

Fernihough, Alan; Colvin, Christopher L.; McLaughlin, Eoin

Working Paper

Mind your language: Explaining the retreat of the Irish language frontier

QUCEH Working Paper Series, No. 24-07

Provided in Cooperation with:

Queen's University Centre for Economic History (QUCEH), Queen's University Belfast

Suggested Citation: Fernihough, Alan; Colvin, Christopher L.; McLaughlin, Eoin (2024) : Mind your language: Explaining the retreat of the Irish language frontier, QUCEH Working Paper Series, No. 24-07, Queen's University Centre for Economic History (QUCEH), Belfast

This Version is available at:

<https://hdl.handle.net/10419/300647>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



QUCHEH WORKING PAPER SERIES
<http://www.quceh.org.uk/working-papers>

MIND YOUR LANGUAGE:
EXPLAINING THE RETREAT OF THE
IRISH LANGUAGE FRONTIER

Alan Fernihough (Queen's University Belfast)
Christopher L. Colvin (Queen's University Belfast)
Eoin McLaughlin (Heriot-Watt University)

Working Paper 24-07

QUEEN'S UNIVERSITY CENTRE FOR ECONOMIC HISTORY
Queen's Business School
Queen's University Belfast
185 Stranmillis Road
Belfast BT9 5EE

July 2024

Mind Your Language: Explaining the Retreat of the Irish Language Frontier*

Alan Fernihough[†] Christopher L. Colvin[‡] Eoin McLaughlin[§]

July 2024

Abstract

Why do we choose one language over another? Rival views see language frontiers as exogenous, driven by policy, or endogenous, determined by social, cultural and economic forces. We study language loss in nineteenth-century Ireland's bilingual society using individual-level data from the 1901 census. Our analysis highlights the intergenerational influence of the education received by a community's elders on subsequent generations' language use. This is consistent with an endogenous demand for English driving language choice because the elder generation's literacy was acquired by attending privately financed voluntary primary schools in a period that predates state-funded compulsory schooling.

Keywords: language loss, bilingualism, education policy, census data, Ireland.

JEL Classification: I24, I28, N33, N93, R12, Z13.

*We thank audiences at the ASREC Europe Conference (London, October 2022), Maynooth University (March 2023), the EHS Annual Conference (Warwick, April 2023), Queen's University Belfast (May 2023), the Education and Development in History Workshop (Valencia, July 2023), Martin Luther University Halle-Wittenberg (July 2023), the World Congress of Cliometrics (Dublin, July 2023), the EHES Conference (Vienna, September 2023), the University of Manchester (October 2023), Heriot-Watt University (November 2023), the University of Galway (January 2024) and the University of Antwerp (May 2024) for comments. Our thanks to Luís Guimarães, Philip Neary, Máire Ní Chiosáin, Niall Ó Ciosáin, Cormac Ó Gráda, Kevin O'Rourke, Rowena Pecchenino and John Turner for comments on previous drafts. Any and all errors remain ours alone. Colvin acknowledges the North-South Research Programme, part of the Government of Ireland's Shared Island Initiative.

[†]Queen's University Belfast, CEPH and CEPR. Email: a.fernihough@qub.ac.uk.

[‡]Queen's University Belfast and CEPH. Email: chris.colvin@qub.ac.uk.

[§]Heriot-Watt University and CEPH. Email: e.mclaughlin@hw.ac.uk.

I'm ashamed I don't speak the language myself. I'm told it's a grand language by them that knows.

Old Mother Grogan says to Haines (James Joyce, *Ulysses*, 1904).

1 Introduction

What drives shifts in language frontiers? Max Weber, writing at a time when new ethnolinguistic nation states were being founded all across central Europe, argued that language is artificial—a political tool used by elites to dominate their society (Weber, 1921). For Weber, language communities are defined in opposition to outsiders, and changes in language use are the outcome of policies of linguistic standardisation and national education (Tada, 2018); language frontiers are determined exogenously through government intervention. Pierre Bourdieu, meanwhile, argued that languages compete in a linguistic market, and language shifts are the consequence of decisions made by individuals and communities in pursuit of an economic power advantage (Bourdieu, 1977). For Bourdieu—who grew up speaking Béarnese, a Gascon dialect from south-west France—language frontiers are endogenous to social, cultural and economic forces (Grenfell, 2011).

Economists have joined this discussion in two ways. The first is empirical and treats language frontiers as exogenous. Such studies evaluate language-shifting policies, typically enacted by sub-national or post-colonial governments to promote indigenous or minority languages through education systems and labour market institutions.¹ The second approach is more closely aligned with Bourdieu in that it explores endogenous reasons for linguistic diversity and language acquisition. These contributions are typically theoretical, relying on mathematical modelling to generate stylised facts about changes in language use across time and place.²

We add to this literature an empirical investigation of a marked change in language use that we argue was not the consequence of a particular government policy intervention, but was rather the confluence of various socioeconomic forces. We track and explain the near-disappearance of Irish, or the solidification of English as the sole language of

¹In their literature review, Ginsburgh and Weber (2020) show most studies find negative economic consequences of such policies. For example, Angrist and Lavy (1997) find a negative impact of a 1983 switch in the language of instruction in Morocco on education attainment; and Chakraborty and Bakshi (2016) measure the negative impact of the 1983 abolition of English language teaching in Indian primary schools on labour market outcomes. A related literature studies the impact of language policy on identity ideology: Aspachs-Bracons et al. (2008) find only compulsory language of instruction policies have any impact on identity formation in a comparison of reforms in Catalonia and the Basque Country introduced in 1983; and Blanc and Kubo (2024) argue the centralised provision of education in the nineteenth century helped to assimilate fragmented communities into a homogenised French nation-state.

²For example, Clingingsmith (2015) uses network theory to explain the emergence of global languages, and John and Özgür (2021) adapt endogenous growth models to capture long-run trends in linguistic diversity. See John (2016) for a discussion of the various strands in this literature.

communication, on the island of Ireland in the late nineteenth century. Among other factors, we highlight how parental demand for primary education, delivered in English, acted as a key driver of the language frontier. Thus, our contribution is an *empirical* investigation of an *endogenous* change to a language frontier—a combination thus far largely absent from the economics literature.

Linguistically, Irish—a Celtic language—is very distant from English, making transitioning between the two more like switching from Basque to Castilian than from Catalan to Castilian.³ Following a century of policies aimed at reviving it, Irish remains a minority language classified as “definitely endangered” by UNESCO (UNESCO, 2003; Moseley, 2010). It was used as the main method of communication by just 115,525 people on the island at the time of the last censuses in 2022, or 1.6 per cent of the population—all of whom also spoke English.⁴ Irish is not even the most commonly spoken second language in Ireland today.⁵

Irish speaking was not always so rare. On the eve of the Great Irish Famine in the mid-nineteenth century, over two million people may have used the language as a means of communication (Hindley, 1990).⁶ The first official estimates of the number of Irish speakers is post-Famine, from 1851, and suggest a stark decline: 1.5 million Irish speakers, or 23.3 per cent of the population (British Parliamentary Papers, 1855, p. xvii). This decline continued apace; by 1901 the share of Irish speakers had halved (FitzGerald, 1984, 2003). Slomanson (2012) frames this development as ‘probably the most dramatic case of a major western European language, being spatially and demographically reduced to an utterly peripheralised minority language in the space of one or two generations’.

We analyse the Irish language’s decline in use with the help of individual-level data from the 1901 census. This high-quality source allows us to track the spatiotemporal change in Irish usage across birth cohorts from the 1820s to the 1890s. This approach is

³Ginsburgh and Weber (2020) discuss the various ways of classifying linguistic difference, and the economic consequences of these differences. Irish differs from English across all measures: cladistic (no common stem in the linguistic tree), lexicostatistic (no common vocabularies), phonetic (uses different sounds), and also in terms of grammar (adopts different structural rules) and syntax (a different sequencing of words). See McMahon and McMahon (2005) for a quantitative study of linguistic distance; and Fabb (2016) for an introduction to the relevant linguistic theory written for economists. There is significant dialectic distance between the Irish spoken in different parts of the island (Kessler, 1995).

⁴While it has been a compulsory part of the Republic of Ireland’s education system since 1934, only 10% of Irish speakers declared they spoke it well in April 2022—and just 71,968 citizens used it daily (Central Statistics Office, 2023a). In Northern Ireland, 228,600 (12%) people over the age of three indicated they had some ability to speak Irish in the last census—and 43,557 (2%) claimed to use it daily (Northern Ireland Statistics and Research Agency, 2022, 2023). Previous research has found that modern Irish speakers enjoy a wage premium (Borooah et al., 2009), although the extent and mechanism of this premium are contested (Watson and Nic Ghiolla Phádraig, 2011).

⁵There are more people who use Polish than Irish at home: 123,968 use Polish daily in the Republic of Ireland (Central Statistics Office, 2023b), versus 71,968 who use Irish (Central Statistics Office, 2023a).

⁶This may be a conservative estimate. Williams and Ford (1992) put the number of Irish speakers at 4 million out of a total population of 5 million at the turn of the nineteenth century. Doyle (2015) reports a range of estimates of Irish speakers, from 45% to 80%, at the start of the nineteenth century. Wolf (2014) puts the pre-Famine headcount at 4.1 million.

superior to using the aggregated data sources employed in previous quantitative analyses of Irish language decline by FitzGerald (1984, 2003) because it enables us to: (1) control for demographic and other individual-level characteristics; and (2) explore different levels of spatial aggregation to ascertain which is the most relevant unit of analysis for the study of language transmission.

Our empirical strategy pitches a set of non-exclusive hypotheses for language decline against one another in a regression model framework. We arrive at these hypotheses by combining the insights of a simple theoretical model we construct to summarise concepts from the literature on the economics of language, with the context-specific historical explanations we derive from the literature on the decline of Irish. This approach incorporates both ideas about the intergenerational transmission of language, and the relative costs and benefits of second language acquisition. While acknowledging the significance of intangible socio-linguistic and cultural factors such as identity, our approach aims to complement these perspectives by illustrating how economic analysis can shed light on language shifts.

Slomanon (2012), following de Fréine (1960), argues that there has been a ‘great silence’ both about explaining Ireland’s language shift itself, but also investigating its impact on the development of Irish society. As a consequence, a simplistic popular narrative has emerged: the decline in Irish was a consequence of English being a colonial language of imposition and coercion.⁷ A less extreme rendition of this view is that government policy deprived Irish a space in the public sphere (in politics, the courts, the army, and schools), reducing its prestige, and dissuading intergenerational transmission. But other explanations abound, some more endogenous in nature, focusing on what economists would recognise as constituting either demand or supply factors. We conclude that it is impossible to delineate from the extant literature what the relative influence was of all these different mechanisms.⁸ Our approach is the first to run the full gamut of explanations against one another in a single quantitative empirical framework, and so the first to arrive at conclusions about the *relative* magnitude of different factors.

It is important to reiterate that we are not evaluating the impact of a specific government policy designed to influence language use. Rather, we are exploring the interactions between different cultural, geographic and socioeconomic factors which can influence language acquisition, none of which come anywhere close to resembling an official language policy intervention. While our contribution is empirical in nature, we are therefore much closer to scholars who use economic theory to explore the dynamics of language frontiers and who focus on endogenous reasons for spatiotemporal change, like Clingingsmith (2015) and John and Özgür (2021). And so we are closer to Bourdieu than Weber

⁷This narrative has influenced works of economics: Hynes (2014) refers to the rise of English in Ireland as an example of “linguistic imperialism”.

⁸Corrigan (2003) warns that many of these rival explanations are anyway a product of the language ideologies of their authors.

in viewing language community as endogenous to economic factors.

We focus our analysis on areas of the island that contained *any* Irish speakers before the Great Irish Famine, a mid-century cataclysm that led to approximately one million excess deaths and a further one million people emigrating. The areas we target are in the provinces of Connacht and Munster, the location of a bilingual society at the island’s language frontier. These are precisely the areas where Irish was “lost” during the subsequent language shift; parts of the island where no Irish was spoken pre-Famine, even as a second language, had no Irish to lose.

We ascertain that the intergenerational transmission of language occurs across multiple geographical contexts. Probit regressions performed on our target population suggest, unsurprisingly, that the Irish language ability among household elders determines language use among the next generation. More surprisingly perhaps, we show that this intergenerational transmission mechanism is just as strong if we expand our definition of the elder generation geographically; Irish prevalence in district elders has as much influence as Irish speaking within the household. This relationship disappears when even larger geographic units are used—a result which helps us identify the optimum geographical unit over which to study factors which affect a language’s communicative value.

Our finding on the influence of literacy is potentially profound and runs counter to explanations that focus on Irish decline being an outcome of government policy. Pre-Famine literacy levels gleaned from the 1841 census capture educational differences *before* the establishment of any national schooling system; Ireland’s historical voluntary education institutions were more influential in determining Irish language loss than the new compulsory schooling that was introduced by the state from 1831. The subsequent English literacy boost provided by national schools was less than half as influential in explaining language loss.

Our study continues by using the relevant literature on the economics of language to set out a conceptual framework in Section 2, and then builds on this to develop a context-specific set of historical hypotheses on language decline in the Irish case in Section 3. Section 4 describes our dataset, while Section 5 outlines our empirical strategy and takes the reader through our key findings for each explanatory category. This section ends by putting all our explanatory factors in one more parsimonious model, to discern their relative importance. Section 6 discusses conclusions, limitations and policy implications. A series of appendices describe our data sources and address various selection issues inherent in our dataset.

2 Conceptual Framework

Language acquisition represents a choice—typically by parents—involving trade-offs that can be represented in an economics framework. Marschak (1965) may have been the first to explicitly incorporate costs and benefits into the study of language, but it is Grin (1992) that offers the most appropriate starting point to theorise about the decline of the Irish language. Grin examines the viability of minority languages within a neoclassical framework, highlighting the pivotal role of “language vitality”—a measure of the positive externalities derived from language use. Once a minority language loses its vitality, its use will decline further. The language may face extinction if the number of speakers falls below a threshold level. Grin’s insights are echoed in Wickström (2005), who reinforces the significance of a critical mass in a bilingual setting, showing that bilingualism’s sustainability relies on parental commitment to raising children in a bilingual environment. In the absence of such commitment, a minority language may succumb to the dominant language.

Clingingsmith’s (2015) exploration of global linguistic diversity corroborates the threshold concept. He finds no correlation between language size and growth once a language reaches a critical mass of 35,000 speakers, with languages that fall below this level being in danger of extinction. He uses network theory to illustrate how this threshold concept appears empirically. The work of John and Özgür (2021) explores the interplay between language use and economic growth and provides an economic logic to changing language thresholds. Using an endogenous growth model, they demonstrate how language use and economic activity co-evolve. Economic expansion increases trade, thereby necessitating and reinforcing common language use. This interdependence implies a consolidation of linguistic networks driven by economic imperatives, offering insight into why language networks emerged and have consolidated in the long run.

The work of linguists confers the language to describe the co-existence of two separate languages within a single community—a phenomenon they refer to as “diglossia” (Ferguson, 1959).⁹ Their framework to study language coexistence typically assigns one as the literary, formal, and “high status” language (H), the other informal, mainly oral, and “low status” (L). There is a degree of specialisation: in certain situations, H is more appropriate, while L dominates in others. In a highly-cited article, Ferguson (1959) outlines nine characterises of diglossia: function (different domain of use for H and L); prestige (higher social prestige for H than for L); literary heritage (greater heritage in H); acquisition (L used in childhood, H in formal settings); standardisation (H is more standardised); stability (unstable and shifting in L); grammar (different grammatical

⁹Ferguson’s (1959) original focus was on two versions of the same language (i.e., dialects), but the framework has since been extended to look at the co-existence of two separate languages (i.e., bilingual societies). See Kandler et al. (2010) for an application of the diglossia model to Welsh and Scottish Gaelic speakers using district-level historical census data.

structures of H and L); lexicon (shared vocabulary in H and L); and phonology (similarity of sounds of H and L).¹⁰

While linguists’ diglossia framework helps to describe a static system with the compartmentalization of two languages, it does not explain the changing dynamics of language acquisition choices across time and place. Indeed, Mufwene (2017) bemoans linguists of a lack of theorising on this issue. It is the contribution of *sociolinguistics* (starting with Bourdieu, 1977), then, that helps us explore these dynamics. Probably the most useful concept from this literature is that languages have a “communicative value”, or Q-value, that captures the potential to connect one speaker with another (De Swaan, 2001).¹¹ Within such a network-type theory, the value of bilingual speakers of a particular language within the global language system is key to explaining that language’s vitality. Individuals have an incentive to “level up” and increase their Q-value by acquiring a more widely-spoken second language. Bilingualism may then represent a transitional stage in a path towards a new monolingualism dominated by that high Q-value language.

In the following, we present a simple model of the Irish language that captures the key elements of the language’s nineteenth-century decline within its bilingual language frontier area. The purpose of this exercise is to demonstrate how this decline can be expressed in a parsimonious economic framework. Readers interested in more comprehensive theoretical perspectives on language acquisition should consult the aforementioned studies.¹² Furthermore, to simplify matters we mostly set aside intangible cultural, political and linguistic forces for language acquisition. We do not seek to undermine their importance; instead, we hope to show how economics can help us understand the decline in a language’s use rather than explain it in its totality.

In this model, parents allocate their children’s language acquisition between English (E) and Irish (I), each represented as a continuum from 0 to 1, where the sum is constrained to unity ($E + I = 1$). The quasi-linear utility function is:

$$U = R_E E + R_I \ln(I) \quad \text{s.t. } E + I = 1 \quad (1)$$

where the language choice is a function of the relative anticipated net “returns”: R_I/R_E . These can include both crude economic returns, such as income, and non-economic concerns, such as the importance of language as a marker of cultural identity. As the nineteenth century progressed, the benefits of English proficiency, including the potential for emigration and improved social status within Ireland, became more pronounced. This

¹⁰The relative status of the two languages is a key modelling parameter for linguists working on language shifts (Kandler and Steele, 2017).

¹¹The Q-value is a function of the prevalence and centrality of a language, where prevalence measures the number of competent speakers and centrality the number of multilingual speakers within a defined space. A hypercentral language has a high Q-value; a peripheral language has a low Q-value.

¹²See, also, Smith (2016), John (2016) and other contributions to the *The Palgrave Handbook of Economics and Language* (2016).

societal shift would likely lead to a decrease in the relative value of R_I/R_E , prompting a reallocation in favour of English (E), as indicated by the model's utility function.

The constraint $E + I = 1$ allows parents to choose a mix of either language. The setup suggests a trade-off between learning English and Irish, but also allows for a spectrum of bilingualism. The quasi-linear utility function allows the languages to be considered as gross substitutes, and thus an older generation's emphasis on Irish will be the determinant of English. Additionally, Irish and English are imperfect substitutes, which reflects the reality that while English and Irish can replace each other in some functions (e.g., as a written language), they are not interchangeable in all aspects because there are unique benefits to speaking Irish that cannot be obtained by speaking English (i.e., benefits linked to cultural identity). Note that the possibility of diglossia is also embedded here. Thus, even when the pecuniary returns to speaking English far exceed those of speaking Irish, there can still be a demand for Irish as an L language, which would be incorporated in R_I . Another realistic feature of this functional form is that English exhibits constant returns to scale, whereas the returns for Irish are decreasing. This captures the importance of English as a *lingua franca* with global utility, while Irish's utility, especially in economic terms, declines as fewer people speak it and is less useful in a globalised economy.

By optimising U with respect to E , we derive the following Walrasian language demand function:

$$E = \begin{cases} 1 - \frac{R_I}{R_E}, & \text{if } R_I \leq R_E, \\ 0, & \text{if } R_I > R_E, \end{cases} \quad (2)$$

where the demand for English is positively related to the language's net returns. In the case where R_I is greater than R_E , a corner solution occurs, and the demand for English becomes zero.

Next, we introduce discrete-time intergenerational effects by assuming that imparting a language is easier when a greater share of individuals from the previous generation speak it. This can be expressed as:

$$R_{I,t} = B_{I,t} + \delta_I I_{t-1}, \quad (3)$$

$$R_{E,t} = B_{E,t} + \delta_E E_{t-1}, \quad (4)$$

where the expected net returns at time t , R_t , are a function of fixed/exogenous net benefits, B_t , and endogenous costs with $\delta > 0$. This implies the greater the number of parents who speak English/Irish, the lower the cost for the next generation to learn English/Irish. Note that here we do not consider network effects explicitly, although they are implicitly incorporated by assuming that each parent is identical, and thus that the

previous generation's language ability is equivalent to the next.¹³

We can see that the English language evolves according to the following law of motion, considering a corner solution when $R_{I,t} > R_{E,t}$:

$$E_t = \begin{cases} 1 - \frac{B_{I,t} + \delta_I I_{t-1}}{B_{E,t} + \delta_E E_{t-1}}, & \text{if } R_{I,t} \leq R_{E,t}, \\ 0, & \text{if } R_{I,t} > R_{E,t}. \end{cases} \quad (5)$$

In practical terms, the presence of a corner solution signifies that when the benefits of speaking English surpass the threshold represented by the $R_{I,t}$ term, there is a shift towards English. This shift not only allows for English speaking to begin, but also sets in motion a compounding effect where reduced costs for the next generation further incentivise the adoption of English, contributing to the Irish language's demise.

Despite its parsimony, the model encapsulates core features that shed light on the dynamic interplay between linguistic choice and socioeconomic incentives. Primarily, the model illustrates how English could supplant Irish within a bilingual community. It specifies how the relative advantages of learning English over Irish, which undoubtedly grew over the nineteenth century, not only fostered a contemporaneous shift towards English speaking, but also significantly influenced intergenerational language transmission. As the relative returns to English proficiency rose, this reduced Irish in the subsequent generation who, in turn, faced a lower cost in transmitting the language to their children, the grandchildren of the previous generation.

Our framework also works with language revival. A heightened enthusiasm for the Irish language could bolster the prevalence of Irish speaking, with government policy directed at increasing R_I . Nevertheless, if the prevailing levels of proficiency in Irish are markedly low (E_{t-1} is close to 1), then the costs associated with language revival may still dominate, thereby impeding endeavours to improve the language's status. This aspect potentially explains the rare day-to-day usage of the Irish language today, despite substantial governmental policy interventions aimed at its revival across the past century.

The model stands apart in demonstrating that the core tenets of Irish language attrition can be adeptly captured within a neoclassical economic framework. This framework incorporates a straightforward trade-off between the costs and benefits associated with language acquisition and a simple mechanism of intergenerational transmission. We incorporate these insights in our empirical approach by including variables that reflect historical elements that would have influenced language choice, as well as broader network-type elements that reflect the spillovers in the language use.

¹³If one wanted to incorporate network effects, these could plausibly enter the $B_{I,t}$ term, akin to Grin's (1992) language vitality parameter.

3 Hypothesis Development

Scholars have proposed six non-mutually exclusive hypotheses for the loss of the Irish language over the nineteenth century: (1) the lack of support provided by Irish nationalists; (2) the influence of the English-language education system; (3) its abandonment by the Catholic Church; (4) demographic changes driven by emigration; (5) the growing influence of the English state; and (6) economic factors increasing the demand for English. Our goal is to translate these qualitative theories into empirically testable hypotheses, which we evaluate using an econometric framework. These six factors can be seen as context-specific influences affecting the costs and benefits of language acquisition and the dynamics of intergenerational language transmission in the model described in Section 2.

3.1 Nationalist Politics

As the main figure in the nationalist movement of the early and mid-nineteenth century, Daniel O’Connell (1775–1848) is frequently identified as a key culprit in the decline of the Irish language for his failure to support it (Wall, 1969). Known as The Liberator for his pivotal role in Catholic emancipation, O’Connell’s negative views of the Irish language are widely quoted:

I am sufficiently utilitarian not to regret its gradual abandonment. A diversity of tongues is no benefit; it was first imposed on mankind as a curse, at the building of Babel. It would be of vast advantage to mankind if all the inhabitants of the earth spoke the same language. Therefore, although the Irish language is connected with many recollections that twine round the hearts of Irishmen, yet the superior utility of the English tongue, as the medium of all modern communication, is so great that I can witness without a sigh the gradual disuse of Irish (quoted in Murphy, 1948; Ó Cuív, 1969; Ó Tuathaigh, 1974).

The main accusation against O’Connell and his supporters was that they did nothing to encourage a Gaelic cultural revival alongside their political activities (Murphy, 1948). While Catholic emancipation was a mass political movement, its medium was English.¹⁴ Ironically, the subsequent generation of nationalists, so-called “Young Irelanders”, advocated for an Irish language under the influence of continental European romanticism which associated language and national identity. However, none of its leaders could speak Irish and their main publication, *The Nation*, was published in English (Doyle, 2015).

¹⁴Doyle (2015) speculates that O’Connell would have been incapable of encouraging Irish even if he wanted to; he speculates that his Irish had not developed beyond what he had learnt as a child. But Wolf (2014) disputes such narratives, arguing that O’Connell was not shy to use Irish in contexts where it aided in communicating his politics.

3.2 Education

Explanations that focus on education normally highlight the influence of the new compulsory national schooling system that emerged in the 1830s. But voluntary primary education was already present on the island. As Corrigan (2003) urges, it is important to disentangle the relative influence of these two institution types because they have different implications for the drivers of language choice.

3.2.1 Hedge Schools

Primary education in early nineteenth-century Ireland was provided by a myriad of institutions known as “hedge schools”. About 9,000 in number, this was a decentralised system without meaningful state involvement (FitzGerald, 2013).¹⁵ There was significant variation in registration and attendance by birth cohort and region, but overall about 44 per cent of boys and 26 per cent of girls attended schools between the ages of six and 13 at the time of the 1821 census (Blum et al., 2017). Children in Irish-speaking areas were marginally less likely to attend school, but this effect is very small in a descriptive regression of school attendance (Ó Gráda, 2013).

The curricula of hedge schools reveal information about the educational preferences of fee-paying parents. The evidence suggests that English was the medium of instruction everywhere (Wall, 1969); parents sent their children to these schools exactly to acquire the language of economic opportunity (Mac Giolla Chríost, 2005). Besides scripture, the other subject taught in these schools was arithmetic (McManus, 2002). Poor parents demanded this broad curriculum to realise motivations for their children to advance socially and acquire higher socioeconomic positions.

3.2.2 National Schools

Ireland was the first polity of the United Kingdom to receive state-funded primary schooling, starting in 1831.¹⁶ The establishment of the National Education Board, the central government agency which set the curriculum in these schools, has long been viewed as a key factor in the decline of the Irish language (e.g., Corcoran, 1925). This is because lessons were directed to be delivered in English, with no allowance for the Irish language until the end of the nineteenth century. The argument goes that national schools were used to promote social and cultural assimilation (Akenson, 1970; Coolahan, 1981; Nic Craith, 2002). Williams and Ford (1992) go as far as arguing that ‘there is no

¹⁵At the time of an 1824 survey into school provisioning, about one-third of schools were managed by Catholic institutions, one-third by Protestants, and a further third by non-denominational charitable groups—although children from one denomination would be taught by schoolmasters from another in denominationally diverse areas (FitzGerald, 2013).

¹⁶National schools did not replace hedge schools overnight; some Catholic dioceses continued to sponsor hedge schools (Barr, 2009).

doubt at all that the education policies of the Board turned an Irish-speaking nation into an English-speaking one' (p. 256). Coolahan (2001) argues that the policy of national schooling was purposefully exclusionary of Irish, but adds that it was not the cause of decline—rather, it was a 'supportive force hastening the decline'.

In the first decade of existence of the new system, the number of national schools was in the range of 1,000; by the turn of the century, the number was close to 9,000 (British Parliamentary Papers, 1851, 1900). The state paid teacher salaries and provided two-thirds of funding towards the building of new schools. The primary school system was intended to provide non-denominational education, but a *de facto* denominational system emerged (Coolahan, 1981). The core curriculum of national schools was reading, writing, and arithmetic (the so-called 3-Rs), with English always being the medium of instruction. National schools provided education to the bulk of the population. Comerford (2003) concludes that 'as the Irish population acquired literacy it did so in English'.

Later revivalists believed that reintroducing compulsory Irish would bring back the language. They achieved little success. Wall (1969) points to the fact that secondary education, which was never state-funded, also made use of an English curriculum. University education too was in English. Understandably, there was little interest in changing the medium of instruction in post-primary settings. The distinct lack of an Irish language print culture, and the dramatic decline in the availability of Irish-language literature from the 1850s described recently in Ó Ciosáin (2024), cannot have helped.¹⁷

3.3 Religion

Historically, Ireland's religious divide was also a linguistic divide; Catholicism was associated with Irish and Protestantism with English (Wolf, 2014). Roman Catholicism was a pillar of Irish nationalist identity, and in the context of an Anglican state also an alternative source of social control (O'Brien, 1989). The fact that this facet of Irish nationalism was arguably indifferent, if not unsupportive, of the Irish language is highly significant.

The post-Famine period saw a shift from 'traditional lax and heterodox practice towards a more formal and rigorous Catholicism' (Jackson, 1999). This resulted in increased mass attendance, with an associated enhancement of the political and social influence of the clergy. Catholic religious services were not through Irish (or, indeed, English); mass itself was typically delivered in Latin (Cullen, 1990). There was little support for the Irish language in the seminary.¹⁸

¹⁷Popular literature accessible to rural folk came in the form of inexpensive, small-format, English-language paperback prints called chapbooks (Ó Ciosáin, 1997). Part of the problem in getting a mass Irish-language print culture was the absence of a uniform orthography, typology and spelling—unlike other Celtic languages (Ó Ciosáin, 2024).

¹⁸Maynooth College, a state-funded Catholic seminary, appointed a cleric to its chair in Irish who had not met the advertised standard and did not develop the language there (McMahon, 2008).

While the Catholic clergy are apportioned the blame for the demise of Irish in popular narratives, Wall (1969) believes this explanation lacks nuance as it overlooks the much earlier change in the status of the language. And Wolf (2014) disputes the narrative that the Catholic church was antipathetic towards the Irish language. Instead he argues that clerics encouraged and facilitated bilingualism.¹⁹

3.4 Demographic Change

Nineteenth-century Ireland has a peculiar demographic history characterised by mass depopulation. Emigration has been a prominent driver of this change.²⁰ The Great Irish Famine intensified existing migration patterns and precipitated an ageing population (Colvin et al., 2024). The Famine’s proximate cause was an exogenous ecological shock—*Phytophthora infestans* (potato blight)—which decimated the main source of food of the majority of the population. Internal and international migration was the main non-governmental responses to famine conditions (Ó Gráda, 1995; Ó Murchadha, 2011).

Ravenstein (1879) raises emigration as a possible contributing factor in the decline of Irish and the shift to English, but discounted this argument as migrants tended to be from English-speaking parts of the country and that Irish migrants in Britain were not Irish speaking.²¹ In pre-Famine Ireland, speaking Irish was correlated with various indicators of poverty (Fernihough and Ó Gráda, 2022). These were the areas which experienced the most acute effects of the Famine and, therefore, had the scope for the largest potential change in language use in *absolute* terms.

Instead, Fitzpatrick (1986) argues that emigration was a key factor in the growth of literacy—which was by definition literacy in English.²² Lee (1989) highlights the fact that many other Europeans migrated to the United States and adopted the English language, but only at the migrant destination; the Irish, by contrast, could become proficient in English at the point of departure.

3.5 The English State

Doyle (2015) documents three distinct historical language communities in Ireland: Gaeil (Irish), Sean-Ghaill (Old English) and Nua-Ghaill (New English). Sean-Ghaill were bilingual; Gaeil and Nua-Ghaill were monoglot Irish and English speakers. The English-speaking frontier in 1500 was a small area, known as the English pale, that was under

¹⁹For example, John Kiely, a Catholic clergyman from Co. Cork, stated that clergy were ‘obliged to catechize and instruct in both [English & Irish] languages’ (British Parliamentary Papers, 1825, p. 320)

²⁰Ireland had one of the highest emigration rates per 1,000 population in Europe (Hatton and Williamson, 1994).

²¹Corrigan (2003) agrees that Famine-induced migration from Irish-speaking areas was quantitatively insufficient to account for the size of the language shift.

²²However, Fitzpatrick’s (1986) argument focuses on explaining female emigration and overlooked male literacy and emigration trends.

Royal control (Doyle, 2015, Figure 3.1). The frontier gradually expanded, helped by the expansion of the Tudor and Stuart states.

The Cromwellian conquest in the 1640s was a large upheaval that precipitated the mass dissemination of English through the arrival of English and Scottish settlers in parts of Ireland beyond the pale (known as plantations). However, the estimated English-speaking frontier, while experiencing growth, was still in the absolute minority on the island (Doyle, 2015, Figure 4.1). Dating the transition from Irish-speaking to English-speaking majority is debated, with Doyle (2015) pessimistically placing it already in the mid-eighteenth century, but Wall (1969) and others one hundred years later.

Williams and Ford (1992) point to the collapse of Gaelic Ireland, which they associate with a set of legal disabilities imposed on the non-Anglican majority known collectively as the penal laws, as a key driver of the language frontier. Wall (1969), meanwhile, argues the decline set in only after the penal laws were largely already repealed. For Lee (1989), it was the post-1800 *growth* of the size of the state in Ireland that was a key factor in the adoption of English, as state bureaucracy was conducted entirely in English.²³

3.6 Changing Demand for English

The final hypothesis to account for language dynamics concerns the growing *economic* demand for English. Doyle (2015) argues it was a gradual shift in demand rather than an abrupt breaking point. Parents came to see English as the language of opportunity that facilitated social mobility. It gave lower and middle-class folk access to socially important professions such as in the army and the church (O’Brien, 1989). English also facilitated overseas migration, even before the Famine. As the older generation of grandparents passed away, the preference for English became more pronounced, making the shift on the surface level seem abrupt without taking into account the generational differences.

It has also been argued that the language shift was a response to the growth in commercial relations with, and dependence on, Britain; English was the language of commerce, while Irish was the language of poverty (Nic Craith, 2002). However, Lee (1989) argues that this explains the adoption of English, but not the abandonment of Irish. Other small countries that competed in the same export markets—such as Denmark—did not abandon their native tongues. Wolf (2014) argues that language was not a barrier to market access as Irish speakers could anyway use bilingual interpreters.

²³One indicator of the state’s reach is the high and increasing level of policing across the island over the nineteenth century (McLaughlin et al., 2021).

4 Data Construction and Sample

We use census data to retrospectively track Irish’s spatiotemporal decline. While FitzGerald (1984, 2003) relied on the published aggregated census returns of 1851, 1861, 1881, and 1911, our primary data source is the full 1901 census return, recorded at household level (held at the National Archives of Ireland). We believe this census source is superior for two reasons. Firstly, we use the full individual returns, meaning we have distinct demographic data for over four million subjects. The full returns allow us to explore various levels of spatial aggregation, unlike FitzGerald (1984), who was restricted to the geographically aggregated Baronial system ($N = 336$). Secondly, while FitzGerald (2003) used more geographically disaggregated district electoral division (DED, $N > 3,400$) spatial units from 1911, this source restricts the pre-Famine birth cohorts to all individuals over 60 years of age. Our 1901 census data are less likely to suffer from age-related sample selection complications.²⁴

The 1901 returns required household heads to specify whether each individual household member spoke Irish, distinguishing bilingual from monolingual speakers. Notably monoglot English speakers did not have to complete this section, highlighting the English language’s dominance. This prevalence of English was reflected in the returns, as 20 per cent of the population 20 years of age and above indicated some proficiency in Irish, but just 1 per cent were returned as “Irish Only” speakers. Given the extremely low number of monoglot Irish speakers, we focus our analysis on both bilingual and monoglot Irish.

Figure 1 displays the Irish language’s retreat over the nineteenth century across pre- and post-Famine birth cohorts, aggregated to the DED level.²⁵ Before the Famine, Irish-speaking majorities scarcely existed outside the most Western and Southern parts of the island. This frontier weakened in the second half of the nineteenth century, with all of the decline—unsurprisingly given that Irish was essentially extinct elsewhere—occurring in formerly Irish-speaking strongholds.

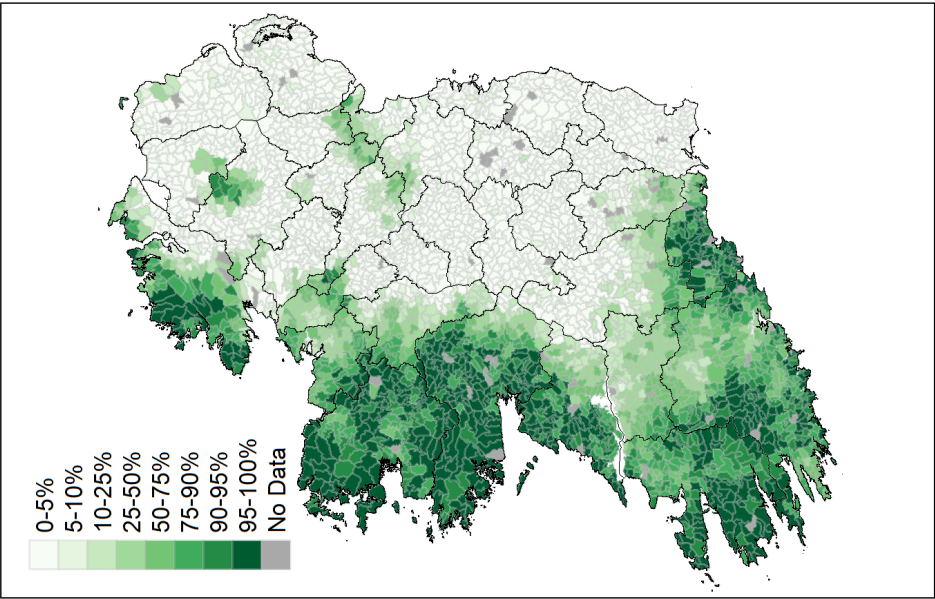
Also evident in Figure 1 is the bimodal distribution of the Irish-speaking share across the island. Districts either had full Irish coverage (over 90 per cent of the cohort population with Irish-speaking ability) or a complete absence (fewer than 10 per cent). For pre-Famine birth cohorts, 20 per cent of districts could be categorised as fully Irish speaking, and Irish was completely absent in nearly half of these areas. Post-Famine, the share of districts where over 90 per cent of the population were Irish speaking fell from

²⁴The 1911 census also suffers from a well-known age-exaggeration issue caused by the introduction of the state pension in 1909 (de Bromhead et al., 2021).

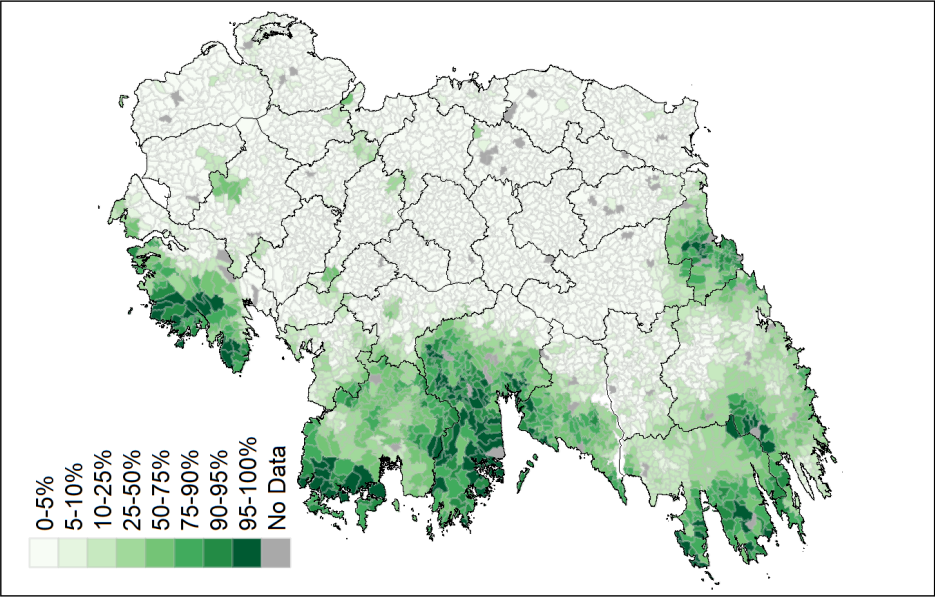
²⁵DEDs were Ireland’s basic territorial unit used for rating and census purposes, and to organise poor relief. They were brought together into larger areas for defining polling districts and constituencies. Devised in the mid-nineteenth century to create equally populated units, they were small (on average 25 km²) but varied in size; they reflected local interests and historical and geographical considerations. They can be thought of as Local Administrative Units regions, one level below the basic NUTS-3 region level used in European statistics gathering today.

Figure 1: Irish Language Ability by DED in the Pre- and Post-Famine Birth Cohorts

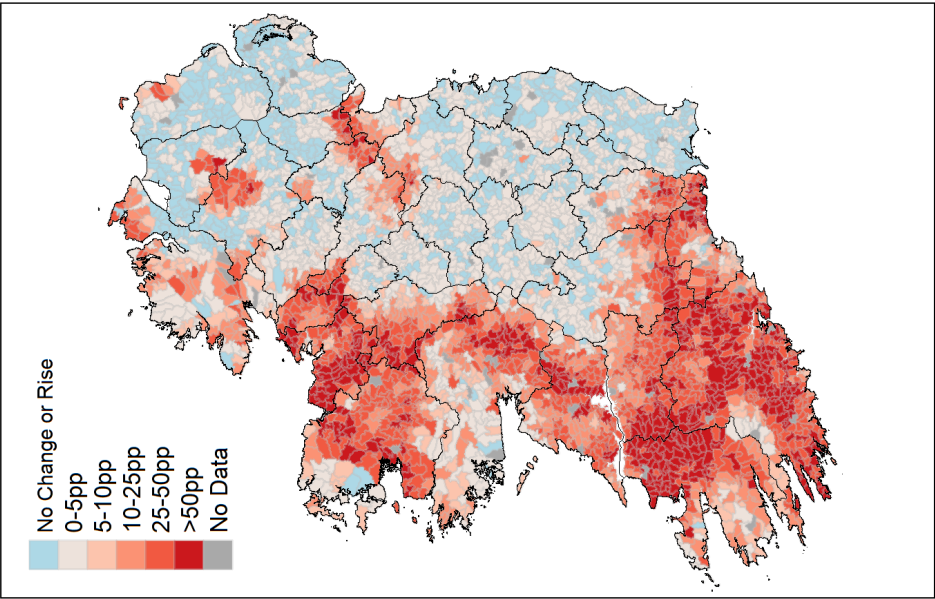
Map A: 1821-41



Map B: 1861-81



Map C: Difference



Note: DED outlines in grey; county outlines in black.

Table 1: Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
Irish Language Ability	290,251	0.495	0.500	0	1
Female	290,251	0.490	0.500	0	1
Year of Birth: 1871-81	290,251	0.605	0.489	0	1
Population Fall: 1841-51	290,251	0.290	0.234	-0.988	1.059
Population Fall: 1851-71	290,251	0.168	0.246	-0.650	2.196
Latitude	290,251	53.026	1.012	51.437	55.351
Longitude	290,251	-8.844	0.743	-10.449	-6.214
Previous Generation Irish-Speaking: Household	228,881	0.818	0.368	0.000	1.000
Previous Generation Irish-Speaking: Townland	289,399	0.787	0.247	0.000	1.000
Previous Generation Irish-Speaking: DED	290,251	0.737	0.201	0.216	1.000
Previous Generation Irish-Speaking: County	290,251	0.651	0.217	0.034	0.893
Literacy, Both Read and Write	289,129	0.864	0.343	0	1
DED Literacy: Pre-National Schools	290,251	17.633	7.113	0.000	37.234
DED Literacy: National School Increase	290,251	23.051	11.738	-11.352	68.481
DED Literacy: Post-Famine Increase	289,129	45.399	10.884	6.178	81.963
Roman Catholic	290,032	0.959	0.198	0	1
Roman Catholic Name Index	275,759	70.576	24.508	0.000	100.000
Traditional Name Index	275,759	26.322	24.100	0.000	69.610
Teachers in DED (Per Capita)	290,251	0.006	0.005	0.000	0.043
Priests in DED (Per Capita)	290,251	0.003	0.006	0.000	0.131
Police in DED (Per Capita)	290,251	0.003	0.004	0.000	0.030
Market Access, 1841	290,251	4,257.650	4,591.274	13.291	45,175.350
Port Access, 1841	290,251	5.743	2.346	1.252	12.788
Market Access Increase: 1841-71	290,251	0.791	0.442	0.098	3.382
Port Access Increase: 1841-71	290,251	0.373	0.164	0.005	0.835
Resident in Birth County	290,251	0.943	0.232	0	1
Share of DED Born in County of Residence	290,251	93.943	7.255	14.188	100.000
Pre-Famine Land Value per Acre	290,251	0.423	0.462	0.008	3.190
Occupation: Irish Speaking	281,702	0.234	0.085	0.000	1.000

Note: Authors' calculations using data described in Appendix A.

20 per cent to 5 per cent, while the equivalent monoglot English share rose from half to two-thirds.

Three scenarios describe how the prevalence of Irish language ability fared over the century. Firstly, Irish remained popular in areas with high initial concentrations of Irish speakers. Secondly, Irish speaking declined in areas where Irish had been popular in past generations. Finally, Irish speaking remained rare, without exception, in areas where Irish had already disappeared. Informed by the information in Figure 1, we compose our analysis sample by creating a data subset consisting of all individuals aged between 20 and 39 (born between 1861 and 1881) resident in pre-Famine Irish-speaking majority DEDs. Segregating our analysis sample means that we target individuals located in areas where a majority of the previous generation spoke Irish—precisely the group where Irish was lost in the subsequent generation because individuals from Irish minority DEDs did not take up Irish and had no language ability to lose.

Table 1 reports the summary statistics of the variables used in our analysis. Appendix A provides a complete outline of the data sources and data generation procedures. We

discard all observations where the return indicated that they were neither the household head nor a relative. This removes people residing in institutions and more transient groups such as boarders and lodgers. The majority of these variables account for either individual-level differences or DED group-level differences.

Overall, around half of the observations indicated they possessed Irish-speaking ability in the 1901 Census. As expected, this share is far greater than the 14.5 per cent in these age cohorts returned as Irish-speaking in the complete census returns because our sample is comprised of individuals resident in formerly Irish-speaking majority districts. The sex ratio is balanced; 49 per cent of the sample is female. The age of the sample skews slightly younger, with three-fifths belonging to the 1871-81 birth cohort compared to the two-fifths in the 1861-71 group. On average, the observations were residents in DEDs that lost around one-third of their population (0.29 log points) in the Famine (between the 1841 and 1851 censuses) and nearly 20 per cent (0.17 log points) in the post-Famine period (between the 1851 and 1871 censuses). The DED's latitude and longitude coordinates feature throughout our regression analysis to control for spatial confounding.

Intergenerational transmission is measured across several geographic strata. All individuals in our data reside in formerly Irish-speaking districts; the average Irish-speaking at the household level among elders (40 years or older and a relative) was around 82 per cent. This means over four-fifths of the sample had an Irish-speaking parent. The sample size drops from around 290k to 229k for this variable, as a minority share of the population did not live with an older relative. The share of the previous generation with Irish ability declines gradually as we increase the geographic scope from the household (82 per cent) to the county (65 per cent) via townlands (79 per cent) and DEDs (74 per cent).

It is worth noting the scale of these geographical units. In 1901 there were nearly 70k townlands with an average townland population of just 64 people. DEDs were larger, with over 3.5k distinct areas recorded and an average population of 772. Counties are far larger still; Ireland's 32 counties had an average population of 138k in 1901. Aggregating neighbourhoods at different resolutions allows us to investigate the relevant geographic level through which intergenerational transmission occurs.

The establishment and expansion of national schools from the 1830s on was an important driver of literacy in nineteenth-century Ireland. This is evident in the statistics on individual literacy in our data, with 86 per cent indicating they could both read and write. However, we can track change over the century more broadly. The 1841 Census Report (British Parliamentary Papers, 1843) provides data on a similar scale to DEDs and allows us to create a measure of literacy that pre-dates national schools. Since this measure of community-level literacy pre-dates the birth of the sample observations, it serves as a proxy for historical levels of schooling. Aggregate pre-national school literacy stood at a mere 18 per cent. Using the 1901 data, we can discern the immediate impact

of national schooling by looking at the literacy difference between those aged 60 and over (born before 1841) and the previous literacy total based on the 1841 official report. Table 1 suggests the introduction of national schools led to a 23-percentage point increase in literacy in the years that immediately followed. This trend continued and the difference between the pre- and post-Famine birth cohorts (the literacy gap between our DED cohorts: 20–39-year-olds and the group aged 60 and above) was 45 per cent, as schooling became almost universal across the island.

We explore the influence of religion on both the extensive and intensive margin. The extensive margin, one’s religious affiliation, is not particularly illuminating because the sample is overwhelmingly (96 per cent) Catholic. However, like Connor (2021) we can use first names to measure religious adherence at the intensive margin. We create a name index that measures names considered distinctly more “Catholic”. A score of zero indicates no Catholics held this name; a score of one is the opposite extreme. We also replicate Connor by constructing a traditional name index that measures historically popular names. The presence of the state and church is potentially important, and we create variables measuring the relative frequency of teachers, policemen, and all religious-based occupations. The latter category was mainly Catholic nuns and priests (94 per cent of religious-based occupations in our sample DEDs).

Geographical remoteness and market access are another plausible source of explanatory power. Here we follow Fernihough and Lyons (2022) and measure both market and port access. The raw market and port access variables are set at their pre-Famine values (i.e., assuming no railways). Market access is the cumulative sum of all land values weighted by their ease of access (i.e., the time and cost it takes to travel between them via land, waterway, or road). Port access measures the ease at which the district can travel to one of Ireland’s seven main ports. The introduction of railways revolutionised nineteenth-century transportation. Since this revolution overlaps with our sample window, we incorporate the influence of rail with variables that measure increases to both access variables. Between 1841 and 1871 market and port access rose by 79 and 37 per cent.

We define all people resident in counties outside their declared county of birth as internal migrants. This variable provides insight into the influence of migration on Irish speaking. 94 per cent of our observations resided in their county of birth at the time of the census. Migration at the broad DED level is also measured with an almost identical level of community migration evident (94 per cent). The low share of internal migrants in our data reflects the nature of our sample, as few migrants were moving into districts with pre-Famine Irish-speaking majorities.

Land values, reported in the 1851 census but surveyed before the Famine, are also included. Note that the value per acre, essentially a fair annual rent, varies considerably. If a substantial number of people dropped Irish to improve their chances in the labour market, then occupational advancement may have contributed to the language’s demise.

For each observation, we take the stated occupation and assign the occupation’s average level of Irish speaking. Not everybody returned a census form with a legible occupation and we only use occupational categories with over 200 respondents. Fortunately, most of the data (97 per cent) were assigned values for this variable.

5 Results

5.1 Empirical Strategy

We estimate various probit regressions that model the Irish-language ability dummy variable as a function of multiple explanatory factors. Equation (1) outlines the basic estimation equation:

$$\Pr[\text{Irish}|X, Z, \text{lon}, \text{lat}] = \Phi[Z\gamma + X\beta + f(\text{lon}, \text{lat})] \quad (6)$$

where Z includes the baseline controls: age cohort, sex, and two variables capturing historical DED population changes, one between 1841 and 1851 and the other between 1851 and 1871. The last two variables control for differences in both Famine depopulation as well as the post-Famine aftershocks experienced by many of the poorest districts. By including Famine depopulation, we address the potentially offsetting effect of sample selection biases, although extensions of our analysis probe the robustness of this approach. The function $f(\text{lon}, \text{lat})$ represents the inclusion of both longitude and latitude coordinates as third-order polynomials with interaction terms, to control for spatial confounding. Finally, the term X includes application-specific explanatory variables.

To ease interpretation, we report all coefficients expressed as marginal effects and rescale all continuous explanatory variables by normalising them to have a zero mean and 0.5 standard deviation. This means that all marginal effects represent the predicted probability change based on a two-standard deviation (2SD) increase in the associated explanatory variable. The advantage of this approach is that it makes all effects comparable—the largest estimated marginal effect thus offers the most explanatory power throughout. Standard errors are cluster-corrected to account for correlation within DEDs, although the conventional markers of statistical significance are (unless explicitly noted) omitted to avoid conflating statistical and economic significance (American Statistical Association, 2016), and in the interests of parsimony.

5.2 Spatio-Linguistic Geography

How does language transfer from one generation to the next? Table 2 reports probit regression marginal effects from several model estimates that link an individual’s Irish-speaking ability to the ability of their household and neighbours. We measure intergen-

Table 2: Spatio-Linguistic Geography: Probit Regression Marginal Effects

	Irish Language Ability						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.001 (0.002)	-0.024 (0.002)	-0.015 (0.002)	-0.020 (0.002)	-0.027 (0.002)	-0.003 (0.002)	-0.024 (0.002)
Year of Birth: 1871-81	-0.151 (0.002)	-0.106 (0.002)	-0.137 (0.002)	-0.130 (0.002)	-0.129 (0.002)	-0.145 (0.002)	-0.108 (0.002)
Population Fall: 1841-51	-0.028 (0.002)	-0.029 (0.002)	-0.040 (0.002)	-0.083 (0.002)	-0.015 (0.002)	-0.063 (0.002)	-0.061 (0.002)
Population Fall: 1851-71	-0.042 (0.002)	-0.048 (0.002)	-0.004 (0.002)	0.047 (0.002)	-0.097 (0.002)	0.042 (0.002)	0.044 (0.002)
Irish-Speaking: Household	0.498 (0.003)					0.302 (0.003)	
Irish-Speaking: Parent		0.478 (0.004)					0.273 (0.005)
Irish-Speaking: Townland			0.483 (0.002)			0.179 (0.003)	0.233 (0.004)
Irish-Speaking: DED				0.482 (0.002)		0.260 (0.003)	0.277 (0.004)
Irish-Speaking: County					0.199 (0.002)	0.001 (0.003)	-0.007 (0.003)
Lat-Lon Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Average Irish Speaking	0.494	0.477	0.495	0.495	0.495	0.494	0.477
McFadden's <i>R</i> -Squared	0.255	0.209	0.274	0.265	0.113	0.361	0.341
Observations	228,817	173,590	289,311	290,163	290,163	228,817	173,582

Note: Columns (1) to (7) present a series of individual-level probit regression marginal effects where an Irish-speaking dummy variable is the outcome of interest. District Electoral Division (DED) cluster-robust standard errors are in parentheses. Conventional statistical significance markers are omitted.

erational transmission by restricting the Irish-speaking environment, whatever the level of aggregation, to represent individuals aged 40 and above. This removes the potential for reverse causality, a common problem in the peer effects literature (Angrist, 2014).

Column (1) of Table 2 measures the intra-household intergenerational transmission of Irish ability. The baseline regressors indicate that those born in 1871–81 were 15 percentage points less likely to speak Irish relative to their 1861–71 counterparts. The Famine and post-Famine population fall is associated with small decreases in Irish speaking as a 2SD movement reduces the probability of Irish speaking by 0.03 and 0.04. Interestingly, the post-Famine population loss provides more explanatory power than the Famine itself. The estimated female marginal effect indicates no sex difference in Irish-speaking ability.

The household transmission marginal effect is 0.5. Given that roughly half of the sample were Irish speakers, this means Irish speaking was absent among residents in households where the previous generation did not speak Irish. But if the previous generation were Irish speaking, this boosts the likelihood of Irish speaking by 50 percentage points. If the next generation wholly adopted, without deviation, the previous generation's language ability, we would expect to find a marginal effect of one. This

suggests that while households are important, they are not the only influence on language acquisition.

Column (2) takes a focussed view on intergenerational transmission as we restrict the estimation sample and only include “sons” and “daughters”, while the previous generation’s Irish ability is exclusively sourced from the designated “household head”. This revision does little to alter the results obtained in column (1), although daughters were slightly less likely to speak Irish relative to sons (by 2.4pp).

In columns (3) to (5) we expand the geographic area through which intergenerational transmission occurs. These widening areas incorporate Irish speaking, for those aged 40 and older, at the townland, DED, and the broad county level. The marginal effects for both townlands (0.48) and DEDs (0.48) are broadly comparable to the household estimates. The effect attenuates substantially (0.20) for county-level transmission, indicating that DED is probably the optimal geographic resolution in this application. These results suggest intergenerational transmission occurs at both the community and household levels. However, we cannot disentangle and isolate the household’s influence because part of the estimated household effect is partly attributable to the indirect community’s effect on the household.

The model in column (6) addresses this shortcoming by including all intergenerational transmission variables. Both sets of marginal effects reveal that transmission occurs across the first three spatial units—households, townlands, and DEDs—but not counties. Households appear to exhibit the most influence with a marginal effect of 0.30, but the conditional importance of DEDs (0.26) is similar in scale—followed closely by Irish-speaking ability in the over-40s at the townland level (0.18). Column (7) shows that once the sample is restricted to just “sons” and “daughters” the results are similar. These results are consistent with a multiple geographies interpretation of language acquisition, whereby language is culturally transmitted across several dimensions, all of which matter, but none of which are sufficient on their own to explain language loss.

5.3 Education

Table 3 explores the role of education, both at the individual level and aggregate DED level over time. We again report probit regression marginal effects, where the explanatory variables are either dichotomous dummy variables or have been rescaled so that a one-unit change reflects a 2SD shift.

Column (1) examines the basic literacy-Irish link and suggests literacy is associated with a 14-percentage point drop in Irish-speaking ability. In column (2), we expand the potential influence of the education system to incorporate environmental differences in literacy. Community-wide variation offers more explanatory power, with historical differences in literacy that pre-date national schooling being the prime mover. A 2SD

Table 3: Education: Probit Regression Marginal Effects

	Irish Language Ability						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.022 (0.002)	-0.020 (0.002)	-0.024 (0.002)	-0.025 (0.002)	-0.019 (0.003)	-0.023 (0.004)	-0.020 (0.002)
Year of Birth: 1871-81	-0.116 (0.002)	-0.119 (0.002)	-0.107 (0.003)	-0.109 (0.002)	-0.109 (0.003)	-0.108 (0.004)	-0.127 (0.002)
Population Fall: 1841-51	-0.013 (0.002)	-0.042 (0.002)	-0.022 (0.002)	-0.043 (0.002)	-0.054 (0.004)	-0.033 (0.004)	-0.031 (0.002)
Population Fall: 1851-71	-0.090 (0.002)	-0.009 (0.002)	-0.071 (0.003)	0.004 (0.002)	-0.038 (0.005)	0.014 (0.005)	0.0002 (0.002)
Fully Literate	-0.135 (0.003)	-0.061 (0.003)	-0.073 (0.004)	-0.014 (0.003)	-0.059 (0.005)	-0.054 (0.006)	-0.064 (0.003)
Literacy in DED: Pre-National School		-0.489 (0.003)		-0.515 (0.003)	-0.501 (0.005)	-0.340 (0.007)	-0.482 (0.004)
Literacy in DED: National School Rise		-0.205 (0.003)		-0.228 (0.004)	-0.173 (0.006)	-0.147 (0.008)	-0.227 (0.004)
Literacy in DED: Post-Famine Rise		-0.080 (0.003)		-0.117 (0.004)	-0.079 (0.007)	-0.071 (0.008)	-0.090 (0.005)
Parent Fully Literate			-0.125 (0.002)	-0.051 (0.002)			
Sample	Full	Full	Sons and Daugh- ters	Sons and Daugh- ters	Teachers in DED: None	Irish- Speaking Teach- ers in DED: Minor- ity	Irish- Speaking Teach- ers in DED: Major- ity
Average Irish Speaking	0.495	0.495	0.477	0.477	0.507	0.286	0.554
McFadden's <i>R</i> -Squared	0.104	0.219	0.119	0.241	0.271	0.184	0.184
Observations	289,041	289,041	172,100	172,100	67,408	51,754	169,879

Note: Columns (1) to (7) present a series of individual-level probit regression marginal effects where an Irish-speaking dummy variable is the outcome of interest. District Electoral Division (DED) cluster-robust standard errors are in parentheses. Conventional statistical significance markers are omitted.

shift (14pp change relative to the variable's mean of 18pp) in pre-national school literacy reduces the probability of Irish speaking in our sample by 0.49. The impact of national schools is important, but the estimated marginal effect of -0.21 is less than half the pre-national school value. Literacy among contemporaries—the post-Famine rise—is the least important contextual effect, although its marginal effect is still more influential than individual literacy (-0.08 compared to -0.06).

The results in column (2) underline the importance of community-level intergenerational transmission, although they omit parent-to-child within-household transmission. Columns (3) and (4) address this by focussing on our subsample of sons and daughters. The third column shows that parental literacy reduces Irish-speaking ability by 13 percentage points, whereas the corresponding reduction for individual literacy is almost half, standing at 7 percentage points. Even within households, education's influence operates on an intergenerational basis. In the fourth column, we reiterate the importance of geographic context as the parental literacy marginal effect diminishes after the three DED literacy variables are included. All three DED literacy measures return similar marginal effects to those presented in the second column. The largest marginal effect is

associated with pre-national school literacy.

National schools are a prime suspect in the Irish language’s decline. If the education system was the vehicle through which the state coercively discouraged bilingualism, we would expect to find that the literacy boost associated with the state’s establishment of this system has the largest marginal effect. However, variation in literacy that pre-dates this establishment is, by far, the principal variable in this regard. While the national schools’ boost to literacy is evident, the magnitude of this marginal effect is secondary. The importance of decentralised voluntary (“hedge”) schooling compared with the emerging state-mandated alternative suggests that (grand-)parents were willing and satisfied to send their children to receive an English-medium education rather than compelled by the state and national education policy.

Columns (5) to (7) explore this theme in detail by splitting the sample according to teachers in the DED and their Irish-speaking ability. If the suppression of the Irish language was through discipline and punishment by teachers, the literacy-related marginal effects should plausibly be greater in districts with none or few Irish-speaking teachers. Our reasoning here is that districts where half or more of the teachers were themselves Irish speakers would have been less likely to punish students for speaking Irish. However, our results are inconsistent with this hypothesis as, if anything, all literacy-related marginal effects are stronger in districts where half or more of the teaching population had Irish speaking ability. This is revealed by comparing columns (6) and (7), as the results in column (5) are difficult to interpret because there is ambiguity surrounding observations in DEDs where no teachers reside. These findings once again downplay education’s role as a coercive force rather than a voluntary one. Language dynamics were driven by endogenous demand, not exogenous policy intervention.

5.4 Church and State Capacity

Our results thus far point to the importance of education operating on an intergenerational basis. Next, we look at the influence of the Catholic church and the state more broadly. The probit marginal effect results are in Table 4. Column (1) shows that Catholicism is a strong predictor of Irish language ability. The estimated marginal effect of 0.47 relative to the Irish-speaking sample mean of 0.5 effectively suggests that the few (around 4 per cent) non-Catholics in our sample were not Irish speakers. This result is unsurprising, but given the complex nature of religious faith and national identity/ethnicity, it would be absurd to argue that these results demonstrate that the church promoted Irish speaking during this period.

Column (2) attempts to address this issue using first names as a marker for Catholicism on the intensive margin. Following Connor (2021), we argue that a person’s first name conveys information about their religious background and upbringing. Those

Table 4: Church and State Capacity: Probit Regression Marginal Effects

	Irish Language Ability				
	(1)	(2)	(3)	(4)	(5)
Female	-0.021 (0.002)	-0.017 (0.002)	-0.016 (0.002)	-0.016 (0.002)	-0.017 (0.002)
Year of Birth: 1871-81	-0.135 (0.002)	-0.134 (0.002)	-0.134 (0.002)	-0.134 (0.002)	-0.139 (0.002)
Population Fall: 1841-51	-0.014 (0.002)	-0.014 (0.002)	-0.021 (0.002)	-0.035 (0.002)	-0.015 (0.002)
Population Fall: 1851-71	-0.081 (0.002)	-0.080 (0.002)	-0.075 (0.002)	-0.065 (0.002)	-0.084 (0.002)
Roman Catholic	0.471 (0.002)	0.458 (0.003)	0.458 (0.003)	0.457 (0.003)	
Roman Catholic Name Index		0.057 (0.002)	0.056 (0.002)	0.055 (0.002)	0.056 (0.002)
Traditional Name Index		0.033 (0.002)	0.032 (0.002)	0.032 (0.002)	0.033 (0.002)
Clergy per Capita			-0.031 (0.002)	-0.015 (0.002)	
Police per Capita				-0.045 (0.002)	
Teachers per Capita				-0.027 (0.002)	
Average Irish Speaking	0.495	0.496	0.496	0.496	0.514
McFadden's <i>R</i> -Squared	0.134	0.135	0.135	0.137	0.109
Observations	289,945	275,479	275,479	275,479	264,733
Catholic Only Subsample	No	No	No	No	Yes

Note: Columns (1) to (5) present a series of individual-level probit regression marginal effects where an Irish-speaking dummy variable is the outcome of interest. District Electoral Division (DED) cluster-robust standard errors are in parentheses. Conventional statistical significance markers are omitted.

with more distinctly Catholic first names were more likely to adhere to the church's philosophical position on various issues. Connor showed that this measure of Catholicism explains fertility behaviour. We also construct a traditional name index that captures the persistence of popular names that might persevere for non-religious reasons.

Interestingly, the marginal effect estimate on the Roman Catholic name index was 0.06, meaning that a 2SD increase in this measure increased Irish-speaking propensity by 6 percentage points. While the Catholic church's attitude in nineteenth-century Ireland was, at best, ambivalent towards the preservation of Irish speaking, the evidence here suggests that this attitude did not transmit to the church's followers. Traditional names, which can indicate more conservative values, also explain the upkeep of Irish

speaking (marginal effect of 0.03). The relatively small magnitude of both name marginal effects cautions against overemphasis in our interpretation here—although given that both variables only serve as rough proxies of religious adherence and traditional values, this implies a degree of measurement error and, arguably, this could downplay their statistical influence. Column (5) reruns the first column’s specification on a Catholic-only subsample. The estimated marginal effects for both name index variables are the same, demonstrating that the small number of non-Catholics does not influence our results.

5.5 Economics and Sociodemographics

The simultaneous decline of the Irish language and demographic upheaval of the mid-to-late nineteenth century is suggestive. In Table 5, we investigate these changes formally. In the first column of Table 5, we add variables that measure the share of Irish speakers in each individual’s occupation and a dummy variable that indicates whether the observation resides in their county of birth. The positive sign on the occupation’s Irish-speaking level is as expected, but the size of this marginal effect (0.11) is small relative to comparable factors. If socioeconomic advancement was a reason to abandon bilingualism, we would expect to find a close alignment between occupational class and Irish speaking.

Figure 2 displays a scatterplot that connects Irish speaking and occupational literacy for the 119 most common occupations in the complete 1901 census returns.²⁶ Assuming that more prestigious occupations have higher levels of literacy, we find only a weak correlation connecting Irish-speaking and literacy (correlation coefficient of -0.3) with a large degree of dispersion evident. Some occupations with public-facing tasks—such as priests, teachers, and shopkeepers—had high shares of both literacy and Irish speaking. The high rate of bilingualism among priests aligns with Wolf’s (2014) assertion that the Catholic Church facilitated bilingualism. Occupations with comparable responsibilities and levels of literacy—wool and linen weaving, for instance—had diverging levels of Irish language abilities reflecting geographic and cultural forces rather than socioeconomic ones.

The relationship between migration and Irish speaking is stronger. Those who remained in their county of birth were 23 percentage points more likely to be Irish speakers. Despite only measuring internal migration, this finding is in keeping with the hypothesis that Irish speaking declined, in part, due to migration’s appeal as parents abandoned the native language to prepare their children for life overseas, most likely elsewhere in the Anglosphere. It also highlights our concerns about sample selection. If non-Irish speakers were more likely to leave, this means our sample overestimates the number of Irish speakers who originate from these communities. We address this concern more

²⁶This dataset was created by aggregating all occupations and classifying all those with 300 or more responses. Non occupations that appeared in this field, such as “scholar” or “retired”, were discarded.

Table 5: Economics and Sociodemographics: Probit Regression Marginal Effects

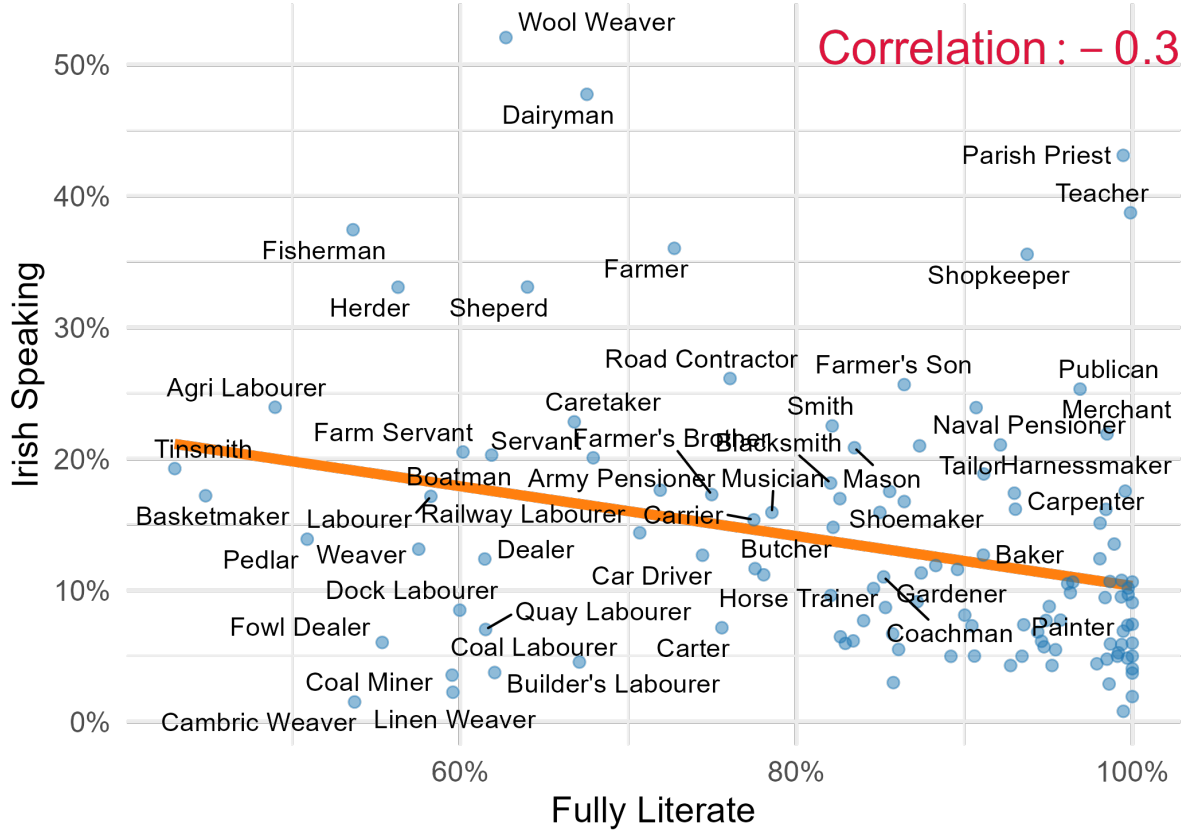
	Irish Language Ability		
	(1)	(2)	(3)
Female	-0.002 (0.002)	-0.003 (0.002)	-0.002 (0.002)
Year of Birth: 1871-81	-0.122 (0.002)	-0.123 (0.002)	-0.124 (0.002)
Population Fall: 1841-51	-0.039 (0.002)	-0.081 (0.002)	-0.075 (0.002)
Population Fall: 1851-71	-0.084 (0.002)	-0.056 (0.002)	-0.036 (0.002)
Occupation: Irish Speaking	0.106 (0.002)	0.095 (0.002)	0.088 (0.002)
Non-Migrant: Resides in Birth County	0.223 (0.004)	0.192 (0.004)	0.192 (0.004)
Non-Migrants in DED		0.043 (0.003)	0.043 (0.003)
Population Density: 1841		0.050 (0.004)	-0.001 (0.004)
Pre-Famine Land Value per Acre		-0.158 (0.005)	-0.057 (0.004)
Market Access: 1841			-0.172 (0.005)
Port Access: 1841			-0.208 (0.009)
Market Access Rise: 1841-71			0.027 (0.002)
Port Access Rise: 1841-71			-0.101 (0.004)
Average Irish Speaking	0.495	0.496	0.496
McFadden's <i>R</i> -Squared	0.134	0.135	0.135
Observations	281,612	281,612	281,612

Note: Columns (1) to (3) present a series of individual-level probit regression marginal effects where an Irish-speaking dummy variable is the outcome of interest. District Electoral Division (DED) cluster-robust standard errors are in parentheses. Conventional statistical significance markers are omitted.

formally in Appendix B and C.

Column (2) of Table 5 introduces several DED-level contextual variables. The share of residents born outside their county of residence in the DED shows a weak positive relationship with Irish speaking. That migration is associated with fewer Irish speakers matches the individual-level result, albeit with substantially less explanatory power. Irish was less likely to fall in areas with a larger pre-Famine population density and more

Figure 2: Literacy-Irish Occupational Connection



Note: Depicted are the 119 most common occupations in the complete 1901 census.

likely to fall in areas with more valuable land. But the strength of these associations is diminished in column (3) as additional market access covariates are included. These market access variables measure the interplay between the transport network and markets. Districts with greater market access found it easier (both in terms of time and money) to travel domestically, whereas port access captures external market forces. We measure the impact of railways via the market and port access changes occurring between 1841 and 1871, following Fernihough and Lyons (2022). We also account for differences in Ireland's pre-rail network by including both access measures calculated based on the 1841 system of roads and navigable waterways.

Market access, a way of capturing geographic remoteness consistent with new economic geography, contributed to the decline in Irish speaking. However, this manifested along the pre-existing transport network as the pre-Famine domestic and foreign access variables yield marked marginal effects of -0.17 and -0.21 . The influence of railway-inspired improvement is less impressive. A 2SD increase in port access growth between 1841 and 1871 is associated with a 10 percentage point fall in Irish speaking, although market access change during this period was associated with a small increase (0.03 marginal effect) in language ability. This finding once again suggests language change is

Table 6: Variable Category Importance Decomposition

	Explanatory Power (per cent)	
	(1)	(2)
Baseline	1.931	4.686
Geographic Coordinates	52.122	
Education	36.035	59.429
Church and State Capacity	5.673	16.079
Economics and Sociodemographic	4.239	19.806

Note: The sum of the squared standardised probit coefficients represents the total explained variation.

more a product of long-run generational elements rather than short-term trends.

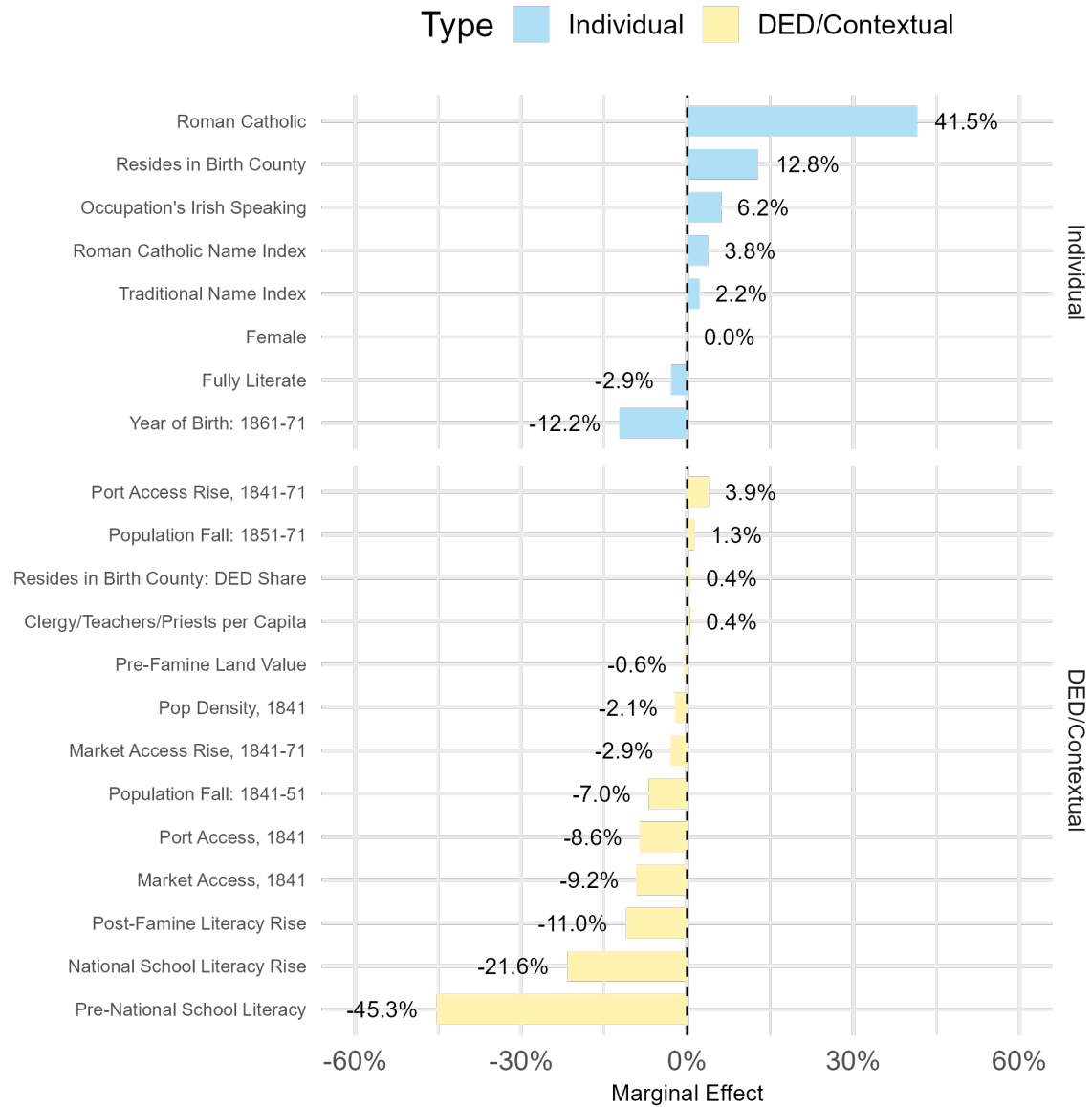
This model also includes latitude and longitude coordinates, which enter as a third-order polynomial function with interactions. Therefore, the marginal effects reported here capture more than just the spatial retreat of the Irish-speaking frontier, although the estimated marginal effects strengthen slightly if the aforementioned latitude-longitude spatial controls are omitted.

5.6 Including All Explanatory Factors

The preceding subsections provided a thematic analysis of the Irish language’s decline. In this subsection, we look at a fully specified model that unifies these themes. Our results, displayed in Figure 3, show the marginal effects stratified by whether the variable is recorded on an individual on a contextual (DED) level. All explanatory variables are again scaled to facilitate comparisons and the bars ordered by the variable’s importance.

Most of the marginal effects shown in Figure 3 are similar to their counterparts shown in Tables 2 to 5. Aside from the expected Roman Catholic effect (since we know the small number of other religions were almost entirely non-Irish speaking), the collection of literacy variables provide the most explanatory power, with variation in pre-national school literacy being the most prominent. A 2SD increase in pre-national school literacy is associated with a 0.45 decrease in the probability that residents several decades later are Irish speaking. The increase attributable to the establishment of national schools is smaller, approximately half (-0.22), but still of great importance. The post-Famine increase in literacy, among contemporaries, still holds relevance and its estimated marginal effect of -0.11 is followed closely in importance by the market and port access measures. We find support for all our main results in the separate thematic analyses. None of the estimated marginal effects dramatically change despite the more saturated specification. Thus, none of our previous conclusions requires revisions based on Figure 3.

Figure 3: Fully Specified Probit Model Marginal Effects



Note: Depicted bars are marginal effects from a 2SD increase.

Table 6 decomposes the explanatory power of each set of regressors by theme by taking the squared standardised probit coefficients and assuming that this sum represents the total explained variation. Thus, columns (1) and (2) both add to 100. The nonlinear function of latitude and longitude coordinates accounts for over half of all the model's explanatory power in column (1). One interpretation is it highlights the relevance of spatial diffusion and contagion effects. Physical location, conditional on our full set of regressors, appears to play a role as the Irish-speaking frontier continued its retreat. An alternative view is the explanatory power offered by geographic coordinates is ambiguous. Different permutations of latitude and longitude coordinates do not, by themselves, affect Irish speaking and may mask the true influence of other factors. For example, if pre-

national school literacy was a spatially diffuse process, then part of this variable’s influence may be channelled through the latitude and longitude coordinates.

Column (2) details the decomposition performed on a probit regression without these spatial effects. Omitting the coordinates increases the relevance of the other hypothesis domains. Baseline characteristics—year of birth cohort and sex—account for less than 5 per cent of explained variation. In line with the results in Figure 3, we find that variables relating to education offer substantial explanatory power and contribute to at least 36 per cent of the explained variation, and possibly as high as 60 per cent. This underscores the importance of the literacy and education channel. The contribution of the other variable sets is secondary, each contributing only around 5 per cent of the determining input—although the reported values are larger, close to 20 per cent each. The discrepancy suggests an extensive overlap with geographic position.²⁷

6 Discussion and Conclusion

Our analysis traces the contours of Irish language loss across the nineteenth century. Ireland’s language shift is striking in its apparent speed and permanency. Bilingualism proved to be a transitional stage towards a new English-language monolingualism. Subsequent policy interventions designed to revive the language have barely halted further loss; areas that the Irish government today recognises as places where the Irish language is a predominant vernacular represent just a small subset of the bilingual districts we analyse in our study. A major report into the modern use of Irish in these regions by Ó Giollagáin and Charlton (2015) warns that the language is on the verge of becoming dysfunctional even there; Irish has a social function only among their older generation, while younger speakers only have a passive or weak knowledge of the language.

Bringing back a language that was so emphatically lost was always going to be a tall order. But part of the problem, we argue, is a failure by Ireland’s policymakers to understand why the island’s indigenous language fell out of favour in the first place. We contend that the decline of Irish was not due to direct exogenous forces. Rather, we find that the language frontier moved further and further west because of socioeconomic forces present on the island. The decline was the outcome of individuals and families making calculated trade-offs between the costs and benefits of language acquisition and retention. Bourdieu, not Weber, explains the Irish case; the shrinking size of Ireland’s bilingual community was endogenous to social, cultural and economic forces rather than the consequence of state intervention.

²⁷This is unsurprising. For example, we would expect variables that measure market access to strongly correlate with both latitude and longitude coordinates. While it is infeasible to adequately disentangle these effects, this does not alter the finding that variation in educational outcomes is the primary source compared to the latter two domains.

In particular, our econometric analysis highlights the role of intergenerational transmission through educational institutions. This was not an exogenous shift imposed through state-funded, centrally planned education policy. Rather, it was the pre-existing education institutions that influenced the Irish language border. Parents enrolled their children in these hedge schools to acquire a second language—English—that would give them access to economic opportunity. These institutions’ decentralised and voluntary nature empirically supports the hypothesis that language boundaries during the modern period—an era marked by the emergence of the homogenised nation-state—were at least partially shaped by endogenous forces.

Another finding worth highlighting here is the relevant spatial resolution for analysing language transmission. Our individual-level census data allow us to discern that vertical transmission within households was important, but it was not the only influence on language acquisition. Horizontal transmission between households within a geographic locale is probably of greater, but certainly of equal, importance for language transmission and retention. Finally, we also highlight the importance of time in the analysis of language change. That the effects of pre-national schooling could manifest themselves in a census conducted over 50 years later underlines that the economic analysis of language necessitates a long-run, historical, perspective.

Our analysis entails several limitations. Methodologically, we offer a unified quantitative approach that traces the impact of contributory sources on Irish-speaking ability. This approach does not, however, fully delineate these explanatory variables. They are not mutually exclusive, and we cannot provide a framework that neatly compartmentalises these various elements. Thus, while our results offer clues about the drivers of language loss, we must speculate on why the decline was so severe and—as argued by Slomanson (2012)—historically unprecedented from a comparative perspective.

The diglossia framework is typically used to describe a stable bilingual society where a language with a high social status can co-exist with another with a lower, more informal status. Our analysis, in contrast, focuses exactly on the dynamics: an unstable society with a marked change in language use over time. Our theoretical model framed the shift in Ireland’s language frontier as being a consequence of parental choice, where costs and benefits of both languages are weighed up, and a decision to pass on Irish to the next generation is made. Our evidence points to the modal decision becoming not to pass on the low-status language to the next generation. An alternative interpretation is that Irish lost its status even as the island’s low-status language and that English supplanted Irish for all the social, cultural and economic uses historically associated with the informal indigenous language within its previously bilingual society.

Europe today has several prominent language frontiers that do not correspond with national boundaries. Additionally, various actively spoken minority languages are used alongside national languages within several regions of Europe’s constituent polities. A

number of these minority languages even straddle political borders. We speculate the Irish case differs from these others because of the nature of the language with which it competed. The communicative value of English, a global language already by the nineteenth century when Irish saw its biggest decline in use, was enormous. Irish simply could not compete; it did not offer its users the full range of functions offered by English.

Where Finnish, say, competed in Finland's bilingual society with a Swedish language originating from a relatively small neighbouring economy (Sweden), Irish competed with the language used not only by its industrial superpower neighbour (Great Britain) but also in the principal countries that hosted its migrant diaspora (United States, Canada and Australia). The comparative fate of the Welsh language is apt. Unlike Ireland, nineteenth-century Wales industrialised, offering rural Welsh speakers economic opportunities within their home nation (Thomas, 1959; Baines, 1986). Conversely, Ireland's lack of industrialisation and economic opportunities meant the Irish had to migrate to where the industrial jobs were located—either elsewhere in the United Kingdom, or overseas in one of its English-speaking offshoots (Ó Gráda, 1995).

What makes the fate of the Irish language unique is the interaction of economic conditions, mass migration, and the rise of the Anglosphere as the dominant cultural and political force of the nineteenth century. Surprisingly, education, not demographic decline, holds the most explanatory leverage in our econometrics. This is not paradoxical, as our models rely on statistical variation at the intensive, not extensive, margin. Thus, the uniform macro-demographic landscape does not translate into population change at the micro level. Here, education is crucial, with communities self-selecting into both education and language loss. In terms of linguistic diglossia, English, the high-status language of law and commerce, now supplanted Irish to additionally function as the low-status language. Schools acted as incubators where English was not only the medium of instruction but also that of daily interaction in the wider community.

Our interpretation is necessarily speculative. Unfortunately, the *relative* communicative value of the Irish language's competing vernacular is not something that can be easily captured in a study of just two languages. An empirical approach involving multiple countries and diverse linguistic experiences, or the application of our methodology to other contexts, could be interesting and a fruitful area for future research.

References

- Akenson, D. H. (1970). *The Irish Education Experiment: The National System of Education in the Nineteenth Century*. London: Routledge & K. Paul.
- American Statistical Association (2016). Statement on statistical significance and p-values. *The American Statistician* 70(2), 131–133.
- Angrist, J. D. (2014). The perils of peer effects. *Labour Economics* 30, 98–108.
- Angrist, J. D. and V. Lavy (1997). The effect of a change in language of instruction on the returns to schooling in Morocco. *Journal of Labour Economics* 15(1), S48–S76.
- Aspachs-Bracons, O., J. Costa-Font, I. Clots-Figueras, and P. Masella (2008). Compulsory language educational policies and identity formation. *Journal of the European Economic Association* 6(2/3), 434–444.
- Baines, D. (1986). *Migration in a mature economy: Emigration and internal migration in England and Wales 1861–1900*. Cambridge: Cambridge University Press.
- Barr, C. (2009). MacHale, John. In *Dictionary of Irish Biography*. Dublin: Royal Irish Academy. URL: <https://doi.org/10.3318/dib.005220.v1>.
- Blanc, G. and M. Kubo (2024). French. Mimeo, University of Manchester. URL: https://www.guillaumeblanc.com/files/theme/BlancKubo_French.pdf.
- Blum, M., C. L. Colvin, L. McAtackney, and E. McLaughlin (2017). Women of an uncertain age: Quantifying human capital accumulation in rural Ireland in the nineteenth century. *The Economic History Review* 70(1), 187–223.
- Borooah, V. K., D. A. Dineen, and N. Lynch (2009). Language and occupational status: Linguistic elitism in the Irish labour market. *The Economic and Social Review* 40(4), 435–460.
- Bourdieu, P. (1977). The economics of linguistic exchanges. *Social Science Information* 16(6), 645–668.
- British Parliamentary Papers (1825). *The Select Committee of the House of Lords appointed to inquire into the state of Ireland, more particularly with reference to the circumstances which may have led to disturbances in that part of the United Kingdom*, Volume (521). London: HMSO: House of Lords.
- British Parliamentary Papers (1843). *Report of the Commissioners Appointed to Take the Census of Ireland for the Year 1841*. London: HMSO.
- British Parliamentary Papers (1851). *The Seventeenth Report of the Commissioners of National Education in Ireland (For the Year 1850.) Vol. I*, Volume 2304. London: HMSO: House of Commons. [1405] [1405-II], H.C., pp 101, 641.
- British Parliamentary Papers (1855). *The Census of Ireland for the Year 1851: Part IV: Report on Ages and Education*. London: HMSO: House of Commons.
- British Parliamentary Papers (1900). *The Sixty-Sixth Report of the Commissioners of National Education in Ireland, Year 1899–1900*. London: HMSO. [Cd.285] [Cd.286] [Cd.287] [Cd. 288] [Cd.289], H.C., pp xxiii, 189, 251, 455, 585, 733.
- Central Statistics Office (2023a). Census 2022 Profile 8 – The Irish Language and Education. CSO Statistical Publication, 19 December 2023, Dublin. URL: <https://www.cso.ie/en/releasesandpublications/ep/p-cpp8/census2022profile8-theirishlanguageandeducation/irishlanguageandthegaeltacht/>.

- Central Statistics Office (2023b). Census of Population 2022 – Summary Results. CSO Statistical Publication, 30 May 2023, Dublin. URL: <https://www.cso.ie/en/releasesandpublications/ep/p-cpsr/censusofpopulation2022-summaryresults/migrationanddiversity/>.
- Chakraborty, T. and S. K. Bakshi (2016). English language premium: Evidence from a policy experiment in India. *Economics of Education Review* 50, 1–16.
- Clingingsmith, D. (2015). Are the world’s languages consolidating? The dynamics and distribution of language population. *Economic Journal* 127(599), 143–176.
- Colvin, C. L., S. Henderson, and E. McLaughlin (2024). Age structure and age heaping: Solving Ireland’s post-Famine digit preference puzzle. *European Review of Economic History* (in press).
- Comerford, R. V. (2003). *Ireland: Inventing the Nation*. London: Bloomsbury.
- Connor, D. S. (2021). In the name of the father? Fertility, religion, and child naming in the demographic transition. *Demography* 58(5), 1793–1815.
- Coolahan, J. (1981). *Irish Education: Its History and Structure*. Dublin: Institute of Public Administration.
- Coolahan, J. (2001). Education as cultural imperialism: The denial of the Irish language to Irish speakers, 1831–1922. *Paedagogica Historica* 37(1), 17–33.
- Corcoran, T. (1925). The Irish language in the Irish schools. *Studies: An Irish Quarterly Review* 14(55), 377–388.
- Corrigan, K. P. (2003). The ideology of nationalism and its impact on accounts of language shift in nineteenth century Ireland. *AAA: Arbeiten aus Anglistik und Amerikanistik* 28(2), 201–229.
- Cullen, L. M. (1990). Patrons, teachers and literacy in Irish: 1700–1850. In M. Daly and D. Dickson (Eds.), *The Origins of Popular Literacy in Ireland: Language Change and Educational Development 1700–1920*, pp. 15–44. Dublin.
- de Bromhead, A., A. Fernihough, and C. Ó Gráda (2021). 175 years of mismeasuring Ireland?: Some traps set by historical data. *Journal of the Statistical Social Inquiry Society of Ireland* LI, 128–155.
- de Fréine, S. (1965/1960). *The Great Silence*. Mercier Press. [1965 translation of the original 1960 Irish edition].
- de Swaan, A. (2001). *Words of the World: The Global Language System*. London: Polity Press.
- Doyle, A. (2015). *A History of the Irish Language: From the Norman Invasion to Independence*. Oxford: Oxford University Press.
- Fabb, N. (2016). Linguistic theory, linguistic diversity and whorfian economics. In V. Ginsburgh and S. Weber (Eds.), *The Palgrave Handbook of Economics and Language*, pp. 17–60. London: Palgrave Macmillan.
- Ferguson, C. A. (1959). Diglossia. *Word* 15(2), 325–340.
- Fernihough, A. and R. C. Lyons (2022). Agglomeration and emigration: The economic impact of railways in post-Famine Ireland. QUCEH Working Paper Series, Paper No. 22-11. URL: <https://www.quceh.org.uk/uploads/1/0/5/5/10558478/wp22-11.pdf>.
- Fernihough, A. and C. Ó Gráda (2022). Population and poverty in Ireland on the eve of the Great Famine. *Demography* 59(5), 1607–1630.
- FitzGerald, G. (1984). Estimates for baronies of minimum level of Irish-speaking amongst successive

- decennial cohorts: 1771–1781 to 1861–1871. *Proceedings of the Royal Irish Academy. Section C: Archaeology, Celtic Studies, History, Linguistics, Literature* 84C, 117–155.
- FitzGerald, G. (2003). Irish-speaking in the pre-Famine period: A study based on the 1911 census data for people born before 1851 and still alive in 1911. *Proceedings of the Royal Irish Academy: Archaeology, Culture, History, Literature* 103(5), 191–283.
- FitzGerald, G. (2013). Irish primary education in 1824. In G. FitzGerald (Ed.), *Irish Primary Education in the Early Nineteenth Century: An Analysis of the First and Second Reports of the Commissioners of Irish Education Inquiry, 1825–6*. Dublin: Royal Irish Academy.
- Fitzpatrick, D. (1986). ‘A share of the honeycomb’: Education, emigration and Irishwomen. *Continuity and Change* 1(2), 217–234.
- Ginsburgh, V. and S. Weber (2020). The economics of language. *Journal of Economic Literature* 58(2), 348–404.
- Grenfell, M. J. (2011). Bordieu, language and linguistics. In M. J. Grenfell (Ed.), *Bordieu, language and linguistics*, pp. 35–63. London: Continuum.
- Grin, F. (1992). Towards a threshold theory of minority language survival. *Kyklos* 45(1), 69–97.
- Hatton, T. J. and J. G. Williamson (1994). International migration 1850–1939: An economic survey. In T. J. Hatton and J. G. Williamson (Eds.), *Migration and the International Labour Market 1850–1939*. London: Taylor & Francis.
- Hindley, R. (1990). *The Death of the Irish Language: A Qualified Obituary*. London: Routledge.
- Hynes, W. (2014). To what extent were economic factors important in the separation of the south of Ireland from the United Kingdom and what was the economic impact? *Cambridge Journal of Economics* 38(2), 369–397.
- Jackson, A. (1999). *Ireland 1798–1998: Politics and War*. Oxford: Blackwell Publishing.
- John, A. (2016). Dynamic models of language evolution: The economic perspective. In V. Ginsburgh and S. Weber (Eds.), *The Palgrave Handbook of Economics and Language*, pp. 101–120. London: Palgrave Macmillan.
- John, A. and O. Özgür (2021). Linguistic diversity in the very long run. *The Economic Journal* 131(635), 1186–1214.
- Kandler, A. and J. Steele (2017). Modeling language shifts. *PNAS* 114(19), 4821–4853.
- Kandler, A., R. Unger, and J. Steele (2010). Language shift, bilingualism and the future of Britain’s Celtic languages. *Philosophical Transactions of the Royal Society B* 365(1559), 3855–3864.
- Kelly, M. and A. S. Fotheringham (2011). The online atlas of Irish population change 1841–2002: A new resource for analysing national trends and local variations in Irish population dynamics. *Irish Geography* 44(2-3), 215–244.
- Kessler, B. (1995). Computational dialectology in Irish Gaelic. In S. P. Abney and E. W. Hinrichs (Eds.), *Seventh Conference of the European Chapter of the Association for Computational Linguistics*, Dublin, Ireland. Association for Computational Linguistics. URL: <https://aclanthology.org/E95-1009>.
- Lee, J. J. (1989). *Ireland 1912–1985: Politics and Society*. Cambridge: Cambridge University Press.
- Mac Giolla Chríost, D. (2005). *The Irish Language in Ireland: From Goídel to Globalisation*. Abingdon: Routledge.

- Marschak, J. (1965). Economics of language. *Behavioral Science* 10(2), 135–140.
- McLaughlin, E., C. L. Colvin, and M. Blum (2021, 12). Anthropometric history: Revisiting what’s in it for Ireland. *Irish Economic and Social History* 48(1), 3–26.
- McMahon, A. and R. McMahon (2005). *Language Classification by Numbers*. Oxford: Oxford University Press.
- McMahon, T. M. (2008). *Grand Opportunity: The Gaelic Revival and Irish Society, 1893-1910*. Syracuse, New York: Syracuse University Press.
- McManus, A. (2002). *The Irish Hedge School and its Books, 1695–1831*. Dublin: Four Courts.
- Moseley, C. (Ed.) (2010). *Atlas of the World’s Languages in Danger* (3rd ed.). Paris: United Nations Educational, Scientific and Cultural Organization. URL: <https://unesdoc.unesco.org/ark:/48223/pf0000187026>.
- Mufwene, S. (2017). Language vitality: The weak theoretical underpinnings of what can be an exciting research area. *Language* 93(4), e202–e223.
- Murphy, G. (1948). The Gaelic background. In M. Tierney (Ed.), *Daniel O’Connell: Nine Centenary Essays*. Dublin: Browne and Nolan Limited.
- National Archives of Ireland (2019). Census of Ireland 1901/1911 and census fragments and substitutes, 1821–51. [Data set]. Census years: 1901. URL: <http://census.nationalarchives.ie/pages/1901/>.
- Nic Craith, M. (2002). ‘The tiger no longer speaks Celtic’: Economic conditioning and the Irish language. In U. Kockel (Ed.), *Culture and Economy: Contemporary Perspectives*, pp. 175–95. Aldershot: Ashgate.
- Northern Ireland Statistics and Research Agency (2022). Main Statistics for Northern Ireland: Language. Statistical Bulletin, 22 September 2022, Belfast. URL: <https://www.nisra.gov.uk/system/files/statistics/census-2021-main-statistics-for-northern-ireland-phase-1-statistical-bulletin-language.pdf>.
- Northern Ireland Statistics and Research Agency (2023). Census 2021 Main Statistics Language Tables. 30 November 2023, Belfast. URL: <https://www.nisra.gov.uk/publications/census-2021-main-statistics-language-tables>.
- Ó Ciosáin, N. (1997). *Print and Popular Culture in Ireland, 1750–1850*. Basingstoke: Macmillan.
- Ó Ciosáin, N. (2024). *Print and the Celtic Languages: Publishing and Reading in Irish, Welsh, Gaelic and Breton, 1700–1900*. Abingdon: Routledge.
- Ó Cuív, B. (1969). Irish in the modern world. In Ó Cuív, B. (Ed.), *A View on the Irish Language*. Dublin: The Stationary Office.
- Ó Giollagáin, C. and M. Charlton (2015). Nuashonrú ar an Staidéar Cuimsitheach Teangeolaíoch ar Úsáid na Gaeilge sa Ghaeltacht: 2006—2011. Technical report. URL: <https://www.rte.ie/documents/news/udaras-na-gaeltachta.pdf>.
- Ó Gráda, C. (2013). School attendance and literacy in Ireland before the Great Famine: A simple baronial analysis. In G. FitzGerald (Ed.), *Irish Primary Education in the Early Nineteenth Century: An Analysis of the First and Second Reports of the Commissioners of Irish Education Inquiry, 1825–6*. Dublin: Royal Irish Academy.
- Ó Gráda, C. (1995). *Ireland: A New Economic History 1780–1939*. Oxford: Oxford University Press.
- Ó Murchadha, C. (2011). *The Great Famine: Ireland’s Agony 1845–1852*. London: Bloomsbury.

- Ó Tuathaigh, G. (1974). Gaelic Ireland, popular politics and Daniel O’Connell. *Journal of the Galway Archaeological and Historical Society* 34, 21–34.
- O’Brien, G. (1989). The strange death of the Irish language, 1780–1800. In G. O’Brien (Ed.), *Parliament, Politics and People: Essays in Eighteenth-Century Irish History*, pp. 149–170. Dublin: Irish Academic Press.
- Ravenstein, E. G. (1879). On the Celtic languages in the British Isles: A statistical survey. *Journal of the Statistical Society of London* 42(3), 579–636.
- Slomanson, P. (2012). On the great silence: A gap in Irish historiography and consequences for language education in Ireland. *Nordic Irish Studies* 11(2), 95–114.
- Smith, A. D. M. (2016). Dynamic models of language evolution: The linguistic perspective. In V. Ginsburgh and S. Weber (Eds.), *The Palgrave Handbook of Economics and Language*, pp. 61–100. London: Palgrave Macmillan.
- Tada, M. (2018). Language, ethnicity, and the nation-state: On Max Weber’s conception of imagined linguistic community. *Theory and Society* 47, 437–466.
- Thomas, B. (1959). Wales and the Atlantic economy. *Scottish Journal of Political Economy* 6(3), 169–192.
- UNESCO (2003). Language vitality and endangerment. International Expert Meeting on UNESCO Programme Safeguarding of Endangered Languages, Paris, 10–12 March 2003. URL: <https://ich.unesco.org/doc/src/00120-EN.pdf>.
- Wall, M. (1969). The decline of the Irish language. In Ó Cuív, B. (Ed.), *A View of the Irish Language*. Dublin: The Stationary Office.
- Watson, I. and M. Nic Ghiolla Phádraig (2011). Linguistic elitism: The advantage of speaking Irish rather than the Irish-speaker advantage. *The Economic and Social Review* 42(4), 437–454.
- Weber, M. (1968/1921). *Economy and Society: An Outline of Interpretive Sociology*. Berkeley: University of California Press. [1968 translation of original 1921 German edition].
- Wickström, B.-A. (2005). Can bilingualism be dynamically stable? A simple model of language choice. *Rationality and Society* 17(1), 81–115.
- Williams, J. E. C. and P. K. Ford (1992). *The Irish Literary Tradition*. Cardiff: University of Wales Press.
- Wolf, N. M. (2014). *An Irish-Speaking Island: State, Religion, Community, and Linguistic Landscape in Ireland, 1770-1870*. Madison, Wisconsin: University of Wisconsin Press.

Appendices

A Data Sources and Data Generation Procedures

Table A1: Data Sources and Generation Procedures

Statistic	Description	Source
Irish Language Ability	A binary indicator assigned a value of one when the observation is either a bilingual or monolingual Irish speaker.	National Archives of Ireland (2019).
Female	A binary indicator assigned a value of one when the observation is female.	Same as above.
Year of Birth: 1871-81	A binary indicator assigned a value of one if the observation's birth year falls between 1871 and 1881, corresponding to a census age of 20-29. The omitted group in our analysis are those born between 1861-71 or aged 30-39.	Same as above.
Population Fall: 1841-51	Represents the population decline associated with the Famine, with aggregated values varying at the district electoral division level.	Kelly and Fotheringham (2011).
Latitude	Latitude coordinate from the district electoral division's polygon center point.	Same as above.
Longitude	Longitude coordinate from the district electoral division's polygon center point.	Same as above.
Previous Generation Irish-Speaking: Household	The share of relatives (mostly parents) in the observation's household who are 40 years old and above who are Irish speaking (bilingual and monolingual).	National Archives of Ireland (2019).
Previous Generation Irish-Speaking: Townland	The share of residents in the observation's townland who are 40 years old and above who are Irish speaking (bilingual and monolingual).	Same as above.
Previous Generation Irish-Speaking: DED	The share of residents in the observation's district electoral division who are 40 years old and above who are Irish speaking (bilingual and monolingual).	Same as above.
Previous Generation Irish-Speaking: County	The share of residents in the observation's county who are 40 years old and above who are Irish speaking (bilingual and monolingual).	Same as above.
Literacy, Both Read and Write	A binary indicator assigned a value of one when the observation can both read and write.	Same as above.

Continued on next page

Table A1 continued from previous page

Statistic	Description	Source
DED Literacy: Pre-National Schools	The share of residents with the ability to both read and write in 1841. These data were areally interpolated (Prenner and Revord, 2019) by superimposing civil parish polygons (the spatial units reporting literacy in 1841) over the district electoral division boundaries used in this application.	British Parliamentary Papers (1843).
DED Literacy: National School Increase	The difference between the share of fully literate aged 60 and above (born in 1841 or before) and the pre-national school literacy variable at the district electoral division level. The vast majority of these individuals would have been taught in the national school system and therefore this difference represents the literacy boost stimulated by national schools.	British Parliamentary Papers (1843) and National Archives of Ireland (2019).
DED Literacy: Post-Famine Increase	The difference between the share of fully literate aged 20-39 (born 1861-1881) and the level of literacy amongst those 60 and above. Again, this variable is measured at the district electoral division level. This variable measures the contemporary level of literacy amongst each observation's peers. We omit each observation from each district's calculation. Thus, someone who is illiterate will live in a more literate district compared to a person in the same district who is literate.	National Archives of Ireland (2019).
Roman Catholic	Binary indicator that takes a unit value when the observation is a Roman Catholic.	Same as above.
Roman Catholic Name Index	An index capturing the likelihood of someone's religious affiliation based on their first name.	National Archives of Ireland (2019) and see Connor (2021).
Traditional Name Index	A complementary index to that outlined above that accounts for the historic prevalence and popularity of certain names.	Same as above.
Teachers in DED (Per Capita)	The district population of teachers aged 40 and above.	National Archives of Ireland (2019).
Priests in DED (Per Capita)	The district population (both men and women, for all religions) aged 40 and above employed in religious occupations.	Same as above.
Police in DED (Per Capita)	The district population aged 40 and above employed in the police force (RIC).	Same as above.
<i>Continued on next page</i>		

Table A1 continued from previous page

Statistic	Description	Source
Market Access, 1841	The sum of all other districts' land values (a proxy for market size/economic activity) weighted by transport costs in 1841.	See Fernihough and Lyons (2022).
Port Access, 1841	The minimum cost of accessing an international port.	Same as above.
Market Access Increase: 1841-71	Market access growth caused by the introduction of railways.	Same as above.
Port Access Increase: 1841-71	International port access caused by railways.	Same as above.
Resident in Birth County	Binary indicator that takes a unit value when the observation is resident in their county of birth. A measure of non-migrant status.	National Archives of Ireland (2019).
Share of DED Born in County of Residence	Share of district population who reside outside their county of birth.	Same as above.
Pre-Famine Land Value per Acre	Poor Law land values, a rough measure of annual land rent.	See Fernihough and Ó Gráda (2022) and Fernihough and Lyons (2022).
Occupation: Irish Speaking	The Irish-speaking share of the individual's profession.	National Archives of Ireland (2019).

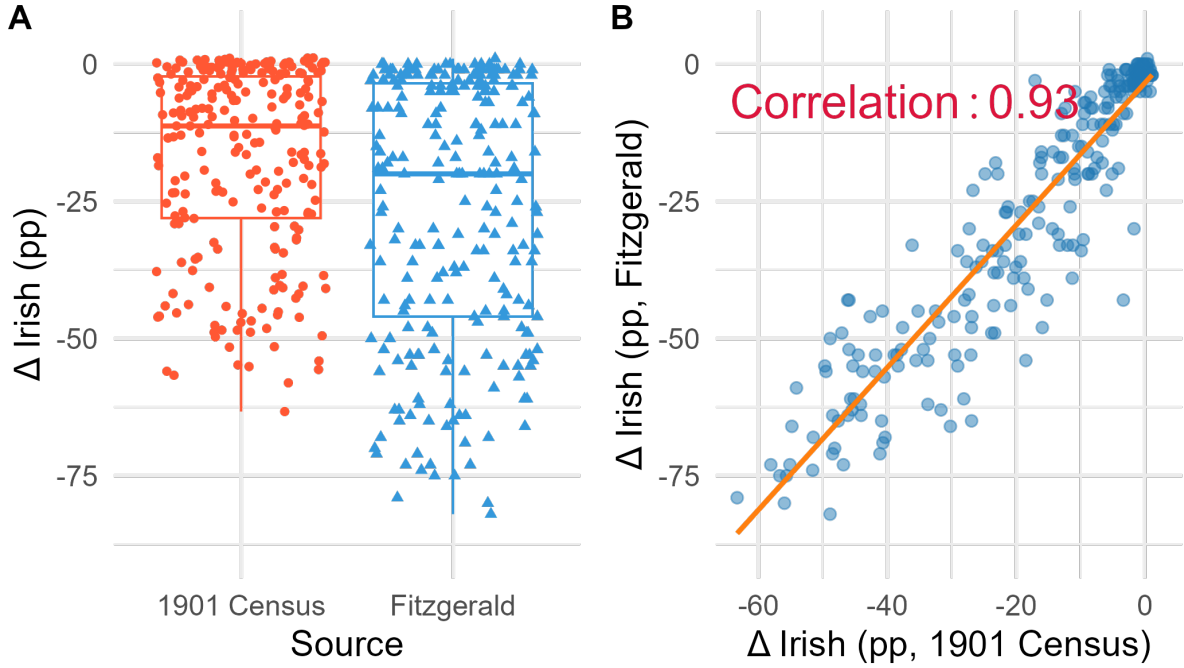
B Comparison with the 1881 Census

The role of sample selection in our previous results is of concern. If those who left previously Irish-speaking districts were most likely to be Irish speakers, this means we are overestimating the proportion of Irish speakers. Of greater worry is that emigrants were systematically selected in a way that is correlated with our regressors of interest, creating bias in our marginal effect estimates.

Garret FitzGerald used baronial-level data to estimate Irish speaking for decennial birth cohorts up to 1861-71 for 247 of Ireland's 336 (74 per cent) baronies (FitzGerald, 1984). The omitted baronies were mostly in the east, with non-existent levels of Irish speaking in the nineteenth century. Whilst not directly comparable, we can reshape our source, the individual 1901 census returns, to match Fitzgerald's data. By superimposing a baronial GIS shapefile, we can accurately approximate baronial-level data. Fitzgerald's primary source was the 1881 census report and this means we do not have data for the 1871-81 cohort, used in our analysis.

To compare the sources, we create a differenced variable measuring the change in Irish speaking between birth cohorts from 1821-31 and 1861-71. The Fitzgerald source for measuring the decline in Irish speaking is less likely to suffer from migration sample selection bias because the latter birth cohort was only 10-19 years of age in his source

Figure A1: Comparison with FitzGerald (1984)



and therefore less likely to have emigrated relative to those aged 30-39 in our 1901 census source.

Figure A1 provides a comparative synopsis of the 247 baronies for both sources. The box plot in panel A shows that our data underestimate the scale of the language shift. On average, Irish fell by 26 percentage points according to FitzGerald, where we only see a 13 percentage point drop. This discrepancy is consistent with our data overestimating the degree of Irish speaking in post-Famine birth cohorts.

Panel B depicts the association between the two sources. This image is more reassuring as a very strong relationship connects the two sources. The estimated correlation coefficient of 0.94 points to an almost one-for-one relationship that implies that the 1901 data are a scaled-down version of the earlier 1881 data. While emigration reduces our estimates of the Irish language decline, these reductions are proportionate thus validating our use of 1901 in regression analysis.

The consequences of using the 1901 source data can be further probed through regression. Table A2 contains the regression results for two model specifications that run parallel with both measures of the Irish language featuring as the dependent variable. To match our earlier models, we trim the sample and focus on the 95 baronies where the majority share has Irish ability in the pre-Famine period. We account for spatial confounding by estimating a Generalized Additive Model (GAM) with latitude and longitude coordinate variables entering as an isotropic smooth (Wood, 2017). We also correct the standard errors for spatial autocorrelation at a distance of 50km (Conley,

Table A2: Baronial Regressions Explaining Irish Language Ability in 1871 (pp)

	Source: 1901 Census		Source: Fitzgerald	
	(1)	(2)	(3)	(4)
Literacy: 1841 Census	−1.297*** (0.388)	−1.562*** (0.307)	−1.273*** (0.448)	−1.368*** (0.444)
Literacy Growth: National Schooling	−4.393*** (1.225)	−2.650** (1.209)	−3.681** (1.531)	−3.827** (1.565)
Literacy Growth: Post-Famine	0.012 (0.842)	1.112 (0.889)	−1.446 (1.166)	−1.576 (1.159)
Famine Loss, 1841-61		−1.659*** (0.501)		−0.070 (0.716)
Market Access Growth, 1841-61		0.268 (1.190)		−1.336 (1.328)
Port Access Growth, 1841-61		−0.352 (1.790)		−0.956 (2.232)
Lat-Lon Controls	Yes	Yes	Yes	Yes
Average Change in Irish Speaking	−30.408	−30.408	−45.284	−45.284
Observations	95	95	95	95

Note: Columns (1)–(2) and (3)–(4) present regression analyses of baronial shares of Irish speakers within the 1861-71 birth cohort, using our dataset and Fitzgerald’s dataset, respectively. The share of Irish speakers is expressed as a percentage point in both cases. All explanatory variables in the regression models have been standardized; therefore, a one-unit change represents a shift of two standard deviations. Conley standard errors, accounting for spatial correlation within a 50km radius using a uniform kernel, are presented in parentheses. *p<0.1; **p<0.05; ***p<0.01.

1999). Again, we have rescaled all of the explanatory variables to represent 2SD changes but, unlike earlier result tables, denote statistical significance in a conventional manner since they are more informative with the smaller estimation sample.

Table A2 presents a comparative analysis of data from 1901, featured in the first two columns, and Fitzgerald’s 1881 source, represented in the last two columns. The central observation indicates that the coefficients derived from Fitzgerald’s data either match or exceed ours in magnitude. This effectively counters the argument that sample selection primarily drives our findings, as the results remain consistent even when using earlier sources with fewer sample selection concerns.

When comparing across all columns, we find that pre-national schooling explains the share of Irish speakers in formerly majority Irish-speaking areas. Baronies where literacy was higher during this period had a lower share of Irish speakers later in the century. Quantitatively, the estimated coefficient is of a similar magnitude regardless of specification or Irish-speaking data source. These results imply that a 2SD (this literacy variable’s mean is 19pp with an SD of 6pp) increase in literacy predicts a 3 percentage

point fall in Irish-speaking ability. According to these results, the literacy boost provided by national schools helped reduce Irish speaking. On average, the introduction of the national school system led to a 22 percentage point increase in literacy. This implies a counterfactual effect of around 4.5 percentage points (as a 22% increase is a 3SD change).

In contrast to individual-level results, literacy growth in the post-Famine period was associated with higher levels of Irish speaking—although the estimated coefficients here are smaller than the other literacy estimates and not always statistically significant. The most likely reason for the positive coefficients is that we know the literacy effects are intergenerationally transmitted and therefore we would not expect to find a significant contemporaneous influence of literacy on Irish speaking. The positive effect may simply reflect the fact that post-Famine literacy gains occurred in areas with lower levels of pre-Famine literacy and that the small sample size here is inadequate to differentiate these influences.

C Selective Migration and Synthetic Cohorts

In the previous appendix, we found that selective migration overestimates the prevalence of Irish speaking in our analysis. Migration was a key feature of nineteenth-century Irish demography, and it appears that those who left were less likely to speak Irish. What would our results look like if we included those who left? Given that our analysis emphasises explanatory factors that differ at a spatially aggregated level (by DED), and that we have accurate measures of cohort depletion in these areas, we can generate synthetic cohorts reflecting the profiles of emigrants. Appending these observations to our data allows us to probe the robustness of our main results.

We produce these synthetic cohorts by the following method. We aggregate our data up to the DED level. Armed with population information from the decennial censuses, we construct two estimates of cohort depletion. The first assumes that the year-of-birth cohorts for each district can be approximated by taking the current cohort present in 1901 and scaling it up so that the true population is the current population plus a mark-up equal to the size of the cohort depletion occurring between late childhood and early adulthood. For example, the Graigabbey district in Co. Galway had 23 observations in our dataset born between 1861 and 1871. The population in the district dropped from 383 to 348 between the 1881 and 1891 censuses, a fall of around 9 per cent. Extrapolating this loss to our sample size implies two missing data points ($-23 \times -0.09 \approx 2$).

The second estimate, less conservatively, boosts the population to counteract all post-1841 population loss. In the small number of DEDs where the population increases (5.8% of the pre-Famine majority Irish-speaking DEDs with the first method and 2.6% with the second), no synthetic observations are generated. After ascertaining the size of these cohorts, we simulate Irish-speaking ability. We account for the likely possibility that Irish-

Table A3: Probit Regression Marginal Effects With Synthetic Cohorts

	Irish Language Ability					
	(1)	(2)	(3)	(4)	(5)	(6)
Year of Birth: 1871-81	-0.125 (0.002)	-0.112 (0.002)	-0.098 (0.001)	-0.125 (0.002)	-0.112 (0.002)	-0.098 (0.001)
Population Fall: 1841-51				-0.039 (0.002)	-0.040 (0.002)	-0.067 (0.002)
Population Fall: 1851-71				-0.008 (0.002)	-0.009 (0.002)	-0.041 (0.002)
Literacy in DED: Pre-National School	-0.490 (0.003)	-0.439 (0.002)	-0.345 (0.002)	-0.497 (0.003)	-0.447 (0.003)	-0.355 (0.002)
Literacy in DED: National School Rise	-0.234 (0.003)	-0.198 (0.003)	-0.170 (0.002)	-0.223 (0.003)	-0.187 (0.003)	-0.147 (0.002)
Literacy in DED: Post-Famine Rise	-0.103 (0.003)	-0.074 (0.003)	-0.067 (0.002)	-0.095 (0.003)	-0.067 (0.003)	-0.054 (0.002)
Average Irish Speaking	0.495	0.467	0.401	0.495	0.467	0.401
McFadden's <i>R</i> -Squared	0.215	0.183	0.133	0.216	0.184	0.137
Observations	289,041	320,277	427,232	289,041	320,277	427,232
Synthetic Observations	0	31,236	138,191	0	31,236	138,191

Note: Columns (1) to (6) present a series of individual-level probit regression marginal effects where an Irish-speaking dummy variable is the outcome of interest. District Electoral Division (DED) cluster-robust standard errors are in parentheses. Conventional statistical significance markers are omitted.

speaking was lower among those who left by using the average Irish-speaking ability of those with the same county of birth as the DEDs, but who resided elsewhere on the island. For example, 74 per cent of Galway's residents born in the county were Irish speakers, compared to 28 per cent of those born in Galway but resident elsewhere. Once we attach a likely "migrant speaking" probability to each county of birth, we use this statistic to simulate draws that match this likelihood. So in the case of Graigabbey, we generate observations where Irish-speaking ability is a 0-1 indicator, where $Pr(Irish = 1) = 0.28$.

The results of this exercise are reported in Table A3. Our most important results pertain to education, particularly educational variables that mapped onto our DED-level information. We do not need to simulate DED-varying explanatory factors as they are fixed. We replicate our main results in column (1), finding that a 2SD increase in pre-national school literacy reduces the probability of Irish speaking by 49 percentage points. The estimated national and post-national school variable marginal effects are consistent with the results in the main body of the paper.

We perform the probit regression analysis on our two samples in columns (2) and (3), including synthetic cohorts. Those who left were less likely to speak Irish, thus inflating the average level of Irish-speaking ability. This is reflected in both supplementary samples as the average Irish-speaking drops from 50 to 47 per cent, and then to 40 per cent. The magnitude of all the aforementioned marginal effects also attenuates, but not substan-

tially. Even if we assume that pre-Famine population levels were maintained, this only reduces the pre-national school variable’s effect from a 49 to a 35-percentage point swing. The more realistic birth-cohort depletion assumption reported in the second column leads to a 44-percentage point marginal effect. Furthermore, the rank order of variable importance is preserved—pre-national schooling always has twice the explanatory power compared to the literacy improvements stimulated by the national schools. Columns (4) to (6) include explanatory variables that measure population changes. Again, the inclusion of synthetic cohorts leaves our main results intact.

Appendix References

- British Parliamentary Papers (1843). *Report of the Commissioners Appointed to Take the Census of Ireland for the Year 1841*. London: HMSO.
- Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics* 92(1), 1–45.
- Connor, D. S. (2021). In the name of the father? Fertility, religion, and child naming in the demographic transition. *Demography* 58(5), 1793–1815.
- Fernihough, A. and R. C. Lyons (2022). Agglomeration and emigration: The economic impact of railways in post-Famine Ireland. QUCEH Working Paper Series, Paper No. 22-11. URL: <https://www.quceh.org.uk/uploads/1/0/5/5/10558478/wp22-11.pdf>.
- Fernihough, A. and C. Ó Gráda (2022). Population and poverty in Ireland on the eve of the Great Famine. *Demography* 59(5), 1607–1630.
- FitzGerald, G. (1984). Estimates for baronies of minimum level of Irish-speaking amongst successive decennial cohorts: 1771–1781 to 1861–1871. *Proceedings of the Royal Irish Academy. Section C: Archaeology, Celtic Studies, History, Linguistics, Literature* 84C, 117–155.
- National Archives of Ireland (2019). Census of Ireland 1901/1911 and census fragments and substitutes, 1821–51. [Data set]. Census years: 1901. URL: <http://census.nationalarchives.ie/pages/1901/>.
- Prener, C. and C. Revord (2019). areal: An R package for areal weighted interpolation. *Journal of Open Source Software* 4(37).
- Wood, S. N. (2017). *Generalized Additive Models: An Introduction With R*. Boca Raton, Florida: CRC Press.