious that the behaviour of the control group is quite different from any of the treatment groups. The outer edges of the box represent the 25th and 75th percentile of the distribution and the line across the middle of the box is the median, or 50th percentile.

It is clear that the control group has a greater median number of investments, wider inter-quartile range, and the distribution has much greater dispersion. This indicates that households with many investments are more likely to be in the control group rather than any of the treatment groups.

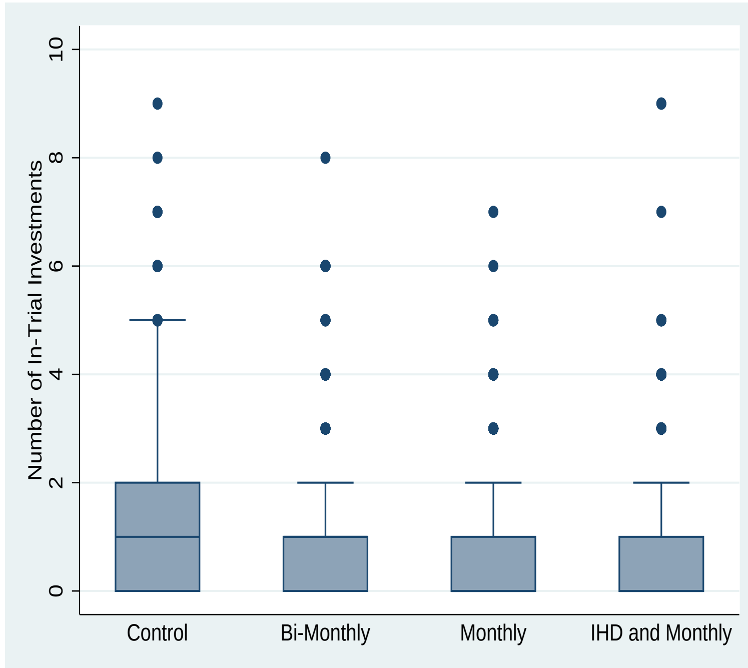


Figure 2.2: Investments boxplot

Focusing now on the difference between control and treatment group means, the results reported in table 2.5, allow us to reject the null hypothesis, that there is no difference between the mean of the control group and the mean of any treatment group[7]. Furthermore, we can conclude that the mean of the control group is greater than the mean of any treatment group[8]. Welch's t-test is used to allow for samples with unequal variance.

Table 2.5: Difference in mean in-trial adoption between control and treatment groups

| Diff | Mean | Std. Err. | [95% Conf. Interval] | $H_A : diff \neq 0$ | $H_A : diff > 0$ |
|---|---|---|---|---|---|
| (Control) - (Treatment 1) | 0.166 | 0.074 | [0.020 - 0.312] | 0.026** | 0.013** |
| (Control) - (Treatment 2) | 0.232 | 0.071 | [0.093 - 0.371] | 0.001*** | 0.001*** |
| (Control) - (Treatment 3) | 0.191 | 0.076 | [0.043 - 0.339] | 0.012** | 0.006*** |

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

### 2.4.2 Regression analysis

The results from the binary regression model further confirm that treatment had a negative effect on investment. We report odds ratios (OR) and use the control group as our reference group in each

[7] We reject $H_0$ at a 1% level of significance for treatment 2, and at a 5% level for treatments 1 and 3.
[8] We reject $H_0$ at a 1% level of significance for treatments 2 and 3, and at a 5% level for treatment 1.

estimation. As can be seen from column 1 of table 2.6 below, the treatment groups are less likely to adopt than the control group ($OR < 1$). Interpreting the odds ratios, we can conclude that the treatment groups were 0.72-0.77 as likely, or 23%-28% less likely than the control group to adopt any energy saving measure over the 12 month trial period. Decomposing the dependent variable, we find the result is driven by the adoption of lagging jackets, attic insulation and double-glazing. No other variables are statistically significant.

Treatment not only reduced the likelihood of investment, but also reduced the number of energy saving investments that households made. The results of the negative binomial regression are reported in column 3 of table 2.6. Reporting the incident rate ratios (IRRs) it is found that being in the treatment group reduced the expected number of energy saving features adopted by 15%-21%.

Table 2.6: Effect of treatment on investment in energy efficiency

| Variable | (1) Binary - OR | (2) Binary - OR | (3) Count-IRR | (4) Count-IRR |
|---|---|---|---|---|
| Bi-monthly Statement | 0.77** | 0.76* | 0.85** | 0.87** |
| | (0.09) | (0.09) | (0.06) | (0.06) |
| Monthly Statement | 0.71*** | 0.68*** | 0.79*** | 0.79*** |
| | (0.08) | (0.08) | (0.06) | (0.05) |
| IHD and Bi-monthly Statement | 0.72*** | 0.72*** | 0.83** | 0.83** |
| | (0.08) | (0.08) | (0.06) | (0.06) |
| Constant | 1.37*** | 1.21 | 1.13** | 0.73 |
| | (0.11) | (0.78) | (0.05) | (0.26) |
| lnalpha | | | -0.34 | -0.5 |
| | | | (0.08) | (0.08) |
| Household level controls | N | Y | N | Y |
| Observations | 2,456 | 2,456 | 2,456 | 2,456 |
| ll | -1693 | -1650 | -3368 | -3313 |
| dfm | 3 | 43 | 3 | 43 |
| chi2 | 12.42 | 99.64 | 11.92 | 131.24 |

Notes: Robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1
Results are reported relative to the control group

For robustness we also ran both of the above models including a range of household level control variables that may influence adoption[9]. As can be seen from columns 2 and 4 of table 2.6 the magnitude of some of the coefficients change slightly but it does not materially alter our results.

[9] These include socioeconomic characteristics, current stock of appliances, current stock of energy efficient measures, total energy usage for 6 month benchmark period, heating type, house characteristics and use of internet.

We tested for equality of coefficients in all of the above models and the treatment groups are not statistically different from each other.

### 2.4.3 Discussion

These results are certainly a perverse outcome given the overall objective of the trial, and it is unclear why this happened. It could be the result of a moral licensing effect, if someone does "something good" (reduce their energy consumption patterns), they might then feel more justified in doing "something bad" (reduce their investment in efficiency), as they feel they have a moral license to do this. Empirical evidence of behavioural change in energy conservation campaigns has previously been ascribed to this effect (Tiefenbeck et al., 2013). In this case households who received information on their water consumption and reduced their usage, also increased their electricity consumption by 5.6 percent compared to a control group. However the authors caution that while their results are consistent with moral licensing, they are unable to confirm that this is the precise psychological mechanism at play.

Another potential explanation would be a priming effect, whereby improved feedback and information may have focused the treatment group on curtailment behaviour, but distracted them from other means of saving energy, such as investing in efficiency. On the other hand the control group were not primed and therefore focused on a broader range of energy saving options.

One could also argue that these individuals are being economically rational, by taking the least-cost option when faced with a number of alternatives. The treatment groups may have seen the benefits of adaptation and undertook the less costly conservation measure. Rather than investing in a more efficient central-heating boiler, they could achieve similar efficiency gains through adaptation, and at a reduced cost, by time-shifting their demand to less expensive periods.

After the trial concluded, the respondents were asked to rate on a Likert-type scale of 1-5, how the trial influenced their level of agreement with a number of statements. Broadly speaking these statements can be categorised into respondents' attitudes towards and interest in electricity reduction; their behaviour over the past 12 months related to electricity reduction; constraints on their ability to reduce electricity consumption even if they wanted to; their understanding of tariffs and information stimuli; their awareness of the cost of their appliance usage; and the effect of the trial on their investment decisions.

We find that the control and treatment groups do not vary much across these categories, but some differences emerge and we will discuss these now in greater detail.

1. The treatment groups were more likely to agree that they had made changes in order to reduce their electricity consumption (71% Treatment vs 60% Control), and they tended to describe these changes as minor (74%) as opposed to major (30%).

2. The control group were less likely to agree that they knew what do do in order to reduce their electricity usage (77% Treatment vs 65% Control).

3. The control group were more likely to agree that they did not know enough about the energy usage of individual appliances in order to reduce their consumption (35% Treatment vs 45% Control).

4. The treatment groups did not seem to have much difficulty understanding tariffs (90% of them reported that they spent less than 1 hour in total over the course of the trial understanding the new tariff structure) or statements, and they overwhelmingly felt that both the tariffs and statements helped them to reduce their usage of electricity (generally in excess of 80% agreement for a range of statements related to whether the tariff or statement helped households reduce their usage).

5. The treatment groups (but not the control) were asked if they felt the trial had any effect on their investments in energy efficiency measures within the home, and a broad consensus did not emerge on this question. (46% agreed, 42% disagreed and the remaining 12% neither agreed nor disagreed). However, critically we cannot identify those in the treatment group who might have adopted, but didn't, nor can we identify those in the treatment group who adopted less than they otherwise might have, were it not for the trial.

In summary, the treatment group certainly did feel they had made a bigger contribution to electricity reduction than the control group; while this may be considered a necessary condition for evidence of a moral licensing effect, it is by no means sufficient.

The treatment group did not seem to have a huge difficulty in understanding the tariffs or statements, however this does not necessarily rule out that a priming effect was at play.

The treatment group felt more in control of their usage than they otherwise would have been, and this may have encouraged some to change their usage through curtailment rather than investment.

This result is quite possibly a complex combination of the above factors and there is no reason to assume that the psychological mechanisms should be homogeneous across individuals or households.

## 2.5 Concluding remarks

We find that being supplied with better information on electricity usage and exposure to time-of-use tariffs resulted in households reducing their investment in energy efficiency measures. It must be stressed that due to the short time-scale of this study, it was only possible to monitor behaviour over a 12 month period. Another interesting element to note is that while this was an electricity smart-metering trial, most of the capital investments related to improving thermal efficiency, not electricity, and only 1% of the sample have electric heating[10].

While we have detailed information on the type of energy efficiency investments made, or not made in the case of the treatment group, it is not possible to quantify whether or not these savings will be offset over a longer period as a result of reduced investment in energy efficient appliances. However, given that the overall electricity reduction relative to the control group was 2.5% (CER, 2011), this result raises the importance of taking a more comprehensive view in the evaluation of energy efficiency interventions.

It also highlights the benefit of allowing a longer time-scale for this type of experiment, as due to the relatively short duration of the study, it is impossible to know the long term behavioural change of households.

Given the widespread implementation of smart-metering trials, this research opens an avenue for others to empirically test whether our results are reproduced in other countries and across different domains.

[10]See Appendix A.

## 2.A Detailed descriptive statistics

Table A1: Pre-trial distribution of socioeconomic characteristics

| Variable | Control | Treatment 1 | Treatment 2 | Treatment 3 | Total |
|---|---|---|---|---|---|
| Gender | | | | | |
| male | 52% | 47% | 52% | 51% | 51% |
| female | 48% | 53% | 48% | 49% | 49% |
| Age | | | | | |
| $18-25$ | 0% | 1% | 0% | 1% | 0% |
| $26-35$ | 8% | 10% | 7% | 9% | 9% |
| $36-45$ | 18% | 20% | 20% | 20% | 19% |
| $46-55$ | 24% | 24% | 25% | 24% | 24% |
| $56-65$ | 21% | 22% | 24% | 23% | 22% |
| 65+ | 28% | 23% | 22% | 22% | 24% |
| refused | 1% | 1% | 1% | 0% | 1% |
| Employment Status | | | | | |
| employee | 43% | 49% | 47% | 47% | 46% |
| self-employed (no employees) | 4% | 5% | 5% | 8% | 5% |
| self-employed (with employees) | 6% | 7% | 8% | 5% | 6% |
| unemployed (seeking work) | 4% | 3% | 5% | 4% | 4% |
| unemployed (not seeking work) | 3% | 4% | 4% | 3% | 4% |
| retired | 38% | 31% | 30% | 32% | 33% |
| carer | 2% | 1% | 1% | 1% | 1% |
| Social Class | | | | | |
| AB | 12% | 14% | 16% | 17% | 15% |
| C1 | 25% | 29% | 27% | 28% | 27% |
| C2 | 18% | 16% | 15% | 17% | 16% |
| DE | 42% | 38% | 38% | 35% | 39% |
| F | 3% | 2% | 3% | 2% | 2% |
| refused | 1% | 1% | 1% | 1% | 1% |
| Education | | | | | |
| no formal education | 1% | 1% | 2% | 1% | 1% |
| primary | 15% | 10% | 11% | 11% | 11% |
| secondary to junior cert | 15% | 19% | 17% | 16% | 17% |
| secondary to leaving cert | 30% | 29% | 28% | 28% | 29% |
| third level | 34% | 37% | 37% | 39% | 36% |
| refused | 5% | 6% | 5% | 6% | 5% |

*Source: Author's calculations using data from CER (2011)*

Table A2: Pre-trial distribution of house characteristics

| Variable | Control | Treatment 1 | Treatment 2 | Treatment 3 | Total |
|---|---|---|---|---|---|
| House type | | | | | |
| apartment | 2% | 1% | 1% | 2% | 2% |
| semi-detached house | 30% | 33% | 29% | 29% | 30% |
| detached house | 27% | 26% | 29% | 27% | 27% |
| terraced house | 14% | 15% | 14% | 14% | 14% |
| bungalow | 27% | 24% | 27% | 27% | 26% |
| refused | 0% | 1% | 0% | 0% | 0% |
| Tenure type | | | | | |
| rent from a private landlord | 2% | 1% | 1% | 1% | 1% |
| rent from a local authority | 5% | 4% | 5% | 4% | 4% |
| own outright | 58% | 56% | 56% | 55% | 56% |
| own with mortgage | 35% | 40% | 38% | 39% | 38% |
| refused | 0% | 0% | 0% | 1% | 0% |
| Dwelling Building Energy Rating certified | | | | | |
| yes | 1% | 1% | 1% | 1% | 1% |
| no | 86% | 88% | 88% | 87% | 87% |
| don't know | 13% | 11% | 11% | 12% | 12% |
| Number of people over 15 | | | | | |
| 1 | 4% | 5% | 5% | 4% | 4% |
| 2 | 62% | 58% | 61% | 61% | 61% |
| 3 | 20% | 20% | 17% | 16% | 18% |
| 4 | 9% | 12% | 12% | 12% | 11% |
| 5 | 3% | 4% | 3% | 6% | 4% |
| 6 | 1% | 1% | 1% | 0% | 1% |
| 7 | 0% | 0% | 0% | 0% | 0% |
| Number of people under 15 | | | | | |
| 0 | 77% | 72% | 72% | 73% | 74% |
| 1 | 10% | 11% | 12% | 13% | 11% |
| 2 | 8% | 11% | 9% | 8% | 9% |
| 3 | 3% | 4% | 5% | 4% | 4% |
| 4 | 2% | 1% | 1% | 1% | 1% |
| 5 | 0% | 0% | 0% | 0% | 0% |
| 6 | 0% | 0% | 0% | 0% | 0% |
| Number of bedrooms | | | | | |
| 1 | 1% | 0% | 0% | 1% | 1% |
| 2 | 11% | 9% | 6% | 10% | 9% |
| 3 | 43% | 46% | 45% | 40% | 43% |
| 4 | 35% | 34% | 35% | 37% | 35% |
| 5 | 10% | 11% | 13% | 12% | 11% |
| 6 | 0% | 0% | 0% | 0% | 0% |
| House age | | | | | |
| 1900 – 1940 | 2% | 2% | 2% | 1% | 8% |
| 1941 – 1960 | 3% | 2% | 3% | 2% | 10% |
| 1961 – 1970 | 3% | 3% | 3% | 2% | 10% |
| 1971 – 1980 | 6% | 4% | 5% | 4% | 18% |
| 1981 – 1990 | 3% | 3% | 3% | 3% | 11% |
| 1990 – 1997 | 2% | 2% | 2% | 2% | 8% |
| after 1997 | 6% | 5% | 5% | 5% | 21% |

*Source: Author's calculations using data from CER (2011)*

Table A3: Pre-trial distribution of appliance stock and heating type

| Variable | Control | Treatment 1 | Treatment 2 | Treatment 3 | Total |
|---|---|---|---|---|---|
| Appliance | | | | | |
| washing machine | 28% | 24% | 25% | 22% | 98% |
| tumble drier | 19% | 16% | 17% | 15% | 68% |
| dishwasher | 18% | 16% | 17% | 15% | 66% |
| electric shower instant | 19% | 16% | 18% | 16% | 69% |
| electric shower pumped | 8% | 7% | 7% | 6% | 29% |
| electric cooker | 22% | 19% | 18% | 17% | 76% |
| stand alone freezer | 14% | 13% | 13% | 12% | 51% |
| water pump | 5% | 4% | 5% | 4% | 19% |
| immersion | 22% | 18% | 19% | 18% | 77% |
| solar panels to heat water | 0% | 0% | 1% | 0% | 2% |
| Heating Type | | | | | |
| electric | 2% | 1% | 1% | 1% | 1% |
| electric (plug-in) | 1% | 0% | 0% | 0% | 0% |
| gas | 25% | 29% | 30% | 28% | 28% |
| oil | 42% | 44% | 43% | 45% | 43% |
| solid fuel | 29% | 24% | 23% | 26% | 26% |
| renewable | 0% | 1% | 1% | 1% | 1% |
| other | 1% | 1% | 0% | 0% | 1% |

*Source: Author's calculations using data from CER (2011)*

# Chapter 3

# Consumer preferences and the influence of networks in electric vehicle diffusion

## 3.1 Introduction

The electrification of transport is considered an important element in de-carbonising our economies. This may have environmental benefits both in direct emission reductions and in helping to better integrate intermittent renewable energy sources into generation portfolios (Sioshansi and Denholm, 2009; Ernst et al., 2011). However, negative externalities can also emerge.

As discussed in Chapter 1, large penetrations of EVs may exacerbate existing peaks in electricity demand, when charging, and geographic clustering of electrical load may put pressure on certain nodes along low-voltage distribution networks. This is an engineering problem with economic determinants, and we are interested in the factors that might give rise to this outcome.

In order to examine this we develop an agent-based, threshold model of innovation diffusion to simulate the adoption of electric vehicles among Irish households. Adoption is modelled as a binary choice, and agents have a threshold beyond which the benefits to adoption exceed the costs. Their utility is a function of their socioeconomic characteristics, environmental behaviour and attitudes. Their utility from adoption increases as their peers adopt and as the innovation gains popularity within the population. Agents are heterogeneous, drawn from a detailed, nationally representative study connected to an Irish electricity smart-metering trial. Agent populations are spatially explicit; we can generate populations of agents to match any geographic area of interest.

In determining the overall diffusion level in the population, we find that network type matters and that it is important who adopts first. When our analysis is applied to populations representative of certain geographic areas we find that even mild peer effects can generate high adoption and clustering in these areas, even if the overall adoption level is quite low.

While our analysis is applied to Irish data and electric vehicle adoption, this methodology could easily be applied to other countries and products, provided there is adequate spatially linked census data that can be combined with a suitable micro-level dataset.

## 3.2 Agent-based models

Agent-based modelling (ABM) is a form of "bottom-up" modelling with its roots in Complex Adaptive Systems (Macal and North, 2009). This field is chiefly concerned with the study of how seemingly complex systems can arise from the interactions of relatively simple components. It is informed by a diverse range of disciplines, initially from sciences such as ecology, physics, computer science and applied mathematics. More recently this methodology has been embraced by economists and sociologists[1].

This type of modelling is useful when describing systems of interacting agents, which exhibit emergent properties not easily deduced by aggregating the preferences of individual agents. Macal and North (2009) consider an ABM to be a computer model in which the objects are self-contained, autonomous entities that interact with each other within an environment. The behaviour of an agent is based on *"the current state of its interactions with other agents and the environment"*. An agent will have rules that define its behaviour and within more sophisticated models an agent may be able to adapt its behaviour and learn from its experiences.

While not strictly an ABM, an early model of social influence and a demonstration of how simple rules can produce complexity was Schelling's model of population segregation in US cities (Schelling, 1971). By using a very simple Markov process to describe each individual's behaviour, the model could produce results that mimicked real segregation patterns remarkably accurately. Another interesting aspect of this was how the emergent dynamics of the system at equilibrium can be vastly different from the individual preferences; *"there is no simple correspondence of individual incentive to collective results"*. In a related example, Granovetter (1978) demonstrated how slight perturbations to initial distributions of preferences can lead to vastly different outcomes in populations that on aggregate appear very similar. Thus underlining the inherent problem with the use of aggregate models that do not consider interactions. While Granovetter was concerned with the decision to riot and Schelling with segregation patterns, similar techniques can also be used to describe the dissemination of new ideas and the adoption of new technologies.

[1] Miller and Page (2009) provide a very useful introduction and a recent paper by Sornette (2014) provides a historical summary of cross-pollination between financial economics and physics.

ABMs have been widely used to study the diffusion of new technology. We will mainly focus on the diffusion of "green" technologies, with a particular interest in threshold models for the purposes of this work.

### 3.2.1 Agent-based adoption models

Within an ABM framework, the diffusion process can take different forms. Linder (2011) use a Bass Diffusion Model to simulate adoption profiles for EVs amongst households in Stuttgart in 2020. Based on microdata they construct different adopter types and simulate various scenarios of spatial diffusion. They find that adopters will concentrate in urban areas and that spatial differences will become quite apparent by 2020.

Tran (2012a) explicitly models social network effects to examine their interaction with individual preferences in innovation diffusion. The author simulates adoption profiles for a variety of competing vehicle types; including petrol, diesel, EV and hybrid electric vehicle (HEV). Using heterogeneous agents, it is found that network influence can be an important factor in driving high levels of adoption, even if agents have low individual incentives to adopt. However, the authors caution that homophily can account for a lot of what first appears to be contagion. Other related work examines the techno-behavioural aspects of diffusion (Tran, 2012b).

Threshold models are also widely used throughout the literature. Eppstein et al. (2011) develop spatially explicit, agent-based consumer choice models of plug-in hybrid electric vehicle (PHEV) adoption to assess the market's sensitivity to various input parameters, such as fuel prices, battery range, purchase price and government subsidies. Agents have characteristics such as age and salary. Social groups are modelled as homophilous networks with fat-tailed degree distributions. An agent's threshold is defined as the proportion of adopters within its social group required for the agent to adopt; this is negatively correlated with their salary - the authors argue that wealthier agents would be less risk averse, thus having a lower threshold.

Hamilton et al. (2009) model technology diffusion in a situation where consumers are uncertain about the performance of a new technology versus the old one. The agent environment is a square lattice of $N$ cells and agents receive electricity from three potential sources; the grid, solar or CHP. Agent's thresholds are normally distributed and heterogeneous. An interesting aspect of this paper is the emergence of "punctuated equilibria". This is a characteristic of a complex system, in which negligible changes in inputs can induce dramatic shifts in system outputs. This feature is observed empirically when the diffusion process does not follow the textbook S-curve.

Other research simulates the adoption of smart-metering technology (Zhang and Nuttall, 2007). This paper has similar features to others mentioned in that agents have local and random interactions on a grid, allowing for networks which exhibit small-world and scale-free properties. An interesting feature of this paper is the finding that a random and dispersed initial seeding can yield a much higher overall diffusion rate than a controlled centralised one. This has interesting policy implications when one considers government interventions to induce adoption of socially beneficial technologies.

A big problem inhibiting the widespread adoption of green technologies is consumer indifference. This is explicitly modelled by Kowalska-Pyzalska et al. (2013), in which the authors examine the diffusion of dynamic electricity tariffs. Again, the environment is modelled as a square lattice. Influence is channelled through nearest-neighbours and media. Indifference is essentially noise in the system and can arise if the product offers both advantages and disadvantages between which the agent is unable to compare. They find that due to high levels of indifference, widespread adoption will be "virtually impossible" in modern societies, highlighting the need for better information in order to overcome this.

The focus of our research broadly follows a number of other papers in which agents have a utility function associated with the adoption of a new product. Agents have a threshold beyond which the benefits of adoption exceed the costs and they adopt. Within this literature thresholds can take various distributional forms, be heterogeneous or be the same for all agents.

Delre et al. (2010) use a simple utility function consisting of two terms, individual preference and social influence. Agents are heterogeneous in their preferences and thresholds are uniformly distributed. The authors experiment with regular lattice and preferential attachment networks. An interesting element of this research is the study of so-called VIP effects. This is created using scale-free, non-symmetric adjacency matrices, with heterogenous weights on the edges to represent social networks. This is a more realistic representation of real-life networks where influence is not bi-directional. Certain nodes have a greater number of connections, but they also have a greater magnitude. Interestingly, they find that when modelling influential nodes it is the number of connections, more so than their weight or magnitude that propagates the diffusion process.

McCullen et al. (2013) extend this approach to multi-parameter models. Along with peer-effects, they include a term for overall diffusion in the population (the S-curve). Thus, agents are influenced by their personal preferences, peer-effects and social-norms. Various network types are constructed; random, small-world and highly clustered community-based networks. They take a dynamical systems, rather than agent-based approach, and run multiple simulations to explore the parameter space. They find that the level of adoption depends strongly on network topology and the relative weightings of parameters.

Palmer et al. (2013) employ a similar methodology to simulate the diffusion of photovoltaic (PV) systems in Italy. Each agent has a unique utility which is a function of the payback period of the investment, the environmental benefit to the agent, their income and peer-influence from other agents within the population. Adopter categories are constructed using Sinus-Milieu data for Italy and this information is also used to construct homophilous social groups. The aggregate adoption level is calibrated to past data, then projections are made. To do this the authors find a common threshold for all agents that best fits the data and the model output is very sensitive to changes in this. The relative weightings of preferences vary by adopter category and these weights are also chosen to calibrate the model to past data.

In terms of the spatial aspect of adoption, Campbell et al. (2012) use a clustering algorithm on Birmingham Census data to determine the areas most likely to have high proportions of early adopters. Others employ a similar methodology using Finnish data (Saarenpaa et al., 2013), but go a little further in that they find the correlation between certain demographic and socio-economic characteristics and early HEV adoption. Sorda et al. (2013) link GIS data with an agent-based simulation to model investment decisions in combined heat and power (CHP) biogas plants in two different German states. This model is then used to access the way in which changes in current support mechanisms may affect the electricity generated from such plants.

We incorporate elements from a range of the above mentioned research in our work; we use clustering algorithms on Census data to determine where we might expect dense concentrations of adopters to form. We then generate agent populations based on the distribution of characteristics within these areas. Following that, we employ a version of McCullen et al. (2013)'s main model equation in our agent-based model. We also include elements of Palmer et al. (2013) in that we decompose personal preferences into economic and environmental categories.

## 3.3 Materials and methods

### 3.3.1 Basic principles

The basic principle examined by this model is the interaction of individual preferences and peer-effects in the adoption or spread of behaviour, attitudes or new technologies.

### 3.3.2 Entities, state variables and scales

Entities, or agents, represent households. Agents are heterogenous and defined by a matrix of static and dynamic characteristics. They are located within a neighbourhood. The neighbourhood is represented by a discrete, homogenous grid. A regular lattice of size $N1 * N2$, where $N1 * N2 = N$ the total number of agents.

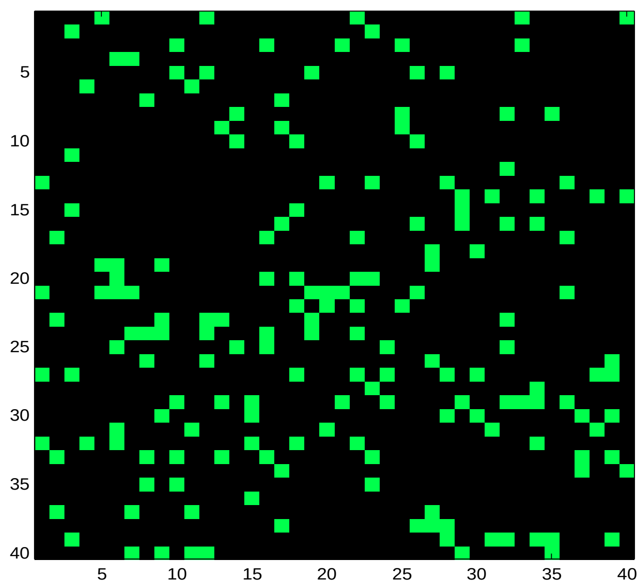


Figure 3.1: Agent environment

Note: $N = 1600$ agents at $t = 15$. Agents are represented as squares on the grid. The green agents have adopted, the black have yet to adopt. Graphics generated using MATLAB code adapted from Lai and Poltera (2009)

The global environment consists of a number of adjacent neighbourhoods based on Irish Census 2011 data. We will mainly work with Electoral Districts (EDs) of about 1,000 households. However, we could easily adapt our environment to cover Small-Areas (approximately 100 households) or much larger regions.

Agents are created using a large, nationally representative micro-data set connected to an electricity metering trial (CER Smart Metering Customer Behavioural Trial 2009). For more information on data used see Appendix 3.A.

Rogers (1995) segmented the population into different classes of adopter; innovator, early adopter, early majority, late majority and laggards. To create our agents we would ideally determine empirically the characteristics that define the innovators and early adopters of electric vehicles. Due to a lack of data this is quite difficult, however, survey information on early adopters does exist.

In a comprehensive review of the literature, Hjorthol (2013) conclude that early adopters of EVs will share broadly similar characteristics with early adopters of hybrid electric vehicles (HEVs). EV owners will predominantly be middle-aged, male, have high levels of education and income, be urban dwellers and own one or more car. HEV owners share most of these characteristics but tend to be older.

Both academic and grey literature in other countries tends to correspond with these findings. In a 2010 survey of US consumers, Deloitte (2010) expect early adopters of EVs to be young, high income individuals, already owning one or more vehicle. They predict adoption will concentrate in areas such as southern California, due to weather, infrastructure and high proportions of these individuals. In another US study, Scarborough Research (2007) again find HEV owners to be older than EV owners, but otherwise have similar characteristics. While research on Prius owners in the UK find them to be predominantly men, aged over 50, in two-car households, earning in excess of £4,000 per month and residing in the south-east (Ozaki and Sevastyanova, 2011).

Agents are described by two attributes, informed by the above literature, which we call Income Utility ($IU$) and Environmental Utility ($EU$). Fig. 3.2 below shows the distribution of these attributes generated from survey data in the full population of 3099 households. We will draw sub-samples from this population for simulation runs.

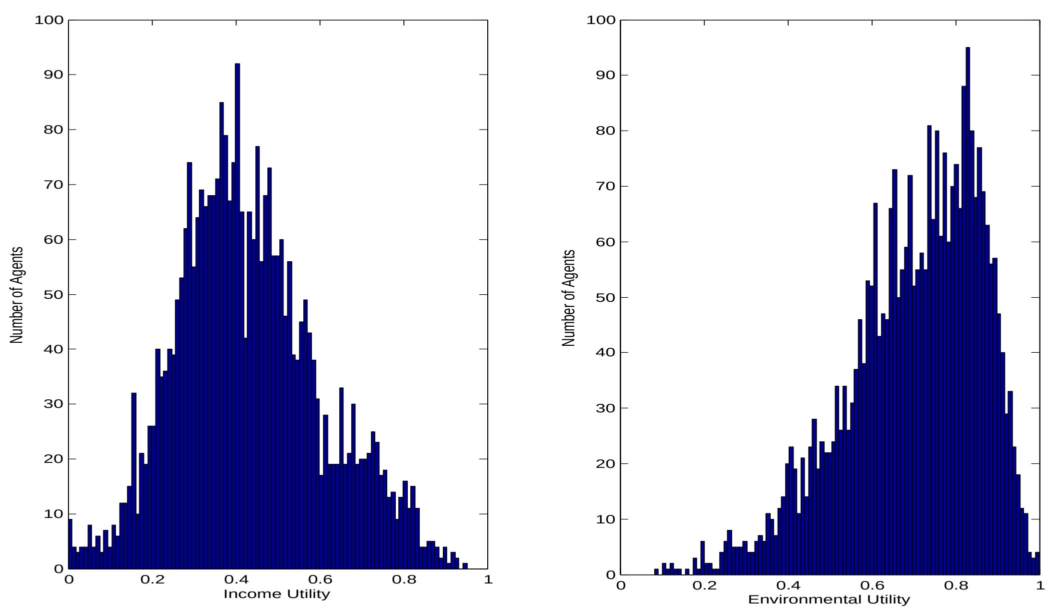


Figure 3.2: Distribution of preferences for full agent population

*IU* is based on an agent's social class, tenure type and age and we would expect this to be highly correlated with their income. Unfortunately, income data was not available. This is an adoption probability which could be considered an implicit budget constraint. Other things being equal, agents with a higher IU are more likely to adopt. To create it, we ranked each category of social class, tenure type and age from 0-1, as can be seen in table 3.1.

To transform the rankings into the distribution shown, we summed up the total for each agent and standardised between 0 and 1. For instance, an agent of social class "C", who's tenure is "own with mortgage" and is aged 25-59 will have a total of $(0.25 + 0.5 + 1) = 1.75$. This total is divided by the highest possible value (3 in this case) to create an $IU = 0.58$ for this agent. Performing this procedure for all agents results in a rather lumpy distribution, so we add a degree of noise by allowing it vary randomly within a bound of $IU \pm 0.1$, resulting in the distribution shown in fig. 3.2.

Table 3.1: Survey answers for Income Utility (IU)

| Social Class | Rank |
|---|---|
| AB | 1.00 |
| C | 0.25 |
| DE | 0.20 |
| F | 0.17 |
| No answer | 0.14 |
| **Tenure** | |
| own outright | 1.00 |
| own with mortgage | 0.50 |
| rent private | 0.33 |
| rent from local authority | 0.25 |
| other | 0.20 |
| **Age** | |
| 25-59 | 1.00 |
| Other | 0.50 |

Source: Survey questions from CER (2011)

Agent's *EU* is based on their previous adoption of energy saving technology within their homes and their attitude toward the environment, $EU \in [0, 1]$. Other things being equal, agents with a higher EU are more likely to adopt.

We use a similar method to calculate EU. The answers to various survey questions listed in tables 3.2 and 3.3 are ranked in order of pro-environmental behaviour and environmental concern. These are then summed up and standardised. In this case we have two categories; behavioural and attitudinal.

We feel behavioural measures, particularly those related to the current stock of energy efficiency measures adopted in the home is a better guide to future adoption than declared environmental attitudes. This largely determines the distribution of $EU$, and is assigned a weight 5 times that of the attitudinal data.

Table 3.2: Survey answers for Environmental Utility (EU) - behavioural

| Lightbulbs | Rank | Attic | Rank |
|---|---|---|---|
| All | 1.00 | Yes, more than 5 years ago | 1.00 |
| 3/4 | 0.80 | Yes, less than 5 years ago | 1.00 |
| 1/2 | 0.60 | Don't know | 0.00 |
| 1/4 | 0.40 | no | 0.00 |
| None | 0.20 | | |
| **Windows** | **Rank** | **External Wall insulated** | **Rank** |
| All | 1.00 | Yes | 1.00 |
| 3/4 | 0.80 | No | 0.00 |
| 1/2 | 0.60 | Don't know | 0.00 |
| 1/4 | 0.40 | | |
| None | 0.20 | | |
| **Lagging** | **Rank** | | |
| Yes | 1.00 | | |
| No | 0.00 | | |

Source: Survey questions from CER (2011). Note: "Lightbulbs" and "Windows" refer to the proportion of energy saving lightbulbs and double-glazed windows in the agent's home

Therefore an agent with $IU \geq 0.7$ and $EU \geq 0.7$ for example, is likely to be aged 25-59, a home owner, of social class AB, have strong environmental preferences and already an adopter of a range of energy saving technologies within their home.

This work is limited by the fact that we do not have access to econometric estimates of these parameters. The weightings we attach to different attitudes or behaviours are estimates. However, we rank them based on characteristics that are prominent amongst early stage adopters of EVs. Much research in this space simply uses some statistical distribution (such as uniform or normal), or various point estimates to describe agent preferences. We feel our work is a useful extension to this as the distributions we use are generated (however imperfectly) from detailed survey data, allowing us to describe preferences with bivariate distributions based on a nationally representative sample. Furthermore, this enables us to link these characteristics with known Census aggregates for particular areas, in order to generate agent populations that match geographic areas of interest.

Table 3.3: Survey answers for Environmental Utility (EU) - attitudinal

| Opportunity to sell back electricity | Rank |
|---|---|
| Very dissatisfied | 1.00 |
| - | 0.50 |
| - | 0.33 |
| - | 0.25 |
| Very satisfied | 0.20 |
| Environmental damage with electricity generation | Rank |
| Very dissatisfied | 1.00 |
| - | 0.50 |
| - | 0.33 |
| - | 0.25 |
| Very satisfied | 0.20 |
| Percentage generated from renewables | Rank |
| Very dissatisfied | 1.00 |
| - | 0.50 |
| - | 0.33 |
| - | 0.25 |
| Very satisfied | 0.20 |

Source: Survey questions from CER (2011)

### Threshold distribution

Each agent has a unique threshold ($\theta_i$). The threshold represents the costs or barriers to adoption for the agent. If the cost of purchasing an EV dropped significantly, this could be modeled by lowering the threshold for all agents, or if a subset of agents were resident in an apartment block, their thresholds would be relatively high as it would be difficult for them to charge their vehicle at home, making them less likely to purchase one.

As discussed above in section 3.2, within these models thresholds can be heterogeneous or homogeneous. They can be informed by data or assumed to take a particular distributional form. We interpret thresholds in a similar manner to Bale et al. (2013) and Bale et al. (2014), however we assume a normal distribution as its functional form, as per Eppstein et al. (2011). The threshold distribution chosen is an important determinant of adoption levels. It is bounded by 0 and 1, $\theta \in [0, 1]$. An agent with a threshold of 1 will never adopt, as the utility from adoption can never exceed the threshold for this agent. An agent with a threshold of 0 will always adopt. Any agent with a threshold greater than 1(less than 0), will be assigned a threshold of 1(0) as per fig. 3.3.

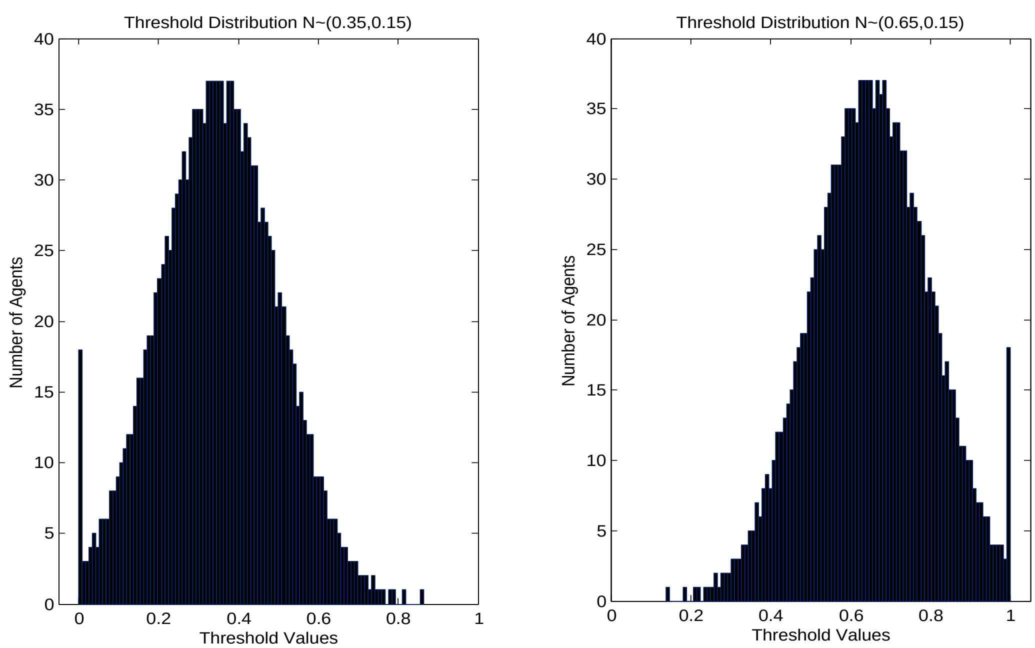


Figure 3.3: Threshold distributions for $N = 1600$ agents

Once we select a distribution to describe the range of thresholds, we can assign each agent their threshold value randomly, or we can assign a threshold based on their characteristics. In related research (Eppstein et al., 2011), the authors use agent's salary to generate cross-correlations for other distributions, including agent's thresholds. They create a threshold distribution that is negatively correlated with salary ($corr = -0.67$). The justification for this is that agents with a higher salary are exposed to less risk when purchasing a vehicle and will have a lower threshold as a result. We follow this research and in all simulations threshold ($\theta$) is negatively correlated with Income Utility ($IU$), with $corr \in [-0.45, -0.55]$. The range of values is due to sampling variability between simulation runs.

The model results are highly sensitive to the choice of threshold distribution. Appendix 3.E explores this in more detail.

### 3.3.3 Collectives

We will mainly experiment with two different types of social network, that have a number of the characteristics of real-life social networks; small-world and preferential attachment.

In order to better describe these, it is first useful to introduce two other network types; regular lattice and random.

Starting with the regular lattice in fig. 3.4(a), each node is connected to its $k$ nearest neighbours, with $k = 2$ in this case. This is a very simple network type and not very realistic as nodes can only have local connections. In the next network, a random graph fig. 3.4(b), nodes do not necessarily have any local connections but are randomly connected to any other node in the network. Again this is an unrealistic characterisation of a real social network.

A small-world network, fig. 3.4(c) contains properties of both a regular lattice and a random graph. They are both highly clustered and have short characteristic path lengths. A network is said to be small-world if the mean distance between any two pairs of nodes is small relative to the total number of nodes in the network. This distance grows proportionally to the log of the number of nodes in the network. These networks were classed as a particular type of random graph in seminal work by Watts and Strogatz (1998). The authors found that by randomly re-wiring a certain proportion of the edges of a regular lattice, they could create a graph which had "small-world" properties. We illustrate this by showing a graph with a re-wiring probability of 0.1. A wide range of networks in both physical and social systems have this property. This includes the neural networks of worms, the electric power grid and collaboration networks in academia and film. For a discussion of their statistical properties, see Amaral et al. (2000).

Preferential attachment networks fig. 3.4(d), exhibit a scale-free property. The connectivity distribution of nodes within these networks decay as a power-law. This essentially means that the majority of nodes have few connections but a small proportion have a very large number of connections, e.g. node v1. As new nodes enter the network, they connect with existing nodes with a

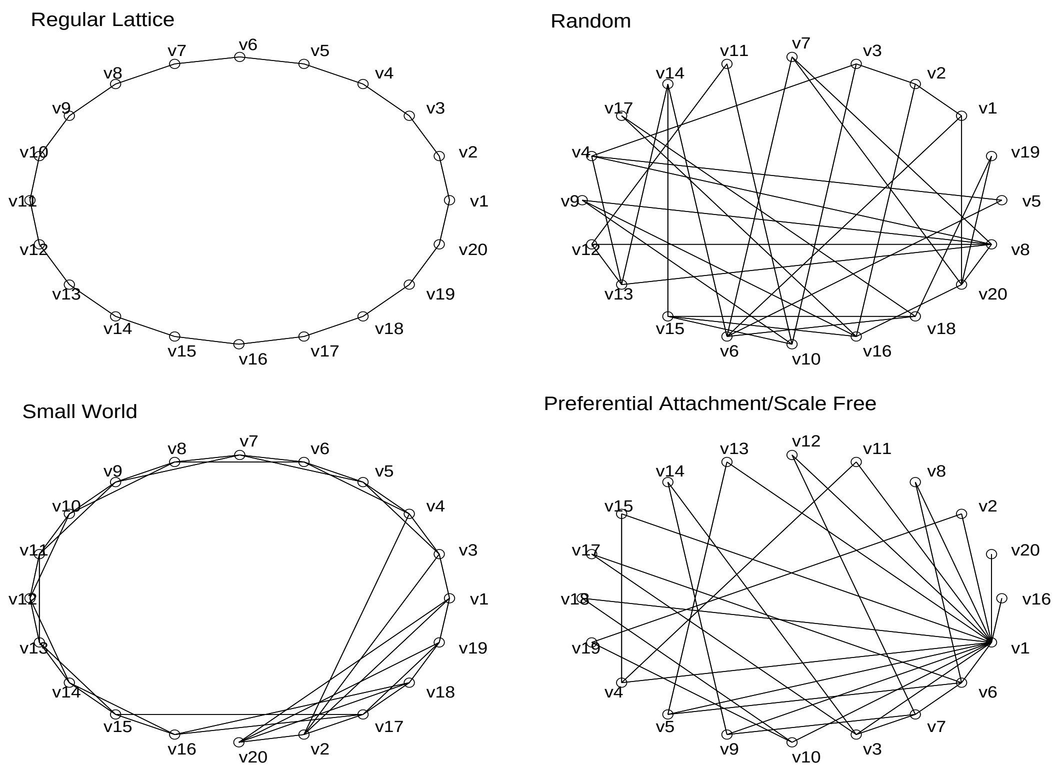


Figure 3.4: Social network diagrams

Note: $N = 20$ nodes or agents in each case. Agents arranged in a circle for illustrative purposes only. Network matrices constructed using CONTEST (Higham and Taylor, 2008). Network graphics using "Matlab Tools for Network Analysis" (MIT, 2011) probability which is an increasing function of the number of connections these nodes currently have. These highly-influential nodes could be considered opinion leaders in a community. Many real-life networks also exhibit this property.

None of the network types described exhibit all of the characteristics of real-life social networks. For instance a right-skew distribution is an important property of preferential attachment networks, i.e. some nodes are very influential, however this network type does not allow for local clustering. The opposite can be said for small-world networks. Another important characteristic of social networks absent from the listed models is homophily; the fact that people form groups with other similar people. Further information on this can be found in Hamill and Gilbert (2009).

### 3.3.4 Process overview and scheduling

The main process in this model is the diffusion of EVs. At every time step, each agent decides to adopt or not. A time step is defined as the length of time it takes to update all agents and thus does not correspond to real time. Agents are updated sequentially on the grid, starting at the top-left corner[2].

Simulations last until all agents are activated or until the model reaches equilibrium. Equilibrium will occur when a gap in either the threshold distribution, or in the agent's individual preferences do not allow any further agents to be activated.

### 3.3.5 The model

#### Agent decision making

The adoption status of each agent is represented as a binary variable; 1 if $agent_i$ has adopted, 0 otherwise. Adoption is an absorbing state - once an adopter it is not possible to switch back. This model is adapted from the work of Delre et al. (2010), McCullen et al. (2013) and Palmer et al. (2013).

[2]To ensure our results are not dependent on the updating sequence, we have re-run the model updating agents randomly as a robustness check. The results hold. Also, each simulation contains a different sample of agents chosen from the agent population and each agent is randomly assigned a place on the grid. New networks are assigned for each simulation run. These procedures also help to ensure that the results are not dependent on the updating sequence.

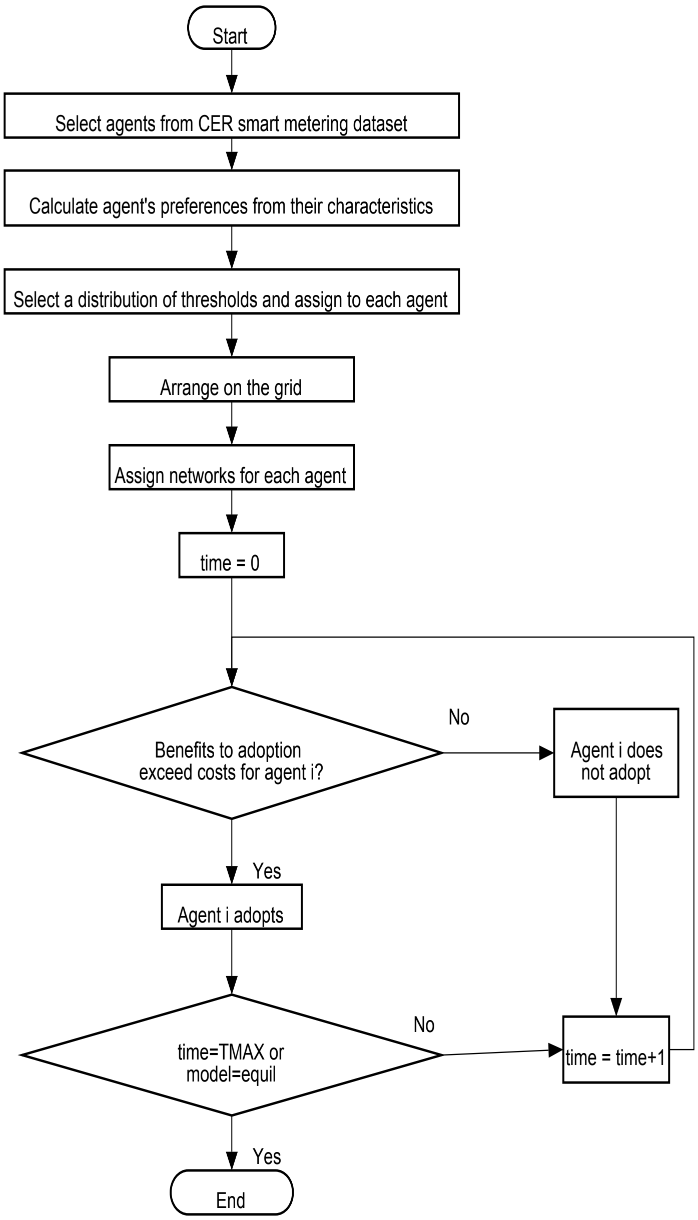


Figure 3.5: Flow-chart of updates for agent$_i$

Note: Procedure is iterated through all other agents

$$x_i(t+1) = \begin{cases} 1, & \text{if } x_i(t) = 1 \\ 1, & \text{if } x_i(t) = 0 \text{ and if equation ( 3.2) is true} \\ 0, & \text{otherwise} \end{cases} \tag{3.1}$$

The adoption submodel is the main engine of the agent-based model. $Agent_i$ adopts with a certain probability once its utility $U_i$ exceeds its threshold $\theta_i$ at time$_t$, $\theta_i$ is time-invariant. The parameter "crit" is a decaying stochastic cost term that accounts for poor products or poor information about the product when in its infancy. This is to account for the inertia that can exist in the early stages of technology adoption. For example if the critical value $crit = 0.1$, adoption will be slow in the early periods but by time $t = 10$, $t * crit = 1$ and the product will have reached maturity, or agents will be fully informed of its benefits[3]. Agent$_i$ adopts if:

$$U_i(t) \geq \theta_i \text{ and if } t * crit \geq rand(0,1) \tag{3.2}$$

Utility is a weighted function of individual preferences $IU$, $EU$, peer effects $G$ and wider social norms $S$.

$$U_i(t) = \alpha_i IU_i + \beta_i EU_i + \gamma_i G_i(t) + \delta_i S(t) \tag{3.3}$$

With $\alpha_i + \beta_i + \gamma_i + \delta_i = 1$

### Individual preferences

Individual Preferences are determined by an agent's income utility (IU) and an environmental utility (EU). These have already been described in detail.

### Peer influences

Agents are modelled as nodes in a network. Influence between nodes is represented by an adjacency matrix A.

$$A_{i,j} = \begin{cases} 1, & \text{if node i influences node j} \\ 0, & \text{otherwise} \end{cases} \tag{3.4}$$

[3]The sensitivity of results to changes in this value is discussed in Appendix 3.E.

To illustrate this, consider the following symmetric adjacency matrix, generated for a network of 10 agents. Each row represents an agent's network. We can see that $agent_1$ is in a network with $agents_{2,4,5,9}$. $Agent_2$ is in a network with $agents_{1,3,4,5,7,8}$

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{3.5}$$

The degree distribution of this network is illustrated in fig. 3.6

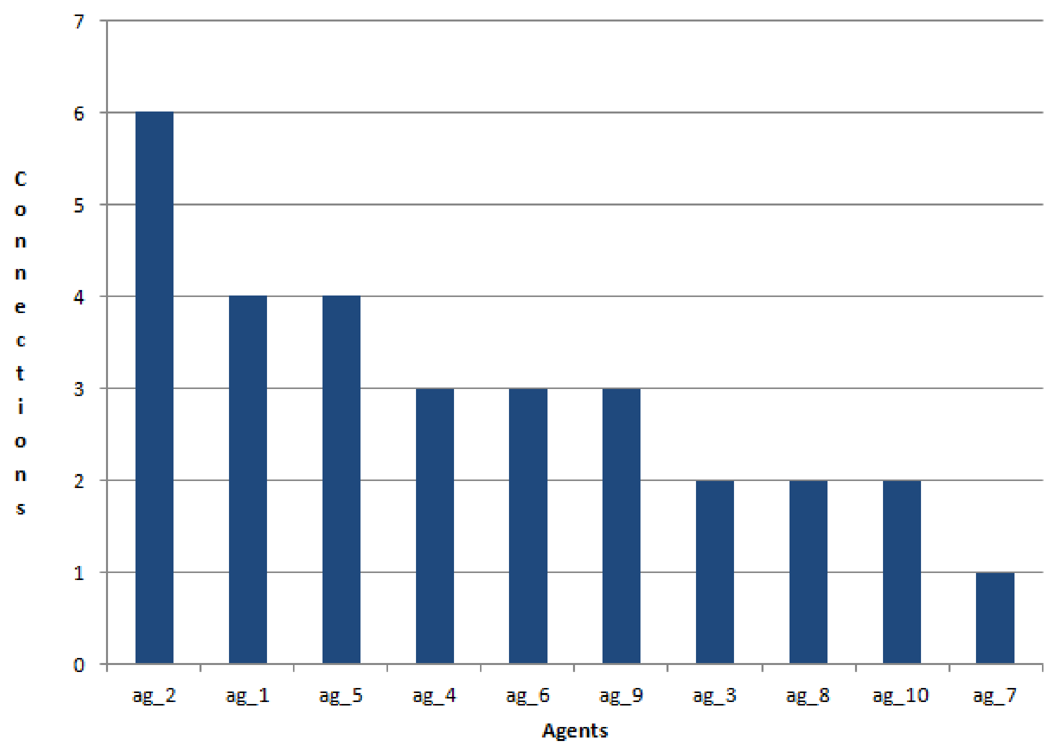


Figure 3.6: Degree distribution for preferential attachment network of $N = 10$ nodes

Each agent's influence is given by their node degree k. This is the number of other agents within their social group. In the above example $k_1 = 4$ and $k_2 = 6$

$$k_i = \sum_{j=1}^{N} A_{ij} \tag{3.6}$$

Group influence is the proportion of adopters within each agent's social group.

$$G_i(t) = \frac{1}{k_i} \sum_{j=1}^{N} A_{ij} x_j(t) \tag{3.7}$$

**Social norms**

This is the total number of adopters in the population.

$$S(t) = \frac{1}{N} \sum_{j=1}^{N} x_j(t) \tag{3.8}$$

**Consumer groups**

As stated above $\alpha_i + \beta_i + \gamma_i + \delta_i = 1$. We can set fixed weights or allow them to vary depending on the agent or product.

Electric vehicles are expensive items and this should be reflected in the decision rule that agents might make when purchasing one. Therefore we feel that peer effects might have a limited impact for certain agents. An individual may be more likely to buy one if a neighbour or friend has one and speaks favourably about it, or someone with very strong environmental concerns may be more inclined to buy one. However, income will constrain the available options. To initially calibrate the model we create four different consumer groups. We segment agents based on their IU and EU. This is described in table 3.4 below.

Table 3.4: Parameter weightings for agent preferences

| Group | Category | Rule 1 | Rule 2 | Income ($\alpha$) | Environmental ($\beta$) | Social Network ($\gamma$) | Population ($\delta$) |
|---|---|---|---|---|---|---|---|
| 1 | Low income | $IU \leq 0.3$ | | rand(0.3-0.7) | $(1-\alpha)/3$ | $(1-\alpha)/3$ | $(1-\alpha)/3$ |
| 2 | Mid-high income environmentalists | $IU \geq 0.3$ | $EU \geq 0.7$ | $(1-\beta)/3$ | rand(0.3-0.7) | $(1-\beta)/3$ | $(1-\beta)/3$ |
| 3 | Mid- high income indifferent | $IU \geq 0.3$ | $0.3 \leq EU \leq 0.7$ | $(1-\gamma)/3$ | $(1-\gamma)/3$ | rand(0.3-0.7) | $(1-\gamma)/3$ |
| 4 | Mid-high income non-environmentalists | $IU \geq 0.3$ | $EU \leq 0.3$ | $(1-\delta)/3$ | $(1-\delta)/3$ | $(1-\delta)/3$ | rand(0.3-0.7) |

- Group 1: Low income agents will be financially constrained regardless of their environmental preferences and thus will place a high importance on IU. This is an implicit budget constraint. We allow $\alpha$ take a random value between 0.3 and 0.7
- Group 2: Mid-high income environmentalists will place a high importance on environmental factors. We allow $\beta$ take a random value between 0.3 and 0.7
- Group 3: Mid-high income indifferent agents may be influenced by word of mouth from their peers. We allow $\gamma$ take a random value between 0.3 and 0.7
- Group 4: Mid-high income non-environmentalists have negative attitudes and behaviour towards environmental issues and will only adopt if a high proportion of the total population adopt. We allow $\delta$ take a random value between 0.3 and 0.7

## 3.4 Simulations and results

### 3.4.1 Individual preferences versus peer effects

Our first round of simulations provide a sensitivity check to better understand the interaction between individual preferences and peer effects in driving adoption. We randomly select a sample of $N = 1600$ agents from our population and connect them using preferential attachment networks. We set a relatively low threshold level of $\theta \sim N(0.45, 0.15)$. This is to allow a small number of innovators to adopt the product initially. We set homogenous weights for agents and run a range of simulations under three broad categories.

Table 3.5: Personal preferences vs peer effects

| Description | Personal weights | Peer weights |
|---|---|---|
| High personal preferences/Low peer effects | $\alpha_i + \beta_i = 0.8$ | $\gamma_i + \delta_i = 0.2$ |
| Moderate personal preferences/Moderate peer effects | $\alpha_i + \beta_i = 0.5$ | $\gamma_i + \delta_i = 0.5$ |
| Low personal preferences/High peer effects | $\alpha_i + \beta_i = 0.2$ | $\gamma_i + \delta_i = 0.8$ |

The results are reported in Appendix 3.C. We find that adoption levels are highest when the weight on personal preferences is moderate ($\alpha_i + \beta_i = 0.5$). A high weight on personal preferences ($\alpha_i + \beta_i = 0.8$) also leads to high adoption rates, while a low weight on personal preferences ($\alpha_i + \beta_i = 0.2$) results in very low adoption levels, completely inhibiting the process in some cases. This is because the system needs some individuals to be driven by their own preferences in order for the diffusion process to take off. Once it reaches a critical mass it requires other agents to place a value on the positive network externality they receive from others in their network adopting. This will in turn induce them to adopt. If agents place too a high weight on this from the start we will not have the innovators to start the process. It is clear that within the first group of simulations ($\alpha_i + \beta_i = 0.8$), adoption is highest for higher values of $\beta$, this is because the underlying distribution of EU is negatively skewed, while that of IU is closer to a normal distribution.

For this threshold level we find that the critical weight on personal preferences is approximately 0.35. A lower weight results in almost no adoption, a higher weight leads to very high adoption levels.

### 3.4.2 The effect of seeding and network type on adoption level

For this set of simulations we again select a random sample of $N = 1600$ agents from our population described in section 3.3.2. We set threshold distribution $\theta \sim N(0.65, 0.15)$ to generate relatively low adoption levels[4].

Something we can examine with a network model is whether it matters who adopts first within the population. In order to do this we seed different groups of agents, i.e. we initialise them as being EV owners. To seed $agent_i$, we set $x_i = 1$ at time $t = 0$. This may be policy relevant for initiatives which target specific consumer groups in order to encourage the mass adoption of EVs. Three different seeding methods are examined.

1. Seed random: We seed random samples of 5% of the agent population
2. Seed early adopter: We seed the 5% of the agent population with the highest probability of adoption, given their personal preferences
3. Seed most influential nodes: We seed the 5% of the agent population with the greatest number of connections. This can be either a random or early adopter seeding. To do this we place a 5% seeding, then re-arrange the adjacency matrix to make these the most connected nodes in the network

Results are reported in fig. 3.7 and tables 3.6 and 3.7 below. We first connect agents with preferential attachment networks, table 3.6. Without intervention average adoption after 15 time periods is just over 9%. We then compare seeding a random group of agents with seeding the early adopters. Interestingly a random seeding yields a higher mean adoption level. We suspect this is because

[4] Results have been verified for a range of threshold distributions.

the early adopters are likely to adopt anyway. However, by activating a random group of agents, adoption is spreading to agents and groups that would otherwise not have adopted. Zhang and Nuttall (2007) found similar results in their ABM of smart-metering technology adoption.

Within a preferential attachment network some nodes have a much greater number of connections than others. We investigate the outcome when the most influential nodes adopt first. The results from this are represented with the dotted lines. We find that seeding is much more effective when the most influential are targeted. Again, a random distribution results in higher adoption than if the high probability agents adopt first. However, in reality it is impossible to fully map the network topology. This simulation is run to illustrate the point that if we aim to encourage high levels of adoption through the targeting of specific consumer groups, we need much better data on the channels through which technology diffuses.

These results also hold for small-world networks as can be seen in table 3.7. We must be cautious when interpreting these results, due to limitations with the model. In particular, the adjacency matrices we use are symmetrical and we do not place different weights on the links between nodes in the network. In reality, as well as having many connections, some individuals are likely to have a greater weight to their connections.

Table 3.6: EV uptake: Preferential-attachment network

| Intervention | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| No seeding | 9.28 | 1.04 | 7.00 | 12.25 |
| Random 5% | 15.93 | 1.04 | 13.44 | 18.63 |
| Early Adopter 5% | 14.18 | 0.89 | 11.50 | 16.63 |
| Random most influential 5% | 23.31 | 1.51 | 18.25 | 26.81 |
| Early Adopter most influential 5% | 21.45 | 1.38 | 18.19 | 24.63 |

Note: Results reported after 100 simulations of $N = 1600$ agents
$\theta \sim N(0.65, 0.15)$ for all simulations

Table 3.7: EV uptake: Small-world network

| Intervention | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| No seeding | 9.25 | 0.95 | 7.00 | 11.44 |
| Random 5% | 15.48 | 1.01 | 12.69 | 17.56 |
| Early Adopter 5% | 14.04 | 0.90 | 11.50 | 16.25 |

Note: Results reported after 100 simulations of $N = 1600$ agents
$\theta \sim N(0.65, 0.15)$ for all simulations

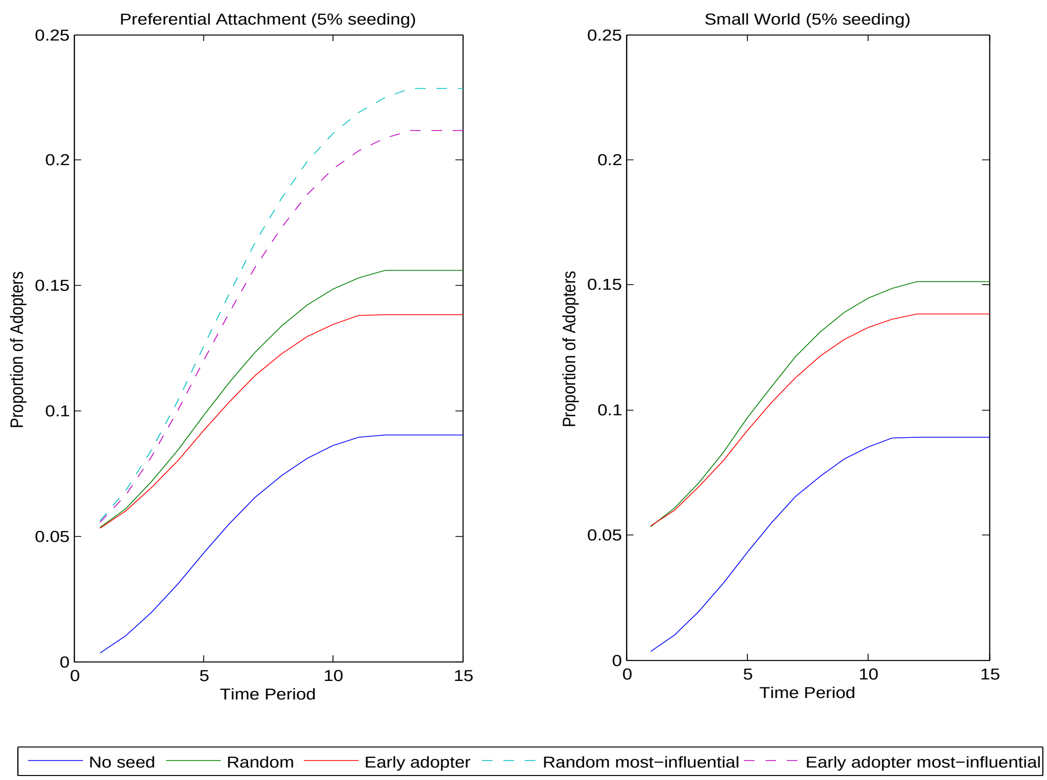


Figure 3.7: EV adoption levels for alternate seeding strategies on random agent sample

Note: Results reported for average uptake after 100 simulations of $N = 1600$ agents. $\theta \sim N(0.65, 0.15)$ for all simulations

In reality though, neighbourhood composition is not random and clusters may develop if certain types of individual self-select into particular areas. We will examine this in the next section.

### 3.4.3 EV adoption for specific neighbourhoods

To motivate this section we use Census 2011 data to generate a "heat-map" of areas with high proportions of likely early adopters in Dublin. This is displayed in fig. 3.8. To rank Census small-areas we used Ward's method, as developed by Ward (1963), a hierarchical clustering algorithm. These results are purely descriptive, do not feed into the simulations and were generated only to graphically represent the level of spatial heterogeneity that may exist.

Dublin is divided into over 4000 small-areas. The objective of Ward's method is to group observations into clusters by minimising within-cluster variance. Each individual area/observation is initialised as a cluster. Clusters are then merged with each other, at each step finding the pair of clusters that leads to lowest within-cluster variance.

We decided after trial and error to allow 10 distinct clusters. We merged on:

1. absocial. The proportion of individuals of AB social class in each area
2. carwork. The proportion of individuals who drive to work
3. house. The proportion of individuals who live in a house

This work is informed by Campbell et al. (2012) who employ this method to determine the spatial location of potential early EV adopters in Birmingham. Descriptive statistics on clusters used can be found in Appendix 3.B.

There is quite a degree of spatial heterogeneity, as can be seen. We can also see areas with significant proportions of potential adopters. This is interesting as even weak peer-effects in these areas may induce very high adoption levels. To simulate this process we generate agent populations that represent the distributions of households within these areas.

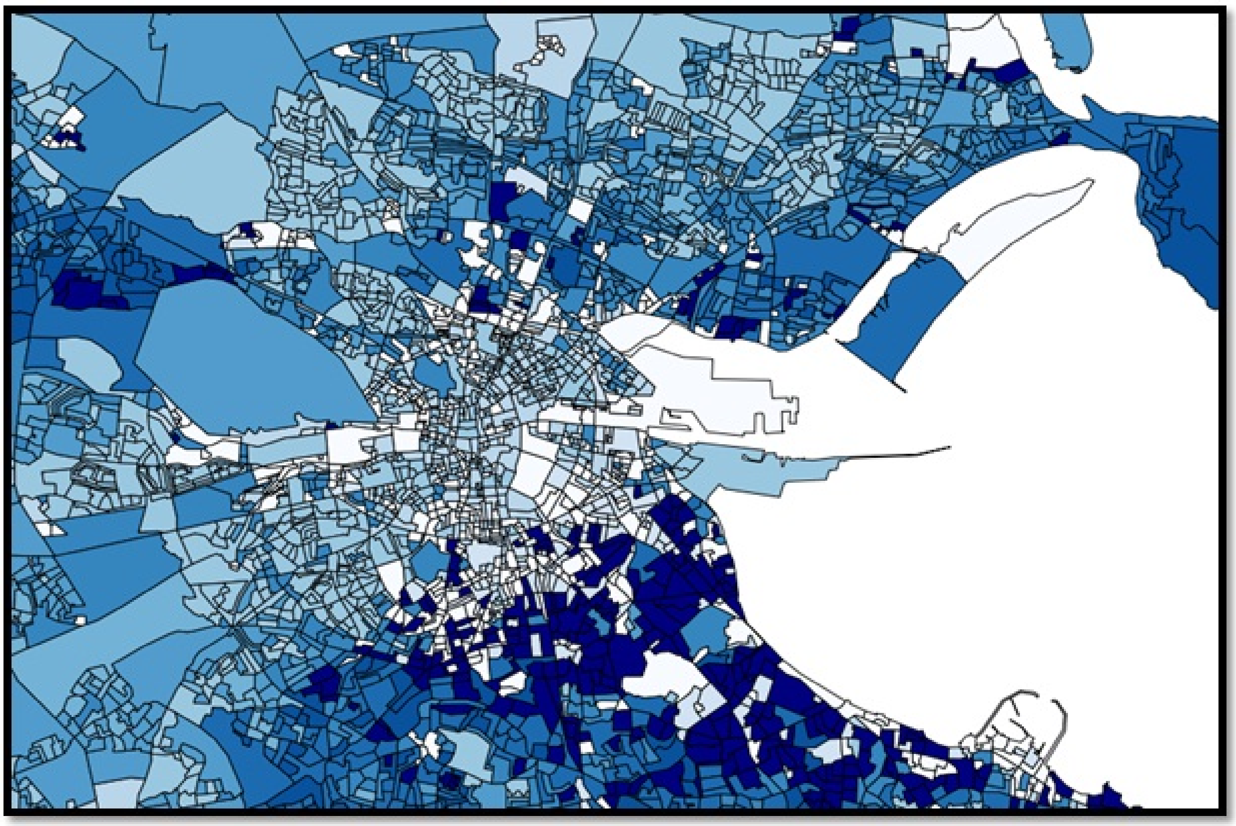

Figure 3.8: Spatial distribution of likely early adopters for Census Small-Areas in Dublin

Note: There are on average 100 households in each area. Areas are ranked 1-10. 1 White(dark blue) indicates a very low(high) proportion

### Adoption levels

In this section we focus on four neighbourhoods, based on Census Electoral Districts (EDs) which vary in their socioeconomic characteristics. The selected EDs are Clonskeagh-Roebuck, Dalkey-Hill, Palmerstown West and Arran Quay C. We select houses based on the social class of their chief income earner and tenure type. Synthetic populations were created using PopGen, an iterative proportional fitting (IPF) algorithm, documented in Ye et al. (2009).

Details on fitness and neighbourhood distributions of characteristics can be found in Appendix 3.D.

Table 3.8: Neighbourhoods included in simulations

| Number | Electoral District (ED) | Reason for inclusion |
|---|---|---|
| 1 | Clonskeagh-Roebuck | Current location of a electricity distribution test network examining the effect of EV charging |
| 2 | Dalkey-Hill | High proportion of social class AB and homeowners |
| 3 | Palmerstown West | High proportion of social class DE and local authority renters |
| 4 | Arran Quay C | Good mix of social class and high proportion of private renters |

Below are the resulting distributions of preferences for each ED.

We use the previous model specification and parameter set up. We split our grid into four quadrants to represent each ED. This process no longer generates the smooth adoption S-curves previously observed when using a random sample, observed in fig. 3.7. This is because we now have groups of similar individuals and their preference distribution can take discrete jumps, potentially causing a number of similar agents to adopt within the same time period as each other.

Results are reported below in fig. 3.10 and tables 3.9 and 3.10. The wealthier neighbourhoods (1,2) with a greater proportion of homeowners have much higher adoption levels than the less wealthy (3,4), as expected. Interestingly, the choice of network used to connect agents now begins to have a greater impact on the results.

Preferential attachment networks lead to average adoption levels of 17-18% in the wealthier neighbourhoods and 0.7-3.2% in the less wealthy neighbourhoods.

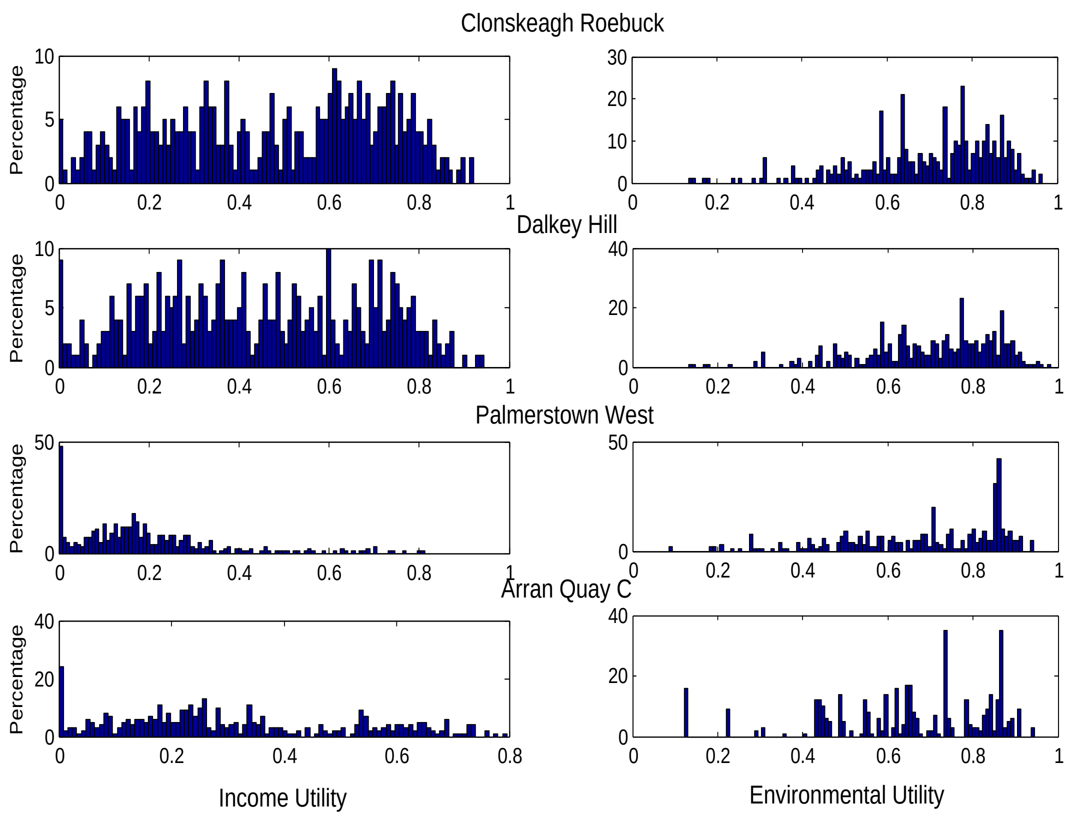


Figure 3.9: Distribution of Income Utility (IU) and Environmental Utility (EU) for different EDs

However, small-world networks generate over 10% higher average adoption levels for the wealthier neighbourhoods than preferential attachment networks. We can easily get to 30% adoption for these areas. This level is very sensitive to changes in threshold distribution and a reduction of threshold mean from 0.65-0.55 can induce uptake of over 50% for these areas. As discussed in section 3.3.2 this lowering of threshold is analogous to the cost or barriers to adoption dropping. For instance, if technological developments allowed price to fall or if charging infrastructure improved for a particular area the cost to adoption for agents in that area would fall. This high adoption in certain areas may have an impact on the distribution network in these areas. We will discuss this in the next section.

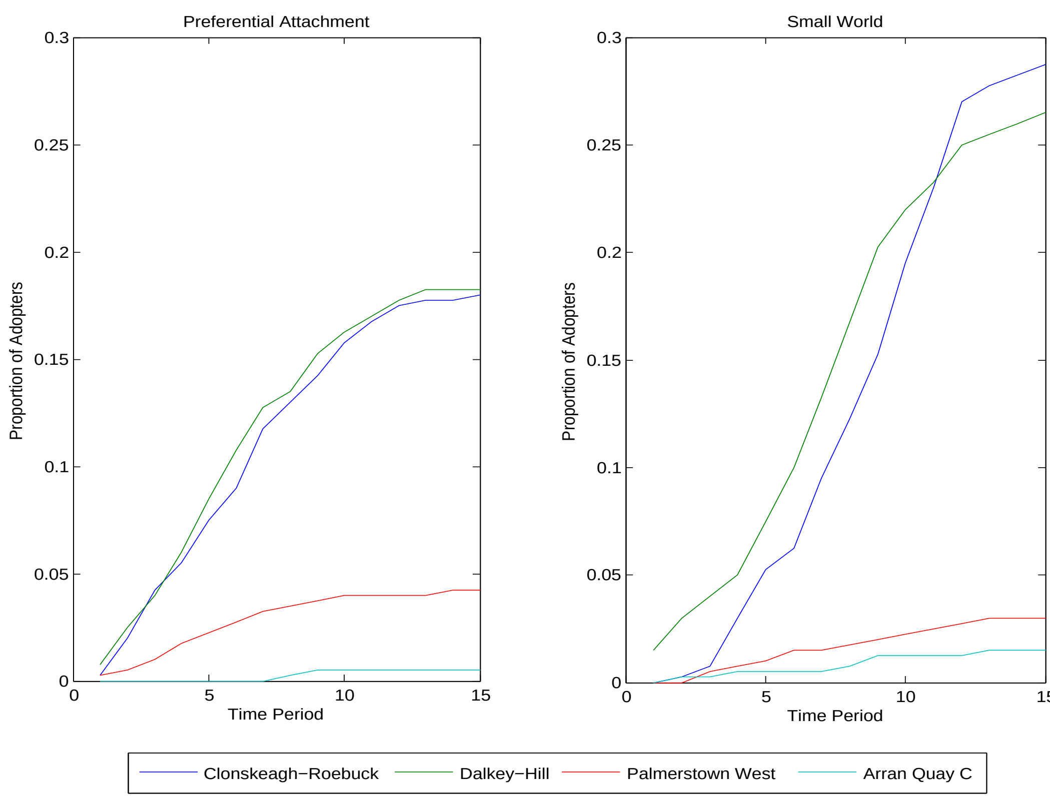


Figure 3.10: EV adoption levels for agent samples based on specific areas

Note: Results reported for average uptake after 100 simulations of $N = 1600$ agents. $\theta \sim N(0.65, 0.15)$

Table 3.9: EV uptake for different neighbourhoods: Preferential-attachment network

| ED | Mean | Std. Dev | Min | Max |
|---|---|---|---|---|
| Clonskeagh-Roebuck | 18.02 | 1.02 | 16 | 21 |
| Dalkey-Hill | 17.61 | 1.1 | 15.5 | 21 |
| Palmerstown West | 3.18 | 0.42 | 2.5 | 4.25 |
| Arran Quay C | 0.69 | 0.38 | 0 | 1.75 |

Note: Results reported after 100 simulations of $N = 1600$ agents
$\theta \sim N(0.65, 0.15)$ for all simulations

Table 3.10: EV uptake for different neighbourhoods: Small-world network

| ED | Mean | Std. Dev | Min | Max |
|---|---|---|---|---|
| Clonskeagh-Roebuck | 29.77 | 0.61 | 27 | 30.5 |
| Dalkey-Hill | 31.79 | 0.33 | 30.75 | 32.5 |
| Palmerstown West | 3.97 | 0.1 | 3.75 | 4.25 |
| Arran Quay C | 1.99 | 0.035 | 1.75 | 2 |

Note: Results reported after 100 simulations of $N = 1600$ agents
$\theta \sim N(0.65, 0.15)$ for all simulations

**Local clustering**

The most interesting results occur when we examine the level of clustering of adopters. We find small-world networks yield a smaller number of clusters than preferential attachment networks. However cluster size increases dramatically if agents are connected locally using small-world networks. If agents are connected to just their two nearest neighbours ($k = 1$), mean cluster size increases from 2 to 12. Cluster size further increases as we increase the number of local connections agents are allowed to have. We can't generalise these results as they may be particular to the areas in question, and our networks are a stylised version of reality. However, we can say that even mild peer-effects could induce dense clusters in areas with high proportions of adopters.

Table 3.11: Clustering results

| Network type | Number of clusters | | | Size of clusters | | |
|---|---|---|---|---|---|---|
| Small-World | Mean | Max | Min | Mean | Max | Min |
| $k = 1$ | 21.06 | 24 | 17 | 12.26 | 295 | 2 |
| $k = 2$ | 21.06 | 24 | 17 | 22.47 | 242 | 2 |
| $k = 3$ | 21.06 | 24 | 17 | 23.78 | 237 | 2 |
| $k = 4$ | 15.32 | 17 | 13 | 33.84 | 236 | 2 |
| $k = 5$ | 21.06 | 24 | 17 | 26.66 | 249 | 2 |
| Preferential attachment | 23.54 | 29 | 16 | 2.36 | 9 | 2 |

Note: Results reported for 100 simulations of $N = 1600$ agents
$\theta \sim N(0.65, 0.15)$

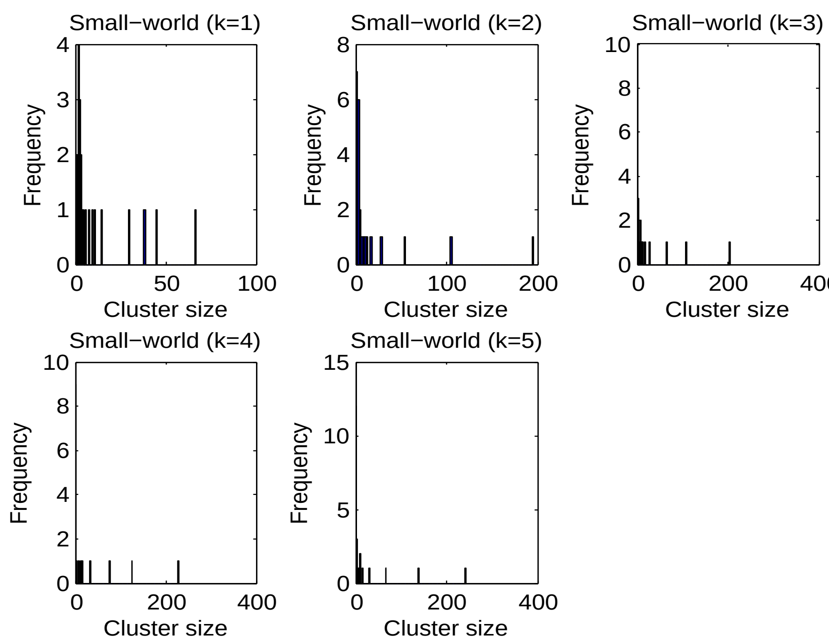


Figure 3.11: Distribution of cluster size for small-world networks

In a study of a test network of 134 households in Dublin, Richardson et al. (2010) found that severe voltage drop and unsafe thermal loading of components can occur when EV penetration reaches 27-44% of the households. Clusters occurring further from the transformer have a greater impact. From the distribution of cluster size we generate, it is quite possible to observe clusters that in reality would be large enough to cause the network to exceed safe operating limits. This problem would be greater if clusters are located towards the end of network line, away from the substation bus.

## 3.5 Concluding remarks

In this paper we present a simple and stylised agent-based model of EV adoption amongst Irish households. The motivation was to create a tool in order to model the drivers of aggregate adoption, the formation of clusters of adopters at a local level and spatially explicit adoption profiles.

This is relevant because we know that spatial dependencies do exist in the adoption of "green" technologies, and if this exists for EVs it may cause negative externalities in the form of reduced life for electricity network assets.

We use an agent-based methodology as there is potential for micro-interactions to generate aggregate outcomes that can't be deduced by merely examining the preferences of individuals. We adapt a methodology used by others and apply it to the diffusion of EVs amongst Irish households.

Agents are created based on detailed, nationally representative survey data. We define simple rules that govern their behaviour, link them with other agents, through different network types and place them within an environment. We then use a spatial microsimulation algorithm to generate agent populations that represent specific geographic areas of interest.

We find that it is important who adopts first in determining the overall diffusion level in the population. We demonstrate that policies targeting so called "early-adopters" are limited unless network topology, the conduit through which technology diffuses, is better understood.

In terms of clustering, we find that mild peer-effects can induce very high adoption levels in certain areas. Even if overall adoption levels are quite low, this could be highly clustered in certain areas and peer-effects within those areas could significantly increase cluster size. This could lead to increased costs for electricity network operators and ultimately for consumers, as the average cost of improvements to the network will be socialised.

There are many limitations to this model. We do not claim to fully understand agent preferences - we merely use survey data to generate distributions that should be considered place-holders in order to calibrate this model. Nor do we claim to create realistic social networks - we experiment with different network types, both of which contain some of the characteristics of actual human social networks. We do not suppose to fully describe the various consumer groups that may exist - we use some simple rules to segment the population into different groups to illustrate how heterogenous groups can alter the dynamics of such a system.

One method of improving this model would be to calibrate the diffusion curves produced to actual data. This would be particularly important if we were trying to predict diffusion rates. This would also allow us to more accurately represent actual elapsed time, rather than time being a function of the number of model iterations. However, for this research we are more interested in the dynamics of the process and as a result this is less of a concern. Also, due to the small number of private EV owners currently in Ireland[5], this would be quite difficult. Another improvement would be to integrate data on the existing charging infrastructure into our spatial model, as proximity to public charging points might influence the decision to purchase an EV.

Clearly a disruptive technology must compete with the incumbents. Driscoll et al. (2013) use revealed preference data to simulate likely market shares for three new EVs, based on the values consumers place on observable vehicle characteristics. They find that in the absence of incredible levels of subsidies, market share for them will remain well below 10%. Our research is different in that we ignore product characteristics but we can change our parameters to allow for any level of market share. However a model which incorporated information on the existing car fleet would be another useful extension.

[5]Approximately 500 in 2015.

## 3.A Data

(1) CER Smart Metering Customer Behavioural Trial. This dataset has been described in detail in Chapter 2. See also CER (2011) for further information.

(2) Census 2011 Small Area Population Statistics (SAPS). This is based on the Irish Census, most recently conducted in 2011. Data is categorised into 46 different tables covering a range of themes, see table A1. This data is disaggregated to Small Area level, of which there are approximately 18,500 across the country; typically these would consist of 80-100 households. We have based our agent populations on Electoral Districts (EDs) of which there are approximately 3400.

Table A1: Census SAPS themes

| Census SAPS Themes | |
|---|---|
| Theme 1: Sex, age and marital status | Theme 9: Social class and socio-economic group |
| Theme 2: Migration, ethnicity and religion | Theme 10: Education |
| Theme 3: Irish language | Theme 11: Commuting |
| Theme 4: Families | Theme 12: Disability, carers and general health |
| Theme 5: Private households | Theme 13: Occupation |
| Theme 6: Housing | Theme 14: Industries |
| Theme 7: Communal establishments | Theme 15: PC and internet Access |
| Theme 8:Principal status | |

Source: Central Statistics Office (CSO), Census 2011

## 3.B Results from Ward's clustering algorithm

Table B1: Clusters generated for Dublin small-areas

| wards10 | | absocial | carwork | house |
|---|---|---|---|---|
| 1 | mean | 0.255 | 0.465 | 0.048 |
| $N = 196$ | sd | 0.118 | 0.096 | 0.047 |
| 2 | mean | 0.392 | 0.234 | 0.060 |
| $N = 244$ | sd | 0.091 | 0.113 | 0.072 |
| 3 | mean | 0.130 | 0.123 | 0.122 |
| $N = 347$ | sd | 0.083 | 0.085 | 0.095 |
| 4 | mean | 0.202 | 0.220 | 0.549 |
| $N = 511$ | sd | 0.127 | 0.092 | 0.139 |
| 5 | mean | 0.332 | 0.472 | 0.427 |
| $N = 231$ | sd | 0.121 | 0.093 | 0.136 |
| 6 | mean | 0.202 | 0.429 | 0.975 |
| $N = 727$ | sd | 0.057 | 0.048 | 0.025 |
| 7 | mean | 0.345 | 0.388 | 0.975 |
| $N = 748$ | sd | 0.063 | 0.071 | 0.031 |
| 8 | mean | 0.334 | 0.364 | 0.757 |
| $N = 404$ | sd | 0.120 | 0.091 | 0.074 |
| 9 | mean | 0.120 | 0.277 | 0.952 |
| $N = 952$ | sd | 0.067 | 0.058 | 0.044 |
| 10 | mean | 0.519 | 0.387 | 0.972 |
| $N = 446$ | sd | 0.065 | 0.060 | 0.041 |
| Total | mean | 0.261 | 0.332 | 0.734 |
| $N = 4806$ | sd | 0.150 | 0.123 | 0.339 |

## 3.C Individual preferences versus peer effects

Note: Values reported for 100 simulations of $N = 1600$ agents at $t = 15$. $\theta \sim N(0.45, 0.15)$

| Weights | Economic ($\alpha$) | Environ-mental ($\beta$) | Social Network ($\gamma$) | Population ($\delta$) | Mean | St Dev | Max | Min |
|---|---|---|---|---|---|---|---|---|
| High personal preferences | 0.7 | 0.1 | 0.1 | 0.1 | 55.7 | 1.6 | 59.6 | 51.8 |
| ($\alpha + \beta = 0.8$) | 0.5 | 0.3 | 0.1 | 0.1 | 73.8 | 1.3 | 76.3 | 69.9 |
| | 0.3 | 0.5 | 0.1 | 0.1 | 85.8 | 0.9 | 87.8 | 82.2 |
| | 0.1 | 0.7 | 0.1 | 0.1 | 91.0 | 0.7 | 92.3 | 88.8 |
| Moderate personal preferences | 0.4 | 0.1 | 0.4 | 0.1 | 42.2 | 7.1 | 60.3 | 25.9 |
| ($\alpha + \beta = 0.5$) | 0.4 | 0.1 | 0.1 | 0.4 | 62.0 | 6.0 | 73.2 | 39.5 |
| | 0.2 | 0.3 | 0.4 | 0.1 | 89.3 | 2.7 | 94.3 | 80.8 |
| | 0.2 | 0.3 | 0.1 | 0.4 | 97.3 | 0.4 | 98.1 | 96.2 |
| | 0.1 | 0.4 | 0.4 | 0.1 | 95.3 | 1.1 | 97.9 | 92.3 |
| | 0.1 | 0.4 | 0.1 | 0.4 | 98.3 | 0.3 | 99.0 | 97.6 |
| Low personal preferences | 0.1 | 0.1 | 0.7 | 0.1 | 2.7 | 2.3 | 8.0 | 0.0 |
| ($\alpha + \beta = 0.2$) | 0.1 | 0.1 | 0.5 | 0.3 | 1.2 | 1.2 | 5.1 | 0.0 |
| | 0.1 | 0.1 | 0.3 | 0.5 | 1.0 | 1.0 | 3.3 | 0.0 |
| | 0.1 | 0.1 | 0.1 | 0.7 | 0.8 | 0.8 | 2.1 | 0.0 |

## 3.D Goodness-of-fit for synthetic agent populations

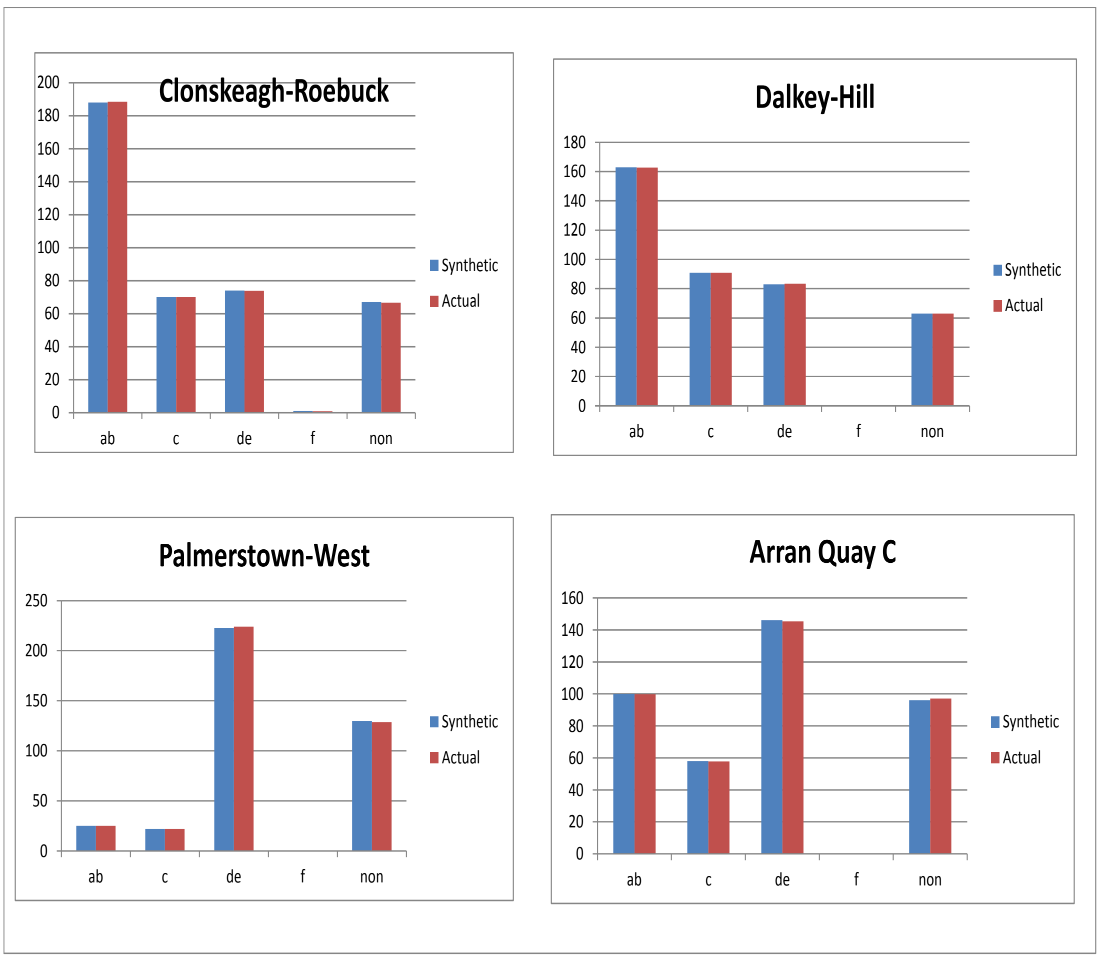


Figure D1: Actual vs synthetic populations: Social class

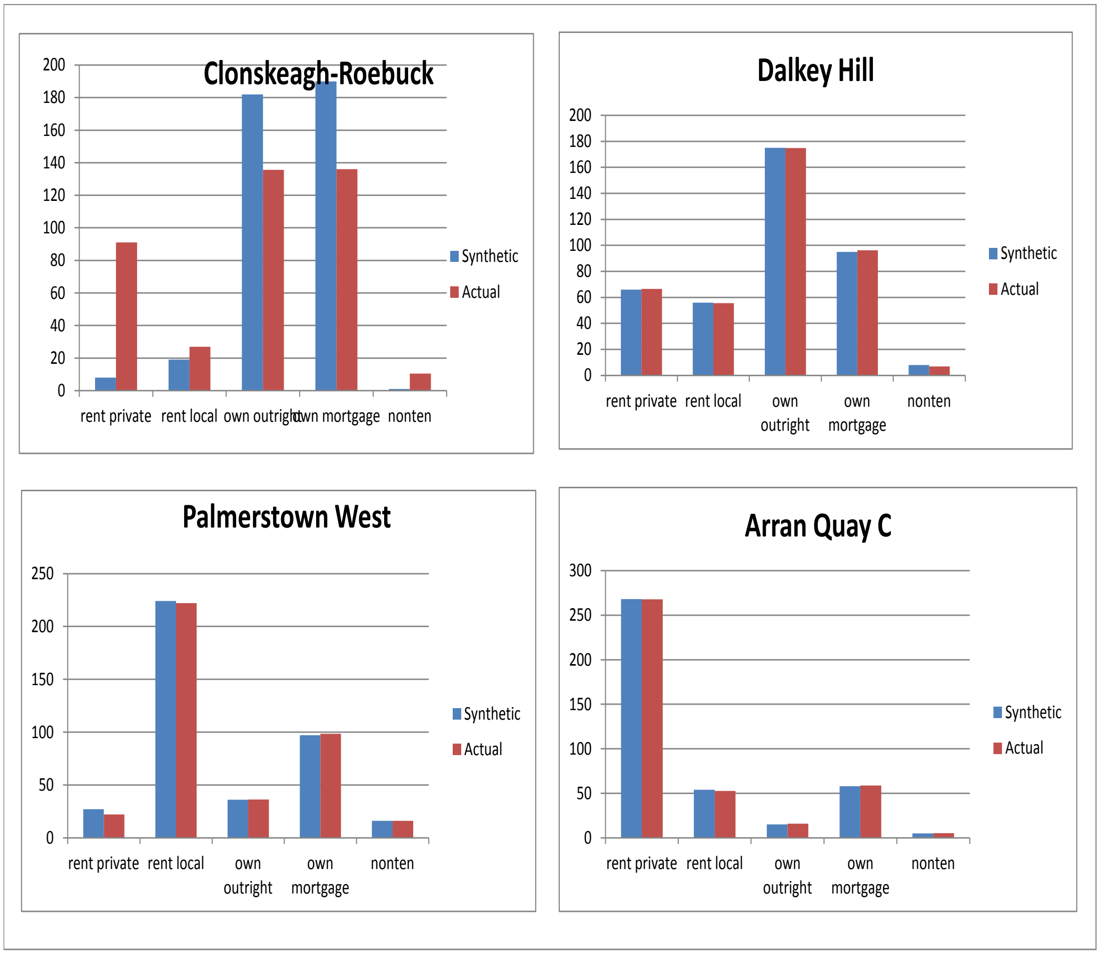


Figure D2: Actual vs synthetic populations: Tenure type

## 3.E Sensitivity analysis

### A Threshold distribution

Due to the model's sensitivity to changes in threshold distribution, It is worth examining this in more detail. We set thresholds to be normally distributed. Both mean and variance have a significant and non-linear influence on the adoption level. An agent with a threshold of 1 will never adopt, regardless of its preferences. An agent with a threshold of 0 will always adopt. As threshold is bounded between 0 and 1, any agent with a threshold greater than 1, will be assigned a threshold of 1. For example, in fig. 3.3.1 in section 3.3.2 we set the threshold mean to 0.35 with standard deviation of 0.15, about 2.5% of the agents will always adopt. The opposite is true in fig. 3.3.2, about 2.5% will never adopt.

To further investigate this effect we run a number of simulations for a variety of threshold distributions. We observe adoption levels of 86% given a mean threshold level of 0.5, however if we increase the mean to 0.6, the adoption level drops to 45%. These non-linearities are in part due to the boundary condition imposed on the distribution. Another finding is that the standard deviation of the threshold distribution has a significant effect on the overall adoption level. This effect is more stable and makes intuitive sense, for any given mean, a wider dispersion results in a greater range of thresholds, allowing more agents to be activated.

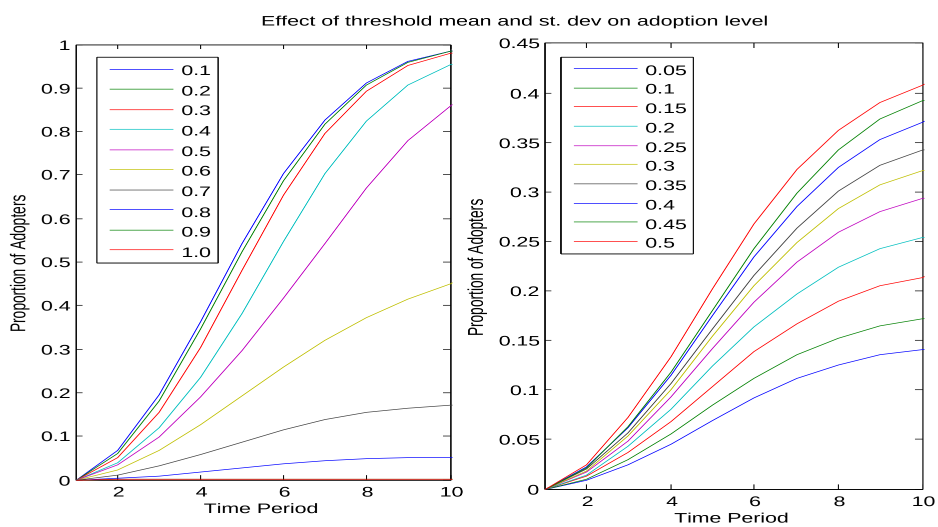


Figure E1: The effect of threshold distribution on adoption level

Note: Average adoption levels reported at time $t = 10$ for 100 simulations of $N = 1000$ agents

## B Initial uncertainty

In this section we will outline the effect of changing the technology/information deficiency parameter "*crit*" that we introduced in subsection 3.3.5. This could be considered a decaying stochastic cost term, or a measure of diminishing risk-aversion. In the initial periods, agents will be reluctant to adopt even if the benefits to adoption outweigh the cost to them.

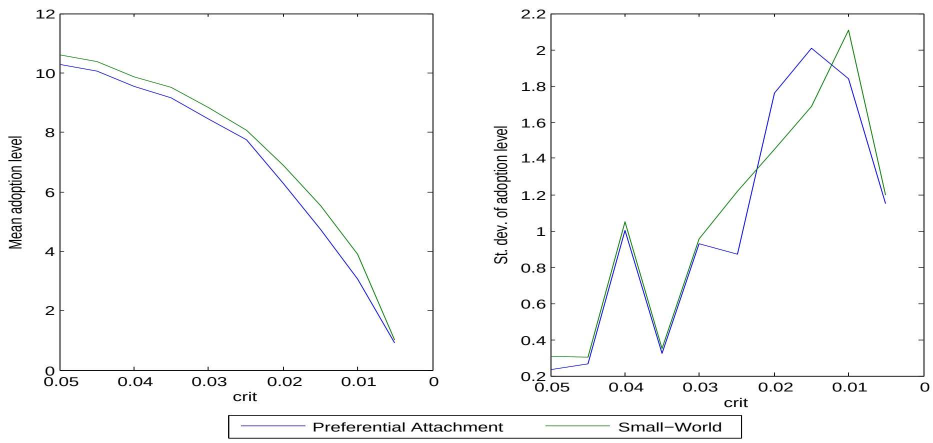


Figure E2: The effect of initial inertia on adoption level

Note: Average adoption levels reported at time $t = 10$ for 100 simulations of $N = 1600$ agents. Uncertainty is inversely proportional to *crit* value. Note inverted scale on x-axis

Fig. E2 illustrates the effect of this term on the mean and standard deviation of the adoption level. We find that high levels of initial uncertainty (low crit value) can completely inhibit adoption, regardless of the average preferences of the population. As we reduce the level of initial uncertainty (increase crit) the variability in adoption levels reduce and begin to converge towards the equilibrium level, for a given population of agents. We observe broadly similar results for both small-world and preferential attachment networks.

# Chapter 4

# The impact of broadband and other infrastructure on new business establishments

## 4.1 Introduction

The focus of this thesis now moves from households to firms, and from technology adoption and ICT in electricity networks, to infrastructure roll-out more generally. Along with other factors, we specifically examine the potential for ICT roll-out and the spatial configuration of electricity networks to influence economic activity.

Regional disparities in economic activity and outcomes become of increased concern to policymakers in times of recession. The previous decade has been no exception in this regard. However, this feature of economic geography has been of interest to economists since Adam Smith (1776). Krugman (1991) describes a phenomenon which is extremely path dependent, and has shown remarkable persistence over time and space.

Within this broader literature, the determinants of new business establishments has emerged as a subject of great interest to economists, policymakers, regional scientists and geographers. In one of the earlier works in this space, Carlton (1983) argued that new business establishments give a better indication of a region's future development than current employment levels, as newly locating plants are responding to current incentives, whereas current employment levels depend primarily on prior location decisions.

Recent advances in GIS modelling and computational power, along with improved spatial data on firm locations and their determinants have enabled researchers to examine this question at increasingly disaggregate geographic scales.

A large body of literature examines this question through the lens of the agglomerative forces which encourage firm entry and boost productivity through the benefits of labour market pooling, efficiency gains through sharing of infrastructure and suppliers, and benefits accruing from knowledge spill-overs. A disaggregated approach is useful in this regard as the geographic span of such agglomerative forces can be quite limited (Duranton and Overman, 2005).

Recent work has explored the diffusion of ICT infrastructure as a factor in explaining regional variation in new business establishments, particularly for knowledge-intensive firms. On the demand side, market access is a widely cited factor, typically characterised as a function of distance to market, which will depend on transport infrastructure or cost, along with other measures of accessibility. Another more recent literature explores the interaction of this with agglomeration economies, as accessibility changes over time.

The determinants of new business establishments will vary by firm and geography. Different types of firm will require different types of infrastructure in their production processes. However, within much of this literature "infrastructure" is generally either ICT, transport infrastructure such as motorways, or some other measure of distance or cost to market. Very little exploration has been undertaken into the relative importance of different types of infrastructure. We contribute to this question in a number of ways.

We examine the factors influencing new business establishments for both domestic and foreign firms in Ireland over a period of significant infrastructural investments. Our data allow us to split firms by sector and by the skill-level of their employees.

Our data cover the introduction of broadband infrastructure in Ireland, and we create a unique panel-dataset on traditional digital subscriber line (DSL) and middle-mile fibre-broadband networks to disentangle the relative impacts of each. Our DSL roll-out data spans 0-100% penetration over our sample period, with substantial local variation. We also access GIS information on a national fibre-optic network roll-out investment along with detailed information on the characteristics of these networks, and when they became operational.

Another feature of this period is the extensive upgrade the motorway network received; 86% of the current network was constructed and average drive times to the nearest motorway junction almost halved between 2002-2011. We have detailed spatial and temporal information on this infrastructure upgrade.

Complementing this, we have a rich dataset of other accessibility measures, such as drive-time to airports, train-stations and third-level institutions, along with detailed GIS maps of the electricity transmission infrastructure. This information is supplemented with other important firm location determinants such as agglomeration economies, human capital and data on the relative cost and quality of the labour force.

Results show that both initial DSL and middle-mile fibre have had a positive impact on firm counts, particularly in the high-tech sectors. Accessibility measures, such as driving times to motorway junctions and airports matter more for high-tech FDI than other firm types. New firms of all types value diversity of skills above specialisation. Human capital and access to third-level institutions are important for all types of firm, but particularly those in high-tech sectors. We address concerns of endogeneity between our outcome and explanatory variables and make a case for the direction of causality of our results.

The electricity network is treated separately in Appendix 4.A, as limitations with the data and difficulties related to model convergence render these results less convincing than the others.

The chapter is organised as follows; Section 4.2 discusses related literature and places this work in context, Section 4.3 describes the methodological approach employed and considerations undertaken; Section 4.4 the data; Section 4.5 outlines the empirical results; Section 4.6 outlines the robustness measures undertaken and Section 4.7 concludes.

## 4.2 Previous literature

The literature on new business establishments is extremely broad, spanning a number of areas. Given this, and the wide range of explanatory factors we are able to account for, our discussion initially covers a number of separate strands, before integrating them together in describing our empirical approach.

### 4.2.1 Broadband infrastructure

High-speed ICT infrastructure is a relatively new development. Even as recently as the early 2000s authors were discussing the imminent communications revolution and the substantial benefits it would bring (Parker, 2000).

The previous communications revolution in telephony brought enormous benefits to rural locations. Indeed it was farmers in isolated locations, rather than urban businesses that reaped the greatest rewards from "*the distance-killing capability of telephones*" (Parker, 2000). It was felt that high-speed ICT would do something similar, and that rural locations without sufficient access to a global economy would suffer an inevitable decline.

From the perspective of the policy-maker, it is hoped that broadband can alleviate the negative externalities associated with rural locations, remove some of the barriers of geography and reduce the cost associated with trade, particularly for firms engaged in ICT intensive activities. As discussed in Chapter 1, policymakers can be quite optimistic about the potential benefits broadband roll-out can bring to rural areas.

From an economic perspective, the relationship between broadband and economic growth is more nuanced. In macro terms, lowering the cost of transmitting data should improve productivity and raise output, and has been shown to have a positive impact on economic growth (Koutroumpis, 2009). However the distributional consequences of this are unclear. If substitution effects dominate, ICT might displace less technically advanced workers and have a negative effect on employment. On the other hand, it may enable higher skill workers to be more productive, complementing employment in these sectors.

From a geographic perspective, we might then expect areas with pre-existing high levels of human capital and knowledge intensive firms better equipped to reap the rewards of ICT roll-out than other areas less endowed with these attributes. Indeed much of the literature finds this to be the case. A number of studies have found that the positive impact of broadband is more pronounced in urban areas and for knowledge-intensive firms (Gillet et al., 2006; Kandilov and Renkow, 2010), and within urban areas the effect can depend on area size and industrial legacy (Mack and Rey, 2014). Malecki (2003) cautions that rural areas *"should not be taken in by the promise of the most recent technology, because newer ones will continue to appear first in urban areas"*.

While some previous literature in this space has examined a broad range of economic indicators, such as rents, wages, house prices, unemployment and new business establishments (Kandilov and Renkow, 2010; Kolko, 2012); we focus our attention on new business establishments (new firm counts). In terms of this specific literature, it has been found that firm size and industry type are important determinants of the relationship between broadband and firm counts (Mack and Grubesic, 2009).

Much of this work cites a potential endogeneity problem between broadband and economic activity. Various techniques such as instrumental variables (Kolko, 2012; Mack and Rey, 2014), and propensity score matching (Kandilov and Renkow, 2010), are used to overcome this potential source of endogeneity.

We will describe our approach in more detail in the methodology section, but will first describe some other important location determinants of new business establishments.

### 4.2.2 Agglomeration economies, the labour market and human capital

Puga (2010) provides an extensive overview of the magnitude and causes of agglomeration economies, discussing both theoretical and empirical work. This work provides a discussion of the main mechanisms through which agglomeration economies operate; sharing, matching and learning. Sharing of facilities, suppliers, and the gains from individual specialisation of skills and labour; sharing a labour pool which facilitates matching and results in a lower probability of unemployment and labour shortages; and learning through knowledge diffusion, the acquisition of skills through worker networks and the idea that cities can foster innovation.

Notable theoretical work in this space includes Ellison and Glaeser (1997), who develop a model-based approach to describe the clustering, or localisation of manufacturing industries in the United States. They account for industry-specific spill-overs, natural advantages and random chance. Krugman (1991) describes the persistence of the US manufacturing belt over time, discussing this phenomenon using a simple theoretical model that incorporates increasing returns, transport costs and demand.

Previous empirical work shows firm and worker productivity to be significantly higher in areas of greater population densities. Combes et al. (2012) describe the magnitude of the elasticity of productivity with respect to city size as ranging from 0.02-0.10, depending on the sector and estimating procedure employed.

A strand of this literature focuses on whether firms, or which type of firms value diversity or specialisation to a greater extent (Reynolds et al., 1995; Feldman and Audretsch, 1999; Acs and Armington, 2004; Barrios, 2006).

Provided agglomerative forces dominate dispersion forces, on average, local agglomerative economies should also increase the entry of new establishments, through firms seeking to take advantage of local external economies. This has been shown to be the case by Guimaraes et al. (2000) and Crozet et al. (2004) amongst others. Interestingly, in a three-country comparison of manufacturing firm location choices (Ireland, Belgium and Portugal), Barrios et al. (2009) shows that these effects can vary considerably by country. Other work demonstrates how the incidence of the effect can vary by geographic scale (Jofre-Monseny et al., 2011).

The importance of the unemployment rate in determining firm birth rates received much attention in the US in the 1990s. Armington and Acs (2002) provides a discussion of this in relation to domestic firms. The direction of the effect is ambiguous in the literature. High unemployment might encourage entrepreneurial activity, with the resulting formation of new firms subsequently reducing the unemployment rate. However, high levels of unemployment might also indicate a reduction in aggregate demand, thus putting downward pressure on new firm formation, particularly if they are selling into local markets.

Storey (1991) found that the direction of this effect can be related to the empirical approach taken, highlighting differences between time-series and cross-sectional estimation results, while in more recent work Carree (2002) does not find any evidence for the existence of unemployment push factors in firm births in US regions.

In our case we examine both indigenous and foreign-owned new business establishments. Much of the FDI in Ireland is export-oriented so we should not expect these firms to be heavily dependent on domestic demand as indigenous firms. Indeed, if high local unemployment rates put downward pressure on wages, we might expect a positive correlation between this and the number of new foreign firms locating in an area, once we control for human capital and labour costs. Coughlin et al. (1991) found this to be the case, indicating that investment might be attracted to locations with excess labour supply.

### 4.2.3 Transport infrastructure and general accessibility

A number of studies look at the impact of transport infrastructure on new firm counts. Rothenberg (2012) find that a large-scale upgrade to the Indonesian motorway network resulted in significant dispersal of manufacturing firms, but these effects are less felt in other sectors and accrue more in urban areas.

Holl (2004b) finds a similar effect in Portugal for the manufacturing sector and also some service sectors. In this work the author also points to an interesting interaction between transport infrastructure and agglomeration economies that has remained relatively unexplored by the literature; as transport and communications infrastructure improve this will interact with the geographic scale over which agglomeration economies operate and a consensus is yet to emerge as to whether this enhances or reduces these effects. The same author finds a similar effect for manufacturing firms in Spain (Holl, 2004a). In this case benefits are concentrated near the new infrastructure and can induce negative spill-overs in more distant locations.

Gibbons and Van Dender (2012) find that transport infrastructure improvements can cause firm entry and discourage exit, but do not effect employment levels at existing firms. Button et al. (1995) highlight important differences between different types of transport infrastructure, finding that road and air links have a greater importance for inward investment than for domestic firms.

### 4.2.4 Firm location in an Irish context

Much of the previous research in a specific Irish context has been descriptive. Egeraat and Curran (2013) look specifically at the Irish pharmaceutical industry and find that governmental intervention and natural endowments have been more important than agglomeration economies in driving spatial concentration.

Large production plants require substantial effluent disposal, fresh water and electricity. The area around Cork Harbour, particularly Little Island and Ringaskiddy was selected by the Industrial Development Authority (IDA) as a location and they purchased land banks and invested in the required drainage infrastructure. Following this investment, considerable effort was directed at drawing plants toward this location.

Other important work in examining the impact FDI has had on the Irish economy includes Barry and Bradley (1997) and Gleeson et al. (2006), while Morgenroth (2009), explores the economic geography of Ireland through the spatial patterns of employment.

More closely related to our own approach, Barrios (2006) employ a conditional-logit specification to study firm location choices at the county level. They find that urbanisation economies are a significant driver for hi-tech foreign location choices, and ascribe this effect to potential knowledge related externalities in these areas.

## 4.3 Empirical approach and econometric considerations

The literature suggests that both foreign and domestic new business establishments in an area will be a function of agglomerative forces, factors relating to human capital, labour market pooling and relative labour costs. Infrastructure provision and market access will also play an important role. Demand based factors should matter more for domestic firms, as foreign firms locating in Ireland largely use it as an export platform. We will discuss these data in detail in Section 4.4, but will first describe our empirical model of new business establishments.

Within the business establishment literature, two main methodologies are employed; choice models or count models[1]. Choice models are based on McFadden's random utility maximisation framework (McFadden et al., 1973). An attraction of this framework lies in its theoretical grounding. However, it has limitations when the choice set is large, and when a substantial number of areas do not gain

[1]For examples of choice models see Pusterla and Resmini (2007); Siedschlag et al. (2013). For count models see Jofre-Monseny et al. (2011); Bhat et al. (2014)

any new firms. Guimaraes et al. (2003) advocate the use of Poisson count models in these cases and demonstrate the circumstances under which they can return identical parameter estimates. Schmidheiny and Brulhart (2011) advise caution in interpretation, as the economic implication of the prediction differs, in that choice models represent a zero-sum world in which one region's loss is another's gain, but this is not the case in a count framework.

We choose to implement a count framework as we have a large proportion of zeros in our dependent variables and a relatively large choice set. New firm counts in a particular area in each time period are modelled as a function of area characteristics, with time effects (year dummy variables) and sector-type effects included also, for sector-specific regressions.

$$y_{ijt} = \alpha_i + \beta_{X_1} X_{it} + \beta_{X_2} X_{ijt} + \beta_{X_3} X_{kt} + \beta_Z Z_i + \gamma_t + \epsilon_{it} \tag{4.1}$$

Where:

$$y_{ijt} = 0, 1, 2, ... \tag{4.2}$$

$y_{ijt}$ denotes the count of new firms of type $j$ in area $i$ at time $t$. $\alpha_i$ are area specific effects, which are included in some specifications. $X_{it}$ is a matrix of explanatory variables that vary by area $i$ and time $t$. $X_{ijt}$ is a matrix of explanatory variables that vary by by area $i$, time $t$ and firm-type $j$. $X_{kt}$ is a matrix of explanatory variables that vary by a more aggregated region $k$ and time $t$. $Z_i$ is a matrix of time-invariant explanatory variables, $\gamma_t$ are time-dummies.

The econometrics literature encourages the use of fixed-effects (FE) models when dealing with panel-data. Fixed-effects panel models allow for a limited form of endogeneity, in that the individual specific effects $\alpha_i$ are permitted to be correlated with the regressors $X_{it}$. In this case it is possible to consistently estimate the coefficient of interest, $\beta$, for the time-varying regressors by appropriate differencing transformations that eliminate $\alpha_i$ (Cameron and Trivedi, 2010).

However in our case, many of our explanatory variables, particularly those relating to the accessibility of our locations are time-invariant. Furthermore, for certain sample splits we have high-proportions of areas which never receive any firms. It is impossible to estimate a fixed-effects specification for either the coefficients on these variables, or for the areas that do not get any firms, as a differencing transformation will eliminate them from the estimations. Given this, we do not use a fixed-effects specification, but a number of other options remain.

A random-effects (RE) specification makes stronger assumptions, in that the individual effects $\alpha_i$ are assumed to be independent of the regressors. As per Cameron et al. (2013) if $f(y_{it}|x_{it}, \alpha_i)$

denotes the density for the $it^{th}$ observation, conditional on $\alpha_i$ and the regressors, the joint density for the $i^{th}$ observation is:

$$f(y_i|X_i) = \int_0^{\infty} [\prod_{t=1}^{T} f(y_{it}|\alpha_i, x_{it})] g(\alpha_i|\eta) d\alpha_i \tag{4.3}$$

where $g(\alpha_i|\eta)$ is the specified density of $\alpha_i$.

Supposing $(y_{it}|x_{it}, \alpha_i)$ is Poisson distributed with mean $\alpha_i\lambda_{it}$, and that $\alpha_i$ is gamma distributed with mean 1, a normalisation, and variance $1/\gamma$. Integrating out $\alpha_i$, the conditional mean can be expressed as:

$$E[y_{it}|x_{it}] = \lambda_{it} = exp(x'_{it})\beta \tag{4.4}$$

with variance:

$$V[y_{it}|x_{it}] = \lambda_{it} + \lambda_{it}^2/\gamma \tag{4.5}$$

In the case of over-dispersion, this can be extended to a Negative-binomial (NB2) model as per Hausman et al. (1984). In this case $y_{it}$ is iid NB2 with parameters $\alpha_i\lambda_{it}$ and $\phi_i$, where $\lambda_{it} = exp(x'_{it})\beta$. $y_{it}$ has mean $\alpha_i\lambda_{it}/\phi_i$ and variance $(\alpha_i\lambda_{it}/\phi_i) * (1 + \alpha_i/\phi_i)$. To obtain a closed-form solution it is assumed that $(1 + \alpha_i/\phi_i)^{-1}$ is a beta-distributed random variable with parameters $(a, b)$. Estimation is generally conducted with maximum likelihood methods.

The random-effects model can be transformed to a pooled or population-averaged estimator, by averaging out the individual effects (Cameron and Trivedi, 2013). In this case:

$$E[y_{it}|x_{it}] = \alpha.exp(x'_{it})\beta \tag{4.6}$$

Whether to use random-effects or population-averaged estimators depends on the inference one would like to make. Each is estimating a different population parameter, but the results tend to be close in practice. Taking the introduction of broadband as an example; the random (or individual) effects estimator will estimate the change in new business establishments for the same area following the introduction of broadband, while the population-average will estimate this effect for the average area.

From a policy perspective, we are more interested in the average effect, and employ population-averaged estimators as a result. Some would argue that to adequately control for unobserved variables and gain insight into the individual-level dynamics, a properly specified random-effects model is required (Neuhaus et al., 1991). Given this concern, we also report results from this specification as a robustness check[2]. All models are run with year dummies and cluster-robust/bootstrapped standard errors.

A number of other econometric issues arise when modelling firm counts over time and across space. These include excess zeros in the dependent variable, spatial dependence and other sources of potential endogeneity.

### 4.3.1 Excess zeros

Related to the issue of over-dispersion is the excess zero problem. This arises when there are a large number of areas with zero counts for firms, or zero counts for firms of a specific type. Zero-inflated and hurdle-count models are widely used to accommodate this problem. Both of these models are finite mixture models with two components, the zero-truncated probability mass function and the untruncated probability mass-function (Cameron and Trivedi, 2013).

A zero-inflated model modifies the probability of the zero outcome using a discrete error distribution (Bhat et al., 2014). Essentially this approach assumes that the zeros and non-zeros derive from different data generating processes.

A hurdle-count model assumes that a certain threshold must be crossed before non-zero entries are observed. If the threshold is not crossed a zero is observed, and if the threshold is crossed a positive count is observed. This model can also be used in cases where the zero counts come from a different data generating process.

A widely used motivating example for the use of these models is in measuring counts of fish caught on a camping trip (Cameron and Trivedi, 2013). In this case the outcome variable is the number of fish caught, but the modeller cannot distinguish between campers who have fished, and campers who have not fished. Clearly, the outcome variable will be inflated with zeros from those campers who have not fished, and two separate data generating processes are at play. A correctly specified model must account for this.

[2]As described in Section 4.5 our data are over-dispersed and we use NB models where possible. Where convergence of this model was not possible a Poisson with cluster robust standard errors was employed. Appendix 4.E provides results from a range of results from alternate model specifications.

In our case, it is not clear that two separate processes are generating the data. It may be the case that firms will only locate in areas once a certain threshold of favourable area characteristics has been passed, in which case a hurdle-count model might be appropriate. Our choice of unit helps to ameliorate this issue, and we will discuss this is more detail in Section 4.4.1. Effectively we only select locations which are at or above the 75th percentile of employment density in Ireland. A plausible argument can be made that all of these areas are legitimate potential locations for new businesses, negating the need to explicitly model the zeros as a separate process.

Also worth noting, is that high proportions of zeros in the data does not necessarily imply an excess-zero problem. This could also potentially be explained by the regressors in the model (Cameron and Trivedi, 2010), and by modelling the zeros as a separate data generating process, one can risk over-specification.

As a robustness check we also report results from zero-inflated Poisson and zero-inflated negative-binomial estimators. The results hold.

### 4.3.2 Spatial dependence

When modelling firm counts at a disaggregate spatial level, spatial dependence could exist in the dependent variable, independent variables, or through unobserved factors that will affect the residuals. These effects are described as spatial spill-over and spatial error effects[3]. Spatial spill-over refers to the fact that many economic and demographic activities are positively correlated with those observed in neighbouring regions. This has been widely observed in the literature on business location choices, see Guimaraes et al. (2004) and Alama-Sabater et al. (2010) for examples.

Both spatial spill-over and spatial error effects are magnified when using highly disaggregated units for obvious reasons, and a trade-off must be made between exploiting the richly observed cross-sectional variation which might exist in measures of economic and social activity, and contaminating the data by choosing a highly disaggregated and sometimes arbitrary unit of observation that imposes boundaries that may not exist in reality.

The spatial econometrics discipline has come under some criticism for, amongst other things, being overly reliant on model diagnostics in determining the appropriate specification, as opposed to the choice being based on a solid theoretical underpinning (Gibbons and Overman, 2012). While others, such as Corrado and Fingleton (2011) consider this criticism a misrepresentation, they point a crucial unresolved issue in spatial econometrics - the identification problem in determining the weights matrix $W$. This can be considered akin to Manski's reflection problem in the identification

[3]Please see Appendix 4.B for a discussion.

of endogenous social effects (Manski, 1993). The matrix elements in $W$ contain an explicit quantification of the magnitude of the spatial effects being transmitted through all locations, but it is impossible to observe this. Furthermore, Corrado and Fingleton (2011) also note that this may lead to erroneous significance in the spatial lag of the dependent variable, when this is actually picking up the effects of omitted spatially dependent explanatory variables.

Aside from the above, and even if we assume that we can correctly characterise the weights matrix, practical difficulties exist in modelling spatial dependence in a non-linear panel setting. Much of the currently available toolboxes[4] deal with spatial dependency in an OLS panel setting, although recent work by Bhat et al. (2014) has developed a spatial multivariate count model in a panel setting, implemented in Gauss. While other work by Bertanha and Moser (2015) has demonstrated that the Poisson conditional fixed-effects maximum likelihood estimator (PCFE) is consistent even if the data are not Poisson-distributed, and are correlated over time - provided the spatial dependence is time-invariant.

Adapting either of the above approaches would be a possibility for us, however, as discussed previously we have important time-invariant explanatory variables which we would like to include, the coefficients on these variables cannot be estimated in a fixed-effects setting.

Rather than using complex estimators to accommodate arbitrary disaggregated boundaries which give rise to spatial dependence, we first try to mitigate this problem, by choosing a specific unit of observation that significantly reduces it. This is described in detail in Section 4.4.1. As a further robustness check models with spatially lagged explanatory variables are also used.

### 4.3.3 Endogenous explanatory variables

There is the potential for an endogenous relationship to exist between new business establishments and a number of our time-varying explanatory variables relating to infrastructure and human capital. Perhaps improved infrastructure encourages economic activity, but equally, areas with large or fast-growing economies are more likely to attract better infrastructure. This results in simultaneity bias in the regressors. Similar arguments can be made for human capital and unemployment. This is primarily a concern for time-varying measures of ICT infrastructure, mororways, human capital and unemployment. As the other time-invariant measures of accessibility and the electricity network do not change, we can be clear enough about the direction of causality for these measures.

The potential endogenous effect of broadband on economic activity has been highlighted by a number of other authors (Van Gaasbeck, 2008; Mack et al., 2011; Kolko, 2012; Mack and Rey,

[4] Such as those in Matlab provided by Elhorst (2012) and LeSage.

2014). Largely this is described as reverse causality between economic activity, such as new business counts, employment, payroll, house rents etc., and broadband availability.

Broadband provision is largely a function of population density and time, in that it goes to the most densely populated areas first, but this effect lessens over time. Admittedly there might well be an endogenous relationship between, for example, existing employment levels and broadband, but by restricting our analysis to new firms in each year, and by controlling for the pre-existing employment levels in each area, we mitigate this problem. Essentially we argue that any endogeneity relates to the stock of firms in an area rather than the flow of new firms to an area.

The nature of our ICT variables also help us to control for a potentially endogenous relationship. We have time-varying data on the existing DSL infrastructure, and data on newly installed fibre-optic broadband networks (MANs). We use MAN-area dummy variables to control for unobserved area characteristics in these MAN locations that might be related to new firm counts, and a time-varying dummy to identify the period from which the areas had fibre enabled. This variation helps to isolate the pure "MAN effect".

Similar problems might exist with the time-varying motorway network data. Again, we argue that any endogenous relationship will be related to the stock rather than the flow of new business establishments.

Local labour markets are represented by the proportion of people with a third level qualification, the unemployment rate and relative employment compensation. An influx of new firms may affect all of these measures. To account for this, we run a robustness check keeping these variables unchanged at their 2002 levels and the results hold.

To further ensure robustness, all explanatory variables are lagged by one period in all estimations, and further robustness checks are undertaken using variables lagged by two periods.

### 4.3.4 Identification strategy

The identification comes from exploiting space and variation in the location of infrastructure and firms. This is illustrated in fig. 4.1. We know what type of firms located in each location and when they first recorded employment. We also know what type of infrastructure was constructed in each location and when it was made operational. We represent infrastructure using the MANs in fig. 4.1, however the following section describes our wide range of data on different infrastructure for each location.

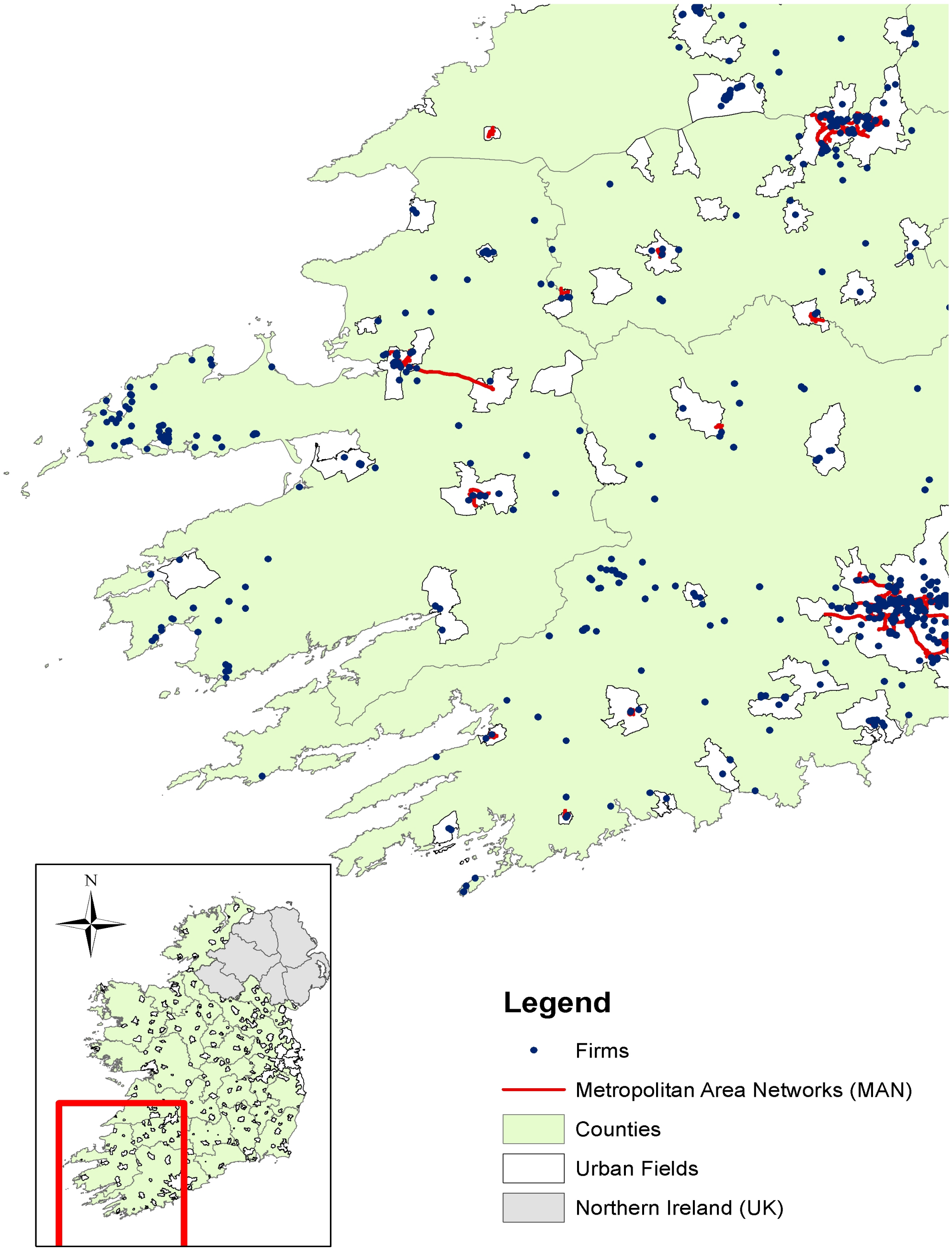


Figure 4.1: Map of Urban Fields and Metropolitan Area Networks for South-East of Ireland

## 4.4 Data

The data used in this analysis comes from a wide range of sources. The period we will examine is from 2002-2011. We select this period because there were significant infrastructural investments in Ireland during this time, and we can exploit this variation in our data in order to evaluate the impact of these changes. We first discuss the unit of analysis used, following this we will describe the dependent variables, and then proceed to describe the range of independent variables included. Please see Appendix 4.D for descriptive statistics on all variables used.

A number of options exist as to what degree of spatial resolution we should choose to use as our unit of observation. This is an extremely important consideration, and we will now discuss in detail our choice of observational unit.

### 4.4.1 Unit of analysis: "Urban Fields"

Ireland is divided into a range of different geographical units for Census purposes. Some of these are detailed in table 4.1. The highest degree of disaggregation is the "Small Area", of which there are over 18,000, and the lowest sub-national level is the Province, of which there are 4. Urban areas are further sub-categorised into "Legal Towns and Cities" and "Towns/Cities (Settlements)".

Table 4.1: Irish Census geographical areas

| Geographical Areas | No of geographic units |
|---|---|
| Provinces | 4 |
| Counties | 34 |
| Electoral Divisions | 3409 |
| Small Areas | 18,488 |

| Urban Areas | No of geographic units |
|---|---|
| Legal Towns and Cities | 85 |
| Towns/Cities (Settlements) | 824 |

As will be detailed in subsequent sections, a number of our explanatory variables are available at the Electoral Division, or ED level. While it is tempting to exploit this degree of cross-sectional variation in our analysis, we feel this degree of disaggregation is not a realistic choice of unit when modelling firm counts in this instance[5].

EDs are quite specific geographic boundaries, artificially imposed for the purposes of population Censuses. Urban areas can consist of a number of contiguous EDs and firms are more likely to choose a location based on the aggregate characteristics of the urban area. As the majority of firms are likely to go to urban areas, an alternative would be to select either of the urban geographies listed in table 4.1.

[5]This might also create severe spatial autocorrelation problems when modelling urban areas consisting of multiple EDs.

However, “Legal Towns and Cities” does not give us sufficient coverage of the population of new business establishments from our firm-level dataset, nor fibre locations from our broadband dataset, while the boundaries of “Towns/Cities (Settlements)” do not map onto ED boundaries making it impossible to aggregate much of the potential explanatory data for these areas.

To overcome these issues, we create new units of analysis which we will refer to as “Urban Fields”. These areas are either single EDs or are aggregations of contiguous EDs which are at or above the 75th percentile of employment density, based on the 2011 Census Place of Work School or College (POWSCAR)[6] and merged using GIS software. We do not consider any EDs which do not fall into this category in our analysis. This selection covers 97 percent of all new foreign firms and 75 percent of all new domestic firms in our sample period.

[6]Detail on the relevant EDs and corresponding Urban Fields available on request.

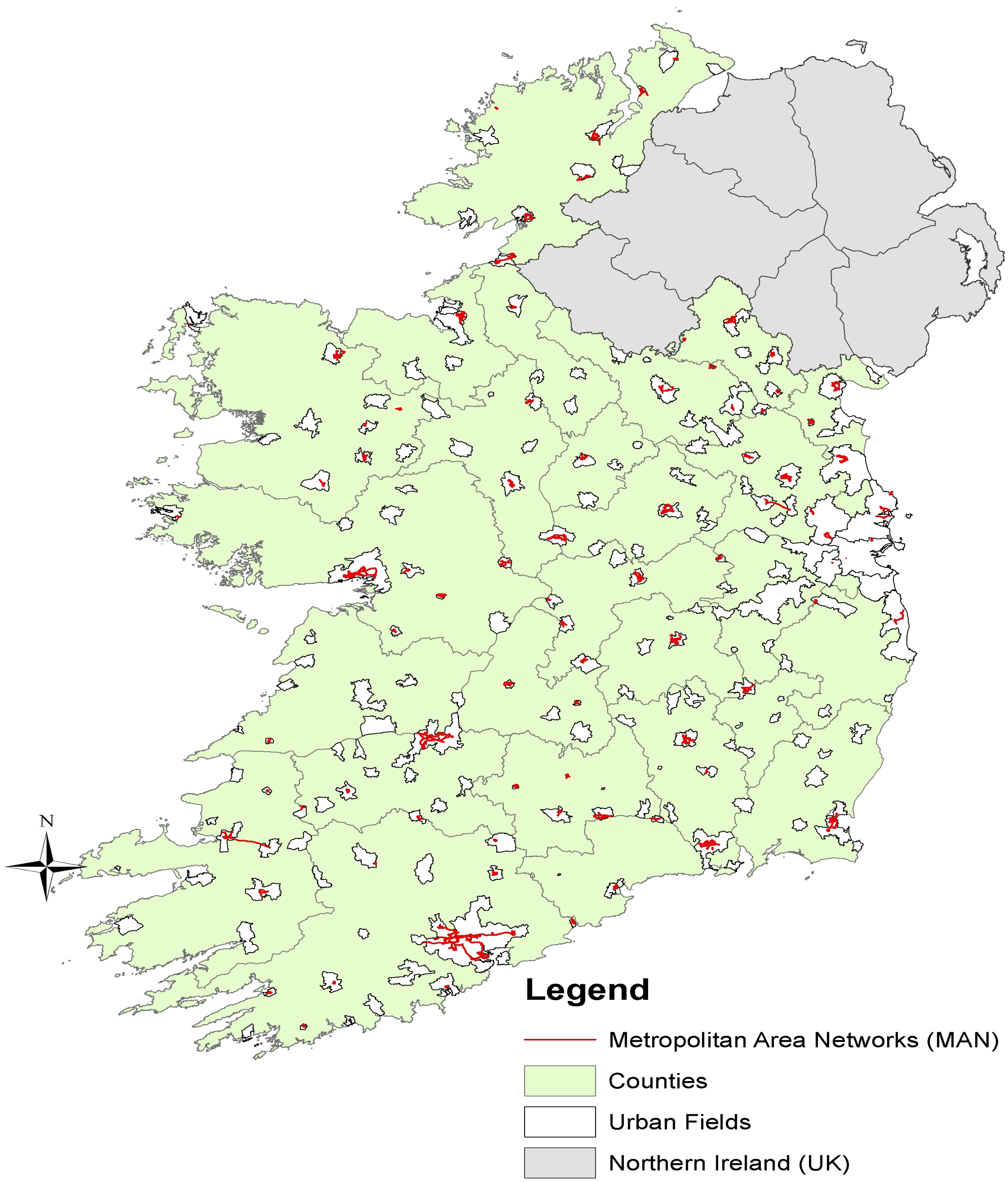


Figure 4.2: Map of Urban Fields and Metropolitan Area Networks

As discussed in the previous section, this choice of aggregation yields a number of benefits. First, it significantly reduces spatial autocorrelation problems, as generally speaking, contiguous EDs with a sufficiently high employment density are considered part of the same Urban Field. Second, it allows us to calculate time-varying agglomeration measures for all areas, see section 4.4.3. Third, it helps alleviate the excess zero problem commonly found in modelling firm counts, as many areas with zero firms are disregarded, and others are aggregated into larger areas. Furthermore, this strategy gives us greater confidence when dealing with the remaining zeros; as all remaining areas are at or above the 75th percentile of employment density, it can be credibly argued that these areas are potential options, even in the event of them never being chosen by new firms. Fourth, it reduces the computational overhead in the modelling process, as we now have 192 units over 10 time periods, as opposed to 3400 units over 10 time periods.

There are some difficulties with this choice of approach however. This is most pronounced on the Eastern seaboard, which contains the capital city, Dublin and has the greatest population density. As can be seen, this area contains a number of contiguous EDs in which the employment density does not fall below the threshold, making the partitioning of the Urban Fields somewhat problematic. In this case we use county boundaries to partition Urban Fields, and we define the Dublin City Urban Field as those EDs that broadly map onto Dublin postcode regions[7]. This area is omitted from all estimations as this is a uniquely attractive area for new firms and very different from the rest of the country in terms of population and employment density, infrastructure and general accessibility.

In total, we create $N = 192$ Urban Fields. The set of all Legal Towns and Cities ($N = 85$) is contained within the set of Urban Fields, while a considerable number of Towns/Cities (Settlements) ($N = 824$) fall outside the Urban Fields as they do not contain sufficient employment density to be included.

### 4.4.2 Firm location variables

Our dependent variables come from the Department of Jobs, Enterprise and Innovation (DJEI) Annual Employment Survey. This is an annual census of employment in all manufacturing and internationally-traded services companies in Ireland, supported by the enterprise development agencies; IDA Ireland, Enterprise Ireland and Údaras na Gaeltachta. This survey has been administered since 1972 by the agency formally known as Forfás, now subsumed within the Department of Jobs, Enterprise and Innovation. This dataset contains firm-level annual data on employment, NACE 4

[7] As can be seen from fig. 4.3 the boundaries of these areas do not align exactly. The author's judgement was used to allocate EDs to either Dublin City or the neighbouring Urban Fields in case where the ED was bisected by the postcode region.

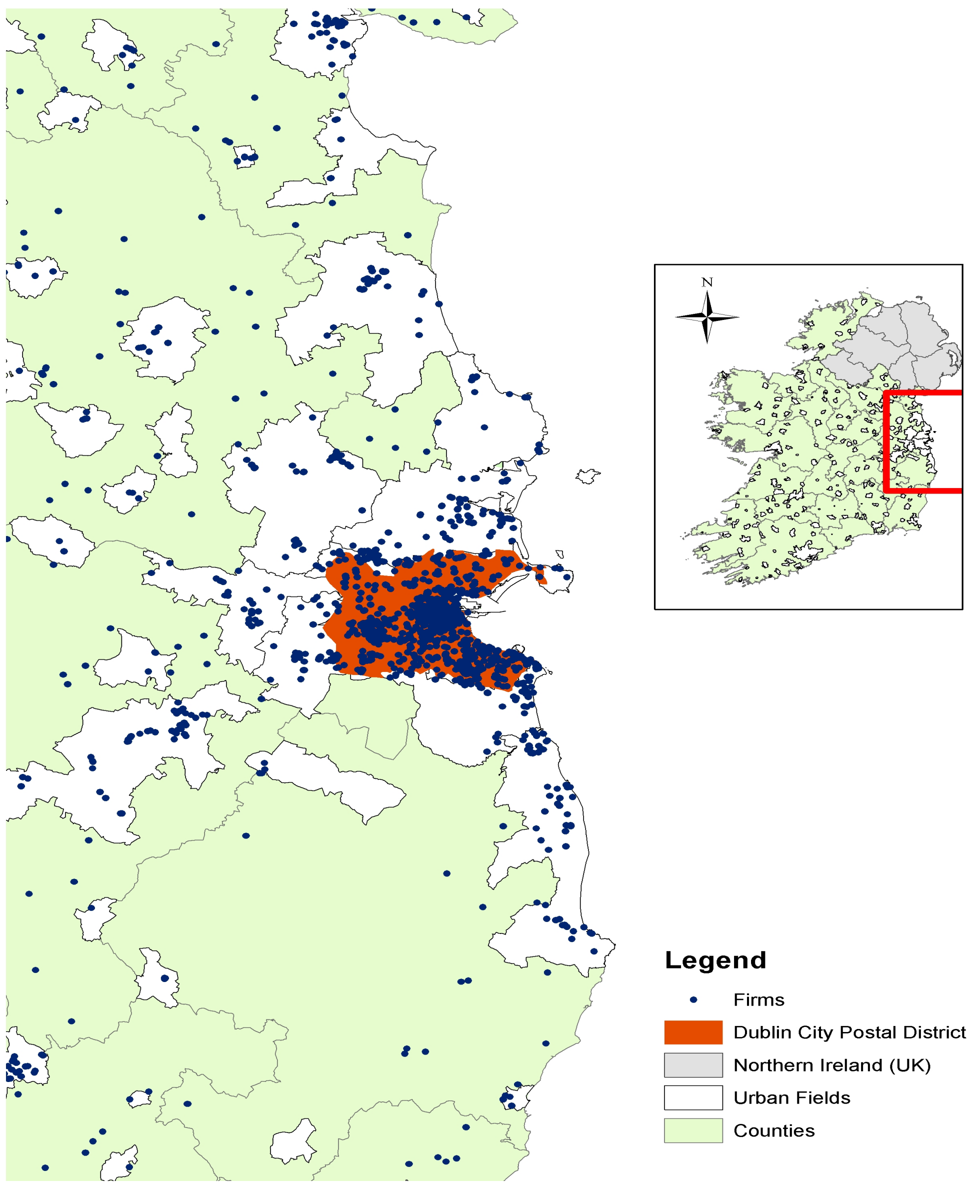


Figure 4.3: Map of East Coast Urban Fields

digit sector, location (geocoordinates), entry/exit, and whether a firm is majority foreign or domestic owned. This survey under-represents the services sector (Lawless, 2012), but contains almost the full population of manufacturing firms (Barrios, 2006), and all agency assisted foreign firms.

This dataset allows us to track the entry of all new firms over our sample period. Entry is recorded as the first time a firm records positive employment numbers in the dataset.

Looking at the foreign firms firstly in table 4.2 we can see that the majority of them are high-tech knowledge intensive services providers. Domestic firms again are predominantly involved in services but there is a greater balance of firms across other sectors within the economy. For a more detailed breakdown see tables C1 and C2 in the appendix.

Table 4.2: New firms by sector

| Sector | Foreign Firms | | Domestic firms | |
|---|---|---|---|---|
| | Count | Percentage | Count | Percentage |
| Hi tech manufacturing | 35 | 8% | 184 | 5% |
| Medium tech manufacturing | 14 | 3% | 240 | 6% |
| Medium-low tech manufacturing | 16 | 4% | 248 | 6% |
| Low tech manufacturing | 25 | 6% | 523 | 14% |
| Knowledge-intensive market services | 7 | 2% | 323 | 8% |
| High-tech knowledge-intensive services | 230 | 52% | 1338 | 35% |
| Other knowledge-intensive services | 5 | 1% | 432 | 11% |
| Less knowledge-intensive market services | 3 | 1% | 202 | 5% |
| Other Less knowledge-intensive services | * | * | 101 | 3% |
| Construction | * | * | 112 | 3% |
| Financial and Insurance | 107 | 24% | 84 | 2% |
| Utilities | * | * | 80 | 2% |
| Total | 443 | 100% | 3867 | 100% |

Note: Sectors are based on a Eurostat aggregation of NACE 2 digit codes.
For details see http://epp.eurostat.ec.europa.eu
* Fields with less than 3 firms not reported to preserve anonymity

Regarding the year of entry in table 4.3, there is a noticeable drop off of new firms in 2010 and 2011. We can see that this is entirely driven by domestic firms and perhaps reflects the prevailing economic conditions in Ireland at this time. We control for time-specific trends in all of our regressions.

Table 4.3: New firms by year of entry

| Firm type | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Foreign Firms | 56 | 34 | 35 | 36 | 46 | 50 | 33 | 39 | 50 | 64 | 443 |
| Domestic firms | 315 | 393 | 448 | 395 | 551 | 387 | 487 | 468 | 230 | 193 | 3867 |
| Total | 371 | 427 | 483 | 431 | 597 | 437 | 520 | 507 | 280 | 257 | 4310 |

The spatial distribution of new foreign firms is quite concentrated in the major urban centres, and along the main motorway arteries, fig. 4.4. A very small proportion of firms go to areas other than the greater Dublin area, Cork, Galway, Limerick and Waterford. These clusters are also apparent for domestic firms, however there is a much greater degree of geographic dispersal. A number of factors might explain this; foreign firms are generally larger and more knowledge intensive, we would expect these firms to favour urban areas to avail of agglomeration economies and better infrastructure. Also, domestic firms, and particularly small domestic firms, may locate in areas in which their owners reside, to this extent the count of new domestic firms might reflect the degree of entrepreneurial activity in an area.

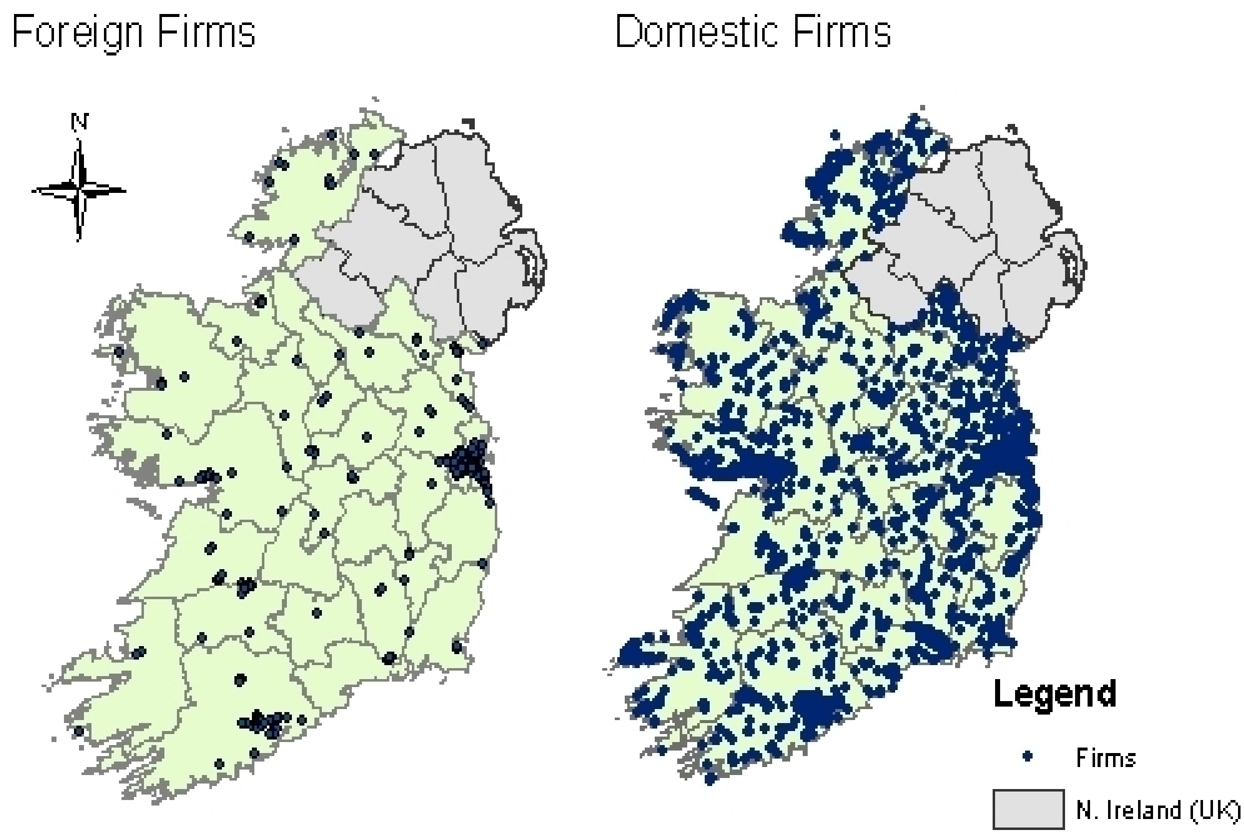


Figure 4.4: Location of new firms 2002-2011

### 4.4.3 Explanatory variables

The following sections document the steps undertaken in creating the dataset of explanatory variables.

**Broadband**

Developing a meaningful and accurate representation of broadband availability at a local level can be difficult. Much of the prior research into how broadband impacts economic activity has been based in the US, and as outlined in Taylor and Schejter (2013), many errors and inconsistencies arise when using Federal Communications Commission (FCC) zip-code level data to measure broadband.

For example, prior to 2005 small providers[8] could voluntarily provide the FCC with information on the number of lines they have in service. Post 2005 the FCC made this compulsory. This had an enormous impact on statistics, and between December 2004 (552 providers) and June 2005 (1270 providers) the number of reported broadband providers more than doubled in the US. This makes any data collected after 2005 non-comparable with previous years.

Other problems exist in determining disconnectedness - those areas with no providers. The FCC must estimate the number of ZIP codes lacking any provider. This information was never publicly reported prior to 2005 and they have never publicly documented how they identified these areas.

There are 3000 more Census zip-code tabulation areas (ZCTAs) than FCC defined zips, and 1000 FCC zip codes that do not match with the Census ZCTAs. This means that any effort to back-out which areas have no connection by subtracting the FCC zips from the universe of Census ZCTAs is *"fatally flawed"*. This also creates difficulty linking Census socioeconomic data with FCC broadband availability data.

Furthermore, in 2008 the FCC changed the unit of observation from zip code to Census Tract (there are roughly twice as many Census Tracts as zip codes) making comparison difficult between the pre-2008 and post-2008 periods.

Given that major changes in broadband availability in the US occurred during the 2005-2008 period, this makes any longitudinal analysis extremely difficult, if not impossible. Quite probably as a consequence of this, much previous US broadband research has focused on the cross-sectional dimension.

We can make a valuable contribution to the literature in this regard, given our data does not suffer from these errors and inconsistencies. Moreover, we are able to take a longitudinal approach to our estimations and our dataset covers the lifetime of broadband availability in Ireland.

In the next sections we will outline the various sources from which we assemble the broadband dataset.

[8] Those with less than 250 high-speed lines. High speed defined as >200kbps in at least one direction.

**Characterisation of broadband access.** To deliver broadband services to businesses and homes, network operators need to put suitable infrastructure in place. These network facilities can be segmented into levels reflecting their cost, technical and competition characteristics. The consumer's premises is connected to an "access network", sometimes referred to as the "last mile". This connection may be provided using various different technologies, e.g. fibre optic cable, copper wire, or radio. Tying together many local access connections, there may be a "middle mile" network such as a local fibre optic ring around a town with local exchanges or nodes in each small area. Finally, longer distance "backhaul" connections are required to bring local traffic together, linking all areas to the internet.

**Local broadband access proxy.** In this paper, the proxy for the quality of local broadband access services in an area is based on data provided by Ireland's former incumbent telecoms operator, Eircom. This panel dataset captures the availability over time of Digital Subscriber Line (DSL) services in 1,060 local telecoms exchange areas. We have data on the proportion of enabled exchanges in each area for each time period and we use this to create dummy variables which identify the period from which DSL was enabled in an area[9]. DSL technology uses traditional telephone lines to deliver broadband services, and during our sample period it was by far the main technology used to deliver local broadband access. The dataset is an updated version of the one employed by Haller and Lyons (2015) to look at broadband adoption and firm productivity in Ireland. Between 2001 and 2010, Eircom rolled out basic broadband services to local exchanges across the country. This proxy is probably more appropriate for SMEs than for larger enterprises; the largest firms would likely have used leased line infrastructure which could have been provided anywhere - at a price. However, roll-out of DSL access still seems a reasonable proxy for the availability of cost-effective broadband service for most of the firms in our sample.

As can be seen from fig. 4.5, the Dublin Urban Fields are demonstrably different from the other Urban Fields, in that enabling happened much more quickly. This is to be expected and is largely a function of population density. In terms of the evolution of enabling over time, in 2002 17% of exchanges were enabled in the Dublin Urban Fields, but most other areas had no enabled exchanges. By 2011 the lowest proportion of enabled exchanges in the Urban Field regions was 95% in the Mid-West. Also included, for comparison's sake, are the non-Urban Field regions. These effectively represent the rest of the country. As one might expect, the enabling was much slower in these more rural areas.

**Metropolitan Area Networks (MANs).** In 2004 the Irish Government awarded an exclusive concession to enet (a private company) to operate a wholesale, open-access telecommunications

[9] Proportions were used in initial estimations but dummy variables proved to have greater explanatory power in this case.

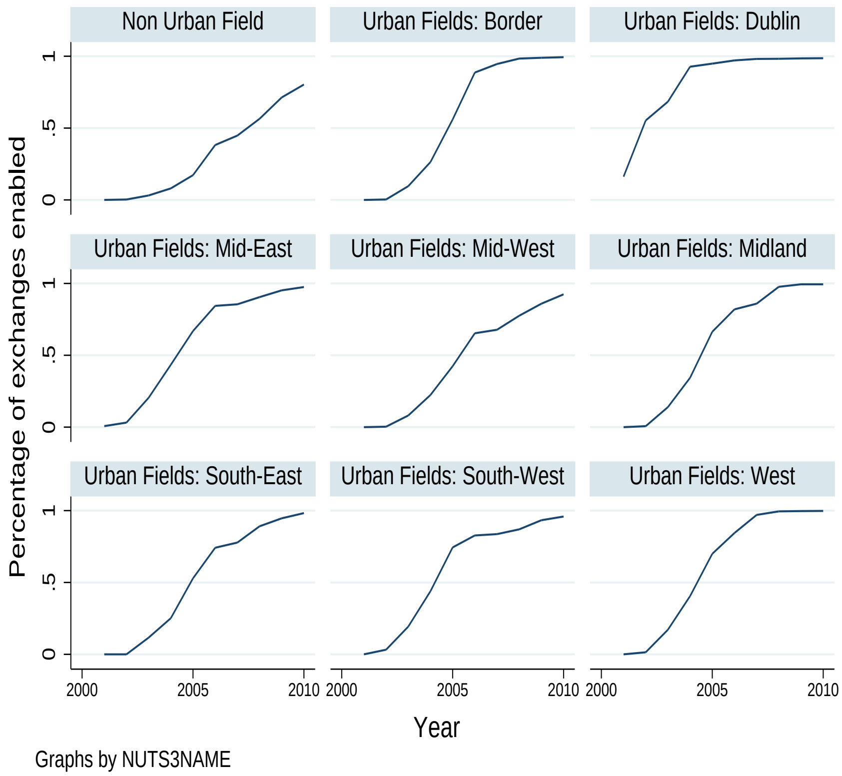


Figure 4.5: Share of enabled exchanges for DSL over time in Urban Fields by NUTS3 regions

infrastructure, known as the Metropolitan Area Networks, or MANs. The MANs are town-level fibre rings, which provide a high bandwidth network to authorised operators, in turn allowing them to sell high capacity broadband services to end-users. The network was rolled out in two phases: phase 1 covered 28 locations and began in 2004, phase 2 covered an additional 66 locations and began in 2009.

Figure 4.2 in the previous section illustrates the MAN locations. The Dublin City region only received two rather small fibre networks. This was because the MANs were conceived as a regional infrastructure intervention. They were to alleviate perceived market failures in regional towns and cities, where commercial broadband providers did not provide an adequate service. It was felt that public intervention in the Dublin City area was not required.

As the MANs are a middle-mile infrastructure, users require local access to the nearest MAN and also a local backhaul infrastructure to connect to the global network.

**Backhaul proxy.** There is no comprehensive public source of mapping data on the development of Ireland's backhaul networks over time. We do have data on the number of backhaul operators and type of backhaul at each MAN, but this does not extend to areas without a MAN. Eircom provided the main backhaul option, covering most areas. To construct a basic proxy for the presence of alternative backhaul provision, we use data provided by BT Ireland.

BT leases a national duct network from CIE (Irish National Rail Network) in which fibre is laid along the railway lines with transmission access points at towns located along the routes. In addition to this, they have metropolitan access fibre networks laid along the roads of the major cities and some of the smaller towns. We have the geo-coordinates of each node in this network, and the installation date of each wholesale On-Net circuit, allowing us to map it spatially and temporally.

### Accessibility variables

The drive times to motorway junctions, airports, railway stations, universities and institutes of technology (IT) were calculated using Microsoft MapPoint in conjunction with the MP MileCharter utility, which can compute travel times and distance between multiple points. Specifically, the shortest travel times are calculated between the centroid of each electoral district (ED) in Ireland and the respective infrastructure. These are then averaged to calculate the drive time from the centroid of each Urban Field. The travel times relate to drive time by car and the route optimisation takes into account the quality of the underlying road infrastructure by allowing the average speeds for different types of roads to differ. For example the average speed in urban streets is assumed to be 32 km/h while that on motorways is assumed to be 104 km/h. The inverse drive-times are then calculated to characterise proximity.

**The motorway network.** Transport infrastructure, and in particular the motorway network underwent significant extensions in Ireland over our sample period. In order to capture these changes we include panel data on the driving time from the centroid of each Urban Field to the nearest motorway junction. Fig 4.6 highlights the expansion over our sample period, in which 86.1% of the current network was constructed.

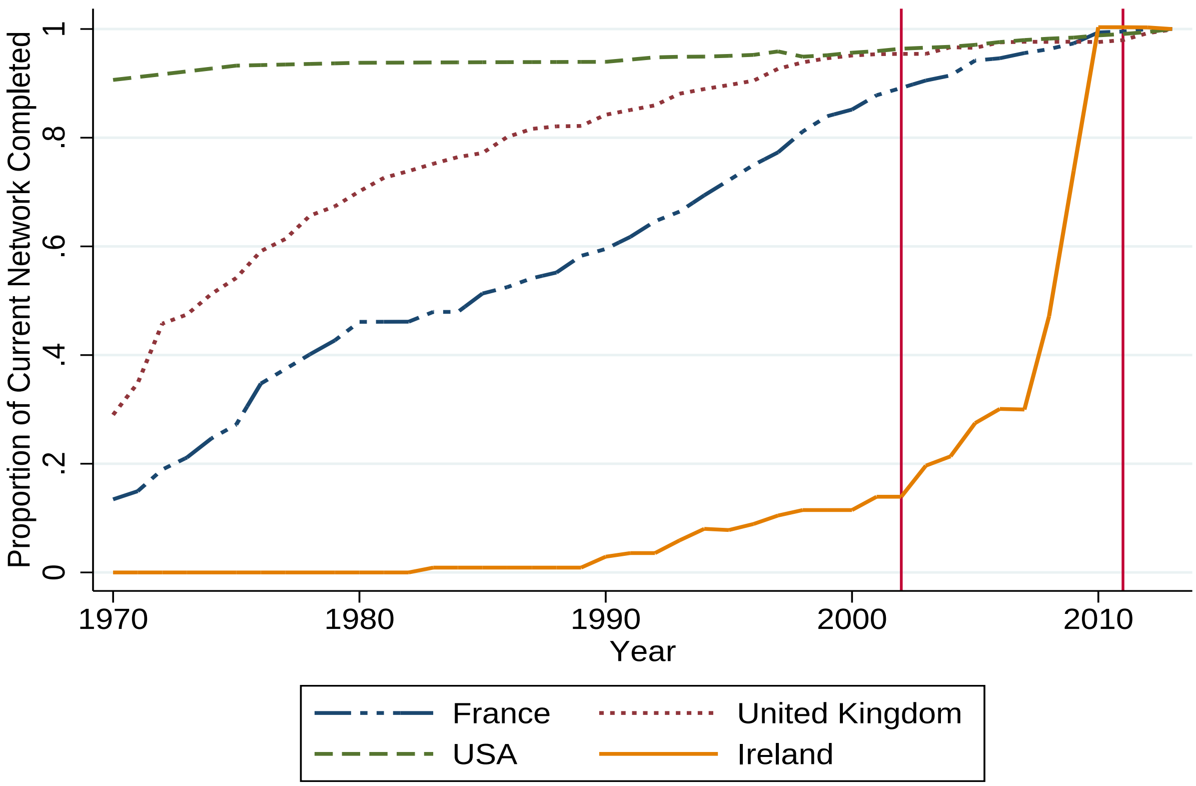


Figure 4.6: International comparison of motorway extension, 1970-2013

Sources: The data for the USA is from the Bureau of Transportation, for the UK from the Office of National Statistics, for France from EUROSTAT and for Ireland from the National Roads Authority (now Transport Infrastructure Ireland).
Note: Vertical lines indicate sample period.

Equally important for our analysis is where the expansion occurred. Fig 4.7 geographically illustrates the change in driving times between 2002 and 2011 in Ireland. During this period the network was extended to link most major urban centres, with the exception of the south-west and north-west, to the capital city, Dublin.

**Other accessibility measures.** These include the driving time in minutes from the centroid of each Urban Field, to the nearest airport, train station, port, university and IT. This data is only available for 2007. However relatively little change occurred in these measures over time.

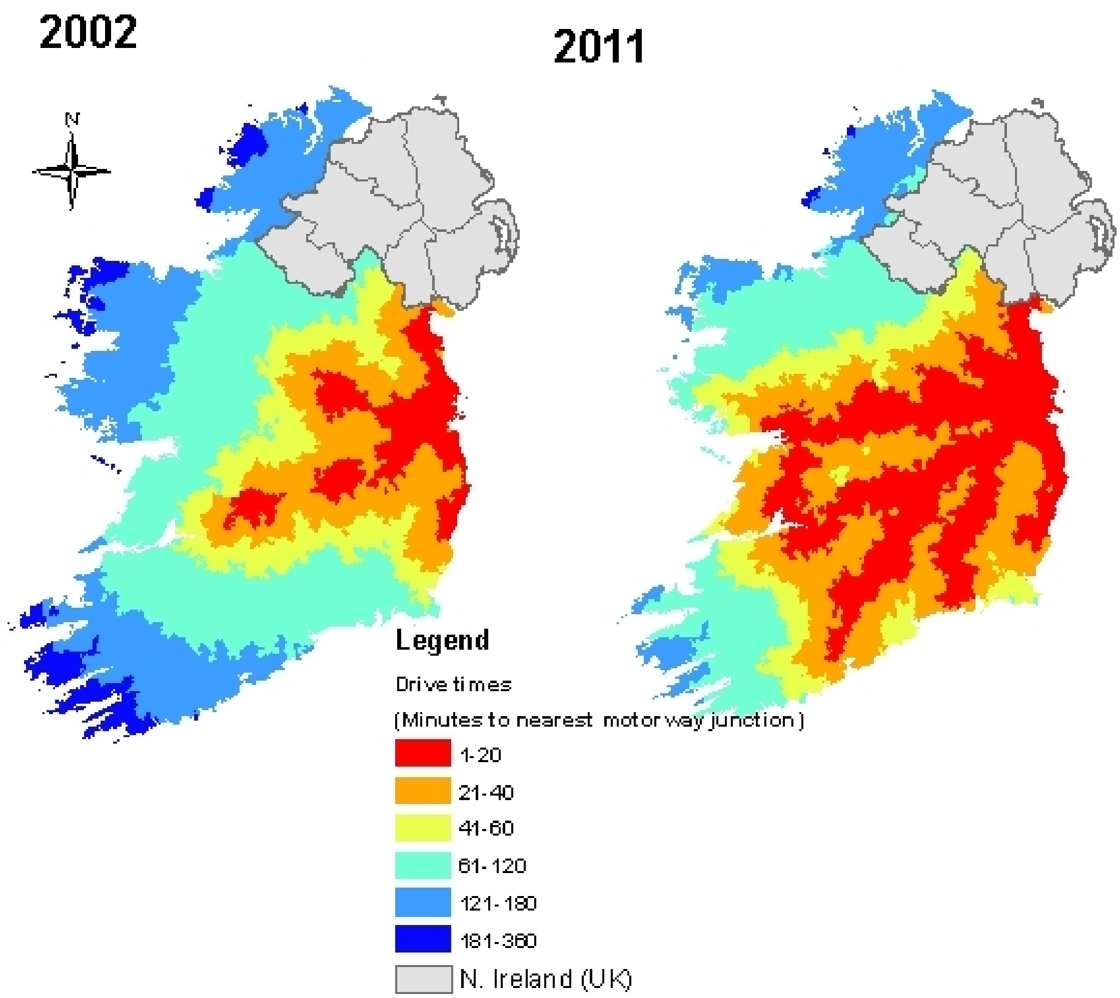


Figure 4.7: Location of motorway extensions, Ireland 2002-2011

### Agglomeration variables

Using the DJEI Annual Employment Survey we calculate a number of alternate agglomeration measures. These include economies of specialisation and diversification, along with a range of measures that reflect employment size and density by sector, skill-level and location of owner.

The sector share of total employment in each Urban Field in each year is defined as:

$$s_{ij}(t) = \frac{E_{ij}(t)}{\sum_{i=1}^{I} E_{ij}(t)} \tag{4.7}$$

Where $E_{ij}(t)$ is employment in sector $i$ in Urban Field $j$ at time $t$, where $i = 1, 2, ..., I$, $j = 1, 2, ..., J$ and $t = 1, 2, ..., T$. and $\sum_{i=1}^{I} E_{ij}(t)$ is total employment across all sectors in Urban Field $j$ at time $t$.

The measure of specialisation we include is calculated as the sum of square of each sectoral employment share in each Urban Field (the spatial Herfindahl-Hirschman index). Low values of this variable indicate a high degree of sectoral diversity in employment, while high values indicate a high degree of concentration of employment in certain sectors. We would expect the coefficient on this variable to be negative if firms value diversity of labour above specialisation. This has been used previously in an Irish context by Barrios (2006); Gleeson et al. (2006) and Morgenroth (2009).

$$u_{ij}(t) = \sum_{i=1}^{I} (l_{ij}(t))^2 \tag{4.8}$$

The sector aggregations in table 4.2 are used to calculate this measure.

Other agglomeration measures used include the employment density, for all firms, by sector and nationality over time. As described in table 4.2, we aggregate NACE 4 digit sectors into a range of categories. To calculate the sector share variables these are then further aggregated into:

- Ownership Type: Foreign, Domestic
- Sector Type: Manufacturing, Services, Financial and Other
- Sector Skill: High-tech, Low-tech and Financial

#### Human capital and demographic variables

A range of other variables are included from the CSO Population Census to capture human capital and demographic factors. These include the proportion of the population with a third-level degree; the proportion with a Ph.D.; the unemployment rate; population in persons; population density (population in persons divided by total area in hectares); population growth and decline. We use data from the 2002, 2006 and 2011 Censuses in creating these variables.

To proxy for the relative wage costs employers face, we also include Compensation of Employees data from the CSO County Incomes and Regional Accounts. These are available on an annual basis at county level.

#### Demand-side variables

Many of the previously mentioned measures represent supply-side factors. Although, relative labour costs, unemployment, human capital and total employment in each Urban Field will also capture local demand. To further characterise access to local markets we include a variable called "centrality" which captures the proximity of each Urban Field to population centres in Ireland.

$$c_i(t) = \sum_{j=1}^{J} \frac{P_j(t)}{D_{ij}} \tag{4.9}$$

$c_i(t)$ is the centrality of Urban Field $i$ at time $t$, , where $i = 1, 2, ..., I$, and $t = 1, 2, ..., T$. $P_j(t)$ is the population of each Electoral Division (ED) $i$ at time $t$, $j = 1, 2, ..., J$. $D_{ij}$ is the drive time in minutes from each Urban Field $i$ to each ED $j$.

## 4.5 Econometric results

In the following section we report our empirical results. We control for time effects using year dummies in all models as the period examined was a particularly turbulent one in Ireland's economic history. The coefficients on these terms are significant and negative from 2006 onwards, consistent with the economic downturn experienced in Ireland during this time. The capital city, Dublin has been excluded from all estimations for reasons discussed earlier.

### 4.5.1 New business establishments by location of owner

The first section of our results focuses on foreign and domestic-owned firms. As stated previously, foreign firms in Ireland tend to be larger, more productive and knowledge-intensive than their Irish counterparts. They are also quite concentrated in relatively few urban areas. The establishment of new domestic businesses in an area is related to entrepreneurial activity, and these firms are also likely to be situated close to where their owners reside. However, both domestic and foreign firms with similar characteristics will have similar requirements and face similar constraints. For instance, high-tech firms will likely require access to high-speed broadband infrastructure and an educated workforce whether they are foreign or domestic owned. Subsequent sections will examine this by pooling foreign and domestic firms by sector and skill-level.

For comparability purposes we initially report both coefficients and average marginal effects (AME) in the below table. For exponential conditional mean models, the coefficients can be interpreted as semi-elasticities in which a one unit increase in the independent variable will result in a corresponding percentage increase in the dependent variable. The AME is the average of the marginal effects taken at each $x = x_i$[10].

Considering the Eircom DSL dummy variable, having a local telephone exchange enabled for DSL broadband in an area results in 43.7% more foreign firm births and 57.5% more domestic firm births (although the coefficient on foreign is not statistically significant). Interpreting the corresponding marginal effect, having an exchange enabled in an area results in 0.043 more foreign firms and 0.588 more domestic firms. Also worth comparing at this point are the coefficients and marginal effects on the specialisation variable. The coefficient on specialisation is larger for foreign firms (-4.178) than for domestic (-1.416), but the reverse is true for the marginal effects, the magnitude is smaller for foreign (-0.132) than for domestic (-0.244). This is explained by differences in the observed count of new business establishments. For foreign firms a larger probability elasticity can result in a smaller marginal effect, because we observe significantly more new domestic firms than foreign firms. The coeff = ME.$\frac{1}{y}$, where $\bar{y}_{foreign} = 0.11$ and $\bar{y}_{domestic} = 0.9$

Regarding the economic interpretation of our results we will focus only on the marginal effects columns, and for all further models we only report significant AME's on factor variables and semi-elasticities on continuous variables. This is for the sake of brevity and ease of interpretation. In all cases, we examine the unit change in $y$ for an associated change in $x$. All broadband variables are binary in which we look at the unit change in $y$ for a discrete change in the base level. For all other

[10]The value of marginal effect will depend on the point at which they are evaluated. For non-linear models the AME is generally larger than the marginal effect at the average (MEM) and in practice can be similar to the coefficients estimated with OLS models. For policy analysis the AME or marginal effect at representative values (MER) are recommended over the MEM. See Cameron and Trivedi (2013) for further details.

variables we report semi-elasticities in which we examine the unit change in $y$ for a proportionate change in $x$. Given this, the magnitude of the effect size for all variables are comparable to each other and reveal which factors have most influence on new business establishments.

Table 4.4: Count of new establishments at Urban Field level 2002-2011, by location of ownership

| Variable Type | Variable | Coefficient | | Marginal effect | |
|---|---|---|---|---|---|
| | | Foreign | Domestic | Foreign | Domestic |
| **Broadband** | Eircom DSL enabled exchange | 0.437 | 0.575*** | 0.043 | 0.588*** |
| | | (0.437) | (0.133) | (0.035) | (0.181) |
| | MAN area dummy | -0.046 | 0.001 | -0.005 | 0.001 |
| | | (0.239) | (0.147) | (0.029) | (0.189) |
| | MAN effect dummy | 0.827*** | 0.231 | 0.103** | 0.297 |
| | | (0.294) | (0.148) | (0.043) | (0.200) |
| **Accessibility** | Motorway | 4.147** | 1.224 | 0.037* | 0.017 |
| **(inverse drive time to nearest...)** | | (2.092) | (0.982) | (0.020) | (0.013) |
| | Airport | 13.117* | 2.152 | 0.067* | 0.015 |
| | | (6.970) | (4.443) | (0.038) | (0.031) |
| | Train Station | 2.822** | 0.593 | 0.044** | 0.019 |
| | | (1.146) | (0.529) | (0.020) | (0.017) |
| **Domestic Market Access** | Centrality | -0.140 | -0.169 | -0.179 | -0.547 |
| | | (0.453) | (0.320) | (0.578) | (1.048) |
| **Agglomeration** | Specialisation | -4.178*** | -1.416*** | -0.132*** | -0.244*** |
| | | (1.090) | (0.330) | (0.034) | (0.061) |
| | Foreign share of employment | 1.739** | 0.493** | 0.127** | 0.104** |
| | | (0.712) | (0.216) | (0.056) | (0.046) |
| | Irish share of employment | -0.290 | 0.551* | -0.011 | 0.024* |
| | | (0.704) | (0.304) | (0.026) | (0.013) |
| | Total employment | 0.065*** | 0.148*** | 0.096** | 0.046*** |
| | | (0.022) | (0.026) | (0.041) | (0.009) |
| **Human Capital** | Inv distance to nearest TL institute | 8.816*** | 8.300*** | 0.062** | 0.058*** |
| | | (3.136) | (1.765) | (0.025) | (0.012) |
| | Pop prop with third level qual | 12.889*** | 9.054*** | 0.479*** | 0.621*** |
| | | (2.345) | (1.620) | (0.094) | (0.101) |
| **Labour Market** | Relative employment comp (county) | 2.559* | 0.481 | 0.304* | 0.136 |
| | | (1.330) | (0.777) | (0.164) | (0.220) |
| | Unemployment | 8.195** | 3.915* | 0.105* | 0.138* |
| | | (4.165) | (2.123) | (0.054) | (0.073) |
| **Other** | Constant | -7.980** | -2.055 | | |
| | | (3.991) | (3.192) | | |
| | Time Dummies | Y | Y | | |
| | N | 1910 | 1910 | | |
| | Wald chi2(24) | 484.240 | 680.552 | | |
| | Prob > chi2 | 0.00 | 0.00 | | |

Notes: NB population-averaged panel estimation with cluster robust standard errors. *** p<0.01, ** p<0.05, * p<0.1
Results vary slightly with alternate specifications. See Appendix 4.E for details. Dublin City omitted
Explanatory variables lagged by one period
Marginal effects: dy/dx for factor levels is the discrete change from the base level
Semi-elasticity: dy/ex reported for all other variables

Considering the broadband variables first, it appears that the availability of first-generation last-mile DSL infrastructure in an area was a statistically significant determinant of new domestic firms but not foreign, resulting in 0.588 more firms in an area. The MAN area dummy is not significant in any estimation. Inclusion of this variable is a robustness check as these locations may have unobserved characteristics that are correlated with new firm counts. By controlling for this we reduce the risk of spurious correlation. The MAN effect dummy is positive and significant for foreign firms but not domestic. On average having a MAN operational in an area is associated with 0.103 more new foreign firms.

The next set of variables considered are the accessibility measures, represented in the model by the inverse drive-time to the nearest motorway junction, airport and train station. All of these measures have a positive sign, indicating that firms value proximity. The relative magnitudes suggest an implicit ranking. Access to airports is valued above access to motorway junctions which in turn is valued above access to train stations. Although weak significance suggests that caution is advisable when drawing inference from these results.

Considering the agglomeration variables next, one of the strongest results in this study emerges: firms value diversity of skills in a location above specialisation. This result applies to both foreign and domestic firms and indeed holds throughout various sample splits employed in subsequent sections. This is consistent with previous research examining firm location choices in Ireland (Barrios, 2006). Areas with pre-existing high proportions of foreign(domestic) employment are associated with increased foreign(domestic) firm birth rates. Density of either foreign or domestic employment is not significant. The pre-existing total employment in each area is included as a scale variable and is an important determinant of new foreign and domestic firms. This variable also helps to account for possible endogeneity between broadband or other time-varying infrastructure measures, such as motorway access, and our dependent variables.

Both inverse drive-time to the nearest third-level institute and the proportion of the population with third-level qualifications are significant factors in new firm formation, for both foreign and domestic firms. Again this holds throughout our analysis and is one of the key messages of this research. Also, the magnitude of the effect on third level distance is considerably larger than any of the other accessibility measures reported. A doubling (or 100% increase) in educational attainment is associated with 0.479 more foreign and 0.621 more domestic firms. Human capital is key for both entrepreneurship and the spatial distribution of FDI. Section 4.5.4 will further explore the interaction of this effect with ICT provision.

Relative labour costs do not have an effect on either foreign or domestic firms. Interestingly, for new foreign establishments there is a positive and significant coefficient on the variable representing the proportion of the labour force who are unemployed. Further estimation reveals that this effect only exists when human capital is controlled for, indicating that firms may be locating where there is an excess supply of skilled labour. Higher unemployment rates may also put downward pressure on wages. This result is consistent with Coughlin et al. (1991). However, significance is weak and any inference must be tempered with caution.

Finally, centrality or access to domestic markets in not significant in any of our estimations.

### 4.5.2 New business establishments by skill-level

Foreign and domestic firms are now pooled and then split by the skill-level of their employees. This is to examine if patterns emerge at this level that are common to both. Firms are allocated into high-tech and low-tech categories, based on the NACE 4 digit aggregations documented in table 4.2. This includes both manufacturing and services firms. Much of services in Ireland is knowledge-intensive and will be included within the high-tech sector.

Table 4.5: Count of new establishments at Urban Field level 2002-2011, by skill-level

| Variable Type | Variable | High-tech | Low-tech |
|---|---|---|---|
| Broadband | 1.Eircom DSL | 0.531*** | 0.142*** |
| | 1.MAN area dummy | - | - |
| | 1.MAN effect dummy | 0.422** | - |
| | 1.MAN increased backhaul | 0.683*** | - |
| Accessibility | Motorway | - | 0.070** |
| (inverse drive time to nearest...) | Airport | - | - |
| | Train Station | - | - |
| Domestic Market Access | Centrality | - | - |
| Agglomeration | Specialisation | -0.448*** | -0.203*** |
| | Foreign/Domestic share of employment | - | - |
| | Foreign/Domestic density of employment | - | - |
| | Total employment | 0.530*** | - |
| Human Capital | Inv dist to nearest third level | 0.314*** | 0.122*** |
| | Pop prop with third level qual | 3.096*** | 0.539*** |
| Labour Market | Relative employment comp (county) | - | - |
| | Unemployment | 0.530*** | - |

Notes: NB population-averaged panel estimation with cluster robust standard errors.
*** p<0.01, ** p<0.05, * p<0.1. Dublin City omitted. Explanatory variables lagged by one period
Results vary slightly with alternate specifications. See Appendix 4.E for details
Marginal effects: dy/dx for factor levels is the discrete change from the base level
Semi-elasticity: dy/ex reported for all other variables

The initial roll-out of DSL has impacted both high-tech and low-tech firms, although the magnitude of this effect is much greater for high-tech firms. We now create another MAN category, where we can examine the impact of increased competition in the backhaul market using the data provided by BT. We find the positive impact of the MANs appears confined to the high-tech sector. There is a slight premium for MANs with access to greater competition in the local backhaul market above those without.

Examining the accessibility variables, motorway access is important for low-tech firms. The agglomeration results again underline the importance of skills diversity as opposed to specialisation in attracting new firms of all types.

Both human capital measures emerge again as significant determinants of both new high-tech and low-tech business establishments. Not surprisingly the size of the effect is much larger for high-tech firms.

### 4.5.3 New business establishments by skill-level and location of owner

Given the differences that have emerged in previous sections, further sample splits are undertaken to disentangle the relative impacts of our explanatory variables in determining various types of firm births. Firms are now split by location of their owner and skill-level.

Table 4.6: Count of new establishments at Urban Field level 2002-2011, by location of ownership and skill-level

| | | High-tech | | Lowtech | |
|---|---|---|---|---|---|
| Variable Type | Variable | Foreign | Domestic | Foreign | Domestic |
| Broadband | 1.Eircom DSL | 0.063*** | 0.455*** | - | 0.143*** |
| | 1.MAN area dummy | - | - | - | - |
| | 1.MAN | 0.177** | 0.309* | - | - |
| | 1.MAN increased backhaul | 0.144** | 0.537*** | - | - |
| Accessibility | Motorway | 0.043** | - | - | 0.069** |
| (inverse drive time to nearest...) | Airport | 0.095** | - | - | - |
| | Train Station | - | - | 0.018*** | - |
| Domestic Market Access | Centrality | - | - | - | - |
| Agglomeration | Specialisation | -0.076** | -0.402*** | -0.023*** | -0.190*** |
| | Foreign/Domestic share of employment | - | - | - | 0.097* |
| | Foreign/Domestic density of employment | - | - | - | - |
| | Total employment | 0.084** | 0.867** | 0.007* | 0.431* |
| Human Capital | Inv dist to nearest third level | 0.055*** | 0.224** | | 0.110*** |
| | Pop prop with third level qual | 0.496*** | 2.515*** | 0.077*** | 0.477*** |
| Labour Market | Relative employment comp (county) | 0.375** | - | - | - |
| | Unemployment | 0.120*** | 0.461*** | - | - |

Notes: NB population-averaged panel estimation with cluster robust standard errors
*** p<0.01, ** p<0.05, * p<0.1. Dublin City omitted. Explanatory variables lagged by one period
Results vary slightly with alternate specifications. See Appendix 4.E for details
Marginal effects: dy/dx for factor levels is the discrete change from the base level
Semi-elasticity: dy/ex reported for all other variables

Initial DSL roll-out is again important for all firm types with the exception of low-tech foreign firms, but much more so for those in the high-tech sector. The impact of middle-mile fibre is again concentrated in the high-tech sectors, and there is a premium where increased competition in backhaul exists.

Motorway access emerges as significant for high-tech foreign and low-tech domestic firms. High-tech foreign firms also value proximity to airports, the effect size is more than double that of motorways. Low-tech foreign firms appear to value proximity to train stations. This variable may also be picking up the fact that most large towns have train stations and foreign firms almost exclusively locate in large urban areas.

Human capital once again emerges as an extremely powerful determinant of location for all firms. This effect is particularly strong for high-tech firms.

High-tech foreign firms appear to be drawn to areas with higher relative wages. The unemployment rate is again significant. The simple correlation of unemployment and new firm births is negligible,[11] and only present once labour cost and quality are controlled for, indicating if anything an excess labour supply effect.

### 4.5.4 Exploring some interactions

A key result emerging in this study is the importance of both fibre broadband and human capital to high-tech knowledge-intensive firms. This section further explores this result by interacting these variables. We do not distinguish between MANs with and without backhaul in this specification, otherwise the model is identical to that used in previous estimations.

Fig. 4.8 illustrates this effect, by comparing the difference in the marginal effects in the presence and absence of a MAN at different levels of educational attainment. For all firms this effect is not significantly different from zero (based on 95% CI) when the proportion of the population with a third-level degree is below approximately 20%. There appears to be a certain degree of non-linearity in the relationship, particularly for high-tech foreign firms. Expected firm counts in the presence of a MAN are higher in areas with greater average educational attainment, and the magnitude of the marginal effect increases as the level of educational attainment increases. For foreign firms, the difference while significantly greater than zero is small in magnitude. For all firms the 95% confidence interval becomes quite wide after an attainment proportion of approximately 35%. This weak significance is largely driven by limited observations above this level.

One particular feature of the relative levels of educational attainment by area is persistence over time. Table 4.7 illustrates this using Spearman's rank correlation. It is clear that there is very

[11] If anything this effect is negative. Correlation with high-tech FDI is -0.02 and high-tech domestic -0.06.

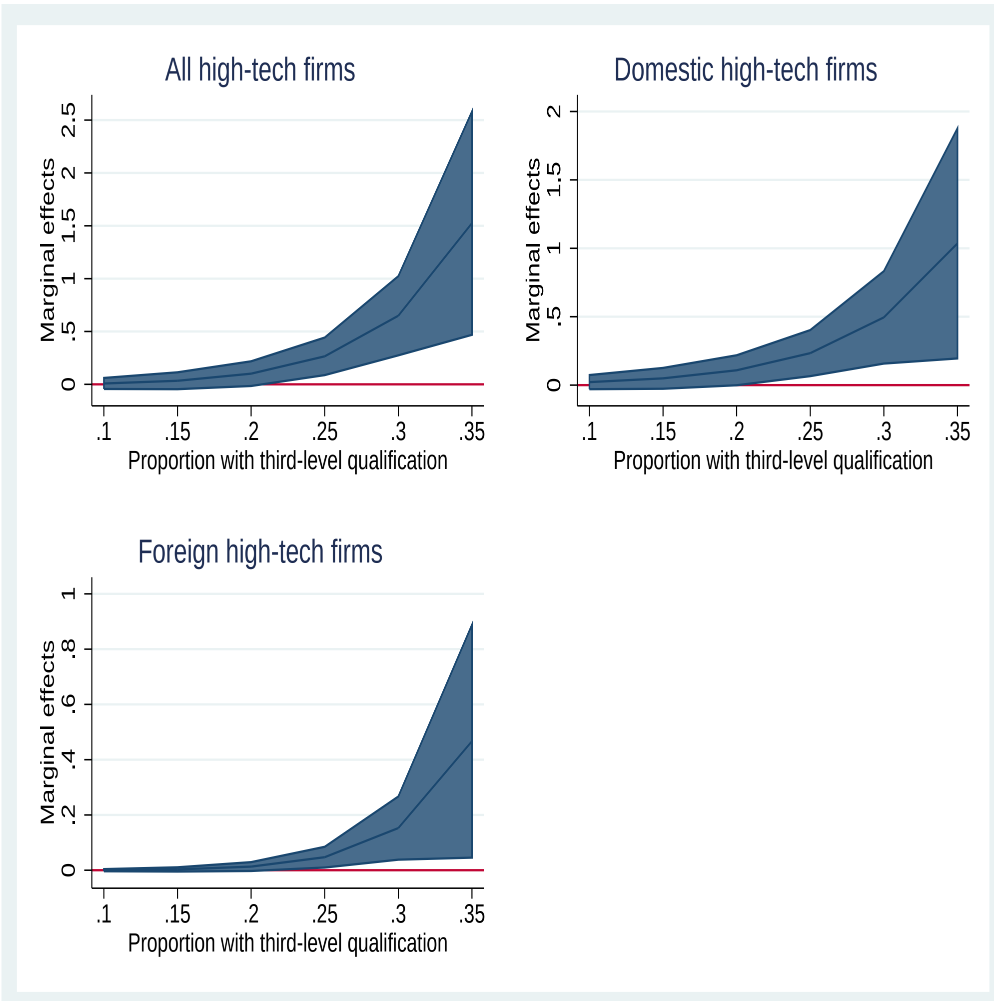


Figure 4.8: Expected firm counts in the presence of MAN at different levels of educational attainment

little change in the relative rankings of areas over our sample period. As figure 4.9 demonstrates, the shape of the distributions remain relatively stable over time, both for urban areas and when considering the country as a whole at a more disaggregated level. The absolute level of third level attainment increased for all areas between 2002 and 2006, but there was little change between 2006 and 2011. However if ICT interacts with third level attainment in the manner discussed above, it might be the case that broadband roll-out exacerbates differences between areas, in their ability to attract new business establishments.

Table 4.7: Spearman rank correlations of educational attainment over time

| | 2002 | 2006 | 2011 |
|---|---|---|---|
| 2002 | 1 | | |
| 2006 | 0.942*** | 1 | |
| 2011 | 0.923*** | 0.953*** | 1 |

Urban Fields with higher educational attainment tend to be those more densely populated regions along the east coast - close to Dublin, and large regional cities and their surrounding Urban Fields, as per figure 4.10. This is perhaps another channel through which the "urban bias" of broadband persists.

## 4.6 Robustness

As discussed in Section 4.3 several potential sources of error exist. By modelling at the Urban Field level we reduce the potential for spatial spill-overs and this also helps reduce the excess zero problem. To further account for these issues we also include other model specifications which explicitly model the zero observations as a separate process. The results are reported in table E3 in Appendix 4.E. Some coefficients change a little in magnitude but in general the sign and significance remain consistent across all models. We run estimations with spatially lagged independent variables in table E2. To account for potential labour market spill-overs we generate spatially lagged variables for the unemployment rate; the proportion of the population with a third-level qualification; and specialisation. The spatial weights matrix is distance-based with the threshold at 50km. The inclusion of these variables does not substantively alter our results.

Endogeneity is a concern for a number of our explanatory factors. We have already addressed this with a particular focus on the broadband variables in section 4.3.3. To further account for this, all explanatory variables are lagged by one period in all estimations, we also report the results of estimations with two-period lags in table E2. Further, we address potential endogeneity between

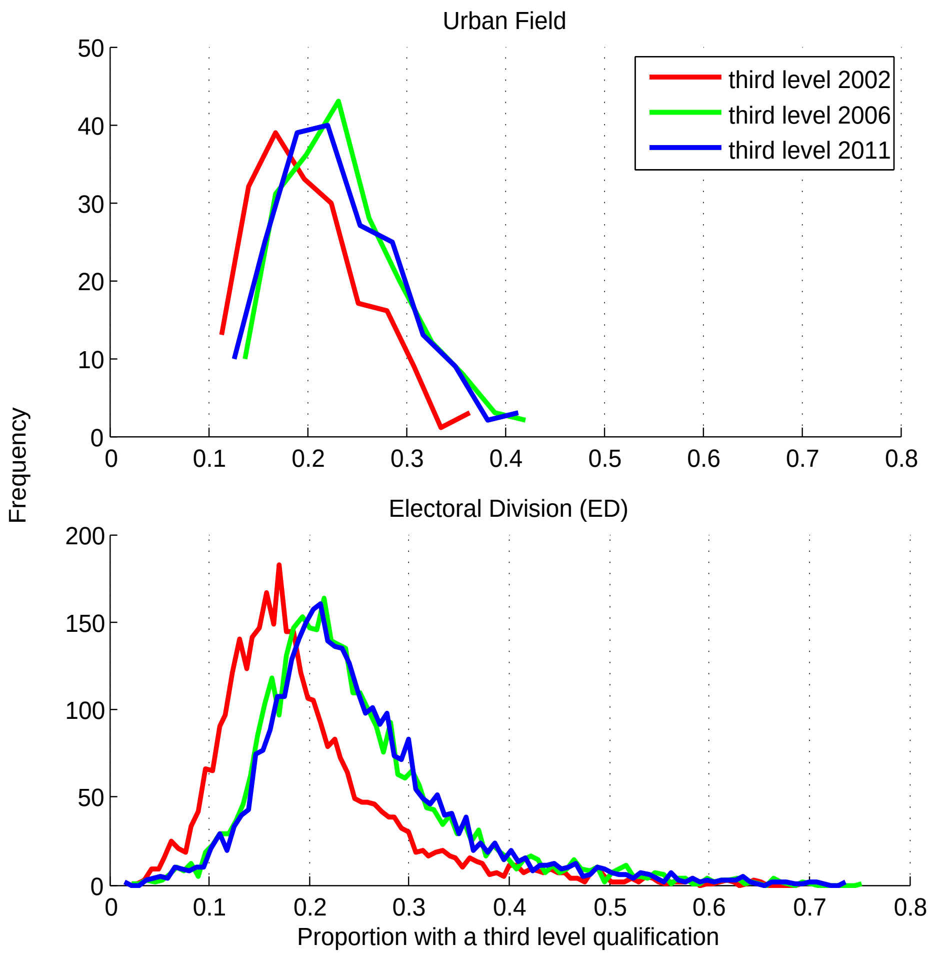


Figure 4.9: Distribution of educational attainment over time

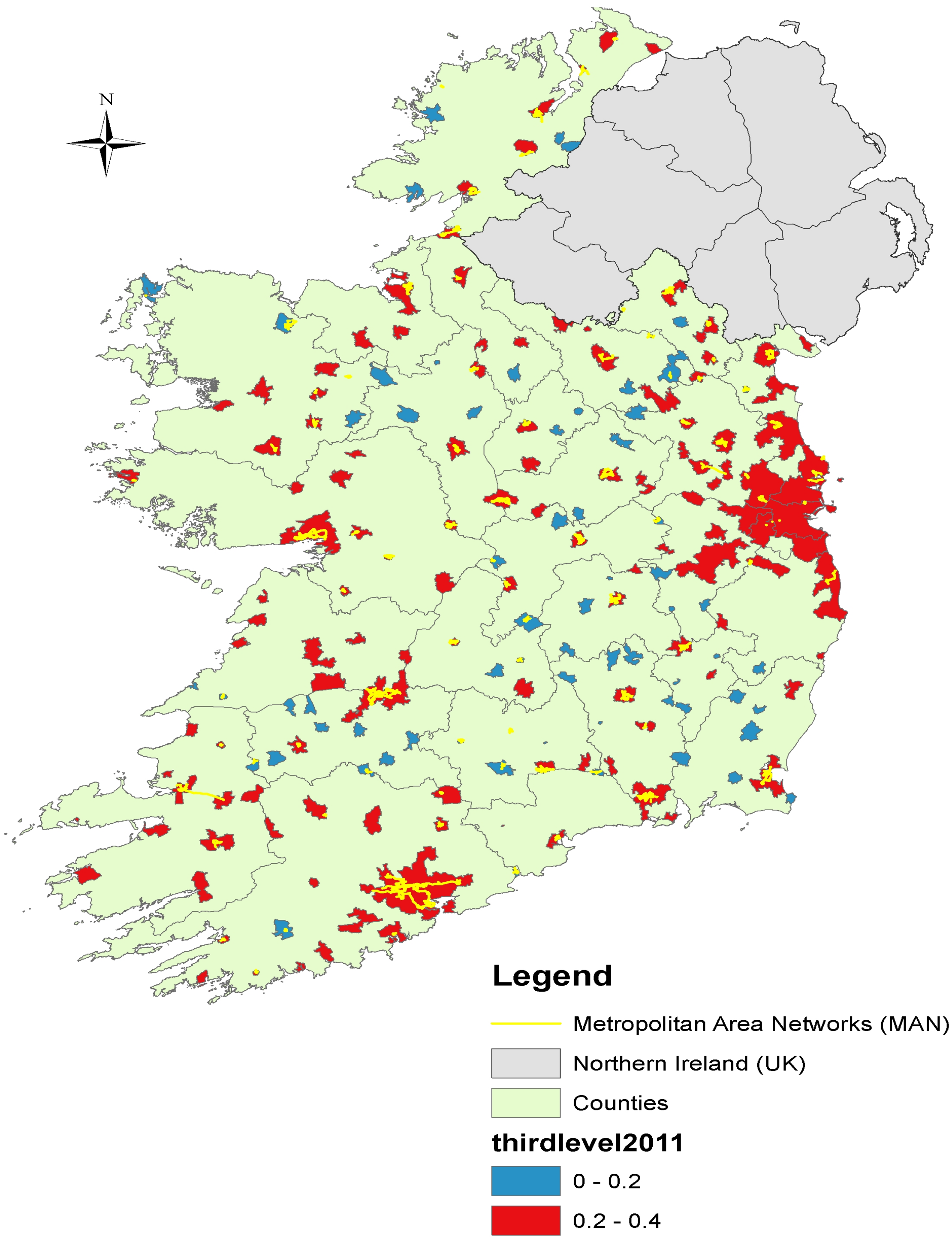


Figure 4.10: Proportion of 2011 population with third-level education and MAN locations

new business establishments and the proportion of educational attainment and the unemployment rate, by re-running estimations using the 2002 level of these variables. As these quantities do not change over time, they cannot possibly be responding to changes in the dependent variable. Again, the results hold, as can be seen from table E2.

Finally the results of several alternative panel specifications are reported in table E1. Ideally fixed-effects might have been used, however given the proportion of zeros and important time-invariant explanatory factors, we do not use this specification. As we have a considerable degree of over-dispersion in the data we employ a NB model with cluster-robust standard errors at the Urban Field level. A range of alternative NB/Poisson population-averaged and random-effects models with cluster-robust and bootstrapped standard errors were estimated. As table E1 shows, results remain reasonably consistent across all specifications, and our reported results are somewhere towards the middle of the range for most variables.

## 4.7 Discussion and concluding remarks

We have examined the factors influencing new business establishments for both indigenous and foreign firms in Ireland over a period of significant infrastructural investment. Our data covers the introduction and recent history of broadband in Ireland and we also capture a period in which 86% of the current motorway network was constructed. Complementing this, we have a rich dataset of other infrastucture, accessibility, human capital and agglomeration measures.

The analysis does not include the Dublin City region, as this area is a uniquely attractive location for new firms in Ireland, much different from the rest of the country in terms of population density, employment density, infrastructure provision and accessibility. We also do not consider areas below the 75th percentile of employment density. Given this, our analysis is an evaluation of how infrastructure roll-out affects regional towns and cities.

On average, the introduction of broadband in an area is associated with increased new firm counts. Other work such as Lehr et al. (2005) has also found this to be the case. Unusually for this literature, we are able to create detailed proxies for the quality of broadband provision. We find that the availability of basic DSL has resulted in increased counts of both high and low-tech firms, while the benefits of middle-mile fibre appear to be concentrated in the high-tech sector. There is an additional effect in areas with greater competition in backhaul availability. The elasticity of new firm counts with respect to broadband is greater for foreign firms than for indigenous, but the marginal effect, in terms of increased new business establishments is smaller as the rate of new foreign owned establishments is much lower than that of indigenous firms. Colombo et al. (2013)

found that the adoption of basic broadband resulted in negligible productivity gains for Italian SMEs, while the benefits of advanced broadband are only realised if relevant to the firm's industry of operations. Given that many of the high-tech firms in our sample are involved in IT services, consultancy and component manufacture, it is likely that a complementarity exists between the adoption of high-speed broadband and their production processes.

Kandilov and Renkow (2010) find evidence of an urban bias when evaluating the impact of broadband on economic activity. Within different urban centres, Mack et al. (2011) finds quite a degree of heterogeneity in the elasticity of new firm counts with respect to broadband provision. It seems certain areas, perhaps related to industrial legacy or geography, have a greater capacity to absorb new technology in a productive manner. We add to this literature by examining how broadband provision interacts with educational attainment in an area for high-tech firms. It appears the benefits of broadband, in terms of increased new businesses, is greater in areas with higher educational attainment, and may not be effective at encouraging new business at all below a certain threshold of educational attainment. This finding echoes previous work, such as Mack (2014), who cautions that while broadband is a key factor in dispersing knowledge intensive firms, it should not be viewed as the only factor.

Accessibility, measured by drive times, appears important to high-tech FDI, but less so for indigenous and low-tech firms, with the relative importance of proximity to airports almost twice that of proximity to motorway junctions. Similarly, Button et al. (1995) found that road and air links have a greater importance for inward investment than for domestic firms, and Mack et al. (2011) found proximity to airports to be important for knowledge-intensive firms.

Previous work, such as Holl (2004b) and Holl (2004a) found that large scale motorway investments in Portugal and Spain, respectively, resulted in a dispersal of manufacturing firms, with the benefits concentrated mostly near the new infrastructure. Given the level of motorway investment during our sample period, it is perhaps surprising that we do not observe a larger effect.

Similar to previous research on firm location choices within Ireland by Barrios (2006), we find that diversity of skills in an area is more important for new business establishments than specialisation. This work is consistent with other research, such as Holl (2004b) and supports the "Nursery City" argument proposed by Duranton and Puga (2001), which suggests that diversity of skills is more important for start-ups, whereas specialisation is more important for subsequent employment growth. The essential argument is that diversified areas act as a nursery for new firms in search of their ideal production processes, and this offsets the comparatively lower production costs they might find in more specialised areas. In Ireland, more specialised areas tend to be lower skilled and rural in which agriculture dominates. Given this and that fact that modern firms tend to require a range of skill-sets, our results in this regard are not surprising.

The local unemployment rate has a positive and significant effect on new business establishments for FDI and the high-tech sector, consistent with Coughlin et al. (1991), perhaps suggesting an effect of greater labour availability.

The most striking results are those related to human capital. Proximity to third level institutions is highly significant for all firm-types, with the exception of low-tech FDI. The level of educational attainment is important for firms of all types, but particularly those in high-tech sectors. Furthermore, pre-existing levels of human capital appear to be an important indicator of an area's ability to absorb new ICT technologies productively. This is an interesting finding and worthy of further exploration. Previous work such as Akerman et al. (2015) has pointed to a skill complementarity between broadband adoption and skilled-labour. Broadband can increase the productivity of skilled graduates, particularly in scientific and technical disciplines, but can act as a substitute for less educated workers, lowering their marginal productivity. Other work by Erber and Madlener (2009) points to a complementarity between ICT investments and medium-skilled labour in determining gross-value growth and productivity growth in the financial sector.

There is scope for future research to investigate how ICT roll-out has affected employment levels in the high-tech sectors. This looks like a fruitful avenue, given the strong results we observe here related to new business establishments. Another attractive element of this would be the greater degree of normality we would expect to observe in the dependent variable, allowing the use of a rich set of linear models which could more explicitly account for spatial dependence.

There are a number of limitations to this work. One would suspect that a potentially endogenous relationship exists between new business establishments and a number of our explanatory variables. We have attempted to address this in a number of ways but cannot completely rule it out. Given we have a panel, fixed-effects models, while less efficient than alternatives, are more robust in the presence of unobserved features that we may not have taken account of, despite the wide range of explanatory factors we consider. This is a trade off we make between potentially reduced robustness and a richer set of time-invariant characteristics. We feel this is a trade-off worth making, such is the richness of our set of explanatory factors, supplementing our quite unique data on ICT roll-out across different geographies over time.

## 4.A Incorporating the electricity network

### A Developing a proxy for network constraints

The electricity network data was provided by ESB Networks, Ireland's distribution network operator. This was converted into a GIS map of the entire transmission infrastructure, containing a detailed map of each line from 38kV up to 400kV along with the substation nodes.

Of interest to us is the degree of congestion that might exist on this network, and the extent to which this might create demand constraints, or demand opportunities for firms, thus influencing their location choice.

We did not have access to the available capacity at each node on the network, so we developed a method for calculating a proxy for demand, supply and from this demand constraints or opportunities.

**Electricity Demand Proxy.** We obtain a map of all residential ($N = 1,506,690$) and all business ($N = 86,982$) addresses from the An Post Geodirectory, a database maintained by the Irish postal service that is intended to contain all Irish addresses. We use data from 2009 and make the assumption that the spatial distribution of residences and businesses has not changed significantly over our sample period.

Using GIS software we calculate the total number of residents and employees in each Urban Field. We use the SEAI Energy Statistics for 2011 to calculate the average Irish electricity consumption per person and per employee. We use this to calculate the ratio of average electricity usage per employee to average electricity usage per residential consumer, as per table A1. From this we can aggregate residents and employees in order to create a variable representing electricity demand in each Urban Field.

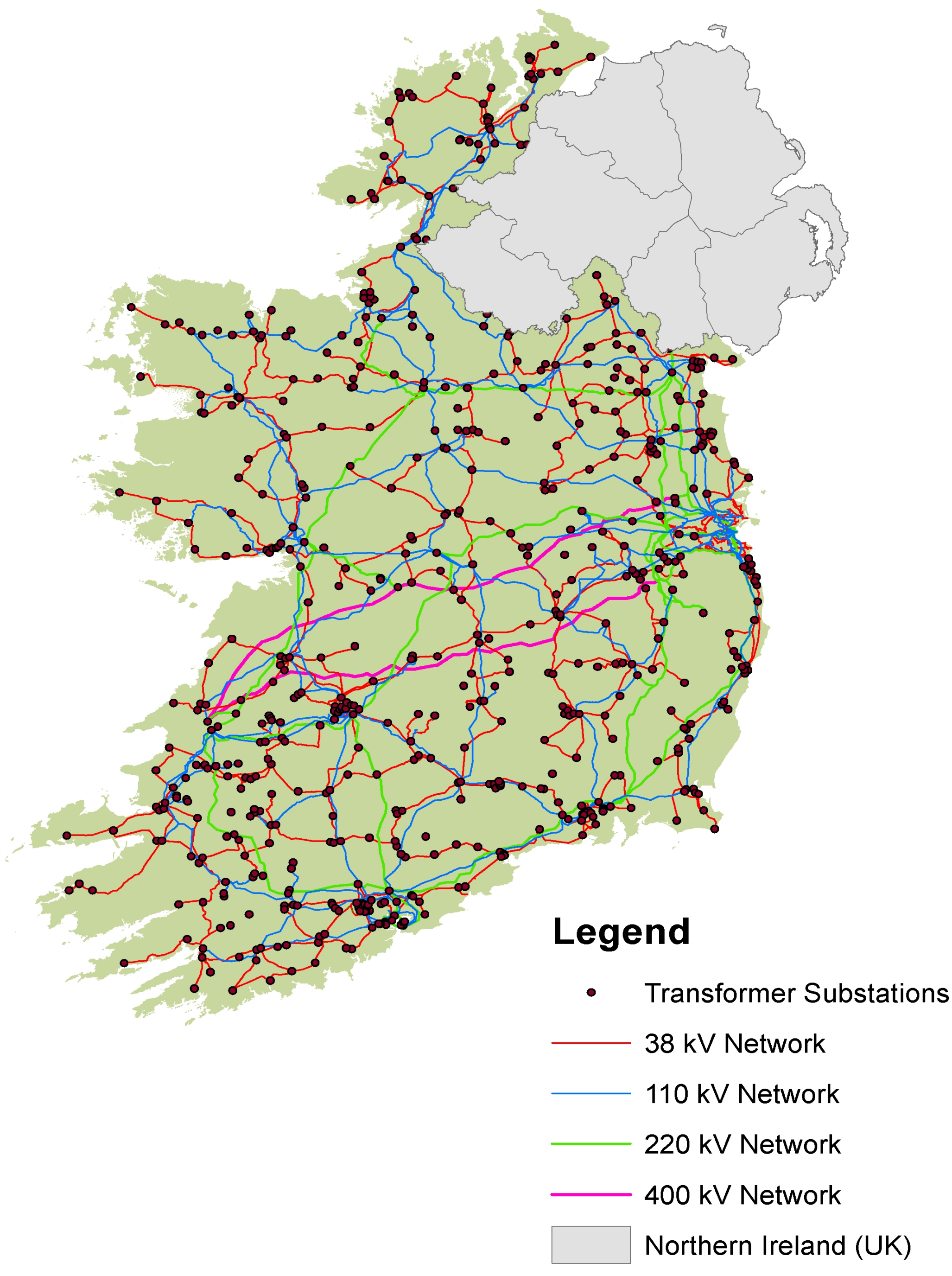


Figure A1: Map of the electricity transmission system

Table A1: Ratio of residential to business electricity usage

| | |
|---|---|
| Final total electricity consumption Ireland 2011 (ktoe) | 2140 |
| | |
| Total Residential | 712 |
| Total persons (SAPS 2011) | 4,588,252 |
| Electricity per person (ktoe) | 0.000155179 |
| Electricity per person (kWh) | 1,805 |
| | |
| Total Business (Industrial, Services, Agriculture and Transport) | 1428 |
| Total jobs (SAPS 2011) | 1,616,967 |
| Electricity per job (ktoe) | 0.000883135 |
| Electricity per job (kWh) | 10,271 |
| | |
| Ratio employee/consumer | 5.7 |

We acknowledge that this is only a proxy for demand and is likely to maintain many inaccuracies. In order to reduce these, we then create a categorical variable from 1-4, each category containing an equal number of Urban Fields, to broadly represent different categories of electricity demand.

Fig. A2 illustrates the heterogeneity of both demand and supply at Urban Field level for the south of Ireland. Of interest is determining the areas where constraints might occur due to insufficient supply.

**Electricity Supply Proxy.** In order to calculate a proxy for supply we use a number of different grid characteristics. The primary measure is available at the substation node level, and then aggregated to all nodes within a 5km radius of each Urban Field. For each Urban Field we calculate:

1. The total number of each type of node (38kV, 110kV, 220kV, 400kV) in each Urban Field
2. The number of transmission lines (38kV, 110kV, 220kV, 400kV) attached to each node
3. The distance of each Urban Field to the nearest node (zero if the node is within an Urban Field boundary)

From these we can calculate the average node type within each Urban Field; the average degree of interconnectedness of the nodes within each Urban Field; the average line capacity category within each Urban Field; the highest line capacity/node category in an Urban Field.

This information allows us to represent the level of transmission infrastructure within or near each Urban Field. From this we can determine which areas might suffer from grid constraints.

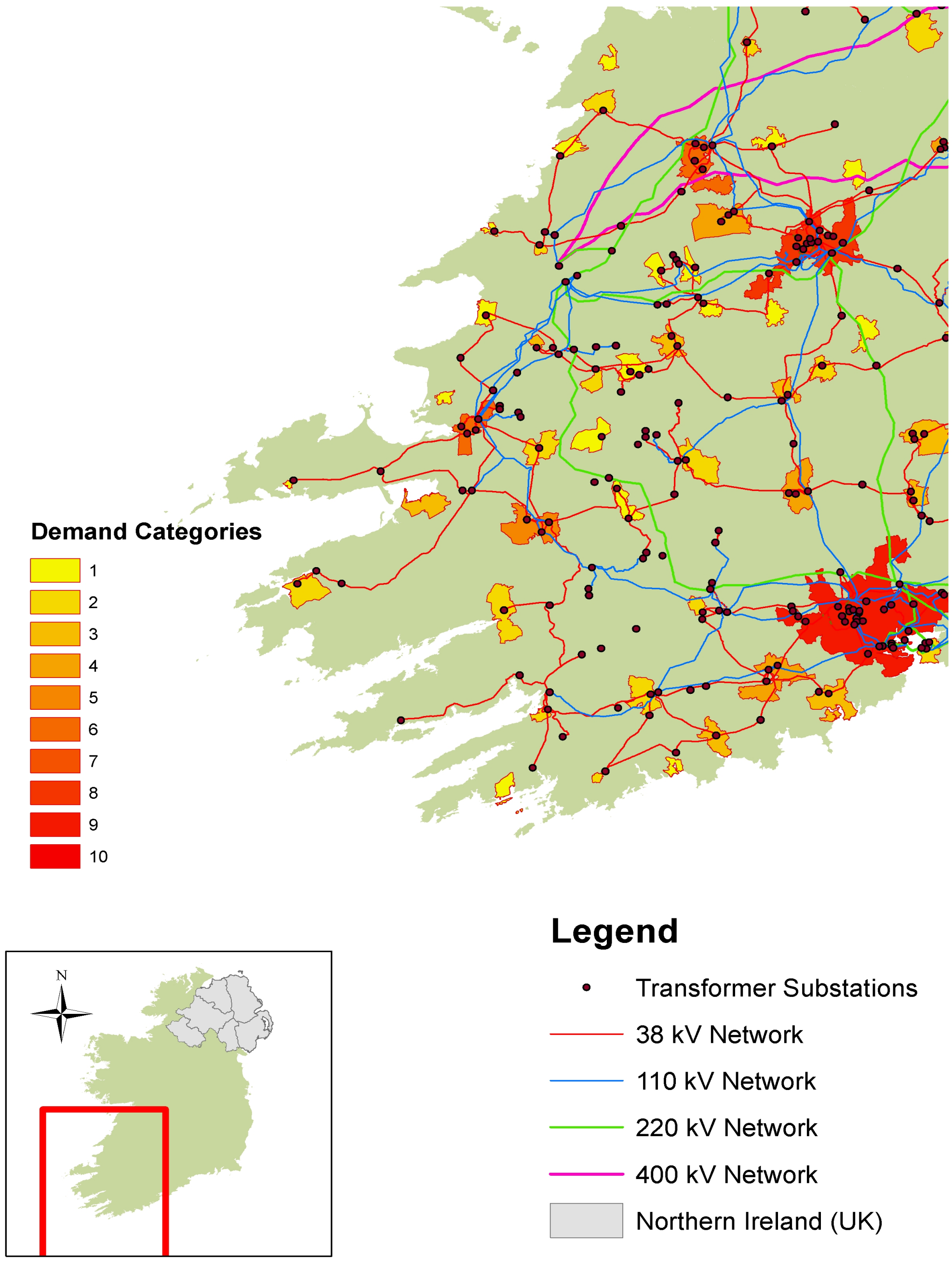


Figure A2: Map of electricity demand categories by Urban Field for the south of Ireland

Note: Colour coded with red(blue) indicating areas of highest(lowest) demand category

## B Electricity results

This section introduces variables which characterise the electricity network into our analysis. There is significant cross-sectional variation in these measures, but data on the transmission network was only available for 2014. This is outside our sample period, which is a clear drawback. However, discussion with network engineers indicate that grid extensions were limited over our sample period, and the primary means of upgrade was replacement of network components. Therefore, the spatial configuration of the transmission grid in 2014 should be broadly representative of the grid in 2002.

As described in Section 4.A, proxies were developed for electricity demand, supply and ultimately network congestion at a spatially disaggregated level. While statistically significant results were obtained for our measures of congestion, inference is problematic, as demand is highly correlated with population and it is unclear whether this or some other omitted variable is driving the results.

Therefore we report results on the electricity supply-side variables. These refer to the level of network provision within 5km of each Urban Field. We create a categorical variable to represent the interconnectivity of the network, included here is a dummy variable representing the lowest category of interconnectivity. This is to capture areas with insufficient grid capacity. We also include the average transformer substation type and a count of the number of substations.

We keep these results in a separate section as difficulties were encountered in the modelling process, in some cases it was impossible to achieve convergence in a panel setting. Also the above mentioned drawback concerning the sample period is a concern. As a result we are less confident drawing inference from them.

Table A2: Count of new establishments at Urban Field level 2002-2011. Impact of electricity variables

| Variable Type | Variable | Foreign | Domestic | High-tech | Low-tech |
|---|---|---|---|---|---|
| Electricity Network | 1. Interconnectivity | -1.822*** | 0.683** | - | 0.241** |
| | Count of substations | - | 1.493*** | 1.465*** | 0.570*** |
| | Average station type | - | - | - | - |

Notes: NB pooled estimation with cluster robust standard errors. *** p<0.01, ** p<0.05, * p<0.1
Only results on electricity variables reported. Dublin City omitted. Exp vars lagged by one period
Marginal effects: dy/dx for factor levels is the discrete change from the base level
Semi-elasticity: dy/ex reported for all other variables

The results show a very large negative correlation between the count of new foreign firms in an area and low network interconnectivity. Domestic and low-tech firm births do not appear to negatively impacted in areas of low network connectivity, indeed a positive effect is observed. The other result that emerges is that areas with station types of higher average capacity are associated with increased firm counts, for all sectors other than FDI.

Table A3: Count of new establishments at Urban Field level 2002-2011. Impact of electricity variables

| | | High-tech | | Low-tech | |
|---|---|---|---|---|---|
| Variable Type | Variable | Foreign | Domestic | Foreign | Domestic |
| Electricity Network | 1. Interconnectivity | -1.165*** | 0.447* | -0.227*** | 0.234** |
| | Count of substations | - | 1.300*** | 0.090* | 0.515*** |
| | Average station type | - | - | - | - |

Notes: NB pooled estimation with cluster robust standard errors. *** p<0.01, ** p<0.05, * p<0.1.
Only results on electricity variables reported. Dublin City omitted. Exp vars lagged by one period
Marginal effects: dy/dx for factor levels is the discrete change from the base level
Semi-elasticity: dy/ex reported for all other variables

Further sample splits indicate that the negative relationship between low network connectivity and foreign firm counts occurs for both high-tech and low-tech firms. Again this effect is positive and significant for domestic firms in both sectors. The positive significant relationship between average station type and new firm counts exists for all sample splits reported, other than high-tech FDI.

It is difficult to draw inference from these results. Foreign firms locate in urban centres which have good electricity provision. Very few of these firms stray outside of large towns, or important motorway arteries. It is more likely that the large negative coefficient on the interconnectivity variable is picking up some other factor, possibly related to the relative remoteness of these areas.

## 4.B Discussion of spatial dependence in modelling firm counts

Spatial dependency is a specific type of cross-sectional dependency in which values observed at one location or region, depend on values observed at other neighbouring locations. This violates the assumption of statistical independence, in which $E(\epsilon_i \epsilon_j) = E(\epsilon_i)E(\epsilon_j) = 0$ (LeSage and Pace, 2009). Tobler (1970) succinctly describes this phenomenon with his first law of geography; *"Everything is related to everything else, but near things are more related than distant things"*.

To illustrate how spatial dependency might be an issue in our case, we adapt the house-price example used by LeSage and Pace (2009) and Gibbons and Overman (2012) and draw on some examples discussed in Bhat et al. (2014).

Consider our outcome variable $y_i$, the count of new firms in a given area $i$, in a purely cross-sectional setting for ease of explanation. This might solely be a function of area characteristics. In which case:

$$y_i = x_i'\beta + \epsilon_i \tag{10}$$

Alternatively, perhaps firms might choose to locate in region $i$ based on the characteristics of region $i$ , but also based on observed economic activity, in this case the location choices of firms in neighbouring regions. The leads to the spatial autoregressive model, and is analogous to its time-series equivalent.

$$y_i = \rho w_i' y + x_i'\beta + \epsilon_i \tag{11}$$

It may also be the case that spatial dependence is transmitted through the explanatory variables, rather then the dependent variable. If for example, neighbouring regions have highly educated workforces or good transport infrastructure, this might induce firms to locate near these regions to avail of skills or accessibility. In this case the location choice is a function of the characteristics of region $i$ and the lagged spatial characteristics of the neighbouring regions.

$$y_i = x_i'\beta + w_i' X\gamma + \epsilon_i \tag{12}$$

Finally it may be the case that unobserved factors from neighbouring locations affect the count of new firms in region $i$. Perhaps to avail of knowledge spill-overs through networks of employees, or other unobserved factors such as public policies or attitudes that might transmit over space, firms will locate in neighbouring regions. In this case a spatial error model is appropriate.

$$y_i = +x_i'\beta + \epsilon_i \tag{13}$$

$$\epsilon_i = \rho w_i' \epsilon + \mu_i \tag{14}$$

## 4.C Detailed firm activity description

Table C1: Foreign firms: Most frequent NACE 4 digit categories by sector

| Sector Type | Sector | No. of firms |
|---|---|---|
| High-tech | Other information technology and computer service activities | 39 |
| | Computer facilities management activities | 31 |
| | Computer programming activities | 28 |
| | Computer consultancy activities | 23 |
| | Manufacture of pharmaceutical preparations | 14 |
| | Manufacture of electronic components | 5 |
| | Data processing, hosting and related activities | 4 |
| | Manufacture of computers and peripheral equipment | 4 |
| | Other Manufacture | 5 |
| Total high-tech | | 171 |
| Low-tech | Manufacture of medical and dental instruments and supplies | 19 |
| | Manufacture of other plastic products | 4 |
| | Manufacture of metal structures, tools and metal treatment | 6 |
| | Other Manufacture | 5 |
| Total Low-tech | | 42 |
| Financial | Fund management activities | 8 |
| | Insurance and other financial services | 7 |
| | Trusts, funds and similar financial entities | 4 |
| Total Financial | | 19 |

Note: Fields containing less than 3 firms merged to preserve anonymity

Table C2: Domestic firms: Most frequent NACE 4 digit categories by sector

| Sector Type | Sector | No. of firms |
|---|---|---|
| High-tech | Computer consultancy activities | 419 |
| | Other information technology and computer service activities | 105 |
| | Other software publishing | 59 |
| | Computer programming activities | 56 |
| | Business and other management consultancy activities | 55 |
| | Artistic creation | 44 |
| | Engineering activities and related technical consultancy | 44 |
| | Data processing, hosting and related activities | 42 |
| | Manufacture of irradiation, electromedical and electrotherapeutic equipment | 35 |
| | Other education n.e.c. | 32 |
| Total High-tech | | 1563 |
| Low-tech | Other personal service activities n.e.c. | 87 |
| | Manufacture of other food products n.e.c. | 51 |
| | Processing and preserving of meat | 40 |
| | Manufacture of other fabricated metal products n.e.c. | 32 |
| | Manufacture of metal structures and parts of structures | 27 |
| | Manufacture of medical and dental instruments and supplies | 24 |
| | Other specialised construction activities n.e.c. | 24 |
| | Manufacture of other furniture | 21 |
| | Recovery of sorted materials | 21 |
| | Manufacture of other plastic products | 20 |
| Total Low-tech | | 1141 |
| Financial | Other financial service activities, except insurance and pension funding n.e.c. | 7 |
| | Insurance, pension funding and leasing companies | 7 |
| | Other credit granting | 5 |
| | Trusts, funds and similar financial entities | 5 |
| | Non-life insurance | 4 |
| | Other activities auxiliary to financial services, except insurance and pension funding | 3 |
| Total Financial | | 33 |

Note: Fields containing less than 3 firms merged to preserve anonymity

## 4.D Descriptive statistics for variables in regressions

Table D1: Dependent variables: Firm counts

| Type of Firm | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Foreign | 1910 | 0.11 | 0.58 | 0 | 8 |
| Domestic | 1910 | 0.90 | 2.95 | 0 | 47 |
| High-tech | 1910 | 0.67 | 2.58 | 0 | 43 |
| Low-tech | 1910 | 0.32 | 0.92 | 0 | 11 |
| Financial | 1910 | 0.02 | 0.18 | 0 | 4 |
| Foreign high-tech | 1910 | 0.09 | 0.48 | 0 | 7 |
| Domestic high-tech | 1910 | 0.58 | 2.24 | 0 | 39 |
| Foreign Low-tech | 1910 | 0.02 | 0.14 | 0 | 3 |
| Domestic Low-tech | 1910 | 0.31 | 0.89 | 0 | 11 |

Table D2: Explanatory variables: Main

| Variable Type | Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|---|
| Broadband | MAN-area dummy | 1910 | 0.372 | 0.483 | 0 | 1 |
| | MAN | 1910 | 0.062 | 0.242 | 0 | 1 |
| | MAN with BT backhaul | 1910 | 0.047 | 0.211 | 0 | 1 |
| | Eircom enabled exchange dummy | 1910 | 0.739 | 0.439 | 0 | 1 |
| Electricity | Number of subsations within 5km | 1910 | 2.555 | 3.358 | 0 | 34 |
| | Number of transmission lines | 1910 | 7.293 | 11.103 | 0 | 101 |
| | Average substation type | 1910 | 1.141 | 0.466 | 0 | 3 |
| Accessibility | Motorway | 1910 | 51.633 | 41.121 | 2.40 | 250.07 |
| | Airport | 1910 | 56.321 | 21.674 | 8.22 | 104.47 |
| | Train Station | 1910 | 21.300 | 16.358 | 1.57 | 82.28 |
| | University | 1910 | 116.208 | 64.990 | 13.45 | 355.35 |
| | IT | 1910 | 63.732 | 28.310 | 5.51 | 181.10 |
| Labour Force | Relative Employment compensation | 1910 | 0.911 | 0.103 | 0.73 | 1.18 |
| | Unemployment rate | 1910 | 0.114 | 0.055 | 0.04 | 0.39 |
| | Total Employment | 1910 | 0.991 | 2.697 | 0 | 29.66 |

Table D3: Explanatory variables: Agglomeration

| Agglomeration Measure | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Specialisation | 1910 | 0.556 | 0.268 | 0 | 1 |
| Total Employment Foreign | 1910 | 542.080 | 1820.260 | 0 | 20097 |
| Foreign share of employment | 1910 | 0.259 | 0.320 | 0 | 1 |
| Foreign density of employment | 1910 | 0.096 | 0.195 | 0 | 1.831 |
| Total Employment Irish | 1910 | 448.814 | 954.849 | 0 | 10124 |
| Irish share of employment | 1910 | 0.679 | 0.358 | 0 | 1 |
| Irish density of employment | 1910 | 0.138 | 0.210 | 0 | 1.743 |
| High-tech share of employment | 1910 | 0.271 | 0.305 | 0 | 1 |
| High-tech density of employment | 1910 | 0.078 | 0.136 | 0 | 0.885 |
| Low-tech share of employment | 1910 | 0.655 | 0.349 | 0 | 1 |
| Low-tech density of employment | 1910 | 0.153 | 0.244 | 0 | 2.493 |
| Services share of employment | 1910 | 0.111 | 0.200 | 0 | 1 |
| Services density of employment | 1910 | 0.031 | 0.085 | 0 | 1.267 |
| Financial share of employment | 1910 | 0.011 | 0.072 | 0 | 1 |
| Financial density of employment | 1910 | 0.003 | 0.020 | 0 | 0.216 |

## 4.E Robustness

Table E1: Comparison of high-tech firm counts from alternative panel model specifications

| | | Reported results | Alternative specifications | | | |
|---|---|---|---|---|---|---|
| Variable Type | Variable | NB, PA vce(robust) | NB, RE vce(boot) | Poisson, PA vce(robust) | Poisson, RE vce(robust) | Poisson, RE vce(boot) |
| Broadband | Eircom DSL enabled exchange | 1.172*** | 0.958*** | 1.096*** | 0.849*** | 0.835*** |
| | | (0.170) | (0.223) | (0.166) | (0.190) | (0.225) |
| | MAN area fixed effect | -0.199 | 0.042 | -0.236 | 0.108 | 0.045 |
| | | (0.171) | (0.237) | (0.185) | (0.224) | (0.225) |
| | MAN no backhaul | 0.453*** | 0.393* | 0.547** | 0.396* | 0.382 |
| | | (0.156) | (0.202) | (0.227) | (0.223) | (0.242) |
| | MAN with backhaul | 0.685*** | 0.533*** | 0.704*** | 0.543*** | 0.531*** |
| | | (0.192) | (0.159) | (0.162) | (0.132) | (0.137) |
| Accessibility | Motorway | 0.888 | 0.097 | 0.151 | -0.255 | -0.540 |
| (inv. drive time to...) | | (1.187) | (1.203) | (1.212) | (1.169) | (1.290) |
| | Airport | 2.501 | 6.721 | 3.298 | 7.824 | 8.579 |
| | | (4.677) | (6.731) | (5.024) | (8.089) | (9.236) |
| | Train Station | 0.385 | 0.405 | 0.610 | 0.776 | 0.725 |
| | | (0.689) | (1.022) | (0.625) | (0.892) | (0.895) |
| Agglomeration | Urbanisation (diversity of employment) | -1.989*** | -1.982*** | -2.511*** | -1.860*** | -1.949*** |
| | | (0.377) | (0.483) | (0.392) | (0.431) | (0.517) |
| | Sector share of employment | 0.352 | 0.655* | 0.453 | 0.698* | 0.561 |
| | | (0.344) | (0.356) | (0.308) | (0.366) | (0.462) |
| | Sector density of employment | 0.305 | -0.347 | 0.177 | -0.700 | -0.345 |
| | | (0.866) | (1.198) | (0.684) | (0.970) | (1.487) |
| | Total employment in island | 0.118*** | 0.088 | 0.076*** | 0.095*** | 0.080 |
| | | (0.023) | (0.057) | (0.015) | (0.037) | (0.098) |
| Human Capital | Inv distance to nearest TL institute | 8.630*** | 9.671** | 8.197*** | 15.230*** | 14.038*** |
| | | (2.160) | (3.901) | (2.299) | (4.611) | (4.926) |
| | Pop prop with third level qual | 11.904*** | 7.881*** | 11.561*** | 6.749*** | 6.944*** |
| | | (1.605) | (1.902) | (0.900) | (1.570) | (2.085) |
| Labour Market | Relative employment comp (county) | 0.394 | 0.961 | 1.592* | 1.461 | 1.066 |
| | | (0.928) | (1.096) | (0.867) | (0.990) | (1.289) |
| | Unemployment | 4.716** | 1.980 | 4.593** | 1.627 | 0.697 |
| | | (2.300) | (2.400) | (2.117) | (2.298) | (1.969) |

Notes: Results from alternative panel-specifications. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.
Dublin City omitted. All explanatory variables lagged by one period

Table E2: Comparison of high-tech firm counts from alternative model specifications

| | | Reported results | Alternative Specifications | | | |
|---|---|---|---|---|---|---|
| Variable Type | Variable | NB, PA vce(robust) | Third-level (2002) | Unemployment (2002) | Spatial Lags | All $x_{t-2}$ |
| Broadband | Eircom DSL | 1.172*** | 0.918*** | 1.187*** | 0.478*** | 1.158*** |
| | | (0.170) | (0.173) | (0.172) | (0.176) | (0.171) |
| | MAN area dummy | -0.199 | -0.226 | -0.226 | -0.181 | -0.228 |
| | | (0.171) | (0.170) | (0.170) | (0.171) | (0.167) |
| | MAN effect | 0.453*** | 0.477*** | 0.428*** | 0.388** | 0.426*** |
| | | (0.156) | (0.158) | (0.155) | (0.183) | (0.164) |
| | MAN with backhaul | 0.685*** | 0.718*** | 0.627*** | 0.552** | 0.665*** |
| | | (0.192) | (0.187) | (0.201) | (0.217) | (0.191) |
| Accessibility | Motorway | 0.888 | 1.584 | 0.690 | 1.360 | 1.034 |
| (inverse drive time to nearest...) | | (1.187) | (1.123) | (1.239) | (1.364) | (1.162) |
| | Airport | 2.501 | 1.860 | 2.922 | 0.725 | 2.945 |
| | | (4.677) | (4.711) | (4.674) | (4.781) | (4.509) |
| | Train Station | 0.385 | 0.527 | 0.348 | 0.838 | 0.460 |
| | | (0.689) | (0.651) | (0.690) | (0.724) | (0.668) |
| Agglomeration | Specialisation | -1.989*** | -1.732*** | -1.940*** | -2.394*** | -1.949*** |
| | | (0.377) | (0.375) | (0.382) | (0.340) | (0.379) |
| | Sector share of employment | 0.352 | 0.280 | 0.370 | 0.670** | 0.374 |
| | | (0.344) | (0.321) | (0.362) | (0.341) | (0.335) |
| | Sector density of employment | 0.305 | -0.056 | 0.385 | 0.046 | 0.271 |
| | | (0.866) | (0.631) | (0.948) | (0.872) | (0.838) |
| | Total employment in island | 0.118*** | 0.125*** | 0.118*** | 0.130*** | 0.124*** |
| | | (0.023) | (0.023) | (0.024) | (0.026) | (0.024) |
| Human Capital | Inv distance to nearest TL institute | 8.630*** | 9.687*** | 7.942*** | 9.082*** | 8.739*** |
| | | (2.160) | (2.188) | (2.100) | (2.361) | (2.353) |
| | Pop prop with third level qual | 11.904*** | | 12.954*** | 12.041*** | 12.880*** |
| | | (1.605) | | (1.462) | (1.669) | (1.696) |
| Labour Market | Relative employment comp (county) | 0.394 | 0.807 | 0.495 | -0.056 | 0.432 |
| | | (0.928) | (0.822) | (0.943) | (0.943) | (0.958) |
| | Unemployment | 4.716** | 6.352*** | | 4.154 | 5.123** |
| | | (2.300) | (2.252) | | (2.609) | (2.189) |
| Robustness | Pop prop with third level qual (2002) | | 19.667*** | | | |
| | | | (2.755) | | | |
| | Unemployment (2002) | | | 8.202*** | | |
| | | | | (2.552) | | |
| | Spatial Lag - unemployment | | | | -0.029 | |
| | | | | | (0.023) | |
| | Spatial Lag - third level prop | | | | -0.030* | |
| | | | | | (0.015) | |
| | Spatial Lag - specialisation | | | | -0.174 | |
| | | | | | (0.520) | |

Notes: NB population-averaged panel estimation with cluster robust standard errors. *** p<0.01, ** p<0.05, * p<0.1.
Dublin City omitted. All explanatory variables lagged by one period

Table E3: Comparison of high-tech firm counts from alternative zero-inflated model specifications

| | | Reported results | Alternative specifications | |
|---|---|---|---|---|
| Variable Type | Variable | NB, PA vce(robust) | Zero-inflated Poisson, vce(robust) | Zero-inflated NB, vce(robust) |
| Broadband | Eircom DSL enabled exchange | 1.172*** | 0.898*** | 0.915*** |
| | | (0.170) | (0.187) | (0.189) |
| | MAN area fixed effect | -0.199 | -0.490*** | -0.261 |
| | | (0.171) | (0.169) | (0.230) |
| | MAN no backhaul | 0.453*** | 0.462*** | 0.414* |
| | | (0.156) | (0.179) | (0.235) |
| | MAN with backhaul | 0.685*** | 0.632*** | 0.588*** |
| | | (0.192) | (0.158) | (0.157) |
| Accessibility | Motorway | 0.888 | 0.486 | -0.489 |
| (inverse drive time to nearest...) | | (1.187) | (1.329) | (1.283) |
| | Airport | 2.501 | 1.231 | 2.839 |
| | | (4.677) | (4.401) | (4.827) |
| | Train Station | 0.385 | 0.753 | 0.925 |
| | | (0.689) | (0.652) | (0.654) |
| Agglomeration | Urbanisation (diversity of employment) | -1.989*** | -2.392*** | -2.297*** |
| | | (0.377) | (0.608) | (0.408) |
| | Sector share of employment | 0.352 | 0.141 | 0.264 |
| | | (0.344) | (0.322) | (0.245) |
| | Sector density of employment | 0.305 | -0.142 | -0.222 |
| | | (0.866) | (0.734) | (0.548) |
| | Total employment in island | 0.118*** | 0.098*** | 0.073*** |
| | | (0.023) | (0.019) | (0.017) |
| Human Capital | Inv distance to nearest TL institute | 8.630*** | 6.545*** | 4.299** |
| | | (2.160) | (2.006) | (1.848) |
| | Pop prop with third level qual | 11.904*** | 11.418*** | 10.823*** |
| | | (1.605) | (1.338) | (0.841) |
| Labour Market | Relative employment comp (county) | 0.394 | 0.448 | 0.924 |
| | | (0.928) | (0.797) | (0.756) |
| | Unemployment | 4.716** | 7.918*** | 6.179** |
| | | (2.300) | (2.935) | (2.861) |

Notes: Results from alternative zero inflated specifications. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Dublin City omitted. All explanatory variables lagged by one period
Zero-inflated models consist of a logit first-stage and count second-stage. Only second-stage reported, identical variable set used in both stages

## 4.F Variables and their sources

Table F1: Variables and their sources

| Variable Type | Variable | Spatial Level | Frequency | Source |
|---|---|---|---|---|
| Firm | Firm Location | Urban Field | annual | DJEI |
| Agglomeration Economies | Localisation (sector share of tot employment) | town and below | annual | DJEI |
| | Specialisation | town and below | annual | DJEI |
| | Firms of own nationality | town and below | annual | DJEI |
| Broadband | MAN data | geocoordinates | annual | enet |
| | Eircom DSL enabled exchange | geocoordinates | annual | Eir |
| | BT backhaul | geocoordinates | annual | BT |
| Accessibility | Motorway network | Urban Field | annual | Author's calculations |
| | Rail Stations | Urban Field | 2007 | Author's calculations |
| | Airports | Urban Field | 2007 | Author's calculations |
| Electricity Network | Network interconnectivity | geocoordinates | 2014 | ESB Networks |
| | Count of transformer substations | geocoordinates | 2014 | ESB Networks |
| | Type of transformer substation | geocoordinates | 2014 | ESB Networks |
| Human Capital | University/IT | urban field | 2007 | Author's calculations |
| | Proportion with third level qualification | Urban Field | 2001, 2006, 2011 | CSO Population Census |
| Demographic | Population in persons | Urban Field | 2001, 2006, 2011 | CSO Population Census |
| | Population density | Urban Field | 2001, 2006, 2011 | CSO Population Census |
| Labour market | Relative labour cost | county | annual | CSO Regional Accounts |
| | Unemployment rate | Urban Field | 2001, 2006, 2011 | CSO Population Census |

# Chapter 5

# Conclusion

Infrastructure financing accounted for half of the European Investment Bank's total lending in the European Union between 2005-2009, and European countries will require significant infrastructual investment over the coming decade (Uppenberg et al., 2011). Given that network infrastructure has the characteristics of a natural monopoly, market failures exist due to its public-good properties, the sunk costs of investment, externalities and market power. To quote Helm (2010), *"Infrastructure poses multiple market failures - and intervention poses multiple government failures"*. The appropriate balance of public and private financing is unclear.

Increased population densities in urban areas might lead to congestion and increased cost in upgrading existing systems, while increased geographic dispersal will increase the fixed cost of extending services. This requires public intervention to ensure universal service in areas where private investors are unwilling to enter. Public investment needs to be allocated in a productive way and a balance achieved between under-provision constraining growth, and over-provision creating no added value.

Making systems more flexible and responsive can help strike this balance. The integration of ICT into previously unreactive, and outmoded systems, such as the electricity grid, has enormous potential to improve performance and reduce costs for operators, suppliers and ultimately consumers. This will be increasingly important as we move to decarbonise electricity systems, through the introduction of variable, intermittent renewable generation and the electrification of heating and transport. As this process develops, unforseen outcomes will emerge with positive and negative consequences. Some of these will relate to behaviour and the way in which economic agents interact with these systems.

Infrastructure roll-out over time provides an interesting lens through which to watch such issues emerge, as both demand and supply factors are uneven in time and across geographies.

This thesis addresses a number of questions relating to the interaction of network infrastructure with the behaviour of households and firms, focusing on, but not confined to, the integration of ICT with electricity networks. Chapter 2 examines a case in which improved information and time-of-use pricing in a smart-metering trial can impact on household investment in energy efficiency. Chapter 3 deals with a range of issues related to the adoption of electric vehicles. Both of these chapters highlight cases in which behavioural externalities may emerge through technology adoption, the first as a result of improved information generated by smart-meters, the second explores a case in which geographic heterogeneities in the location of households, and their emergent behaviour may put pressure on low-voltage distribution networks. Finally, Chapter 4 develops the theme of spatial heterogeneity, this time in infrastructure roll-out. In this chapter a unique dataset of Irish infrastructure is assembled, including ICT and electricity networks, and this is used to examine the way in which the geographic and temporal roll-out of networks can impact the location choices of new businesses.

Focusing on **Chapter 2**, this research examines a way in which households respond to improved information and time-of-use pricing in a smart-metering trial. The results show that this can have the unintended effect of reducing household investment in energy efficiency measures, for households who also reduced their electricity consumption. As far as we are aware, no other research has demonstrated this before in this particular domain. However, a wide psychology literature examines behavioural spill-overs, and other researchers have found similar effects when examining spill-overs between domestic electricity and water usage (Tiefenbeck et al., 2013). It is generally accepted that policies targeting efficiency behaviour have greater potential to reduce energy consumption than those targeting curtailment behaviour. It is therefore possible that the electricity reduction induced through increased curtailment, may be offset by a corresponding increase in consumption, due to reduced investment in efficiency. It was not possible for us to examine this question due to data limitations, but this is an area worthy of further research. This points to a potential miss-measurement of outcomes which may partly compromise smart-metering trials.

Of key interest are those households in the treatment group who might have invested, but didn't, and those in the treatment group who invested less than they otherwise might have, were it not for the trial. Developing this line of questioning, if one could then disentangle the psychological motivations behind these decisions, this would both add to the literature and enhance the potential of such interventions to achieve their goals.

It is hoped that this work might encourage researchers and policymakers to give more consideration to the design of survey instruments in smart-metering trials and behavioural interventions more generally, as interventions targeting one specific behaviour can affect a wider range of outcomes than those specifically targeted. It would be useful to get a more complete picture of the decision making processes involved. Given that current EU directives[1] aim to achieve 80% penetration of smart-meters amongst European consumers by 2020, if our findings are replicated by other studies, and people substitute investment for curtailment behaviour in energy efficiency, this could have implications for carbon-reduction strategies.

Another potential behavioural externality is examined in **Chapter 3**. In this case EV adoption and the potential for clustering of electrical load in low-voltage distribution networks. This work expands on the previous chapter, using the same smart-metering dataset to create a nationally representative population of agents, and assign them socioeconomic and environmental characteristics. This work then links the agent population with Census area aggregates, through the use of a microsimulation algorithm, creating spatially explicit agent populations that represent areas of interest.

A stylised agent-based model is then used to simulate adoption profiles, and to examine some dynamics related to the diffusion process. Our model suggests that it is important who adopts first in determining the overall diffusion level in the population. We demonstrate that policies targeting so called "early-adopters" may be limited unless network topology and social influence, two of the factors which drive technology diffusion, are better understood.

In terms of clustering, we find that mild peer-effects may induce very high adoption levels in certain areas. Even if overall adoption levels are quite low, take-up could be highly clustered and peer-effects could significantly increase cluster size within these areas. Granted, this model is a simplistic proxy for reality, but without making strong assumptions about the nature of peer-effects, it highlights how they can lead to high concentrations of adopters in areas with low financial and attitudinal barriers to adoption. This could lead to increased costs for electricity network operators and ultimately for consumers, as the average cost of improvements to the network will be socialised.

A large techno-economic literature focuses on developing control systems to optimise vehicle charging and to integrate distributed generation and load, but more research could be undertaken to examine the economic, social and geographical factors that may generate such problems in the first place. We know very little about the social channels through which adoption is influenced, but a number of studies have identified geographic clustering, such as those related to solar PV and high-voltage air-conditioning systems outlined in the introduction to this thesis. Other, more anecdotal tales also exist. In a fascinating account of the mass-marketing of the Reva G-Wiz electric vehicle in the UK, Keith Johnston, the former Managing Director of its service provider GoinGreen, describes plotting the adopters on a map of London and watching clusters emerge from the data *"literally people in the same street or area buying as a result of a conversation"* (Johnston, 2011). A better understanding of the dynamics which drive such phenomena should be of interest to academics, policymakers and the private sector.

[1] Such as EU Directive 2009/72/EC.

There are a number of ways in which we could improve this work. Improved parameterisation might involve surveying early adopters of EVs in Ireland in order to better capture agent preferences and network effects, potentially calibrating the output to observed diffusion curves. A more sophisticated model might also account for vehicle characteristics, perhaps by linking this work to revealed preference data, such as that obtained in Driscoll et al. (2013).

As a tool to aid grid planning, it would also be interesting to obtain spatial maps of charging infrastructure, both existing and proposed, along with electricity demand at the substation level, and integrate this into a more spatially explicit model, perhaps for a large metropolitan area, such as Dublin City.

**Chapter 4** continues with the spatial focus, collating a wide range of information on various network infrastructures, including GIS maps which display their spatial configuration at different points in time, and other characteristics, such as proxies for the quality of infrastructure. These are assembled to create a unique panel-dataset of Irish network infrastructure, with a particular focus on ICT roll-out, the electricity transmission system and measures of accessibility. An application is then developed which examines how different forms of network infrastructure impact on economic outcomes in complex ways. The application in question is the spatial distribution of new business establishments in Ireland.

This work is the most comprehensive attempt to map multiple infrastructures in Ireland, and we are unaware of any international efforts which have succeeded in drawing together such a rich set of data for any other country.

Results show that ICT roll-out is important in distributing firms more evenly across geographies, particularly for firms in the high-tech sectors. Also emphasized is the importance of human capital, and proximity to third level institutions. The key result in this research is related to the interaction between ICT roll-out and educational attainment. It appears that an area's ability to productively absorb ICT, in the form of new firm formation, increases for higher levels of education and, below a certain threshold, may not be significantly different from zero.

An obvious extension to this work would be to investigate how ICT roll-out has affected employment levels in the high-tech sectors, given the impact it has had on new high-tech business establishments. This work would include employment levels in new establishments along with existing firms.

Further work might also involve collating maps of other network infrastructure such as gas and water networks, as evidence suggests the location of certain sectors, such as the pharmaceutical cluster in Cork was largely driven by the availability of waste-water treatment (Egeraat and Curran, 2013). A richer set of explanatory factors would be useful to explore more deeply the location determinants of more specific firm types.

Another interesting area would be to further investigate the relationship between population and employment density, the various electricity supply variables we have obtained, and demand constraints or opportunities at the substation node level. Spatial heterogeneities in demand will not always be reflected by adequate supply, but increasing use and availability of spatially linked micro data can enable better planning and public infrastructure provision. Further research might link this work with more sophisticated engineering models of optimal power flow to determine locations of network constraints which might inhibit economic activity.

Our results in this chapter have wider implications for regional development policy and universal service provision. Decisions regarding the provision of distinct infrastructures are rarely taken together. This is perhaps due to particular government departments having authority over certain areas, such as transport or broadband networks. An objective of policymakers within each department might be to ensure a minimum level of infrastructure provision in their particular domain. Take for example the Irish National Broadband Plan, a stated objective of which is to *"ensure that all citizens and businesses have access to high speed broadband no matter where they live or work"* (Department of Communications, Energy and Natural Resources, 2012). This focus on a comprehensive roll-out of one service in particular may be suboptimal. If there are complementarities between infrastructures as suggested in this thesis the optimal pattern of deployment (and public subsidies or cross-subsidies) requires information on all linked infrastructures.

Relating this back to Newbery (2012)'s discussion of suboptimal infrastructure investment, in developing an understanding of the complementarities involved, our work can help policymakers and the private sector allocate investment in a more productive way. Clearly new business establishments are only one of many economic outcomes that may be of interest, but by assembling this dataset a range of other outcomes can now be examined. Unusually in Ireland, the same government department[2] has responsibility for both energy and communications. Given the current and expected integration of these networks, this could prove important as these technologies further intertwine.

[2] The Department of Communications, Energy and Natural Resources (DCENR)

# Bibliography

Abrahamse, W., Steg, L., Vlek, C., and Rothengatter, T. (2005). A review of intervention studies aimed at household energy conservation. *Journal of Environmental Psychology*, 25(3):273–291.

Acs, Z. and Armington, C. (2004). Employment growth and entrepreneurial activity in cities. *Regional Studies*, 38(8):911–927.

Akerman, A., Gaarder, I., and Mogstad, M. (2015). The skill complementarity of broadband internet. Technical report, National Bureau of Economic Research.

Alama-Sabater, L., Artal-Tur, A., and Navarro-Azorinn, J. M. (2010). Industrial location, spatial discrete choice models and the need to account for neighbourhood effects. *The Annals of Regional Science*, 47(2):393–418.

Allcott, H. and Greenstone, M. (2012). Is there an energy efficiency gap? *Journal of Economic Perspectives*, 26(1):3–28.

Allcott, H. and Mullainathan, S. (2010). Behavioral science and energy policy. *Science*, 327(5970):1204–1205.

Amaral, L. A. N., Scala, A., Barthelemy, M., and Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences*, 97(21):11149–11152.

Armington, C. and Acs, Z. J. (2002). The determinants of regional variation in new firm formation. *Regional Studies*, 36(1):33–45.

Axsen, J. and Kurani, K. S. (2012). Interpersonal influence within car buyers social networks: applying five perspectives to plug-in hybrid vehicle drivers. *Environment and Planning A*, 44(5):1047–1065.

Axsen, J., Mountain, D. C., and Jaccard, M. (2009). Combining stated and revealed choice research to simulate the neighbor effect: The case of hybrid-electric vehicles. *Resource and Energy Economics*, 31(3):221–238.

Bale, C. S., McCullen, N. J., Foxon, T. J., Rucklidge, A. M., and Gale, W. F. (2013). Harnessing social networks for promoting adoption of energy technologies in the domestic sector. *Energy Policy*, 63:833–844.

Bale, C. S., McCullen, N. J., Foxon, T. J., Rucklidge, A. M., and Gale, W. F. (2014). Modeling diffusion of energy innovations on a heterogeneous social network and approaches to integration of real-world data. *Complexity*, 19(6):83–94.

Barrios, S. (2006). Multinationals' location choice, agglomeration economies, and public incentives. *International Regional Science Review*, 29(1):81–107.

Barrios, S., Bertinelli, L., Strobl, E., and Teixeira, A. C. (2009). Spatial distribution of manufacturing activity and its determinants: A comparison of three small european countries. *Regional Studies*, 43(5):721–738.

Barry, F. and Bradley, J. (1997). FDI and trade: The Irish host-country experience. *The Economic Journal*, 107(445):1798–1811.

Bertanha, M. and Moser, P. (2015). Spatial errors in count data regressions. *Journal of Econometric Methods*, 0(0).

Bhat, C. R., Paleti, R., and Singh, P. (2014). A spatial multivariate count model for Firm location decisions. *Journal of Regional Science*, pages 462–502.

Blumstein, C., Krieg, B., Schipper, L., and York, C. (1980). Overcoming social and institutional barriers to energy conservation. *Energy*, 5(4):355–371.

Bollinger, B. and Gillingham, K. (2012). Peer effects in the diffusion of solar photovoltaic panels. *Marketing Science*, 31(6):900–912.

Brook-Lyndhurst (2011). Exploring catalyst behaviours. A research report completed for the Department for Environment, Food and Rural Affairs. November 2011.

Button, K. J., Leitham, S., McQuaid, R. W., and Nelson, J. D. (1995). Transport and industrial and commercial location. *The Annals of Regional Science*, 29(2):189–206.

Cameron, A. C. and Trivedi, P. K. (2010). Microeconometrics using Stata.

Cameron, A. C. and Trivedi, P. K. (2013). *Regression analysis of count data*, volume 53. Cambridge University Press.

Cameron, A. C., Trivedi, P. K., et al. (2013). Count panel data.

Campbell, A. R., Ryley, T., and Thring, R. (2012). Identifying the early adopters of alternative fuel vehicles: A case study of birmingham, united kingdom. *Transportation Research Part A: Policy and Practice*, 46(8):1318–1327.

Carlton, D. W. (1983). The location and employment choices of new firms: An econometric model with discrete and continuous endogenous variables. *The Review of Economics and Statistics*, pages 440–449.

Carree, M. A. (2002). Does unemployment affect the number of establishments? A regional analysis for US states. *Regional Studies*, 36(4):389–398.

Carroll, J., Lyons, S., and Denny, E. (2014). Reducing household electricity demand through smart metering: The role of improved information about energy saving. *Energy Economics*, 45:234–243.

CER (2011). Commission for Energy Regulation, 2011b. Results of electricity cost-benefit analysis, customer behaviour trials and technology trials. Smart metering information paper 4. Available: http://www.cer.ie.

Colombo, M. G., Croce, A., and Grilli, L. (2013). ICT services and small businesses productivity gains: An analysis of the adoption of broadband internet technology. *Information Economics and Policy*, 25(3):171–189.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., and Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, 80(6):2543–2594.

Corrado, L. and Fingleton, B. (2011). Where is the economics in spatial econometrics? *Journal of Regional Science*, 52(2):210–239.

Coughlin, C. C., Terza, J. V., and Arromdee, V. (1991). State characteristics and the location of foreign direct investment within the United States. *The Review of Economics and Statistics*, 73(4):675.

Crozet, M., Mayer, T., and Mucchielli, J.-L. (2004). How do firms agglomerate? A study of FDI in France. *Regional Science and Urban Economics*, 34(1):27–54.

Darby, S. (2001). Making it obvious: Designing feedback into energy consumption. In *Energy efficiency in household appliances and lighting*, pages 685–696. Springer.

Deloitte (2010). Gaining traction: A customer view of electric vehicle mass adoption in the US automotive market. Available: http://www.deloitte.com.br.

Delre, S. A., Jager, W., Bijmolt, T. H. A., and Janssen, M. A. (2010). Will it spread or not? The effects of social influences and network topology on innovation diffusion. *Journal of Product Innovation Management*, 27(2):267–282.

Department of Communications, Energy and Natural Resources (2012). Delivering a Connected Society: A National Broadband Plan for Ireland. Available: http://www.dcenr.gov.ie/communications.

Di Cosmo, V., Lyons, S., and Nolan, A. (2014). Estimating the impact of time-of-use pricing on Irish electricity demand. *The Energy Journal*, 35(2).

Diffney, S., Lyons, S., and Malaguzzi Valeri, L. (2013). Evaluation of the effect of the Power of One campaign on natural gas consumption. *Energy Policy*, 62:978–988.

Dolan, P. and Galizzi, M. M. (2015). Like ripples on a pond: Behavioral spillovers and their implications for research and policy. *Journal of Economic Psychology*, 47:1–16.

Driscoll, Á., Lyons, S., Mariuzzo, F., and Tol, R. S. (2013). Simulating demand for electric vehicles using revealed preference data. *Energy Policy*, 62:686–696.

Duranton, G. and Overman, H. G. (2005). Testing for localization using micro-geographic data. *Review of Economic Studies*, 72(4):1077–1106.

Duranton, G. and Puga, D. (2001). Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review*, pages 1454–1477.

Egeraat, C. and Curran, D. (2013). Spatial concentration in the Irish pharmaceutical industry: The role of spatial planning and agglomeration economies. *Tijdschrift voor economische en sociale geografie*, 104(3):338–358.

Elhorst, J. P. (2012). Matlab software for spatial panels. *International Regional Science Review*, 37(3):389–405.

Ellison, G. and Glaeser, E. L. (1997). Geographic concentration in US manufacturing industries: A dartboard approach. *Journal of Political Economy*, 105(5):889–927.

Eppstein, M. J., Grover, D. K., Marshall, J. S., and Rizzo, D. M. (2011). An agent-based model to study market penetration of plug-in hybrid electric vehicles. *Energy Policy*, 39(6):3789–3802.

Erber, G. and Madlener, R. (2009). Productivity in the financial sector: Brains are more important than computers. *DIW Berlin, Weekly Report*, 5(9):55–61.

Ernst, C.-S., Hackbarth, A., Madlener, R., Lunz, B., Uwe Sauer, D., and Eckstein, L. (2011). Battery sizing for serial plug-in hybrid electric vehicles: A model-based economic analysis for Germany. *Energy Policy*, 39(10):5871–5882.

European Commission (2009). Directive 2009/72/EC of the European Parliament and of the Council of 13 July 2009 concerning common rules for the internal market in electricity and repealing Directive 2003/54/EC. Technical report, European Commission.

European Commission (2012). Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on energy efficiency, amending Directives 2009/125/EC and 2010/30/EU and repealing Directives 2004/8/EC and 2006/32/EC. Technical report, European Commission.

European Commission (2014). Cost-benefit analyses & state of play of smart metering deployment in the EU-27, swd(2014) 188 final. Technical report, European Commission.

Farhangi, H. (2010). The path of the smart grid. *IEEE Power and Energy Magazine.*, 8(1):18–28.

Faruqui, A., Sergici, S., and Sharif, A. (2010). The impact of informational feedback on energy consumption. A survey of the experimental evidence. *Energy*, 35(4):1598–1608.

Feldman, M. P. and Audretsch, D. B. (1999). Innovation in cities: Science-based diversity, specialization and localized competition. *European Economic Review*, 43(2):409–429.

Festinger, L. (1962). *A theory of cognitive dissonance*, volume 2. Stanford University Press.

Foley, A., Tyther, B., Calnan, P., and  Gallachir, B. (2013). Impacts of electric vehicle charging under electricity market operations. *Applied Energy*, 101:93–102.

Gardner, G. T. and Stern, P. C. (1996). *Environmental problems and human behavior.* Allyn & Bacon.

Gibbons, L. and Van Dender, K. (2012). Road transport: The effects on firms. *Paris School of Economics, available online: http://www. parisschoolofeconomics. eu/IMG/pdf/Gibbons-Lyytikainen-Overman-RUES-2012june.pdf.*

Gibbons, S. and Overman, H. G. (2012). Mostly pointless spatial econometrics? *Journal of Regional Science*, 52(2):172–191.

Gillet, S., Lehr, W., and Sirbu, M. (2006). Measuring Broadbands Economic Impact. Final Report. National Technical Assistance, Training, Research, and Evaluation Project.

Gleeson, A.-M., Ruane, F., and Sutherland, J. (2006). Public policy, sectoral specialisation and spatial concentration: Irish manufacturing 1985-2002. *Statistical and Social Inquiry Society of Ireland.*

Golove, W. H. and Eto, J. H. (1996). Market barriers to energy efficiency: A critical reappraisal of the rationale for public policies to promote energy efficiency. *LBL-38059. Berkeley, CA: Lawrence Berkeley National Laboratory.*

Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420.

Guimaraes, P., Figueirdo, O., and Woodward, D. (2003). A tractable approach to the firm location decision problem. *Review of Economics and Statistics*, 85(1):201–204.

Guimaraes, P., Figueiredo, O., and Woodward, D. (2000). Agglomeration and the location of foreign direct investment in Portugal. *Journal of Urban Economics*, 47(1):115–135.

Guimaraes, P., Figueiredo, O., and Woodward, D. (2004). Industrial location modeling: Extending the random utility framework. *Journal of Regional Science*, 44(1):1–20.

Gungor, V. C., Sahin, D., Kocak, T., Ergut, S., Buccella, C., Cecati, C., and Hancke, G. P. (2011). Smart grid technologies: Communication technologies and standards. *IEEE Trans. Ind. Inf.*, 7(4):529–539.

Haller, S. A. and Lyons, S. (2015). Broadband adoption and firm productivity: Evidence from Irish manufacturing firms. *Telecommunications Policy*, 39(1):1–13.

Hamill, L. and Gilbert, N. (2009). Social circles: A simple structure for agent-based social network models. *Journal of Artificial Societies and Social Simulation*, 12(2):3.

Hamilton, D. J., Nuttall, W. J., and Roques, F. A. (2009). Agent based simulation of technology adoption. Technical report, EPRG Working Paper 0923, Electricity Policy Research Group, University of Cambridge.

Hausman, J., Hall, B. H., Griliches, Z., et al. (1984). Econometric models for count data with an application to the patents-r&d relationship. *Econometrica*, 52(4):909–38.

Helm, D. (2010). Infrastructure and infrastructure finance: The role of the government and the private sector in the current world. *EIB papers*, 15(2):8–27.

Higham, D. and Taylor, A. (2008). CONTEST (CONtrollable TEST matrices). University of Strathclyde.

Hirst, E. and Grady, S. (1983). Evaluation of a Wisconsin utility home energy audit program. *Journal of Environmental Systems*, 12(4):303–320.

Hjorthol, R. (2013). Attitudes, ownership and use of electric vehicles–a review of literature. Technical report, 1261/2013, Institute of Transport Economics, Norwegian Centre for Transport Research.

Holl, A. (2004a). Manufacturing location and impacts of road transport infrastructure: Empirical evidence from Spain. *Regional Science and Urban Economics*, 34(3):341–363.

Holl, A. (2004b). Transport infrastructure, agglomeration economies, and firm birth: Empirical evidence from Portugal. *Journal of Regional Science*, 44(4):693–712.

Jaffe, A. B. and Stavins, R. N. (1994). The energy-efficiency gap. What does it mean? *Energy Policy*, 22(10):804–810.

Jofre-Monseny, J., Marin-Lopez, R., and Viladecans-Marsal, E. (2011). The mechanisms of agglomeration: Evidence from the effect of inter-industry relations on the location of new firms. *Journal of Urban Economics*, 70(2-3):61–74.

Johnston, K. (2011). Telling tales: How to sell an electric car in a petrol market. Available: https://www.smashwords.com/extreader/read/117214/2/telling-tales-how-to-sell-an-electric-car-in-a-petrol-market.

Kandilov, I. T. and Renkow, M. (2010). Infrastructure investment and rural economic development: An evaluation of USDAs broadband loan program. *Growth and Change*, 41(2):165–191.

Kolko, J. (2012). Broadband and local growth. *Journal of Urban Economics*, 71(1):100–113.

Koutroumpis, P. (2009). The economic impact of broadband on growth: A simultaneous approach. *Telecommunications Policy*, 33(9):471–485.

Kowalska-Pyzalska, A., Maciejowska, K., Sznajd-Weron, K., and Weron, R. (2013). Going green: Agent-based modeling of the diffusion of dynamic electricity tariffs. Technical report, Hugo Steinhaus Center, Wroclaw University of Technology.

Krugman, P. (1991). History and industry location: The case of the manufacturing belt. *The American Economic Review*, pages 80–83.

Lai, M. and Poltera, Y. (2009). Lecture with computer exercises: Modelling and simulating social systems with MATLAB. Technical report, Swiss Federal Institute of Technology.

Lawless, M. (2012). Job creation and destruction in recession. *Central Bank of Ireland, Economic Letter Series*, 1.

Lehr, W. H., Osorio, C. A., Gillett, S. E., and Sirbu, M. A. (2005). Measuring broadbands economic impact. *Broadband properties*, page 12.

LeSage, J. P. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*, volume 196. Chapman & Hall/CRC.

Linder, S. (2011). Spatial diffusion of electric vehicles in the German metropolitan region of Stuttgart. In *ERSA conference papers*, number ersa11p557. European Regional Science Association.

Macal, C. M. and North, M. J. (2009). Agent-based modeling and simulation. In *Winter Simulation Conference*, pages 86–98. Winter Simulation Conference.

Mack, E. A. (2014). Broadband and knowledge intensive firm clusters: Essential link or auxiliary connection? *Papers in Regional Science*, 93(1):3–29.

Mack, E. A., Anselin, L., and Grubesic, T. H. (2011). The importance of broadband provision to knowledge intensive firm location. *Regional Science Policy & Practice*, 3(1):17–35.

Mack, E. A. and Grubesic, T. H. (2009). Broadband provision and firm location in ohio: An exploratory spatial analysis. *Tijdschrift voor economische en sociale geografie*, 100(3):298–315.

Mack, E. A. and Rey, S. J. (2014). An econometric approach for evaluating the linkages between broadband and knowledge intensive firms. *Telecommunications Policy*, 38(1):105–118.

Malecki, E. J. (2003). Digital development in rural areas: Potentials and pitfalls. *Journal of Rural Studies*, 19(2):201–214.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542.

McCullen, N. J., Rucklidge, A. M., Bale, C. S. E., Foxon, T. J., and Gale, W. F. (2013). Multiparameter models of innovation diffusion on complex networks. *SIAM Journal on Applied Dynamical Systems*, 12(1):515–532.

McFadden, D. et al. (1973). Conditional logit analysis of qualitative choice behavior.

Merritt, A. C., Effron, D. A., and Monin, B. (2010). Moral self-licensing: When being good frees us to be bad. *Social and Personality Psychology Compass*, 4(5):344–357.

Midden, C. J., Meter, J. F., Weenig, M. H., and Zieverink, H. J. (1983). Using feedback, reinforcement and information to reduce energy consumption in households: A field-experiment. *Journal of Economic Psychology*, 3(1):65–86.

Miller, D. T. and Effron, D. A. (2010). Chapter three-psychological license: When it is needed and how it functions. *Advances in Experimental Social Psychology*, 43:115–155.

Miller, J. H. and Page, S. E. (2009). *Complex Adaptive Systems: An Introduction to Computational Models of Social Life: An Introduction to Computational Models of Social Life*. Princeton University Press.

MIT, S. E. R. G. (2011). Matlab tools for network analysis (2006-2011).

Morgenroth, E. (2009). Exploring the economic geography of Ireland. *Statistical and Social Inquiry Society of Ireland*.

Neuhaus, J. M., Kalbfleisch, J. D., and Hauck, W. W. (1991). A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review/Revue Internationale de Statistique*, pages 25–35.

Newbery, D. (2012). Energy and infrastructure. *submission to the LSE Growth Commission, available at http://tinyurl. com/c8qtahl*.

Noonan, D. S., Hsieh, L.-H. C., and Matisoff, D. (2013). Spatial effects in energy-efficient residential HVAC technology adoption. *Environment and Behavior*, 45(4):476–503.

Ozaki, R. and Sevastyanova, K. (2011). Going hybrid: An analysis of consumer purchase motivations. *Energy Policy*, 39(5):2217–2227.

Palmer, J., Sorda, G., and Madlener, R. (2013). Modeling the diffusion of residential photovoltaic systems in italy: An agent-based simulation. Technical report, E.ON Energy Research Center, Future Energy Consumer Needs and Behavior (FCN).

Parker, E. B. (2000). Closing the digital divide in rural America. *Telecommunications Policy*, 24(4):281–290.

Parks, K., Denholm, P., and Markel, A. J. (2007). Costs and emissions associated with plug-in hybrid electric vehicle charging in the Xcel Energy Colorado service territory. Technical Report NREL/TP-640-41410, National Renewable Energy Laboratory.

Poortinga, W., Steg, L., Vlek, C., and Wiersma, G. (2003). Household preferences for energy-saving measures: A conjoint analysis. *Journal of Economic Psychology*, 24(1):49–64.

Puga, D. (2010). The magnitude and causes of agglomeration economies. *Journal of Regional Science*, 50(1):203–219.

Pusterla, F. and Resmini, L. (2007). Where do foreign firms locate in transition countries? An empirical investigation. *The Annals of Regional Science*, 41(4):835–856.

Research, S. (2007). Scarborough hybrid vehicle owner consumer profile. Technical report. Available: http://scarborough.com/.

Reynolds, P. D., Miller, B., and Maki, W. R. (1995). Explaining regional variation in business births and deaths: Us 1976–88. *Small Business Economics*, 7(5):389–407.

Richardson, P., Flynn, D., and Keane, A. (2010). Impact assessment of varying penetrations of electric vehicles on low voltage distribution systems. In *Power and Energy Society General Meeting, 2010 IEEE*, pages 1–6. IEEE.

Rogers, E. M. (1995). *Diffusion of Innovations, Fourth Edition*. Free Press.

Rothenberg, A. D. (2012). Transport infrastructure and firm location choice in equilibrium: Evidence from indonesias highways. Technical report, Working Paper, The RAND Corporation.

Saarenpaa, J., Kolehmainen, M., and Niska, H. (2013). Geodemographic analysis and estimation of early plug-in hybrid electric vehicle adoption. *Applied Energy*, 107:456–464.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, 1:143–186.

Schmidheiny, K. and Brulhart, M. (2011). On the equivalence of location choice models: Conditional logit, Nested logit and Poisson. *Journal of Urban Economics*, 69(2):214–222.

Schneider, K., Gerkensmeyer, C., Kintner-Meyer, M., and Fletcher, R. (2008). Impact assessment of plug-in hybrid vehicles on pacific northwest distribution systems. In *Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*, pages 1–6. IEEE.

Schuitema, G. and Jakobsson Bergstad, C. (2012). Acceptability of environmental policies. *Environmental Psychology: An introduction*.

Shao, S., Pipattanasomporn, M., and Rahman, S. (2009). Challenges of phev penetration to the residential distribution network. In *Power & Energy Society General Meeting, 2009. PES'09. IEEE*, pages 1–8. IEEE.

Siedschlag, I., Zhang, X., and Smith, D. (2013). What determines the location choice of multinational firms in the information and communication technologies sector? *Economics of Innovation and New Technology*, 22(6):581–600.

Sioshansi, R. and Denholm, P. (2009). Emissions impacts and benefits of plug-in hybrid electric vehicles and vehicle-to-grid services. *Environmental Science & Technology*, 43(4):1199–1204.

Smith, A. (1776). *The Wealth of Nations*.

Sorda, G., Sunak, Y., and Madlener, R. (2013). An agent-based spatial simulation to evaluate the promotion of electricity from agricultural biogas plants in Germany. *Ecological Economics*, 89:43–60.

Sornette, D. (2014). Physics and financial economics (1776–2014): Puzzles, Ising and agent-based models. *Reports on Progress in Physics*, 77(6):062001.

Staats, H., Harland, P., and Wilke, H. A. (2004). Effecting durable change a team approach to improve environmental behavior in the household. *Environment and Behavior*, 36(3):341–367.

Storey, D. J. (1991). The birth of new firms, does unemployment matter? A review of the evidence. *Small Business Economics*, 3(3):167–178.

Taylor, R. D. and Schejter, A. M. (2013). *Beyond Broadband Access: Developing Data-based Information Policy Strategies*. Fordham Univ Press.

Thogersen, J. (2004). A cognitive dissonance interpretation of consistencies and inconsistencies in environmentally responsible behavior. *Journal of Environmental Psychology*, 24(1):93–103.

Thogersen, J. and Olander, F. (2003). Spillover of environment-friendly consumer behaviour. *Journal of Environmental Psychology*, 23(3):225–236.

Tiefenbeck, V., Staake, T., Roth, K., and Sachs, O. (2013). For better or for worse? Empirical evidence of moral licensing in a behavioral energy conservation campaign. *Energy Policy*, 57:160171.

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography*, pages 234–240.

Tran, M. (2012a). Agent-behaviour and network influence on energy innovation diffusion. *Communications in Nonlinear Science and Numerical Simulation*, 17(9):3682–3695.

Tran, M. (2012b). Technology-behavioural modelling of energy innovation diffusion in the UK. *Applied Energy*, 95:1–11.

Uppenberg, K., Strauss, H., and Wagenvoort, R. (2011). Financing infrastructure: A review of the 2010 EIB Conference in Economics and Finance. Annual Economic Conference and Publication.

Van Gaasbeck, K. A. (2008). A rising tide: Measuring the economic effects of broadband use across California. *The Social Science Journal*, 45(4):691–699.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature*, 393(6684):440–442.

Whitmarsh, L. and O'Neill, S. (2010). Green identity, green living? The role of pro-environmental self-identity in determining consistency across diverse pro-environmental behaviours. *Journal of Environmental Psychology*, 30(3):305–314.

Ye, X., Konduri, K., Pendyala, R. M., Sana, B., and Waddell, P. (2009). A methodology to match distributions of both household and person attributes in the generation of synthetic populations. In *88th Annual Meeting of the Transportation Research Board, Washington, DC*.

Zhang, T. and Nuttall, W. J. (2007). An agent based simulation of smart metering technology adoption. Technical report, EPRG Working Paper 0727, Electricity Policy Research Group, University of Cambridge.