University of Genova
PhD Program in Economics and Quantitative Methods

ESSAYS IN SPATIAL ECONOMICS

Author
Nicolò Badino

Supervisor
Gabriele Cardullo

I'm living for the first time.

— Anonymous

ACKNOWLEDGMENTS

This dissertation is the end of a long path, more than eight years of study. More than once I found myself wondering how I ended up, in the first place, studying economics, and then writing a thesis that sits at the intersection of spatial economics, economic history, industrial organization, and development economics.

The reasons, of course, lie in my own life, and reflect the time I have spent with the people who influenced the path I chose to take. To them I owe a sincere thank you for what I have become and for what I have done.

The curiosity that brought me here is something I owe to my family. From the bottom of my heart, thank you to Ilaria, my sister, and to my parents. As I write these lines, it strikes me that what I ended up doing, economic geography, is precisely the combination of what my parents studied and, evidently, passed on to me: economics, from my father, and geography, from my mother.

My passion for history clearly comes from the stories in my family, from my uncles to my grandparents, who, like so many Europeans of the last century, lived through one of the most historically dense periods of the last millennium. A special thought goes to my grandmother Anna, who, instead of fairy tales when I was a child, used to tell me about Frederick Barbarossa's deeds; and to my great-uncle Paolo, who told me about the time he was a partisan during the Resistance against Nazi-fascism.

I also want to thank my PhD colleagues, with whom I spent entire days sharing ideas and opinions. A special thanks to Marc-Antoine Faure, who introduced me to spatial economics; to Sanmoy Mukherjee, with whom I discussed these topics extensively and shared conference trips. Thanks also to (in alphabetical order) Michele Badolato, Mario Bonfrisco, Manuel Delfino, Giacomo Gaggero, Diletta Migliaccio, Federico Mij, and to all the PhD students from other cohorts with whom I lived some of the best moments of my life, and who have certainly become more friends than colleagues.

Thank you as well to Laura, my partner for almost ten years, with whom I have shared and discussed my research countless times. Even though she is entirely outside economics, her wisdom and the way she sees the world have undoubtedly had one of the greatest impacts on who I am today, and on what I wrote.

A thank you also to my friends, Elisa, Federico, Giorgia, Luca, Rosa, Sabina, Stefano, Virginia, who helped shape the person I am today.

Finally, thank you to Antonella Primi, whose memory will not be forgotten.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

This thesis is the result of two years of work in spatial economics. What emerges immediately is that geography matters.

Economists have long been concerned with the importance of institutions, mechanisms, and incentives, and have only partially studied the geographic component of the economy. Yet spatial economics, although less popular, rests on the work of very important scholars. Among the first is Hotelling, who in 1929 understood that geography is a crucial component of product differentiation, and that prices are a function of sellers' and, I would argue, buyers' locations.

Many other great economists studied these issues over the years, and several revolutions occurred as new mathematical tools and computational advances became available. In the 1990s Paul Krugman, later awarded the Nobel Prize, started the so-called New Economic Geography, studying more closely the forces of agglomeration and dispersion, and going deeper into the impact of trade and regional disparities. More recently, a new branch of the literature has developed with contributions, among others, by Rossi-Hansberg, Redding, Allen, and Arkolakis, combining richer geographical data with more sophisticated methods, made feasible by higher computing power.

My work was profoundly inspired by this latter branch of the literature, in particular after some extremely insightful discussions I had with Marc-Antoine Faure about how geography shapes why some countries, across different ages, succeeded or failed.

This thesis is therefore about three different topics, all connected by a spatial economic setting.

The first chapter is an economic development paper that studies the interaction between geography and climate change in the second millennium AD. The question I pose is whether the well-known divergence in timing and magnitude we observe around the Industrial Revolution is driven by exogenous explanations rather than only by the current hypotheses about institutions, religion, or colonialism. I propose a novel mechanism through which the Little Ice Age—starting in the thirteenth century and ending in the nineteenth—may have had different consequences across European countries depending on their location. While colder countries could have benefited from climatic cooling, creating stronger incentives to develop new techniques to maintain stable agricultural output, warmer countries may have faced weaker incentives to change, keeping them closer to Malthusian stagnation.

The second chapter is an economic history paper that studies the diffusion of the Industrial Revolution across Europe and posits that industrialization—proxied by patent production—may have been facilitated by spatial determinants. I find that patent production in the nineteenth century follows a southeastward direction, and that preexisting roads, waterways, and orography shaped how regions were exposed to innovation: areas that were better connected to already industrialized regions innovated earlier than more relatively isolated regions.

The final chapter is an industrial organization paper that generalizes the Hotelling model to a two-dimensional setting. The original reason for this exercise was to build a framework suitable for more realistic empirical investigations with geography. Through simulations, my setting is able to generate dispersion, agglomeration, and several other interesting properties.

What I learned in these two years of work is that spatial economics needs to be studied more deeply when we talk about inequality across countries and different development paths. I hope that this thesis has contributed, at least to some extent, to this matter.

1 OUT OF THE COLD: THE LITTLE ICE AGE AND THE AGRICULTURAL REVOLUTION

ABSTRACT

Between the seventeenth and nineteenth centuries, Northern Europe overtook Southern Europe, a reversal that preceded modern economic growth. I propose an underexplored channel for this transition: the Little Ice Age, a prolonged period of cooling and heightened climate volatility, whose economic impact was geographically uneven. Because agricultural productivity is non-linear in temperature, the same cooling tightened the subsistence constraint much more in colder regions. In a Malthusian economy this forced households to work more, and higher labor input accelerated learning-by-doing and induced innovation, raising agricultural productivity. I formalize this mechanism in a subsistence labor–leisure model with endogenous knowledge accumulation and derive testable predictions. Calibrated simulations show that regions facing greater climatic stress accumulate technology faster and escape the Malthusian trap earlier, reproducing the key features of the Little Divergence.

1.1 INTRODUCTION

For most of human history, pre-industrial economies experienced near-stagnant living standards. Small productivity gains were absorbed by population growth, keeping consumption close to subsistence, the classic "Malthusian trap" (Ashraf and Galor, 2011; Clark, 2008). The persistence of this equilibrium is striking: Milanovic, Lindert, and Williamson (2011) notes that the GDP per capita of England and Wales in 1290 was roughly comparable to that of the Roman Empire in 14 AD, with similar results for other European countries. The vast majority of the population worked in the agricultural sector, and most production was dedicated to meeting basic caloric needs.

Labor was abundant, but leisure was also relatively high, since producing beyond subsistence merely accelerated population growth, which in turn eroded any gains.

Yet, starting from the seventeenth century, a decisive break occurred. Parts of Northern Europe—especially England, the Netherlands, and Scandinavia—began to experience sustained agricultural productivity growth and later began to industrialize. Meanwhile, Southern Europe largely remained stagnant until the late nineteenth century

(Allen, 2000; Malanima, 2011; Olsson and Svensson, 2010). This North–South timing gap has been termed Europe's "Little Divergence."

The puzzle is twofold. First, why did the escape from the Malthusian trap occur at all, given that it had persisted for millennia? Second, why did it occur first in Northern Europe, when Mediterranean regions had been among the wealthiest and most urbanized areas of Europe as late as the sixteenth century?

Many economic historians agree that the Agricultural Revolution was a precondition for later industrialization (Overton, 1996), but the origins of this agricultural take-off remain contested. Previous explanations have emphasized institutional factors (notably the enclosure acts in England), the expansion of commercial exchange, the Protestant work ethic, and the colonization of the New World. However, a growing branch of the literature has begun to investigate how these developments may themselves have been responses to an exogenous cause: the climatic stress of the Little Ice Age (Parker, 2013; Waldinger, 2022). This paper proposes a novel explanation: the Little Ice Age drove the pre-industrial agricultural innovation that later fueled the Agricultural Revolution. The mechanism is simple. Because agricultural productivity is non-linear in temperature, the same cooling pushed colder regions further away from the agronomic optimum than warmer regions. In a subsistence economy, this asymmetric hit tightened the caloric constraint and forced households in the North to supply more labor to survive. Higher labor input, in turn, accelerated learning-by-doing and induced innovation, so that knowledge accumulated under climatic stress translated into a lasting productivity advantage. I formalize this channel in a Malthusian labor–leisure toy model with a subsistence constraint and endogenous knowledge accumulation, derive testable predictions, and confront them with calibrated simulations and historical evidence. An implication is a reinterpretation of the "Protestant work ethic": rather than the root cause of higher labor supply, it may have been a cultural reinforcement of behavior that climatic necessity had already made optimal.

The climatic shock: the Little Ice Age

Between approximately 1250 and 1850 AD, Europe experienced what climatologists describe as likely the coldest period of the last 8,000 years: the so-called Little Ice Age (LIA) (Wanner, C. Pfister, and Neukom, 2022). While there is no consensus on the exact beginning of the temperature decrease, multiple climate reconstructions suggest a gradual cooling process starting around 1300, with particularly cold phases between 1645 and 1715, the so-called *Maunder Minimum* (Luterbacher et al., 2004). In the European context, the onset of this transition is often dated to a period of abrupt cooling between 1275 and 1300, likely triggered by a cluster of massive volcanic eruptions and sustained by sea-ice/ocean feedbacks (Miller et al., 2012). Paleoclimatic reconstructions indicate that European winter average temperatures during 1500–1900 were reduced by approximately 0.5°C compared to the twentieth century, with the winter of 1708/1709 standing as one of the coldest of the past millennium (Luterbacher et al., 2004).

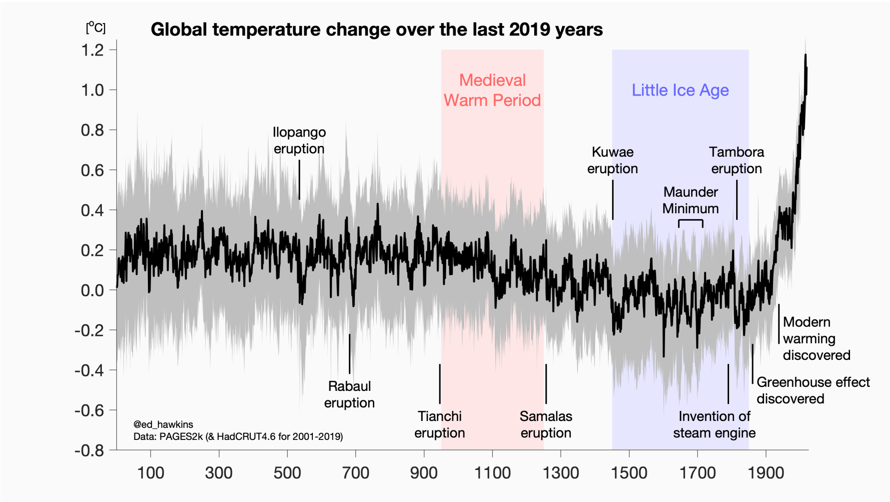


Figure 1: Temperature reconstructions by Ed Hawkins using data from PAGES2k showing the Little Ice Age cooling period.

The Little Ice Age was not, however, a period of uniform cooling. There is a growing consensus in the literature that it was rather a period of local and temporary cooling characterized by increased climate variability and extreme weather events. Neukom et al. (2019) show that there is no evidence for globally coherent cold epochs during the Common Era; rather, the LIA manifested as a complex mosaic of regional anomalies. Even within Europe, the cooling was asymmetric. Proxy data reveal that while England and the Low Countries experienced winters 0.8–1.2°C colder than the modern average, the Eastern Mediterranean saw a milder anomaly of around 0.5–1.0°C (Xoplaki, Luterbacher, et al., 2024). Similarly, summer temperatures in the Central Eastern Alps dropped by up to 1.5°C during the coldest decades (Larocque-Tobler et al., 2019), whereas Southern European summers remained broadly stable. In parts of the Mediterranean, the cooling was modest enough that growing-season conditions were not substantially impaired; some areas even recorded slightly enhanced spring precipitation and warmer early-summer temperatures during the LIA (Luterbacher et al., 2004; Wanner, C. Pfister, and Neukom, 2022), a contrast that will matter for understanding the asymmetric agricultural responses below.

Wanner, C. Pfister, and Neukom (2022) and Luterbacher et al. (2004) emphasize that the European LIA was primarily characterized by extreme seasonality: while winter average temperatures were lower, summer conditions were highly variable, oscillating between cold, wet seasons and periods of drought. This spatiotemporal variability meant that the economic impact was geographically uneven. The cooling was particularly pronounced in the North Atlantic and Alpine regions, whereas parts of Eastern Europe and the Mediterranean experienced different hydrological regimes (Gaceu et al., 2025; Hughes, 2014). For the agrarian economies of Northern and Central Europe, the climatic shift was profound; for the South, it was comparatively more modest.

Region	Summer JJA (°C)	Annual (°C)
Anatolia	20.99	12.47
Balkans	20.01	11.34
Eastern_Europe	16.94	7.75
Iberia	19.46	13.09
Italy	21.07	13.01
Scandinavia_S	15.09	5.67
UK	13.83	8.36

Table 1: Average temperatures (1000–1990 CE) by region. *Summer JJA*: mean June–August temperature from the Updated EuroMed2k reconstruction (Ljungqvist et al., 2019). *Annual*: mean annual temperature from the PAGES 2k EU2k2 BHM reconstruction, computed as ensemble-median anomaly plus CRU TS3.22 1961–1990 baseline.

It is unsurprising, therefore, that the climatic cooling of the seventeenth century often coincided with agrarian and social crises, especially in areas with pre-existing harsh climatic conditions. A famous example comes from the Norse colonies in Greenland: settlements established during the preceding Medieval Warm Period faced increasingly difficult conditions (Dugmore et al., 2012). After around 1250, the climate became much colder and stormier, making agriculture marginal. Paleodietary sources indicate that by around 1300, settlers had strongly reduced their consumption of agricultural products, deriving over 75% of their food from seal hunting and marine resources. By the mid-fourteenth century, trade with Europe had diminished and the Norse colony was in decline; isolated by sea ice and lacking sufficient local resources, the population abandoned the island or perished by the early fifteenth century.

Temperature and agricultural productivity

To understand why the same climatic event, the Little Ice Age, had heterogeneous effects across regions, we must consider the relationship between temperature and agricultural productivity. This relationship is non-linear and hump-shaped, sometimes described as an "inverted U-shape" or asymmetric curve: productivity increases with temperature up to an optimal point, then decreases. Agronomical research by Schlenker and Roberts (2009) shows that agricultural production is strongly dependent on temperature and the length of the growing season, with yields declining sharply when climatic conditions fall outside the optimal range. Specifically, for modern agriculture, they identify critical temperature thresholds beyond which productivity plummets: 29°C for maize, 30°C for soybeans, and 32°C for cotton. While pre-industrial Europe relied on different staples, the underlying physiological principle is consistent: there is a minimum heat accumulation, often measured in Growing Degree Days (GDD), required for a crop to mature. As Burke, Hsiang, and Miguel (2015) shows, on a macroeconomic scale, per capita production peaks at around 13°C. This non-linearity implies that the effect of a temperature change on the agricultural sector depends critically on the initial climatic condition. The regional summer temperature averages in Table 1 illustrate this asymmetry starkly: Mediterranean regions (Italy 21°C JJA, Iberia 19°C JJA; annual means 13°C) sit far above this threshold, while the UK (13.8°C JJA; 8.4°C annual) and southern Scandinavia (15.1°C JJA; 5.7°C annual) are already near or at the lower margin of viable cultivation.

Recent work by Hu et al. (2024) reinforces this perspective to analyze crop yield responses. Their findings highlight that while extreme heat negatively impacts yields, moderate temperature increases in certain contexts can be beneficial, particularly when interacting with precipitation patterns. As shown in Figure 2, the response of crop yields to temperature is non-linear and highly dependent on the baseline climate, supporting the hypothesis that temperature reduction affects mostly regions on the left of the curve, where the temperature is at or below the optimal temperature.

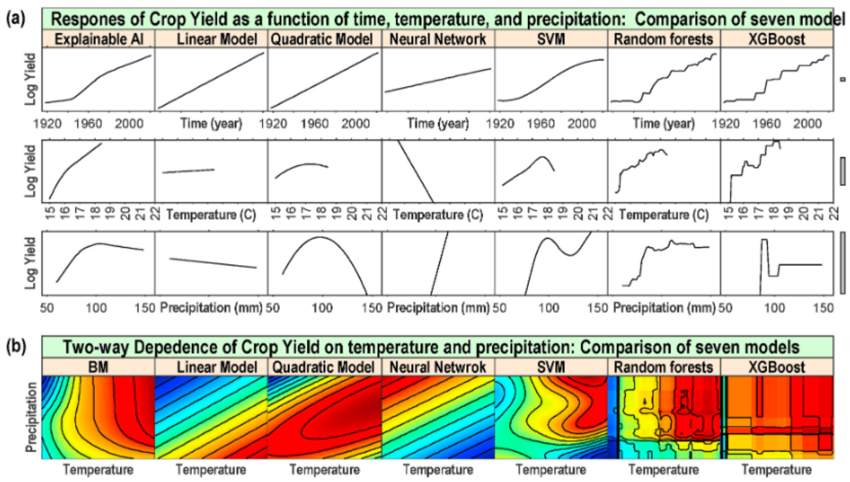


Figure 2: Responses of Crop Yield as a function of time, temperature, and precipitation. Source: Hu et al. (2024).

Following this logic, we can divide the European regions into two groups:

- **Regions above the optimum** (e.g., Mediterranean Southern Europe, with mean summer temperatures of 19–21°C; see Table 1): LIA cooling had paradoxical and ambiguous effects here. Warmer spring temperatures accelerated crop maturation, allowing wheat to complete its growth cycle before the late-summer Mediterranean drought (Oddo and Erdkamp, 2025). At the same time, the LIA increased summer precipitation in parts of the region (Luterbacher et al., 2004; Wanner, C. Pfister, and Neukom, 2022), further attenuating typical heat-and-drought stress. Both mechanisms buffered agricultural output from the cooling shock.

- **Regions near or below the optimum** (e.g., Northern Atlantic Europe, with UK at 13.8°C JJA / 8.4°C annual and southern Scandinavia at 15.1°C JJA / 5.7°C annual; see Table 1): These regions were precariously positioned on the lower tail of the productivity curve. Any further cooling reduced GDD below the minimum threshold for reliable cultivation, leading to sharp declines in agricultural product.

Climatic data for this period are scarce and lack fine spatial granularity. Nonetheless, two considerations support an asymmetric reading. First, existing proxy reconstructions suggest that the LIA cooling was, if anything, more pronounced in Northern Europe than in the Mediterranean (Luterbacher et al., 2004; Neukom et al., 2019), making a uniform-decline assumption conservative. Second, even under that conservative assumption, the non-linear temperature–productivity relationship documented above implies that Northern regions, already operating near the lower boundary of viable cultivation, would have borne a disproportionate share of the productivity loss. This logic motivates what I call the "caloric squeeze" hypothesis: a severe tightening of the margin between agricultural output and subsistence needs (Figure 3).

The caloric squeeze and the subsistence constraint

Caloric intake has a hard biological floor: prolonged deficits lead to starvation, disease, and death. This constraint fundamentally shapes the response to productivity shocks. In pre-industrial societies, the Little Ice Age cooling imposed a double burden. On one hand, it reduced crop yields; on the other, it increased the biophysical cost of survival. Martínez-González et al. (2020) note that during the Maunder Minimum, colder temperatures raised the basal metabolic rate (BMR) required for thermoregulation by up to 17–50%, effectively increasing the subsistence threshold $\bar{y}$ while simultaneously lowering land productivity.

Consider a simple production function:

$$y = f(\text{labor}, \text{land}, \text{climate}).$$

Land is fixed. When the climate deteriorates, reducing effective productivity and increasing caloric needs, households face a stark choice: either reduce consumption below subsistence and perish, or increase labor input to maintain output and survive. This reaction to exogenous shocks pushes toward induced innovation, where scarcity and crises act as catalysts for change (Boserup, 1965; Hicks, 1932).

We call this adaptation "forced industriousness." Historical evidence shows that starting in the 16th century, many European countries experienced a significant increase in labor input, typically attributed in part to the Protestant work ethic (Weber, 1905). However, current quantitative literature suggests that the massive increase in labor input observed in pre-industrial England was not a voluntary choice driven by religion, cultural shifts, or consumerism, but a survival-led necessity. Allen and Weisdorf (2011) demonstrate that between 1600 and 1800, rural workers drastically increased their annual working days, often doubling them from 150 to nearly 300, not to acquire luxury goods, but to maintain a stable subsistence level amidst falling real wages and rising food prices. This "hardship-led industriousness" (Humphries and Weisdorf, 2019) suggests that the compression of the caloric margin was the primary engine forcing households in colder, more vulnerable regions to extract more output from the same land, with labor as the only available margin.

This response requires the climatic pressure to be manageable: too severe a shock, such as in the Greenland case, leads to collapse; too mild, as in the Mediterranean core, reduces the incentive to innovate. We argue that regions like England or the Netherlands faced a pressure strong enough to break the stagnation but not catastrophic enough to cause collapse. That intermediate pressure, significant but survivable, was what produced the industrious response.

Learning-by-doing: from crisis to innovation

The increase in labor hours had an important side effect: it accelerated the accumulation of productive knowledge. Following Arrow (1962), this is learning-by-doing: technical progress arises from experience accumulated through production. The more hours farmers work, the more they experiment, adapt, and discover new techniques. In the framework I am going to present in the next section, the rate of technological accumulation is directly linked to labor effort: $A_{t+1} = A_t[1 + \rho h_t^{\gamma} - \delta_A]$, where h_t represents hours worked and ρ the learning rate. This mechanism implies that the regions forced to work hardest during the Little Ice Age, the colder regions, accumulated technology faster. The harsh climate, by compressing leisure and forcing continuous agricultural effort, inadvertently created a hothouse of innovation. Farmers developed new rotations (including fodder crops like clover and turnips) to sustain livestock through longer winters (Tello et al., 2017), better drainage systems (Skoglund, 2023), improved fertilization techniques, and more resilient crop varieties. As Costello, Kearney, and Gearey (2023) documents, in pastoral regions like Scotland and Ireland, farmers engaged in small-scale cereal cultivation on uplands specifically to buffer against the risk of fodder shortages due to harsher winters. Similar patterns of land-use adaptation and diversification have been documented in southern Sweden, where drainage improvements and the introduction of new crop varieties (such as two-row barley) increased resilience to excess moisture (Skoglund, 2023). Tello et al. (2017) argue that the English agricultural revolution was "more a discovery than an invention, induced by a combination of climate challenges and market incentive". English farmers reacted to lower temperatures by increasing the flow of organic matter and manure into the soil; when temperatures rose again, the faster mineralization of soil organic matter led to bountiful yields that reinforced these innovative strategies. Colder temperatures and shorter growing seasons also disrupted ecological equilibria: prolonged cooling led to a decline in soil bacterial activity, reducing nitrogen release and diminishing land fertility (Tello et al., 2017). Several empirical studies support this dynamic. Dreu and Dijk (2018) show, through an analysis of 500 years of European data, that prolonged cold climate periods coincided with higher rates of scientific and technological innovation, even after controlling for population. Specifically, a 0.5°C drop in mean temperatures is associated with a 0.3–0.6 standard deviation increase in the rate of recorded innovations. The authors find that this link runs through economic pressure: colder temperatures raised grain prices, which in turn incentivized the search for new solutions. It is not that cold "inspires" innovation directly, but it generates the scarcity that makes innovation necessary. Martínez-González et al. (2020), analyzing England during 1645–1740, find that climate change during the Maunder Minimum negatively affected agricultural yields but also accelerated structural change in the agrarian sector, potentially contributing to the onset of the English Agricultural Revolution and the broader European divergence. The induced innovation hypothesis from climate shocks finds a parallel in prehistory. Matranga (2024) demonstrates that increased climatic seasonality—greater contrast between seasons, with long periods of scarcity—drove the Neolithic revolution and the adoption of sedentary agriculture among hunter-gatherer populations. The same logic applies to pre-industrial Europe: where climatic stress was most acute, the incentive to innovate was strongest, and where it was absent, so was the urgency to change. The warmer South exemplifies this contrapositive. Mediterranean farmers, operating closer to the climatic optimum, had less urgency to intensify their labor or experiment with new techniques.

Heterogeneous responses: trade, adaptation, and social crises

Different societies responded to the Little Ice Age in different ways, depending on institutional and geographic conditions. Interregional grain trade grew substantially during the cold period, as grain moved from surplus to deficit regions. Waldinger (2022) provides econometric evidence that trading cities were able to offset the negative effects of the cooling, whereas isolated regions suffered significantly more. This suggests that market integration and commercial networks cushioned the worst effects of climatic stress.

Moreover, the Southern European agricultural model exhibited structural resilience. Oddo and Erdkamp (2025) show that in Tuscany, the Little Ice Age manifested through harsh winters and increased hydrological instability, particularly in fragile landscapes like the *Crete Senesi* where excessive rainfall caused significant soil erosion. However, within this generally cooler regime, "warmer" springs and summers were actually beneficial for wheat maturation. By accelerating the growth cycle, higher temperatures allowed the grain to be harvested in early summer, effectively "escaping" the peak drought of July and August (Oddo and Erdkamp, 2025). Institutions like the *Annona*, the public grain administration, acted as shock absorbers, stabilizing supply and prices during harvest failures. Unlike the North, where cold meant yield failure, the South capitalized on a diversified agricultural portfolio. The cultivation of olives, vines, and legumes alongside cereals acted as natural insurance: specific climatic conditions detrimental to wheat might spare or even favor other crops. This biodiversity, combined with established market networks and institutional buffers, reduced the volatility of the food supply in the Mediterranean relative to the precarious monocultures of the North.

These differential outcomes are, in aggregate, the early signature of what historians call the Little Divergence. Northern economies that faced the sharpest climatic pressure responded with the most rapid adaptation. The Dutch Republic prospered during the coldest decades of the Little Ice Age (U. Pfister, 2022); Allen (2003) classifies the Dutch and English economies as high-wage exceptions precisely in this period. Degroot (2018) documents how Dutch merchants exploited harvest failures elsewhere and inventors adapted to the cold: iceboats for frozen canals, improved windmills. English farmers increased their labor input substantially and pioneered new agricultural rotations (Allen and Weisdorf, 2011). In Scandinavia, drainage improvements and new crop varieties were adopted in direct response to harsher growing conditions (Skoglund, 2023). The pattern is consistent: where climatic pressure was greatest, adaptation came earliest.

Conversely, in regions where these buffers failed, the stress of agricultural failure has been linked to darker social phenomena. Iyigun, Nunn, and Qian (2017) document a strong quantitative correlation between long-term cooling trends and the frequency of conflict in Europe and Asia. The erosion of the agricultural surplus weakened state capacity and intensified competition for shrinking resources, feeding a cycle of instability. Behringer (1999) similarly connects intense witch-hunting episodes to the desperation caused by climate-induced agricultural failures. As Zhang et al. (2011) documents, the coldest periods were associated with significant declines in agricultural production and bio-productivity, precipitating a sequence of social crises—including wars, famines, epidemics, and migration—which contributed to a population collapse around 1650.

From crisis to surplus: post-LIA innovations and demographic-industrial growth

A related question is whether the innovations developed during the Little Ice Age had lasting effects: whether the techniques and institutional changes adopted out of necessity also generated a productive surplus once the climate improved. The thesis is that they did: adaptations born of crisis laid the groundwork for the agricultural and demographic expansion of the 18th and 19th centuries. Historical evidence supports this reading. With the gradual improvement in climate during the 18th century, regions that had innovated early could capitalize on new techniques: more efficient rotations, organic fertilization (through nitrogen-fixing legumes like clover), improved drainage and irrigation systems, and the spread of high-yield imported crops raised agricultural output beyond pre-crisis levels. Tello et al. (2017) argue that investments in soil fertility made to survive the cold—such as intensive manuring and legume rotations—created a "delayed reward": when temperatures rose, the faster mineralization of the accumulated organic matter led to bountiful yields that reinforced these innovative strategies. This agricultural surplus had far-reaching consequences.

First, it supported rapid demographic growth: increased per capita caloric availability reduced famine incidence and increased fertility and survival, leading European population to unprecedented growth between 1750 and 1850. Quantitative estimates indicate that a single innovation—the widespread introduction of the potato—was responsible for a significant share of this growth. According to Nunn and Qian (2011), the adoption of the potato imported from the Americas accounts for a substantial share of Old World population growth and urbanization in the 18th and 19th centuries. In Scandinavia, Skoglund (2023) confirms that the introduction of the potato and new rotations allowed agricultural production to outpace demographic growth, definitively breaking the Malthusian constraints that had previously limited expansion.

A second consequence of the agricultural surplus was the activation of proto-industrial growth. With more food available at relatively lower prices, an increasing share of the population could turn to non-agricultural activities: emigration to cities supported early industrialization, while rural areas developed small-scale manufactures such as domestic textile work. This was a two-way relationship. Agricultural surplus enabled urban growth, but growing urban demand also pulled agricultural production forward. Following a Thünen logic, Kopsidis and Wolf (2012) and U. Pfister (2022) show that proximity to urban markets pushed farmers toward more intensive techniques, reinforcing the interdependence between city and countryside. The agricultural surplus ensured that this transition occurred without the Malthusian bottleneck: in short, north-western Europe entered the 19th century with a larger, better-nourished population more free to engage in emerging sectors.

North–South divergence: the role of climate in European asymmetry

Together, these elements point toward a climate-based explanation of intra-European divergence. The North–South divergence thesis postulates that countries in northern and Atlantic Europe, experiencing a more severe impact from the Little Ice Age, were forced to adapt earlier and more radically, developing agricultural innovations and modern institutions already in the pre-industrial era; conversely, southern (Mediterranean) and more temperate regions felt less the bite of cold climate and therefore delayed abandoning traditional systems. Comparative historical data support this reading. In countries like Sweden or England, the agricultural sector in the late 18th century was already growing faster than population, signaling continuous productivity gains (Olsson and Svensson, 2010). By contrast, Mediterranean countries like Italy remained in a stagnant condition until well into the 19th century, with agricultural output per worker even lower than in the medieval era (Federico and Malanima, 2004). This “little divergence” within Europe, preceding the 19th-century Great Divergence on a global scale, can partly be explained by geo-climatic differences: cold zones had to face marginal harvests and greater risks (late frosts, short summers, etc.), while warm zones like Southern Europe experienced more stable conditions that reduced the urgency of change. For example, England and the Netherlands, characterized by cool oceanic climates and strong seasonal oscillations, precociously adopted reforms such as enclosures and four-field rotation, and improved agricultural markets and property rights to incentivize productive investments. Such measures, as noted by Olsson and Svensson (2010), were decisive in enabling farmers to increase production and react to shocks; they also reveal the greater institutional flexibility of Northern Europe relative to the South. In the South, by contrast, traditional agrarian models (such as extensive cereal latifundia, widespread use of fallow, scarce varietal innovation) persisted longer, partly because the climatic pressure that, in our framework, acts as the primary trigger of agrarian change was comparatively absent. A milder and more stable climate reduced the cost of maintaining traditional systems.

Reassessing the Protestant work ethic

The thesis advanced in this work has important implications for one of the most influential explanations of the Great Divergence: Max Weber's hypothesis that the Protestant Reformation, with its emphasis on hard work and worldly success as signs of divine favor, was a primary driver of Northern European economic development. Weber's argument has been highly influential, and the correlation between Protestantism and early industrialization is undeniable. However, our framework suggests a different interpretation: the greater labor effort observed in Protestant Northern Europe may not have been primarily caused by religious ideology, but rather by climatic necessity. The Reformation spread most successfully in Northern Europe, precisely the regions most severely affected by the Little Ice Age during the seventeenth century (England, the Netherlands, Scandinavia). This geographic overlap is unlikely to be coincidental. It can be argued that the observed "work ethic" of these regions was, at its core, a rational response to environmental pressure rather than a cultural or theological innovation. Farmers in England, the Netherlands, and Scandinavia worked longer hours because they had to; their survival depended on it.

An important example is provided by the removal of holy days following the Reformation. Allen and Weisdorf (2011) document that in 1536, England abolished 49 saints' days on which work had been forbidden. Weber might interpret this as evidence of Protestant asceticism; our framework suggests a different reading. The removal of these days coincided with a period when rural workers needed to increase their annual working days simply to maintain subsistence consumption in the face of falling real wages and rising food prices. The Reformation thus provided the *theological justification* (abolition of the cult of saints) for what was already an *economic necessity* (working more days to avoid starvation).

This interpretation does not imply that the Reformation was irrelevant. Religious and cultural factors likely served as catalysts and reinforcers of the industrious behavior that climate had already made necessary. Protestant theology may have legitimized hard work, making it socially acceptable and morally virtuous. However, the fundamental impetus, the first-order reason why these populations worked more in the first place, may have been environmental and not ideological.

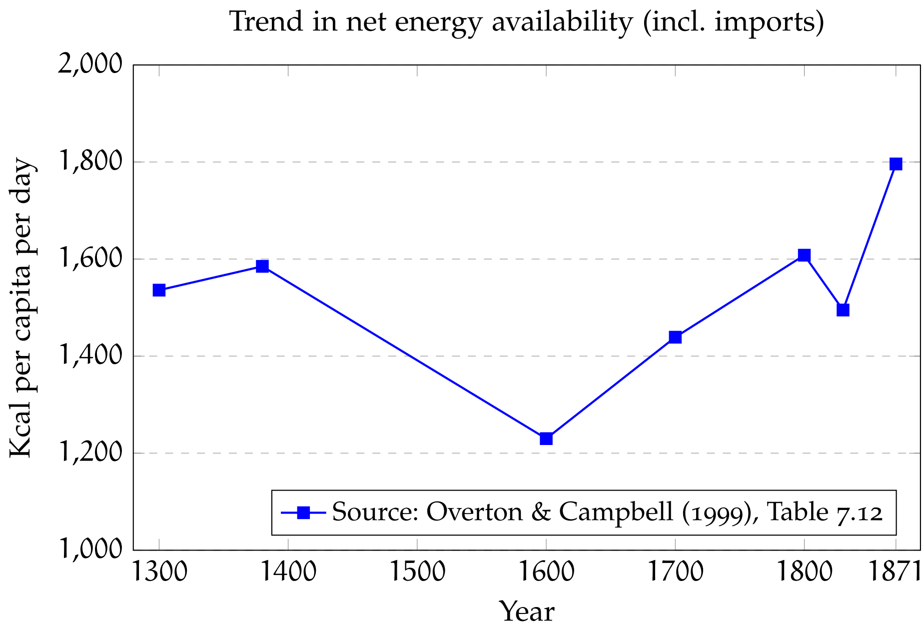


Figure 3: Estimated per-capita kilocalories available (grain and potatoes) in England, 1300–1871. Data include imports and are net of seed, losses, and feed.

Plan of the paper

I develop the arguments just explained, formally and empirically. Section 1.3 presents a model of Malthusian dynamics with induced innovation driven by climate shocks. The model formalizes the intuition that adverse climatic conditions, by forcing households to work harder to maintain subsistence, generate technological accumulation through learning-by-doing. I derive comparative statics showing how labor supply responds to climatic pressure and demonstrate conditions under which regions facing greater stress accumulate technology faster. Section 1.4 extends the analysis to long-run dynamics, showing how the technological lead accumulated during the Little Ice Age translates into an earlier escape from the Malthusian trap and an earlier demographic transition. I demonstrate that the same climatic shock that caused short-run hardship becomes the engine of long-run divergence. The model is calibrated to match pre-industrial European conditions, and numerical simulations are presented in Section 5. I compare two stylized regions, a "Cold" North and a "Warm" South, subject to the same physical temperature shock but experiencing asymmetric productivity impacts due to their different positions on the temperature-productivity curve. The simulations reproduce the key features of the Little Divergence: the North's earlier agricultural take-off, its accumulation of technological advantage, and the eventual reversal of fortune. Before the formal model, I review empirical evidence on the LIA's agricultural effects and the induced innovation literature.

1.2 MODEL

Environment and timing

Time is discrete, $t = 0, 1, 2, \ldots$. A representative agrarian economy is described by three state variables:

$$X_t \equiv (A_t, L_t, T_t),$$

where

- $A_t > 0$ is a stock of production-enhancing knowledge or technology;
- $L_t > 0$ is population size;
- $T_t > 0$ represents an exogenous climatic shifter (e.g., mean temperature or length of the growing season). For parsimony, T_t is interpreted as the *effective* climate term in production and can absorb both yield effects and climate-driven shifts in subsistence needs (e.g., thermoregulatory costs).

Land $R > 0$ is fixed and uniformly distributed across the population. Within each period t the sequence of events is:

1. State (A_t, L_t, T_t) is given.
2. Households choose labor $h_t \in [0, 1]$ (hours as a share of time endowment); leisure is $\ell_t = 1 - h_t$.
3. Output and consumption per capita c_t are realized from production.
4. Births and deaths occur; A_{t+1} and L_{t+1} are determined; the climate shifter T_{t+1} is realised.

Technology and production

Production is organized at the household level with a simple Malthusian technology. Effective labor per capita is given by h_t, and land enters via a congestion term. We assume

$$\begin{aligned} c_t &= y_t \\ &= A_t\, T_t\, \kappa(L_t)\, f(h_t), \end{aligned} \tag{1}$$

where

$$\kappa(L_t) \equiv \left(\frac{R}{L_t}\right)^{1-\alpha}, \qquad \alpha \in (0,1),$$

and $f : [0,1] \to \mathbb{R}_+$ is a reduced-form productivity of hours satisfying only:

$$f(0) = 0, \quad f'(h) > 0, \quad f''(h) \leqslant 0 \quad \text{for all } h \in (0,1]. \tag{2}$$

The term $\kappa(L_t)$ captures Malthusian congestion due to land scarcity, with $1-\alpha$ governing its strength. We abstract from physical capital accumulation, focusing instead on knowledge capital A_t.

For some results we will use the benchmark functional form

$$f(h) = h^{\eta}, \qquad \eta \in (0,1), \tag{3}$$

where η governs the diminishing returns to labor hours. The term $\kappa(L_t) = (R/L_t)^{1-\alpha}$ captures Malthusian congestion. Per-capita output scales with population as $c_t \propto L_t^{-(1-\alpha)}$; higher α implies weaker congestion.

Assumption (Atomistic Households). Each household is atomistic and does not internalize the effect of its labor choice on aggregate knowledge accumulation A_{t+1} (which depends on aggregate labor input $H_t \equiv h_t L_t$). Thus, the optimization problem is static.

Preferences and the subsistence constraint

Households value consumption and leisure. We impose a strict subsistence requirement $\bar{y} > 0$:

$$U(c_t, \ell_t) = u(c_t) + v(\ell_t), \qquad \ell_t = 1 - h_t, \tag{4}$$

subject to the feasibility constraint:

$$c_t \geqslant \bar{y} \quad \text{if feasible.} \tag{5}$$

Formally, if the maximum possible production $y_{max} = A_t T_t \kappa(L_t) f(1)$ is less than $\bar{y}$, the household works $h_t = 1$ to minimize the caloric deficit (Famine regime). If $y_{max} \geqslant \bar{y}$, the household maximizes utility subject to $c_t \geqslant \bar{y}$.

Standard properties apply: $u'(c) > 0, u''(c) < 0$ and $v'(\ell) > 0, v''(\ell) < 0$.

The static problem within a period

Given (A_t, L_t, T_t), let $\Pi_t \equiv A_t T_t \kappa(L_t)$ denote total factor productivity. The household solves:

$$\max_{h_t \in [0,1]} \; u(\Pi_t f(h_t)) + v(1-h_t) \quad \text{subject to} \quad \Pi_t f(h_t) \geqslant \bar{y} \text{ (if } \Pi_t f(1) \geqslant \bar{y}).$$

(6)

Define the minimal labor required to meet subsistence (if feasible):

$$h_t^{\text{need}}(A_t, L_t, T_t) \;=\; f^{-1}\!\left(\frac{\bar{y}}{A_t T_t \kappa(L_t)}\right). \tag{7}$$

Assumption (Survival Rule). If $\Pi_t f(1) < \bar{y}$ (subsistence infeasible), households supply $h_t = 1$ to maximize caloric intake.

Lemma 1.1 (Optimal Labor Choice)**.** *The optimal labor choice h_t^* is characterized by two regimes based on productivity Π_t:*

1. ***Infeasible Regime (Famine):*** *If $\Pi_t f(1) < \bar{y}$, by the Survival Rule:*

   $$h_t^* = 1.$$

2. ***Feasible Regime:*** *If $\Pi_t f(1) \geqslant \bar{y}$, the household chooses $h_t^* \in [h_t^{need}, 1]$. The solution is:*

   $$h_t^* = \max\left\{h_t^{need}, h_t^{unc}\right\},$$

   where h_t^{unc} is the solution to the unconstrained first-order condition:

   $$u'(\Pi_t f(h))\Pi_t f'(h) = v'(1-h).$$

Proof. Part (i) follows directly from the Survival Rule. Part (ii): The constraint $c \geqslant \bar{y}$ binds $h \geqslant h^{\text{need}}$. The objective function is strictly concave on this domain. If the unconstrained peak h^{unc} falls below h^{need}, the constraint binds ($h^* = h^{\text{need}}$). If $h^{unc} \geqslant h^{\text{need}}$, the interior solution is optimal. □

Corollary 1.2 (Corner vs. Interior Solution)**.** *In the feasible regime:*

1. ***Subsistence Trap:*** *Households are stuck at $c_t = \bar{y}$ ($h_t^* = h_t^{need}$) when productivity is low enough that the income effect of leisure dominates:*

   $$u'(\bar{y}) \cdot \Pi_t f'(h^{need}) \leqslant v'(1 - h^{need}).$$

2. ***Interior Growth:*** *Households choose $c_t > \bar{y}$ ($h_t^* > h^{need}$) when productivity is sufficiently high.*

Remark 1.3 (The Greenland Paradox). One caveat: our model predicts innovation only if the shock is survivable ($\Pi_t f(1) \geqslant \bar{y}$). If climatic conditions deteriorate to the point where even maximum effort cannot sustain the population, as was the case for the Norse settlements in Greenland during the LIA, the result is collapse, not induced innovation. The "industrious response" is conditional on the feasibility of the subsistence constraint.

Demographic Dynamics: A Reduced-Form Approach

To close the model, we specify the dynamics of population L_t. The demographic block is deliberately kept in reduced form. Our goal is to import the main qualitative insights of Galor and Weil (2000) without adding a full quality–quantity optimization layer that would complicate the model without altering its central mechanism. The three Galor–Weil dynamics we wish to preserve are: (i) a positive Malthusian link between income and fertility; (ii) an endogenous brake on fertility as technology accumulates, through the quality–quantity trade-off; and (iii) a demographic transition once productivity crosses a critical threshold. We implement all three through simple functional forms on birth and death rates. Population evolves as:

$$L_{t+1} \;=\; L_t \,\exp(n_t - d_t), \tag{8}$$

where n_t is the birth rate and d_t is the death rate. This exponential formulation ensures $L_t > 0$ for all t.

The mortality rate d_t depends negatively on the subsistence ratio $x_t \equiv c_t/\bar{y}$:

$$d(x_t) = \mu\, x_t^{\epsilon_d}, \tag{9}$$

with $\mu > 0$ and $\epsilon_d < 0$.

The fertility rate n_t is governed by a reduced-form function:

$$n(x_t, A_t) \;=\; \bar{n} \cdot x_t^{\epsilon_n} \cdot \frac{1}{1+\omega A_t}. \tag{10}$$

1

1.3 INDUCED INNOVATION VIA LABOR-LEISURE CHOICE

The results of induced industriousness

Building on the labor choice problem, we derive how effort responds to the three state variables: climatic pressure (T_t), technology (A_t), and population (L_t).

Proposition 1.4 (Climate, technology and Malthusian congestion)**.** *In the feasible region where the subsistence constraint binds (i.e.* $c^* = \bar{y}$*), the optimal labor choice* h_t^* *is given by:*

$$h_t^* = f^{-1}\left(\frac{\bar{y}}{A_t T_t \kappa(L_t)}\right). \tag{11}$$

1 Following Galor and Weil (2000), we assume that technological progress raises the return to human capital, thereby increasing the opportunity cost of raising children (the quality-quantity trade-off). This effect operates even at the subsistence level, reflecting the increased time cost of child-rearing in a more complex technological environment.

It follows that h_t^* *is strictly decreasing in* A_t *and* T_t*, and strictly increasing in* L_t*:*

$$\frac{\partial h_t^*}{\partial A_t} < 0, \qquad \frac{\partial h_t^*}{\partial T_t} < 0, \qquad \frac{\partial h_t^*}{\partial L_t} > 0.$$

Proof. By (2), f is strictly increasing and hence f^{-1} is strictly increasing. The term $\bar{y}/(A_t T_t \kappa(L_t))$ is strictly decreasing in A_t and T_t, and strictly increasing in L_t (because $\kappa(L_t)$ falls in L_t). Monotonicity of f^{-1} delivers the sign results. □

Proposition 1.4 provides the core economic intuition of our model. In a subsistence economy, households work as little as possible to secure $\bar{y}$. Consequently, any exogenous stressor, be it a cooling climate ($T_t \downarrow$) or a larger population ($L_t \uparrow$), forces households to increase their labor input h_t to survive.

Remark 1.5 (The Little Ice Age vs. The Black Death). It is instructive to contrast the Little Ice Age with a demographic shock like the Black Death. The Black Death was a sudden labor supply shock that increased the land-to-labor ratio ($R/L \uparrow$), thereby raising per-capita productivity and allowing more leisure. In contrast, the Little Ice Age was a productivity shock ($T \downarrow$). Although the LIA also led to a population decline due to starvation, the climatic deterioration was the primary driver of the crisis. Households were forced to work harder ($h \uparrow$) because the drop in T outweighed the marginal benefit of reduced congestion from lower L.

Learning-by-doing and the innovation feedback

Knowledge accumulation is linked to the time spent in production. We assume the following law of motion for A_t:

$$A_{t+1} = A_t\left[1 + \rho\, H_t^{\gamma} - \delta_A\right], \qquad \rho > 0,\ \gamma \in (0, 1],\ \delta_A \geqslant 0. \tag{12}$$

The term ρH_t^{γ} represents the learning-by-doing rate, explicitly grounding technical change in the accumulation of experience through labor conceptualized by Arrow (1962). This specification allows for a mild scale effect in learning: a larger population generates more aggregate experience, but with diminishing returns governed by γ. δ_A represents knowledge depreciation.

To ensure the model captures the historical reality of pre-industrial stagnation, we impose that in the absence of climatic shocks (normal times), the knowledge generated by subsistence labor roughly balances depreciation:

$$\rho(H_{subsistence})^{\gamma} \approx \delta_A. \tag{13}$$

Here $H_{subsistence}$ denotes aggregate labor input in the subsistence steady state. This condition formally defines the "trap": without the exogenous stimulus of the Little Ice Age, the net growth of technology is negligible ($\Delta A_t \approx 0$), and the economy remains indefinitely in the Malthusian steady state.

Through (12), the industrious response to climate shocks has a permanent effect on the technological trajectory.

Proposition 1.6 (The Escape Mechanism)**.** *Assume the functional forms* $u(c) = \frac{c^{1-\sigma}}{1-\sigma}$, $v(\ell) = \psi\frac{(1-h)^{1-\nu}}{1-\nu}$ *and* $f(h) = h^{\eta}$*. In the* ***Interior Regime*** *($c^* > \bar{y}$), assume the parameter conditions $\eta(1-\sigma) < 1$ and $\nu > 0$. If the income effect is sufficiently weak relative to the substitution effect (governed by the curvature parameter $0 < \sigma < 1$), then optimal hours h_t^* are strictly increasing in total factor productivity Π_t (and thus in A_t):*

$$\frac{\partial h_t^*}{\partial \Pi_t} > 0.$$

Proof. In the interior regime, the first-order condition is $u'(c)\Pi f'(h) - v'(1-h) = 0$. Substituting the functional forms:

$$\eta\Pi^{1-\sigma}h^{\eta(1-\sigma)-1} - \psi(1-h)^{-\nu} = 0.$$

Define the LHS function $G(h, \Pi) \equiv \eta\Pi^{1-\sigma}h^{\eta(1-\sigma)-1} - \psi(1-h)^{-\nu}$. By the Implicit Function Theorem, $\frac{\partial h^*}{\partial \Pi} = -\frac{\partial G/\partial \Pi}{\partial G/\partial h}$. First, differentiate with respect to Π:

$$\frac{\partial G}{\partial \Pi} = \eta(1-\sigma)\Pi^{-\sigma}h^{\eta(1-\sigma)-1}.$$

Since $\sigma < 1$, this term is strictly positive ($\frac{\partial G}{\partial \Pi} > 0$). Second, differentiate with respect to h:

$$\frac{\partial G}{\partial h} = \eta\Pi^{1-\sigma}[\eta(1-\sigma)-1]h^{\eta(1-\sigma)-2} - \psi\nu(1-h)^{-\nu-1}.$$

The second term $(-\psi\nu(1-h)^{-\nu-1})$ is strictly negative since $\nu > 0$. The first term is negative if $\eta(1-\sigma) < 1$. Given our assumption (and since $\eta < 1$ and $\sigma < 1$), both terms are negative, so $\frac{\partial G}{\partial h} < 0$. Thus, $\frac{\partial h^*}{\partial \Pi} = -\frac{(+)}{(-)} > 0$. □

Remark 1.7 (Industriousness and the Substitution Effect)*.* Proposition 1.6 provides the formal underpinning of what Voth (2001) and Vries (2008) call the "Industrious Revolution." Once productivity is high enough to push the economy into the interior regime, a low curvature $\sigma < 1$ in $u(c)$ ensures that the substitution effect dominates the income effect, sustaining high labor input as A grows. This dynamic allows the technological lead accumulated under climatic stress to persist and compound after the climate recovers. The specific value of σ is discussed in the calibration (Section 1.5).

The Mechanism of the Trap and the Escape

The results above describe a specific dynamic in pre-industrial growth, which we call the "Leisure Trap."

In our framework, as long as households are bound by the subsistence requirement, technological improvements ($A_t \uparrow$) do not automatically translate into higher consumption. Instead, they allow households to meet their subsistence needs with fewer hours of work. Since agents value leisure, they respond to "good times" by reducing labor effort. Crucially, because technological progress is driven by learning-by-doing, this reduction in labor effort acts as a brake on innovation. Thus, in a benign environment, the economy tends to settle into a low-work, low-growth equilibrium: prosperity limits the very accumulation of knowledge needed to sustain it.

This mechanism implies that autonomous improvements are insufficient to trigger a transition. The system requires a shock strong enough to break the preference for leisure.

The Little Ice Age provides exactly this shock. The climatic deterioration ($T_t \downarrow$) acts as a negative productivity shock. In the short run, households must work significantly harder ($h_t \uparrow$) just to maintain subsistence consumption. However, this "forced industriousness" has a dynamic benefit. The surge in labor effort accelerates the creation of knowledge via learning-by-doing, forcing the economy to accumulate technology at a rate that would not be chosen voluntarily in a stable climate.

If the shock persists, the stock of technology A_t accumulates to a critical level. Once productivity is sufficiently high, the incentive structure flips: the economy crosses into the "Interior Solution" regime where higher productivity makes consumption more attractive relative to leisure. This sets off a self-reinforcing process of rising hours, rising consumption, and sustained technological growth that eventually breaks the Malthusian trap.

This escape requires a "critical mass" of accumulated knowledge. If the Little Ice Age were to end prematurely, before A_t has reached the takeoff threshold, the incentive to innovate would vanish. Without the pressure of the cold climate, households would revert to their preference for leisure, causing labor hours h to drop. If h falls below the level required to offset knowledge depreciation (δ_A), the economy would gradually slide back into stagnation. European geographical divergence can thus be read as a product of the LIA's varying duration and intensity: only where climatic stress lasted long enough did a structural shift occur and the agricultural revolution take root permanently.

Testable Hypotheses

The model generates three qualitative predictions. Each finds support in the empirical literature reviewed above; the simulations in Section 5 show that the calibrated model reproduces them quantitatively.

1. **Timing of Divergence:** Regions experiencing more severe or persistent cooling during the Little Ice Age should exhibit earlier signs of agricultural productivity growth compared to milder regions. This is consistent with the historical record for Northern Europe relative to the Mediterranean (Martínez-González et al., 2020; Olsson and Svensson, 2010).

2. **Intensity of Shocks:** There should be a positive correlation between the magnitude of the climatic shock and the subsequent rate of innovation, provided the shock does not exceed the survival threshold. Dreu and Dijk (2018) provide direct econometric evidence for this relationship across five centuries of European data.

3. **Reversal of Fortune:** Areas that were marginally less productive in the short run due to climatic stress should eventually overtake climatically stable regions in the long run through accumulated induced innovation, the early signature of the Little Divergence documented by U. Pfister (2022) and Allen (2003).

1.4 LONG-RUN DYNAMICS AND THE GREAT DIVERGENCE

The Takeoff Threshold

Combined with the demographic laws in Section 1.2, the accelerated growth of A in regions hit by the LIA leads to an earlier demographic transition.

Proposition 1.8 (Existence of a Transition Threshold)**.** *Consider the fertility function* $n(x, A)$ *given in* (10)*.*

1. ***Subsistence Regime:*** *If the economy is constrained at subsistence* ($c_t = \tilde{y}$)*, fertility* n_t *is strictly decreasing in technology* A_t*.*

2. ***Interior Regime:*** *In the interior solution regime* ($c_t > \tilde{y}$)*, there exists a critical technological level* A^* *such that the relationship between fertility and technology is hump-shaped:* n_t *increases with* A_t *for* $A_t < A^*$ *(if the income effect dominates) and decreases for* $A_t > A^*$*.*

Proof. We analyze the partial equilibrium response of fertility n_t to an increase in technology A_t, holding L_t constant (and thus $\kappa(L_t)$ constant).

CASE 1: SUBSISTENCE REGIME. Consider the regime defined in Lemma 1.1 where optimal consumption is corner-constrained: $c_t^\star = \bar{y}$. In this case, the relative income $x_t \equiv c_t^\star/\bar{y} = 1$ is constant and independent of A_t. Substituting this into the fertility function (10):

$$n_t|_{c=\bar{y}} = \bar{n} \cdot (1)^{\epsilon_n} \cdot \frac{1}{1+\omega A_t} = \frac{\bar{n}}{1+\omega A_t}.$$

Differentiating with respect to A_t:

$$\frac{\partial n_t}{\partial A_t} = -\frac{\omega \bar{n}}{(1+\omega A_t)^2} < 0.$$

Thus, in the strict subsistence regime, technological improvements lead solely to a decline in fertility (due to the pure substitution/complexity effect).

CASE 2: INTERIOR SOLUTION REGIME. Consider the regime where technology is sufficiently high such that households choose an interior solution $c_t^\star > \bar{y}$. Recall from Lemma 1.1 that in the interior regime h_t^* satisfies $u'(c_t^\star)\Pi_t f'(h_t^*) = v'(1-h_t^*)$, which implies optimal consumption $c_t^\star(A_t)$ is strictly increasing in A_t. Let $x_t(A_t) = c_t^\star(A_t)/\bar{y}$. Taking the natural logarithm of (10):

$$\ln n_t = \ln \bar{n} + \epsilon_n \ln x_t(A_t) - \ln(1+\omega A_t).$$

Differentiating with respect to A_t:

$$\frac{\partial \ln n_t}{\partial A_t} = \epsilon_n \frac{1}{x_t}\frac{\partial x_t}{\partial c_t}\frac{\partial c_t}{\partial A_t} - \frac{\omega}{1+\omega A_t} = \epsilon_n \frac{1}{c_t}\frac{\partial c_t}{\partial A_t} - \frac{\omega}{1+\omega A_t}.$$

Let $\varepsilon_{c,A} \equiv \frac{\partial \ln c_t}{\partial \ln A_t} > 0$ denote the elasticity of consumption to technology. We can rewrite the derivative as:

$$\frac{\partial \ln n_t}{\partial A_t} = \frac{1}{A_t}\left[\epsilon_n \, \varepsilon_{c,A} - \frac{\omega A_t}{1+\omega A_t}\right].$$

Fertility increases if and only if the breakdown of the Malthusian trap (income effect) dominates the complexity cost:

$$\epsilon_n \, \varepsilon_{c,A} > \frac{\omega A}{1+\omega A}.$$

For low levels of A (but high enough to be in the interior regime), the term $\frac{\omega A}{1+\omega A}$ is small, allowing the income effect to dominate ($\partial n/\partial A > 0$). As $A \to \infty$, the term $\frac{\omega A}{1+\omega A} \to 1$. A necessary condition for the demographic transition is that the income elasticity of births is not too high. Specifically, as technological progress makes consumption more efficient ($\varepsilon_{c,A} \to 1$), we require $\epsilon_n < 1$ for the fertility rate to eventually decline. This condition is consistent with empirical evidence for pre-industrial economies, which typically exhibits birth-income elasticities significantly below unity (Clark, 2008). The threshold A^* is defined as the point in the interior regime where this inequality flips. □

This threshold defines the escape from the Malthusian trap. In our climate experiment, the colder region (North) is forced into higher h and thus higher accumulation of A. If the shock is sufficiently persistent, A_{North} crosses A^* while A_{South} remains trapped in stagnation.

1.5 SIMULATION RESULTS

This section presents numerical simulations of the model calibrated for the period 1000–1900 AD. We compare two regions: a "Warm" region (South) and a "Cold" region (North). We posit that region-specific agricultural productivity $T_{r,t}$ is a non-linear function of physical temperature $\mathcal{T}_{r,t}$:

$$T_{r,t} = g(\mathcal{T}_{r,t}). \tag{14}$$

We interpret $g(\cdot)$ as globally hump-shaped in temperature, but we work on the *cold margin* (the left tail), where $g'(\cdot) > 0$ and the response to cooling is steep and non-linear. Thus, while both regions experience the same physical cooling $\Delta\mathcal{T}$ during the Little Ice Age (approx. 1275–1850), the *effective* productivity shock is asymmetric:

- **North (Cold Margin):** $T_{\text{North},t}$ drops significantly ($T_{shock} \approx 10\%$) due to the non-linear response of crop yields to shorter growing seasons.
- **South (Warm Core):** The same cooling has a negligible effect on $T_{\text{South},t}$ as the region remains well within the optimal growing envelope.

To keep the experiment transparent while matching the historical timing, we implement climate forcing as a reduced-form multiplicative wedge on productivity. For the North, $T_{\text{North},t}$ declines linearly from 1 to $1 - T_{shock}$ between 1275 and 1645, remains at $1 - T_{shock}$ during the Maunder Minimum (1645–1715), and then recovers linearly to 1 by 1850. For the South, we set $T_{\text{South},t} = 1$ throughout. This yields a "similar climatic event" that generates divergent economic incentives due to different initial geographic conditions.

Parameter Calibration

The model is calibrated to represent the technological and demographic conditions of Europe over 1000–1900 AD. Table 2 summarizes all parameter values together with their empirical source or calibration target. The calibration strategy follows standard practice in quantitative macroeconomic history (Galor and Weil, 2000; Lagerlöf, 2006): a small number of parameters are grounded directly in the empirical literature, while the remaining ones are set to match well-documented historical moments. Each internally calibrated parameter is pinned down by a single target, and no parameter is freely adjusted to improve the fit of the simulations.

Table 2: Parameter calibration (values used in the simulations)

Parameter	Description	Value	Source / target
Empirically grounded			
α	Land share (congestion)	0.50	Allen (2000), Crafts (1985), and Voigtländer and Voth (2006)
T_{shock}	LIA productivity shock (North)	0.10	Derived from Luterbacher et al. (2004) and Mann et al. (2009)
ϵ_n	Birth elasticity	0.50	Wrigley and Schofield (1981); see text
ϵ_d	Death elasticity	−0.40	Galloway (1988); see text
Internally calibrated (moment-matching)			
σ	CRRA coefficient	0.30	Match rising h 1750–1800 (Voth (2001)); $\sigma < 1.25$ (Chetty (2006))
ψ	Leisure weight	0.20	Match $h \approx 0.76$ (Voth (2001))
ρ	Learning rate	0.0021	Pre-LIA stagnation (Eq. 13)
δ_A	Knowledge depreciation	0.0025	Stagnation: $\rho H^{\gamma} \approx \delta_A$ in pre-LIA steady state
γ	Learning curvature	0.50	Bounded A growth without explosive takeoff
$\bar{n}$	Baseline birth rate	0.0215	Near-zero population growth in pre-LIA steady state
μ	Baseline death rate	0.0155	Near-zero population growth in pre-LIA steady state
ω	Q–Q trade-off	0.25	Match demographic transition timing (late 19th c.; Coale and Watkins (1986))
Normalization and functional forms			
$\bar{y}$	Subsistence requirement	0.90	Normalization
η	Labor elasticity in $f(h) = h^{\eta}$	0.70	Standard concavity
ν	Leisure curvature	1.00	Log-leisure benchmark
A_0	Initial technology	1.50	Normalization

Empirically Grounded Parameters

LAND SHARE ($\alpha = 0.50$) In pre-industrial agriculture, land was the binding factor of production. Allen (2000) estimates factor shares for early modern England using probate records and national accounts, finding that land's share in total factor payments ranged between 40% and 55%. Crafts (1985) provides similar estimates for the pre-industrial period. Voigtländer and Voth (2006) employ $\alpha = 0.40$ in their Unified Growth model of the British Industrial Revolution. Our choice of $\alpha = 0.50$ lies within this empirically grounded range and is a standard benchmark in Malthusian models.

CLIMATIC FORCING ($T_{shock} = 0.10$) During the Maunder Minimum (1645–1715), paleoclimatic reconstructions suggest unusually cold European winters; for Europe as a whole, winter average temperatures during 1500–1900 were about 0.5°C lower than the 20th-century average (Luterbacher et al., 2004; Mann et al., 2009). Schlenker and Roberts (2009) document that, in modern agriculture, a 1°C drop in summer temperature reduces yields by approximately 5%. Studies of medieval English agriculture find that increased summer rainfall (often associated with cooler conditions) reduces yields by approximately 10% per standard deviation. Parker (2013) and Campbell (2016) document widespread harvest failures and agrarian crises during the coldest phases of the LIA, consistent with yield reductions in the 10–15% range. We implement the Little Ice Age as a peak 10% reduction in effective agricultural productivity for the North ($T_{shock} = 0.10$). The South is assumed to be unaffected ($T_{South,t} = 1$).

DEMOGRAPHIC ELASTICITIES ($\epsilon_n = 0.50$, $\epsilon_d = -0.40$) The birth elasticity ϵ_n captures the responsiveness of fertility to changes in real income, the core of Malthus's "preventive check." In pre-industrial England, this operated primarily through the age of first marriage: higher wages allowed earlier marriage and thus higher fertility. Wrigley and Schofield (1981) document that roughly half of pre-industrial fertility variation was driven by marriage timing, which was itself wage-elastic. Our value of $\epsilon_n = 0.50$ is consistent with this evidence.

The death elasticity ϵ_d captures the "positive check", the inverse relationship between income and mortality. When real wages fall, mortality rises due to malnutrition, disease susceptibility, and famine. Galloway (1988) provides one of the most systematic cross-country estimates of the elasticity of mortality to grain prices in pre-industrial Europe, finding absolute values in the range 0.2–0.5 depending on time period and region (e.g., a cumulative elasticity of 0.38 for non-infant mortality in England during 1675–1755). Galloway estimates the elasticity with respect to grain prices, which serve as a direct inverse proxy for real income in pre-industrial economies where food dominated the consumer price index. Our value of $|\epsilon_d| = 0.40$ falls within this range.

Internally Calibrated Parameters

Internally calibrated parameters are set to match well-documented historical moments, as described below.

CRRA COEFFICIENT ($\sigma = 0.30$) The parameter σ governs the curvature of the utility function and determines the balance between income and substitution effects. For the "Industrious Revolution" documented by Voth (2001) and Vries (2008) to occur, wherein households increase labor supply as productivity rises, the substitution effect must dominate. Voth (2001) documents that annual working hours in England rose from approximately 2,500 hours in 1750 to approximately 3,300–3,500 hours by 1800, before declining to around 2,700 hours by 1870. The value $\sigma = 0.30$ is chosen to replicate this positive correlation between technology and labor input during the take-off period. Chetty (2006) shows that an upward-sloping labor supply curve requires $\sigma < 1.25$, with most labor supply studies implying values in the range 0.15–1.78 (mean ≈ 0.71). Our value lies within this range and satisfies the theoretical bound.

LEISURE PREFERENCE ($\psi = 0.20$) The leisure preference parameter ψ is calibrated to match the approximately 2,500 annual hours observed in mid-18th century England (Voth, 2001). In the simulations, one period is one year and h_t is mapped to annual hours by normalizing the time endowment to 3,300 hours, so that 2,500 hours corresponds to $h \approx 0.76$. This is a standard approach in macroeconomic models: ψ is chosen to match the observed allocation of time between work and leisure.

LEARNING AND STAGNATION ($\rho = 0.0021$, $\delta_A = 0.0025$, $\gamma = 0.50$) The pre-industrial era was characterized by extremely slow productivity growth. Crafts and Harley (1992) estimate annual TFP growth at approximately 0.1% for 1760–1800. Antràs and Voth (2003) confirm this using an independent dual method based on factor prices, estimating TFP growth at 0.1–0.2% per year for 1770–1860. We calibrate (ρ, δ_A, γ) jointly so that, in the pre-LIA steady state, the net growth term in (12) is close to zero (the stagnation condition (13)): $\rho H^\gamma \approx \delta_A$. The curvature $\gamma = 0.50$ ensures diminishing returns to learning-by-doing and prevents explosive knowledge takeoff. Given that learning depends on aggregate labor input $H_t = h_t L_t$, the parameter ρ should be interpreted as the productivity of aggregate experience rather than a per-capita learning rate.

BASELINE VITAL RATES ($\bar{n} = 0.0215$, $\mu = 0.0155$) The baseline vital rates are chosen to generate near-stationary population dynamics in the pre-LIA steady state, once the income-dependent checks (ϵ_n, ϵ_d) and the technology-dependent quality–quantity term (ωA_t) are taken into account in (10). For comparison, Wrigley and Schofield (1981) report substantially higher crude vital rates for England; in our reduced-form formulation, $\bar{n}$ and μ should be interpreted as *effective* per-period rates rather than literal crude birth/death rates.

QUALITY–QUANTITY TRADE-OFF ($\omega = 0.25$) The parameter ω scales the impact of technological progress on the opportunity cost of children, following the theoretical framework of Galor and Weil (2000). In our model, higher ω implies that technological advances more strongly incentivize parents to substitute child "quality" (human capital) for "quantity." We calibrate ω so that the demographic transition, the sustained decline in fertility, is triggered in the mid-to-late 19th century, consistent with the timing documented by Coale and Watkins (1986) for Western Europe, where the Total Fertility Rate fell below 4 children per woman between 1870 and 1900.

Normalization and Functional Forms

The subsistence constraint is normalized to $\bar{y} = 0.90$; all output and consumption variables are expressed relative to this threshold. The labor elasticity $\eta = 0.70$ governs diminishing returns to hours at the intensive margin ($f(h) = h^\eta$) and is a conventional functional form assumption. The leisure curvature $\nu = 1.00$ adopts the log-leisure benchmark. Initial technology $A_0 = 1.50$ normalizes the starting point of the simulations.

Main Results: From Stagnation to the Little Divergence

The calibrated model is simulated over the period 1000–1900 AD for two stylized regions: a "Cold" North (subject to a 10% productivity shock during the LIA) and a "Warm" South (negligible shock). The results illustrate the full chain of the proposed mechanism: climatic stress $\rightarrow$ forced industriousness $\rightarrow$ accelerated learning $\rightarrow$ technological divergence $\rightarrow$ escape from the Malthusian trap.

Labor supply response

Figure 4 displays the labor supply response. In the pre-LIA steady state, both regions exhibit similar hours worked ($h \approx 0.76$), consistent with a subsistence-bound equilibrium. As climatic conditions deteriorate after 1275, the North is forced to increase h_t sharply to maintain output above subsistence, precisely the "forced industriousness" predicted by Proposition 1.4. At the peak of the Maunder Minimum (1645–1715), Northern households work approximately 10% more hours than their Southern counterparts. Even after the climate recovers, the North maintains higher labor supply: the accumulated technology has pushed the economy into the interior regime where, by Proposition 1.6, rising productivity sustains high labor input through the substitution effect.

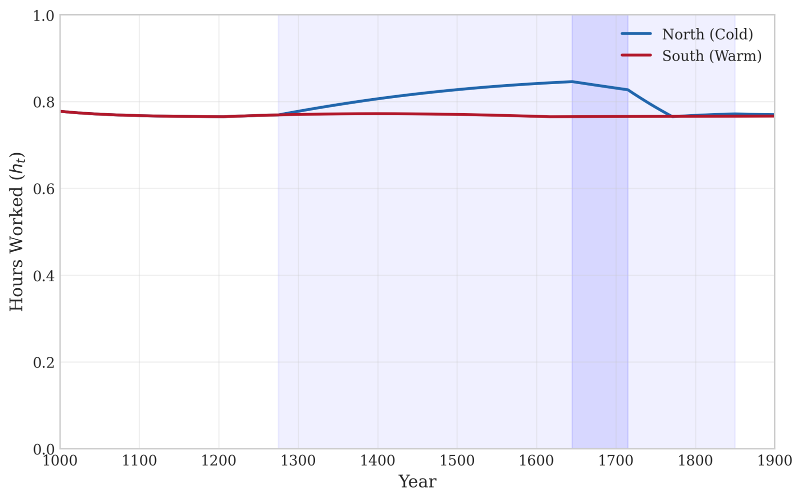


Figure 4: **Labor Supply.** The climatic shock forces Northern households to increase hours worked (h_t) to maintain subsistence. After recovery, the substitution effect sustains high labor input in the North. The shaded area marks the Little Ice Age (1275–1850), with the darker band indicating the Maunder Minimum (1645–1715).

Technological divergence

Figure 5 shows the resulting technological trajectories. Both regions start at $A_0 = 1.5$ and initially follow nearly identical paths. As the LIA intensifies, the higher labor input in the North accelerates learning-by-doing (Eq. 12), causing the Northern technology stock to grow faster. By 1715 the North has accumulated a measurable lead ($A_N/A_S \approx 1.03$), and the gap remains persistent through the post-LIA period (by 1900, $A_N/A_S \approx 1.03$).

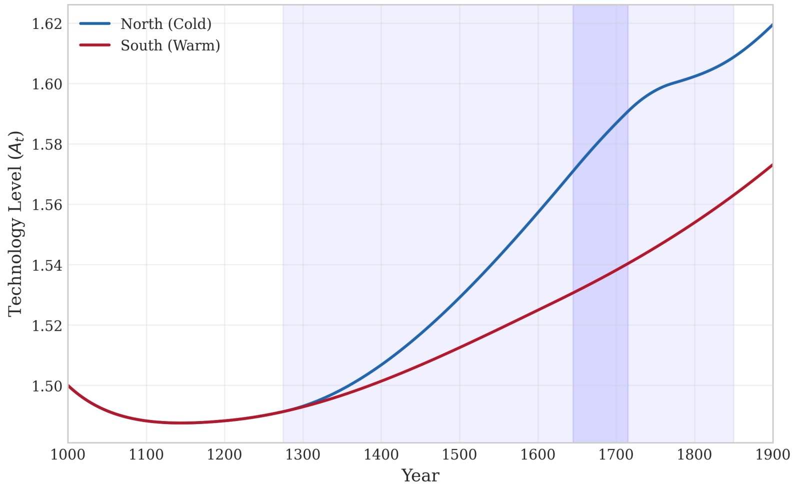


Figure 5: **Technological Divergence.** The North (Cold) accumulates technology faster than the South (Warm) due to higher labor input during the Little Ice Age, generating a persistent productivity advantage.

Escape from the subsistence trap

Figure 6 presents per capita consumption. Both regions begin near subsistence ($c \approx \bar{y} = 0.90$). During the LIA, Northern consumption drops slightly as the climate shock reduces effective productivity, but the accumulated technology eventually lifts consumption above the Southern level. This is the mechanism of the "escape": the forced accumulation of knowledge, though painful in the short run, generates a long-run surplus. The North breaks away from the subsistence floor, while the South remains trapped.

1.6 LIMITATIONS AND DISCUSSION

Our model focuses on the closed-economy labor-leisure channel and abstracts from several factors. On trade: while the LIA likely spurred grain flows from South to North, our mechanism emphasizes local production responses and does not model inter-regional exchange. On institutions: we treat institutions as endogenous to technology, but the LIA likely interacted with pre-existing frameworks, such as

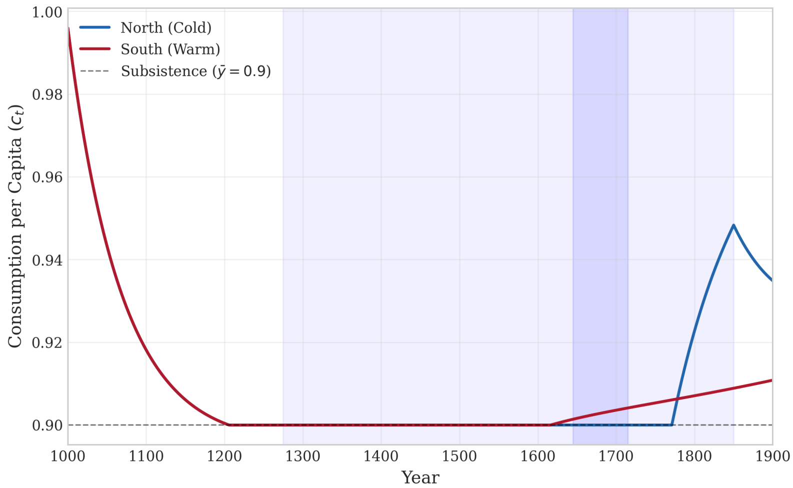


Figure 6: **Escape from the Trap.** Consumption per capita in the North (Cold) eventually exceeds the South (Warm). The dashed line marks the subsistence threshold $\bar{y}$.

the Enclosures, that are not explicitly modeled. On epidemics: we distinguish the LIA from the Black Death, but the interaction between climate, nutrition, and disease is complex and is simplified here into the survival rule.

1.7 CONCLUSION

This paper examines the climatic origins of the Agricultural Revolution and the Great Divergence using a subsistence labor–leisure model. The prolonged stress of the Little Ice Age, we argue, drove institutional and technological innovation in Northern Europe. Unlike previous models that treat technological progress as purely exogenous or density-dependent, our framework identifies environmental adversity as the primary trigger of the "Industrious Revolution."

Our simulations demonstrate that the North's early lead was not accidental. The necessity of maintaining subsistence in a colder climate forced households into a high-labor, high-innovation regime. This industriousness, combined with a technological lead that eventually triggered the demographic transition, allowed the North to escape the Malthusian trap while the South remained stagnant. Whether the same mechanism can explain within-country variation—between Scotland and England, or between Castile and Aragon—remains an open question.

REFERENCES

Allen, Robert C. (2000). "Economic structure and agricultural productivity in Europe, 1300–1800." In: *European Review of Economic History* 4.1, pp. 1–25.

Allen, Robert C. (2003). "Progress and poverty in early modern Europe." In: *The Economic History Review* 56, pp. 403–443.

Allen, Robert C. and Jacob Weisdorf (2011). "Was there an "industrious revolution" before the industrial revolution? An empirical exercise for England, c. 1300–1830." In: *The Economic History Review* 64.3, pp. 715–729.

Antràs, Pol and Hans-Joachim Voth (2003). "Factor Prices and Productivity Growth during the British Industrial Revolution." In: *Explorations in Economic History* 40.1, pp. 52–77.

Arrow, Kenneth J. (1962). "The Economic Implications of Learning by Doing." In: *The Review of Economic Studies* 29.3, pp. 155–173.

Ashraf, Quamrul and Oded Galor (2011). "Dynamics and Stagnation in the Malthusian Epoch." In: *American Economic Review* 101.5, pp. 2003–2041.

Behringer, Wolfgang (1999). "Climatic change and witch-hunting: the impact of the Little Ice Age on mentalities." In: *Climatic Change* 43.1, pp. 335–351.

Boserup, Ester (1965). *The Conditions of Agricultural Growth: The Economics of Agrarian Change Under Population Pressure*. London: Routledge. 124 pp. ISBN: 978-1-315-07036-0.

Burke, Marshall, Solomon M. Hsiang, and Edward Miguel (2015). "Global non-linear effect of temperature on economic production." In: *Nature* 527.7577, pp. 235–239.

Campbell, Bruce M. S. (2016). *The Great Transition: Climate, Disease and Society in the Late-Medieval World*. Cambridge: Cambridge University Press.

Chetty, Raj (2006). "A New Method of Estimating Risk Aversion." In: *American Economic Review* 96.5, pp. 1821–1834.

Clark, Gregory (2008). *A Farewell to Alms: A Brief Economic History of the World*. Princeton: Princeton University Press.

Coale, Ansley J. and Susan Cotts Watkins (1986). *The Decline of Fertility in Europe*. Princeton: Princeton University Press.

Costello, Eugene, Kevin Kearney, and Benjamin Gearey (2023). "Adapting to the Little Ice Age in pastoral regions: An interdisciplinary approach to climate history in north-west Europe." In: *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 56.2, pp. 77–96.

Crafts, N. F. R. (1985). *British Economic Growth during the Industrial Revolution*. Oxford: Oxford University Press.

Crafts, N. F. R. and C. Knick Harley (1992). "Output Growth and the British Industrial Revolution: A Restatement of the Crafts–Harley View." In: *The Economic History Review* 45.4, pp. 703–730.

Degroot, Dagomar (2018). *The Frigid Golden Age: Climate Change, the Little Ice Age, and the Dutch Republic, 1560–1720*. Cambridge: Cambridge University Press.

Dreu, Carsten K. W. De and Matthijs J. van Dijk (2018). "Climatic shocks associate with innovation in science and technology." In: *PLOS ONE* 13.1, e0190122.

Dugmore, Andrew J., Thomas H. McGovern, Orri Vésteinsson, Jette Arneborg, Richard Streeter, and Christian Keller (Mar. 6, 2012). "Cultural adaptation, compounding vulnerabilities and conjunctures in Norse Greenland." In: *Proceedings of the National Academy of Sciences* 109.10, pp. 3658–3663.

Federico, Giovanni and Paolo Malanima (2004). "Progress, decline, growth: product and productivity in Italian agriculture, 1000–2000." In: *The Economic History Review* 57.3, pp. 437–464.

Gaceu, Ovidiu Răzvan, Tudor Caciora, Ștefan Baias, Cezar Morar, Mihai Dudaș, Marius Stupariu, and Maria Maxim (Feb. 12, 2025). "Reconstruction of climatic events from the 16th century in Transylvania: interdisciplinary analysis based on historical sources." In: *Frontiers in Climate* 6. ISSN: 2624-9553.

Galloway, Patrick R. (1988). "Basic Patterns in Annual Variations in Fertility, Nuptiality, Mortality, and Prices in Pre-industrial Europe." In: *Population Studies* 42.2, pp. 275–303.

Galor, Oded and David N. Weil (2000). "Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond." In: *American Economic Review* 90.4, pp. 806–828.

Hicks, J. R. (1932). *The Theory of Wages*. London: Palgrave Macmillan UK. ISBN: 978-1-349-00191-0 978-1-349-00189-7.

Hu, Tongxi, Xuesong Zhang, Sami Khanal, Robyn Wilson, Guoyong Leng, Elizabeth M. Toman, Xuesong Wang, and Yang Li (2024). "Climate change impacts on crop yields: A review of empirical findings, statistical crop models, and machine learning methods." In: *Environmental Modelling & Software* 179, p. 106119.

Hughes, Philip D. (Dec. 2014). "Little Ice Age glaciers in the Mediterranean mountains." In: *Méditerranée. Revue géographique des pays méditerranéens / Journal of Mediterranean geography* 122. ISBN: 9782853999663 Number: 122, pp. 63–79. ISSN: 0025-8296.

Humphries, Jane and Jacob Weisdorf (2019). "Unreal wages? Real income and economic growth in England, 1260–1850." In: *The Economic Journal* 129.623, pp. 2867–2887.

Iyigun, Murat, Nathan Nunn, and Nancy Qian (2017). *The Long-Run Effects of Climate Change on Conflict, 1400-1900*. Working Paper 23033. NBER.

Kopsidis, Michael and Nikolaus Wolf (2012). "Agricultural Productivity Across Prussia during the Industrial Revolution: A Thünen Perspective." In: *The Journal of Economic History* 72.3, pp. 634–670.

Lagerlöf, Nils-Petter (2006). "The Galor–Weil model revisited: A quantitative exercise." In: *Review of Economic Dynamics* 9.1, pp. 116–142.

Larocque-Tobler, Isabelle et al. (2019). "The Little Ice Age signature in a 700-year high-resolution chironomid record of summer temperatures in the Central Eastern Alps." In: *Climate of the Past* 15, pp. 1291–1309.

Ljungqvist, Fredrik Charpentier, Andrea Seim, Paul J. Krusic, Jesper Björklund, Kristina Seftigen, Jan Esper, and Eduardo Zorita (2019). "European warm-season temperature and hydroclimate since 850 CE." In: *Environmental Research Letters* 14.8, p. 084015.

Luterbacher, Jürg, Daniel Dietrich, Elena Xoplaki, Martin Grosjean, and Heinz Wanner (2004). "European Seasonal and Annual Temperature Variability, Trends, and Extremes Since 1500." In: *Science* 303.5663, pp. 1499–1503.

Malanima, Paolo (2011). *The Long Decline of a Leading Economy: GDP in Central and Northern Italy, 1300–1913*. European Review of Economic History 15(2): 169–219. European Review of Economic History.

Mann, Michael E., Zhihua Zhang, Scott Rutherford, Raymond S. Bradley, Malcolm K. Hughes, Drew Shindell, Caspar Ammann, Greg Faluvegi, and Fenbiao Ni (2009). "Global Signatures and Dynamical Origins of the Little Ice Age and Medieval Climate Anomaly." In: *Science* 326.5957, pp. 1256–1260.

Martínez-González, José Luis, Jordi Suriñach, Gabriel Jover, Javier Martín-Vide, Mariano Barriendos-Vallvé, and Enric Tello (2020). "Assessing climate impacts on English economic growth (1645–1740): an econometric approach." In: *Climatic Change* 160.2, pp. 233–249.

Matranga, Andrea (2024). "The Ant and the Grasshopper: Seasonality and the Invention of Agriculture." In: *The Quarterly Journal of Economics* 139.3, pp. 1467–1504.

Milanovic, Branko, Peter H. Lindert, and Jeffrey G. Williamson (2011). "Pre-Industrial Inequality." In: *The Economic Journal* 121.551, pp. 255–272.

Miller, Gifford H. et al. (2012). "Abrupt onset of the Little Ice Age triggered by volcanism and sustained by sea-ice/ocean feedbacks." In: *Geophysical Research Letters* 39.2. ISSN: 1944-8007.

Neukom, Raphael, Nathan Steiger, Juan José Gómez-Navarro, Jianghao Wang, and Johannes P. Werner (July 1, 2019). "No evidence for globally coherent warm and cold periods over the preindustrial Common Era." In: *Nature* 571.7766, pp. 550–554. ISSN: 1476-4687.

Nunn, Nathan and Nancy Qian (2011). "The Potato's Contribution to Population and Urbanization: Evidence from a Historical Experiment." In: *Quarterly Journal of Economics* 126.2, pp. 593–650.

Oddo, Matteo and Paul Erdkamp (2025). "Climate change and institutional resilience in early modern Tuscany: The role of the Annona." In: *The Economic History Review*. Early View.

Olsson, Mats and Patrick Svensson (2010). "Agricultural growth and institutions: Sweden 1700–1860." In: *European Review of Economic History* 14.2, pp. 275–304.

Overton, Mark (1996). *Agricultural Revolution in England: The Transformation of the Agrarian Economy 1500–1850*. Cambridge Studies in Historical Geography. Cambridge: Cambridge University Press.

Parker, Geoffrey (2013). "Global Crisis: War, Climate Change and Catastrophe in the Seventeenth Century." In: *Yale University Press*.

Pfister, Ulrich (2022). "Economic Growth in Germany, 1500–1850." In: *The Journal of Economic History* 82.4, pp. 1071–1107.

Schlenker, Wolfram and Michael J. Roberts (2009). "Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change." In: *Proceedings of the National Academy of Sciences* 106.37, pp. 15594–15598.

Skoglund, Martin (2023). "Climate and Agriculture in Scandinavia during the Little Ice Age: Agrometeorological Relationships ca. 1500–1900." PhD thesis. Uppsala: Swedish University of Agricultural Sciences. ISBN: 978-91-8046-164-1.

Tello, Enric, José Luis Martínez, Gabriel Jover-Avellà, José Ramón Olarieta, Roberto García-Ruiz, Manuel González de Molina, Marc Badia-Miró, Verena Winiwarter, and Nikola Koepke (2017). "The Onset of the English Agricultural Revolution: Climate Factors and Soil Nutrients." In: *The Journal of Interdisciplinary History* 47.4, pp. 445–474.

Voigtländer, Nico and Hans-Joachim Voth (2006). "Why England? Demographic factors, structural change and physical capital accumulation during the Industrial Revolution." In: *Journal of Economic Growth* 11.4, pp. 319–361.

Voth, Hans-Joachim (2001). *Time and Work in England 1750–1830*. Oxford: Oxford University Press.

Vries, Jan de (2008). *The Industrious Revolution: Consumer Behavior and the Household Economy, 1650 to the Present*. Cambridge: Cambridge University Press.

Waldinger, Maria (2022). "The Economic Effects of Long-Term Climate Change: Evidence from the Little Ice Age." In: *Journal of Political Economy* 130.9, pp. 2275–2314.

Wanner, Heinz, Christian Pfister, and Raphael Neukom (July 2022). "The variable European Little Ice Age." In: *Quaternary Science Reviews* 287, p. 107531. ISSN: 02773791.

Weber, Max (1905). *The Protestant Ethic and the Spirit of Capitalism*. London: Routledge.

Wrigley, E. A. and R. S. Schofield (1981). *The Population History of England 1541–1871*. Cambridge: Cambridge University Press.

Xoplaki, Elena, Jürg Luterbacher, et al. (2024). “The Little Ice Age in the Eastern Mediterranean, 14th–17th Centuries.” In: *Climate Change in the Mediterranean*. Brill.

Zhang, David D., Harry F. Lee, Cong Wang, Baosheng Li, Qing Pei, Jane Zhang, and Yulun An (2011). “The causality analysis of climate change and large-scale human crisis.” In: *Proceedings of the National Academy of Sciences* 108.42, pp. 17296–17301.

2 THE SPATIAL DIFFUSION OF THE BRITISH INDUSTRIAL REVOLUTION

ABSTRACT

This study traces the spatial diffusion of formalized inventive activity across Europe using newly available patent data from the PatentCity database, treating the geography of patents as a proxy for the spread of industrialization. We introduce a spatiotemporal extension of kernel density estimation (KDE) to construct diffusion waves of patenting intensity that proxy the timing and location of industrial take-off at the regional level. Combined with a gravity model of knowledge spillovers and a Cox proportional hazards model, our analysis yields three main findings. First, the industrialization of continental Europe followed a wave-like diffusion process with a strong spatial component: proximity to the British epicenter, measured through historical transport networks, was the dominant predictor of both innovation intensity and timing. Second, structural estimation reveals that knowledge spillovers operated with diminishing returns to accumulated knowledge and were primarily a "push" phenomenon driven by source-region population, while receiving-region size played no amplifying role.

2.1 INTRODUCTION

The causes and consequences of the British Industrial Revolution have been studied in depth, but the spatial dynamics of its diffusion across continental Europe have received comparatively less attention. This study investigates the geographical spread of formalized inventive activity, as captured by patent filings, from Great Britain to continental Europe during the 19th century, interpreting patent geography as a proxy for the broader process of industrialization. The Industrial Revolution coincides with the period in which European patent systems became more observable and systematically documentable; patents are therefore a particularly useful source for spatiotemporal analysis of innovation, even though patent statutes existed earlier (e.g., the Venetian Statute of 1474 or the English Statute of Monopolies of 1624). We focus primarily on the European case, positing that the diffusion of registered inventive knowledge followed a spatial-dependent pattern originating from Great Britain with a distinct south-eastward direction. To substantiate this claim, we introduce a spatiotemporal extension of kernel density estimation (KDE) that generates diffusion waves from patent densities above specified thresholds (building on the tradition of using patents as innovation proxies; see Moser, 2005; Sokoloff, 1988). We complement this descriptive approach with a gravity model of knowledge spillovers and a Cox proportional hazards model for the timing of industrialization. On the continent, the first observable clusters of patenting activity appear in northern France and Belgium, the regions geographically closest to the British epicenter, before spreading southward and eastward, tracing a path consistent with proximity-driven knowledge diffusion from Great Britain. The remainder of this paper is structured as follows. Section 2 provides an overview of the industrialization process in continental Europe. Section 3 reviews the relevant literature. Section 4 presents our data and descriptive evidence on the spatial diffusion patterns, including a spatiotemporal kernel density analysis that documents the wave-like spread of innovation. Section 5 develops the theoretical model and estimation strategy. Section 6 presents the empirical results. Section 7 discusses the implications of our results and offers potential explanations for the observed patterns. Section 8 concludes.

2.2 THE INDUSTRIALIZATION OF THE EUROPEAN CONTINENT

The Industrial Revolution, originating in Great Britain in the late 18th century, was preceded by significant developments in agriculture and urbanization: from 1500 to 1750, British per capita agricultural production tripled (Broadberry et al., 2015) and the urban share rose from 7% to 23%. Technological innovations in textiles, iron smelting, and steam power, most notably Watt's improved steam engine (1769), transformed manufacturing and transportation, driving a two-fold increase in British real wages by 1850 (R. C. Allen, 2015).

The diffusion of these innovations across continental Europe was far from uniform. France and Belgium industrialized earliest on the continent; Germany followed after 1850, accelerating sharply after political unification in 1871. The Habsburg Empire, Italy, and Spain came later, each with sharp contrasts between an industrializing core and an agricultural periphery (Pollard, 1981). We exploit this variation below.

This uneven pattern motivates our quantitative analysis of spatial diffusion, which we measure through the geography of patent filings.

2.3 LITERATURE REVIEW

This paper draws on three bodies of literature: the economic history of the Industrial Revolution, models of technology diffusion, and spatial economics.

2.3.1 *Economic history and the Industrial Revolution*

The Industrial Revolution remains one of the most studied events in economic history. Traditional scholarship has extensively debated its causes, with explanations ranging from geography and coal availability (Fernihough and O'Rourke, 2021; Pomeranz, 2000; Wrigley, 2013) to institutions (Acemoglu, Johnson, and Robinson, 2001; North, 1990) and culture (Mokyr, 1992). A fundamental debate concerns the standard of living during the transition. While early optimistic views suggested a rapid rise in welfare, more recent quantitative work offers a nuanced picture. R. C. Allen (2015) argues that high British wages induced labor-saving innovation, yet the aggregate real wage growth during the classic period (1780-1850) remained modest compared to output growth (R. C. Allen, 2009). In this context, patent data have become a standard proxy for measuring innovative activity, despite their known limitations (Griliches, 1990). Bottomley (2014) and MacLeod (1988) document how the British patent system, though costly, became increasingly accessible, while Moser (2005) shows that patents captured only a fraction of total innovation, with significant sectoral heterogeneity. Nevertheless, recent efforts like PatentCity (Bergeaud and Verluise, 2024) and long-run datasets (Berkes, Chen, and Tranchero, 2026) reaffirm the utility of patents for tracking the geography of formalized innovation.

2.3.2 *Technology Diffusion*

Understanding how technology spreads matters as much as understanding its origins. Rogers (2003) conceptualized diffusion as a social process, a framework adapted by economists to model knowledge flows. Keller (2004) establishes that technology diffusion is geographically localized, with knowledge spillovers decaying rapidly with distance. Jaffe, Trajtenberg, and Henderson (1993) provide strong empirical evidence of this localization using patent citations, showing that inventors are significantly more likely to cite peers from the same region. At the macro level, Lucas Jr (2009) models the diffusion of the Industrial Revolution as a learning process mediated by trade, while Comin, Dmitriev, and Rossi-Hansberg (2012) provide a unified theory where technology adoption lags increase with spatial distance from the frontier. We test these mechanisms in the historical context of 19th-century Europe.

2.3.3 *Spatial Diffusion and Economic Geography*

While standard growth models often treat diffusion as a country-level phenomenon, economic geography emphasizes the continuous nature of space. The early "wave-diffusion" models of Hagerstrand

(1968) described innovation spreading from a center like ripples in a pond, a pattern we explicitly test. This spatial perspective was formalized in the New Economic Geography literature (Krugman, 1991), which models how centripetal forces (agglomeration) and centrifugal forces (congestion) shape economic activity. Bottazzi and Peri (2003) and Moretti (2021) extend this to innovation, showing that knowledge spillovers are bounded by distance but can jump between major urban centers. Our study applies these spatial concepts to the Industrial Revolution. Unlike Comin, Dmitriev, and Rossi-Hansberg (2012), who focus on cross-country technology lags, we document the *intra-continental* geographic movement of the patenting frontier.

2.4 DATA AND DESCRIPTIVE EVIDENCE

2.4.1 *Patent Data*

The use of patents as a proxy for innovation has a long-standing tradition in economic literature. They provide a tangible, geographically referenced measure of inventive activity across time and space. Scholars have employed various patent-related metrics to quantify innovative output and its economic impacts (Aghion, Van Reenen, and Zingales, 2013). While the causal relationship between patents and innovation remains debated (Boldrin, Levine, et al., 2008), this study acknowledges the strong correlation between areas of high intellectual production and elevated patent density. Our approach treats patents as an imperfect but informative indicator of inventive activity that can reveal patterns of technological change and diffusion. Buzard et al. (2020) show that proximity to high-density patent production zones correlates strongly with further patent publications in neighboring areas. This spatial dimension is directly relevant to our focus on the geographical diffusion of industrialization across 19th-century Europe. Our study uses the recently published PatentCity database (Bergeaud and Verluise, 2024), which offers open access to spatial and temporal granularity for historical patent data. This database maps patents from the United States, United Kingdom, Germany, and France to specific locations based on inventors' residences, with coverage extending back to the 19th century. The creation of this database involved Natural Language Models to digitize available patent scans, followed by data cleaning and harmonization processes. We queried the PatentCity database for patents published in the British and U.S. patent offices by European patentees prior to January 1, 1900. This selection captures innovation activity across Europe during the peak period of industrial diffusion, while accounting for the international nature of patent filings. Inventors from countries lacking robust historical patent systems often registered their inventions in prominent foreign patent offices, which broadens the geographic coverage of our dataset. A potential concern with this approach is selection bias: our dataset captures only inventors who filed abroad, which may overrepresent cosmopolitan inventors with international connections. While PatentCity also covers domestic French and German patents, these records are limited in number and, since France and Germany already fall within our continental analysis, they would not extend the geographic coverage. We therefore restrict attention to cross-border filings at the UK and US offices to ensure a consistent institutional framework across all observations. This choice implies that our measure captures the diffusion of *frontier* innovation, inventions deemed commercially valuable enough to warrant the cost of international filing, rather than the full universe of inventive activity. If anything, this selection should strengthen the spatial signal, since frontier inventors are precisely those most likely to participate in cross-regional knowledge flows. When patents list multiple inventors, each inventor residence enters as a separate location. As a result, multi-inventor patents contribute multiple inventor-location points, which is appropriate for our focus on the geography of inventors. To analyze the temporal evolution of patent activity, we divided the data into six cumulative groups, representing all patents published before 1850, 1860, 1870, 1880, 1890, and 1900, respectively (i.e., the final group includes patents published before 1900, through 1899). Figure 13 illustrates the spatial distribution of patentees for each of these periods and tracks how the geography of innovation evolved across Europe. The data reveal distinct patterns in the spatial distribution of patents. We observe early clusters of patent activity in northern France and Belgium, followed by a gradual expansion towards southern Germany and northern Italy. Our dataset comprises more than 120 thousand patents filed by European inventors in the British and U.S. patent offices between 1835 and 1899 (i.e., published prior to 1900). The distribution of these patents across time is not uniform: patenting accelerated in the latter decades of the 19th century, likely due to both faster industrial development and changes in patenting practices and institutions.

2.4.2 *Directional Distribution*

To analyze the spatial distribution of patents, we employ the directional distribution method. This approach applies standard deviation calculations to two-dimensional space and can quantify directional trends in patent diffusion. The directional distribution is represented by a standard deviational ellipse, which captures the spatial spread and orientation of the data points. The method rests on calculating spatial standard deviations:

$$\sigma_{1,2} = \left(\frac{\left(\sum_{i=1}^{n} \tilde{x}_i^2 + \sum_{i=1}^{n} \tilde{y}_i^2\right) \pm \sqrt{\left(\sum_{i=1}^{n} \tilde{x}_i^2 - \sum_{i=1}^{n} \tilde{y}_i^2\right)^2 + 4\left(\sum_{i=1}^{n} \tilde{x}_i \tilde{y}_i\right)^2}}{2n} \right)^{1/2}$$

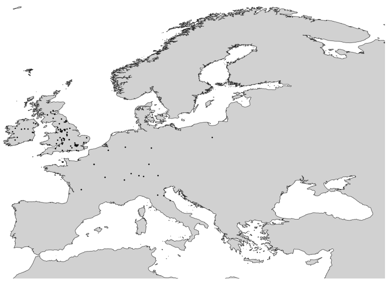


Figure 7: 1850

Figure 8: 1860

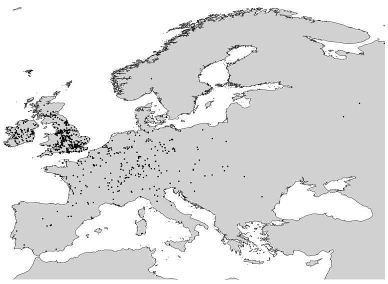


Figure 9: 1870

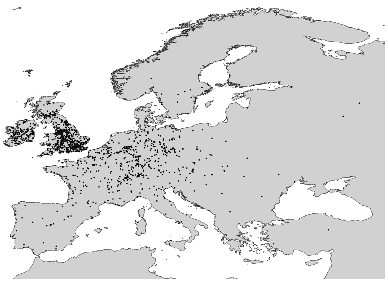


Figure 10: 1880

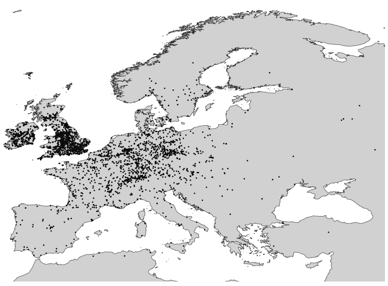


Figure 11: 1890

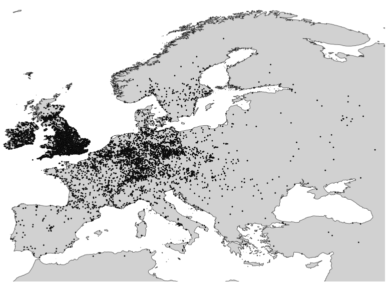


Figure 12: 1900

Figure 13: Spatial Evolution of Patent Activity (1850-1900). Each panel displays the geographical locations of patentees who filed patents in the UK or US during the respective cumulative period (patents published before the year shown). The maps illustrate the progressive expansion of innovative activity from the core regions of Northern France and Belgium towards Southern Germany and Northern Italy.

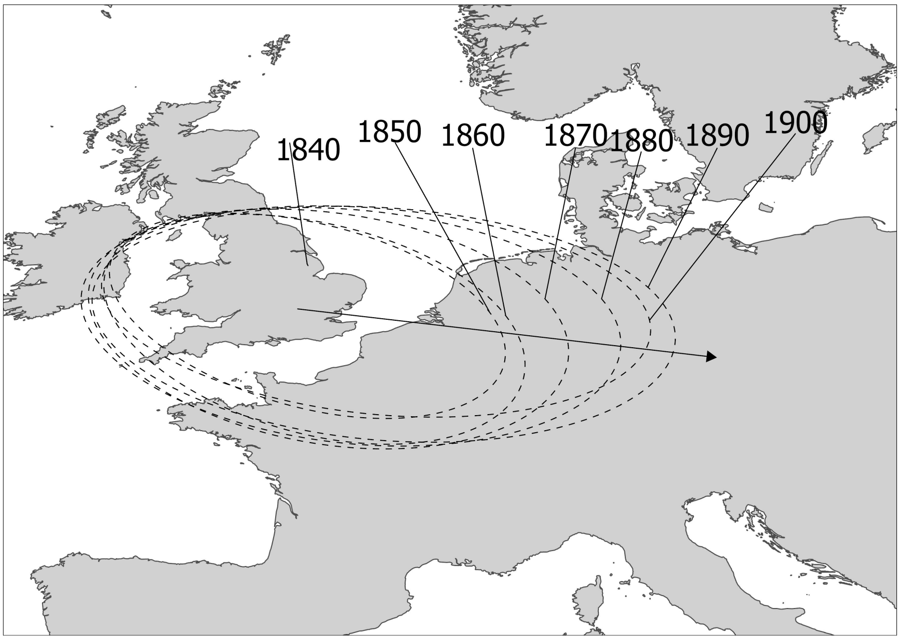


Figure 14: Directional Distribution of Patent Activity (Standard Deviational Ellipses). The ellipses summarize the directional trend of patent locations across the cumulative periods ending in 1850, 1860, . . . , 1900, computed at one standard deviation from the mean center. The elongation and rotation of the ellipses over time demonstrate a consistent south-eastward diffusion of the innovation frontier.

(15)

Where x and y represent the coordinates of each patent location, and the tilde values denote the difference between each coordinate and its mean. We use this to construct standard deviational ellipses (at one standard deviation) that summarize dispersion and orientation. We applied this directional distribution analysis to each of our cumulative subgroups of patents. The results in Figure 14 reveal a clear pattern in the diffusion of patent activity across Europe during the 19th century. In each successive cumulative period, we observe a clear south-eastward elongation of the ellipses: innovation spread from northwest to southeast across the continent. The ellipse for the 1835-1890 group is the first to depart from this trend, reducing its south-eastward stretch. This change suggests a shift in spatial dynamics: the clear south-eastward directionality of earlier decades gave way to a more even spread across the continent. The directional trajectory overlaps suggestively with the urban-industrial corridor later conceptualized as the “Blue Banana” (Brunet, 1989), a discontinuous arc of urbanization stretching from Northwest England to Northern Italy.

2.4.3 *Spatiotemporal Diffusion Patterns*

To characterize the spatiotemporal evolution of patent activity across Europe, we adapt the standard kernel density estimation (KDE) framework by incorporating a temporal component with cumulative time windows. This adaptation lets us document the spatial and temporal dimensions of the innovation diffusion process simultaneously, through density surfaces that show how patent activity spread across the continent over time. The standard kernel density estimation is a non-parametric method for estimating probability density functions. In spatial analysis, KDE interpolates the density of point patterns across a continuous surface. We adapt this method to our historical context, using patent locations as point data to estimate the intensity of innovative activity across Europe. The standard KDE formulation at location i with a quartic kernel is expressed as:

$$\mathrm{KDE}_i = \frac{1}{b^2} \sum_{j=1}^{n} \left[\frac{3}{\pi} \cdot A_j \left(1 - \left(\frac{\delta_{ij}}{b} \right)^2 \right)^2 \right], \quad \text{for } \delta_{ij} < b$$

Where:

- b is the bandwidth or search radius
- A_j is a weight associated with observation j
- δ_{ij} is the Euclidean distance between location i and observation j
- n is the total number of observations

Note that the condition $\delta_{ij} < b$ ensures that only observations within the bandwidth contribute to the density estimation.

To incorporate the temporal dimension of the diffusion process, we modify the formulation to include time and condition the estimation on cumulative time windows. The modified equation becomes:

$$\mathrm{KDE}_{i,t}(t \leqslant T) = \frac{1}{b^2} \sum_{j=1}^{n} \left[\frac{3}{\pi} \cdot A_j \cdot I(t_j \leqslant T) \left(1 - \left(\frac{\delta_{ij}}{b} \right)^2 \right)^2 \right], \quad \text{for } \delta_{ij} < b \tag{16}$$

Where:

- t represents the current time point
- T is the upper bound of the cumulative time window
- $I(t_j \leqslant T)$ is an indicator function that equals 1 if the time of observation j (t_j) is less than or equal to T, and 0 otherwise

This formulation estimates the kernel density at location i and time t using only observations up to time T. By varying T, we can create moving cumulative time windows to analyze the diffusion of patenting activity over time. The application of this model to our dataset yields a matrix of location densities across the European space. This approach has two practical advantages for studying historical industrialization. First, it interpolates point data over a continuum of space without relying on predetermined administrative boundaries, useful for 19th-century Europe, where borders frequently changed. Second, the method computes densities within a constant area size, which makes comparisons between regions with varying local administrative divisions more reliable. We set the bandwidth to $b = 250$ km and evaluate the KDE at each NUTS2 centroid using an EPSG:3035 equal-area projection. To determine the patenting intensity threshold, we compute the 75th percentile of the cross-sectional distribution of positive KDE densities in 1850, the earliest decade in which a substantial number of regions exhibit non-zero patenting activity. A region is classified as having crossed the patenting intensity threshold, which we interpret as a proxy for industrial take-off, in the first decade in which its KDE density exceeds this value. Figure 15 shows the resulting diffusion waves across Europe and how high-density patent areas expanded over time. This approach yields a proxy variable for industrialization at the regional level. Regions that never cross the KDE threshold by 1899 are treated as right-censored at 1900. The resulting event definition provides the dependent variable for the Cox proportional hazards model estimated in Section 2.6.7.

The diffusion waves in Figure 15 reveal a clear, spatially dependent pattern of innovation spread: activity originates near the British epicenter and radiates outward with a pronounced south-eastward drift. These stylized facts raise a natural question: what mechanisms drive this wave-like diffusion? In the following section, we develop a gravity-based model of knowledge spillovers to provide a structural explanation for the patterns documented here.

2.4.4 *Descriptive Statistics*

Table 3 presents summary statistics for the main variables used in our econometric analysis. The dataset covers European NUTS2 regions over the decade endpoints 1840–1900 (patents published prior to 1900).

2.5 MODEL AND ESTIMATION

We construct a parsimonious econometric model that draws on the gravity trade literature. The underlying idea is that, similarly to trade, idea flows between two locations depend positively on the population

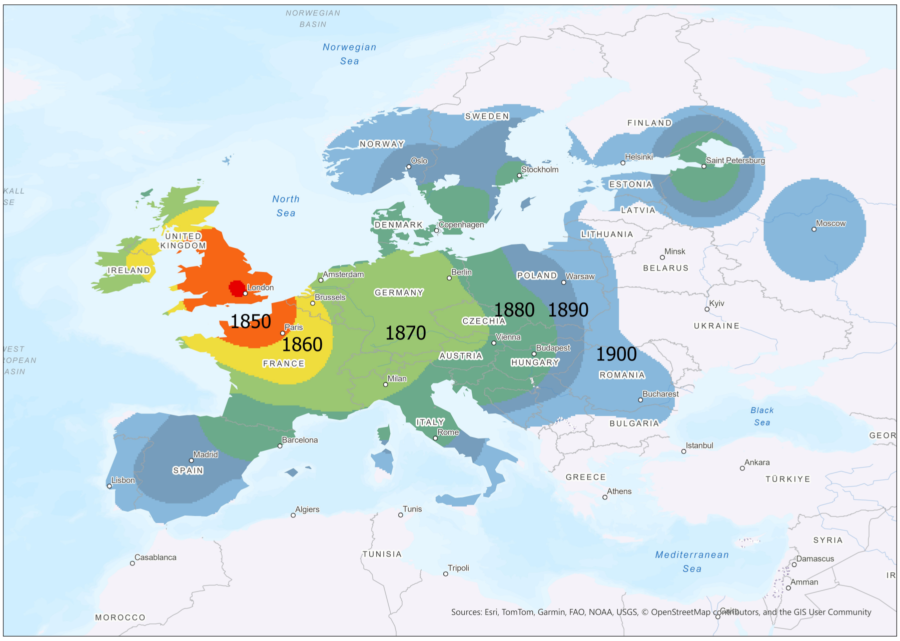


Figure 15: Identification of Patenting Diffusion Waves. The map visualizes the spatial diffusion of inventive activity defined by patent density thresholds; these waves are interpreted as a proxy for the timing of industrial take-off. Regions are classified based on the decade they first crossed the specified patent density threshold; the resulting pattern shows wave-like propagation originating from the industrial core.

Table 3: Descriptive Statistics

Variable	Mean	SD	Min	Max	N
Patents (count)	12.83	74.56	0.00	1,806.00	2,023
Patent Stock (count)	22.13	143.89	0.00	4,231.00	2,023
Population (thousands)	173.81	270.24	1.00	2,953.00	2,023
Spillover Index (billions)	839.96	31,983.24	0.00	1,425,912.21	2,023
Coal Cost (index)	96.63	74.07	3.17	402.42	2,023
Distance from London (km)	1,182.74	564.39	228.20	3,217.55	2,023

size of each location and negatively on the communication cost between them. We argue that this mechanism can explain the European industrialization waves observed in the data, and in particular the "nearby-diffusion" and "jump" diffusion dynamics discussed above.

One of the main challenges of studying economic history is the scarcity of high-resolution data. We do not have data on bilateral trade flows or migration among regions at the subnational level. The model's parsimony reflects both the complexity of the phenomenon and the constraints of the available data.

2.5.1 *The model*

The flow of ideas

Consider a geography composed of N locations, denoted by subscripts. Time is discrete, and at time t, two locations i and j communicate following the gravitational approach.

$$C_{ij,t} = G \cdot \frac{Pop_{i,t}^{\alpha} \cdot Pop_{j,t}^{\beta}}{T_{ij,t}}, \tag{17}$$

where

- $C_{ij,t}$ is the communication intensity between locations i and j at time t
- $Pop_{i,t}$ and $Pop_{j,t}$ are the respective populations of location i and j at time t
- $T_{ij,t}$ is the bilateral communication cost between the two locations at time t
- G, α, and β are parameters

Each period, individuals contribute to the industrial knowledge of their location, which depends on previous cumulated knowledge, current population, and ideas arriving from other locations through bilateral connections.

The knowledge generation process

Formally, the industrial knowledge of location i in each period is a function of past cumulated knowledge, current population, knowledge received through interactions with other locations, and technological input prices.

$$ik_{i,t} = IK_{i,t-1}^{\gamma} Pop_{i,t}^{\eta} \left(\sum_{j \neq i} C_{ij,t} IK_{j,t-1} \right)^{\varepsilon} \cdot IP^{-\theta} \tag{18}$$

- $ik_{i,t}$ is the industrial knowledge of location i at time t

- $IK_{i,t-1}$ is the cumulated industrial knowledge of location i until the previous period
- $\sum_{j \neq i} C_{ij,t} IK_{j,t-1}$ is the knowledge gained from interactions with other locations
- IP is the input price index
- γ, η, ε, and θ are parameters

The interaction component is stronger for nearby and populated locations than for small and distant ones, a parsimonious formulation of spatial spillovers that can account for the diffusion patterns observed in the data.

Substituting equation (17) into (18), we obtain a single knowledge creation function:

$$ik_{i,t} = IK^{\gamma}_{i,t-1} Pop^{\eta}_{i,t} \left(\sum_{j \neq i} G \cdot \frac{Pop^{\alpha}_{i,t} \cdot Pop^{\beta}_{j,t}}{T_{ij,t}} IK_{j,t-1} \right)^{\varepsilon} \cdot IP^{-\theta} \quad (19)$$

2.5.2 *Adjusting the model to the data*

We apply the model to the European continent at the NUTS2 level. Implementing the model requires addressing several unobservable quantities. First, we proxy innovation and knowledge with patents. Second, we parameterize bilateral communication costs using a digitization of main roads from a widely used 18th-century map. Third, given the central role of coal in the first Industrial Revolution, we use proximity to known coal mines as a measure of input price access.

Proxying knowledge and innovation

Industrial knowledge ($ik_{i,t}$) and cumulated industrial knowledge ($IK_{i,t}$) are not directly observable. We proxy them using patent data. We define cumulated patents as

$$CP_{i,t} = \sum_{\tau=0}^{t} P_{i,\tau} \quad (20)$$

and assume a monotonic relationship between cumulated patents and cumulated industrial knowledge: $IK_{i,t} \approx f(CP_{i,t})$, where f is increasing. Under a power-law specification $IK_{i,t} = \delta \cdot CP^{\phi}_{i,t}$, this implies that the structural spillover in equation (19) becomes

$$S_{i,t}(\alpha, \beta, \phi) = \sum_{j \neq i} \frac{Pop^{\alpha}_{i,t} \cdot Pop^{\beta}_{j,t} \cdot CP^{\phi}_{j,t-1}}{T_{ij}} \quad (21)$$

Figure 16: *Map of Europe - Drawn from all the best surveys and rectified by astronomical observation by* Arrowsmith (1798)

Estimating transport costs

Although much of the literature uses Euclidean distance, a more precise measure that accounts for existing roads and waterways better captures the actual communication costs faced by 19th-century agents. To compute travel time we create a dataset of roads and waterways using multiple sources of information. For roads, we use the *Map of Europe - Drawn from all the best surveys and rectified by astronomical observation by* Arrowsmith (1798).

According to Bricker and Tooley (1989) (page 98) "The Arrowsmith output was prodigious: clarity and accuracy were his aims — and he achieved them as no English cartographer was able to before him. His maps are still essential to historians who want to delve into the history of 18th-century exploration in the Pacific."" The Arrowsmith map is one of the most spatially comprehensive cartographic sources for 18th-century European road networks available for our period of analysis. A second source of travel routes is waterways. While maritime and river transport offered the lowest friction for bulk goods, they were not necessarily the fastest for information or passengers before steam. However, they remained essential for heavy industrial inputs. As industrialization progressed, many canals were constructed and rivers were canalized. To avoid endogeneity, we restrict attention to rivers that were navigable prior to industrialization, excluding canalized or newly constructed waterways. We construct a GIS dataset of navigable waterways using multiple sources. Finally, although data on existing railways during this period are available (Bogart, 2014), we exclude them to avoid endogeneity concerns.

The travel cost computation rests on two assumptions:

1. Travel cost is symmetric: travelling from location i to j costs the same as travelling from j to i.

2. Travel cost is constant over the entire period of analysis (1835–1899). This is the strongest assumption, as railway construction and canal improvements substantially reduced travel costs during this period. We impose this restriction to avoid endogeneity concerns (railways were themselves a product of industrialization) but acknowledge it as a limitation to be addressed in future work with time-varying transport networks.

In a rich geography problem, given a matrix of instantaneous speed representing the existing travel network, computing the travel cost between locations i and j requires finding the shortest path connecting them. We implement this using a Dijkstra shortest-path algorithm on a 10 km resolution grid covering continental Europe. We assign base speeds of 4.5 km/h for main roads, 8.0 km/h for sea, 6.0 km/h for navigable waterways, and 2.0 km/h for off-road terrain. These base speeds are then adjusted for terrain slope using the Tobler hiking function, which modulates walking speed as $v(s) = 6\exp(-3.5|s + 0.05|)$ km/h, where s is the slope gradient. For each grid cell, we compute a harmonic-mean Tobler multiplier using both uphill and downhill directions to obtain a direction-independent friction surface. Sea-to-land transitions incur an additional overhead of 24 hours to capture port handling costs. The resulting network yields travel times in hours between all NUTS2 centroid pairs. These speed parameters are informed by the historical transport literature. The off-road speed of 2 km/h reflects the effective pace of goods transport over unimproved terrain: while unloaded walkers reach 5 km/h on flat ground (Tobler, 1993), loaded ox-drawn carts averaged roughly 2–3 km/h on rough surfaces (Rodrigue, 2020). The on-road speed of 4.5 km/h captures a blend of freight haulage (3–4 km/h for horse-drawn wagons) and individual travel, consistent with the observation that pre-turnpike stagecoaches averaged 2–3 km/h and improved to about 6 km/h by the 1830s (Rodrigue, 2020). For comparison, T. Allen (2023) assigns 5 km/h off-road and reduces travel time by 40% on Roman roads following Davey, Hayes, and Norman (1994), while Özak (2018) uses the Tobler function directly. The sea speed of 8 km/h (≈4.3 knots) falls in the lower range of sailing speeds documented in the CLIWOC logbook data for the period 1750–1850 (Kelly and Ó Gráda, 2019). The waterway speed of 6 km/h represents an average over upstream and downstream navigation on major European rivers. The slope adjustment follows the Tobler hiking function (Tobler, 1993), which is widely adopted in spatial cost analysis (T. Allen, 2023; Özak, 2018).

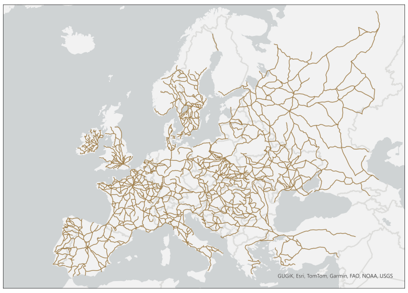


Figure 17: Digitized roads

Figure 18: Digitized navigable waterways

Figure 19: Historical transport network used to construct the friction surface. Roads are digitized from the Arrowsmith (1798) map; waterways include only rivers documented as navigable prior to industrialization.

Input prices and Proximity to coal mines

Transport costs were a major factor influencing coal prices in Europe. For instance, Licio (2022) reports that in 1870, French coal delivered to Turin had a total price of 35 Italian Lire per ton, with transport costs accounting for a significant 20 Lire, more than half the total price. The impact of transport costs is even more pronounced when considering British coal. Due to the additional maritime shipping expenses, British coal reached a final price of 55 Lire per ton in Turin. Of this total, 43 Lire was attributed to transportation: 28 Lire for sea shipping from Cardiff to Genoa, and an additional 15 Lire for rail transport from Genoa to Turin. These examples show that transport costs, both maritime and overland, determined a large share of coal's final delivered price across European locations. With some notable exceptions, such as Licio (2022) for Italy, no high-resolution harmonized coal price dataset exists for all of Europe. Given the importance of transport costs in determining coal prices, we proxy input costs using proximity to coal mines. For each region i, we compute the minimum travel cost to any coalfield in the sample:

$$IP_i = \min_{m \in \mathcal{M}} TC_{im} \tag{22}$$

where $\mathcal{M}$ is the set of coalfield polygons identified in the data. This specification captures the effective input cost faced by a representative firm: producers source coal from the nearest accessible deposit, so the relevant price is the minimum across all mines. As a robustness check, we replace IP_i with the mean travel time across all coalfields; Section 2.6.6 shows that the results are insensitive to this choice.

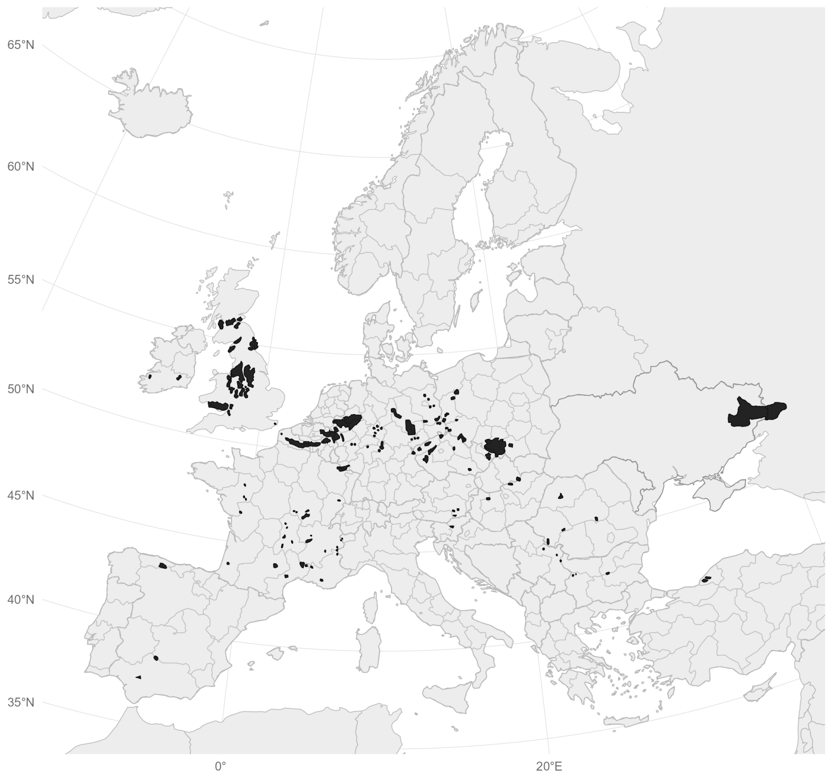


Figure 20: Major coalfields in Europe. Source: author's elaborations from Fernihough and O'Rourke (2021) data. Shapefile from Porto Tapiquén (2015).

Population estimates

We estimate population at the NUTS2 level using a dataset of European city populations compiled by Buringh (2021). For our period of analysis, this dataset provides estimates at three benchmark dates: 1800, 1850, and 1900. To compute regional population aggregates, we sum the city populations within each NUTS2 boundary. While this approach captures only urban population, it provides a consistent measure across regions and time periods.

2.5.3 *Estimation strategy*

From theory to estimating equation

The knowledge generation function in equation (19) cannot be estimated directly because its dependent variable, industrial knowledge ($ik_{i,t}$), is unobservable. However, under the assumption that patent production is proportional to knowledge generation, $P_{i,t} \propto ik_{i,t}$, we can substitute the structural spillover (21) into equation (19) and proxy IK with CP to obtain an estimable reduced-form patent equation:

$$\begin{aligned} P_{i,t} = \exp\big(&c_0 + c_1 \ln CP_{i,t-1} + c_2 \ln Pop_{i,t} \\ &+ c_3 \ln S_{i,t}(\alpha, \beta, \varphi) + c_4 \ln IP_i + \mu_c + \delta_t\big) \cdot \eta_{i,t} \end{aligned} \tag{23}$$

where μ_c and δ_t are country and time fixed effects, and the reduced-form coefficients are composites of the underlying structural parameters: for instance, c_1 absorbs both the persistence of knowledge (γ) and the patent–knowledge elasticity, while c_3 captures the overall strength of spatial spillovers (ε). Country fixed effects absorb time-invariant institutional and geographical characteristics, while time fixed effects capture common shocks such as continent-wide business cycles or technological paradigm shifts.

Identification of structural spillover parameters

While the composite coefficients $c_0, \ldots, c_4$ do not allow the recovery of all individual structural parameters, the three parameters governing the *shape* of the spillover function—α (destination population elasticity), β (origin population elasticity), and φ (knowledge stock elasticity)—enter the model nonlinearly through the spillover index in equation (21) and can be estimated by concentrated least squares.

Two-phase estimation

We employ a two-phase estimation strategy:

1. **Reduced-form estimation.** We first estimate equation (23) with a pre-specified linear spillover index ($\alpha = 0$, $\beta = 1$, $\phi = 1$), yielding $S_{i,t} = \sum_{j \neq i} \text{Pop}_{j,t} \cdot \text{CP}_{j,t-1}/T_{ij}$. This provides baseline estimates of the reduced-form coefficients using both OLS and Poisson Pseudo-Maximum Likelihood (PPML) with country and year fixed effects, following the recommendation of Santos Silva and Tenreyro (2006) for gravity equations with heteroskedastic errors and zero-valued dependent variables.

2. **Structural estimation via concentrated least squares.** We then relax the restrictions on (α, β, ϕ) and recover their values by concentrated least squares. The procedure nests two optimization loops: the outer loop searches over (α, β, ϕ) using the L-BFGS-B algorithm, while for each candidate triplet the inner loop recomputes the spillover $S_{i,t}(\alpha, \beta, \phi)$ using the full bilateral travel cost matrix and re-estimates the linear coefficients by OLS. The outer loop minimizes the sum of squared residuals. We initialize the search with a grid over a coarse parameter space to mitigate the risk of converging to local minima.

Finally, we complement these intensity models with a Cox proportional hazards model for the *timing* of industrialization, using time-invariant regional covariates. Section 2.6.8 reports targeted sensitivity checks.

2.6 EMPIRICAL RESULTS

We adopt a progressive estimation strategy. We begin with reduced-form and structural gravity estimates of patenting intensity, then assess identification and robustness. We then complement these intensity results with a Cox proportional hazards model for the timing of industrialization.

Our dataset consists of a balanced panel of 296 NUTS2 regions observed at decadal frequency from 1840 to 1900, where years denote decade endpoints (e.g., 1900 corresponds to the 1890s). After removing observations with missing population or coal cost data, the estimation sample comprises 2,023 region-decade observations.

2.6.1 *Reduced-Form Gravity Model*

We begin by estimating equation (23) under the baseline restrictions $\alpha = 0$, $\beta = 1$, $\phi = 1$, which yield the linear spillover index $S_{i,t} = \sum_{j \neq i} \text{Pop}_{j,t} \cdot \text{CP}_{j,t-1}/T_{ij}$. We estimate this specification using both OLS on the log-linearized form and Poisson Pseudo-Maximum Likelihood (PPML) with country and year fixed effects, following the recommendation of Santos Silva and Tenreyro (2006) for gravity equations with heteroskedastic errors and zero-valued observations.

Table 4: Gravity Model Results (OLS & PPML with Fixed Effects)

Dependent Variables:	log_patents	patents
Model:	(1)	(2)
	OLS	Poisson
Variables		
Lagged Patent Stock (log)	0.841***	0.785***
	(0.015)	(0.009)
Population (log)	0.144***	0.323***
	(0.013)	(0.012)
Spillover index (log)	0.074***	0.020***
	(0.022)	(0.005)
Coal cost (log)	-0.018	-0.070***
	(0.019)	(0.011)
Fixed-effects		
country_id	Yes	Yes
year	Yes	Yes
Fit statistics		
Observations	2,023	1,974
Squared Correlation	0.85790	0.94597
Pseudo R^2	0.56545	0.93073
BIC	3,383.6	9,748.9

IID standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 4 presents the results. The PPML estimates (Model 2) show that all four covariates are highly significant. The elasticity of patents to the lagged knowledge stock is 0.79, confirming the cumulative nature of the innovation process. Population enters with an elasticity of 0.32, implying diminishing returns to scale in knowledge production. The spillover index is positive and significant (+0.02, $p < 0.01$), consistent with spatial knowledge diffusion. Coal access carries a negative coefficient (−0.07, $p < 0.01$), consistent with the hypothesis that proximity to energy sources was a determinant of innovation capacity during the Industrial Revolution.

2.6.2 *Structural Estimation*

The reduced-form specification imposes restrictive assumptions on the structure of knowledge spillovers. In particular, it constrains the elasticities of origin population (α), destination population (β), and past knowledge (ϕ) to fixed values that may not correspond to the true data-generating process described in equation (19).

To recover these structural parameters, we employ a concentrated least squares approach. The reduced form of our three-equation system (after substituting equation (18) into the patent production equation) yields

$$\begin{aligned} \ln P_{i,t} = c_0 + c_1 \ln CP_{i,t-1} + c_2 \ln Pop_{i,t} \\ + c_3 \ln S_{i,t}(\alpha, \beta, \phi) + c_4 \ln IP_i + \mu_c + \delta_t + \epsilon_{i,t} \end{aligned} \tag{24}$$

where the structural spillover is now

$$S_{i,t}(\alpha, \beta, \phi) = \sum_{j \neq i} \frac{Pop_{i,t}^{\alpha} \cdot Pop_{j,t}^{\beta} \cdot CP_{j,t-1}^{\phi}}{T_{ij}} \tag{25}$$

and the reduced-form coefficients $c_0, \ldots, c_4$ are composites of the underlying structural parameters $(\gamma, \eta, \varepsilon, \theta, \lambda, \mu, \omega, \delta)$. The nonlinear parameters (α, β, ϕ) are, however, separately identified because they enter the model inside the summation operator.

The concentrated least squares procedure works by nesting two optimization loops. In the outer loop, we search over the structural parameters (α, β, ϕ) using the L-BFGS-B algorithm. For each candidate triplet, we recompute the spillover index $S_{i,t}(\alpha, \beta, \phi)$ using the full bilateral travel cost matrix and then estimate the linear parameters $(c_0, \ldots, c_4)$ by OLS in an inner loop. The outer loop minimizes the sum of squared OLS residuals. We initialize the optimization using a grid search over a coarse parameter space to avoid local minima. As a robustness check, we also report PPML estimates of the reduced-form patent equation evaluated at the spillover index implied by the estimated structural parameters.

	OLS (log patents)	PPML (patents)
Lagged Patent Stock (log)	0.84***	0.80***
	(0.02)	(0.01)
Population (log)	0.17***	0.29***
	(0.02)	(0.01)
Structural spillover index (log)	0.47***	−0.11***
	(0.08)	(0.04)
Coal cost (log)	0.02	−0.11***
	(0.02)	(0.01)
Num. obs.	1463	1414
Num. groups: country_id	35	30
Num. groups: year	7	7
R^2 (full model)	0.88	–
R^2 (proj model)	0.71	–
Adj. R^2 (full model)	0.88	–
Adj. R^2 (proj model)	0.71	–
Deviance	–	5437.36
Log Likelihood	–	−3925.20
Pseudo R^2	–	0.93

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5: Structural Gravity Model (Concentrated Least Squares)

The structural estimation requires computing the bilateral spillover for every region pair, which restricts the sample to the subset of regions present in both the panel dataset and the travel cost matrix. Regions for which the structural spillover evaluates to zero (typically peripheral regions with no patenting neighbors within reach) are also excluded because the log transformation is undefined. These restrictions reduce the estimation sample relative to the reduced-form specification.

Table 5 reports the estimates of equation (24) evaluated at the structural spillover index $S_{i,t}(\hat{\alpha}, \hat{\beta}, \hat{\phi})$. Column (1) reports OLS estimates of the log-linear specification (with country and year fixed effects), while Column (2) reports PPML estimates of the corresponding count specification, evaluated at the same spillover index.

The concentrated least squares outer loop yields estimates of the spillover-shape parameters in equation (25): $\hat{\alpha} \approx 0$, $\hat{\beta} = 1.81$ (SE $=$ 0.51), and $\hat{\phi} = 0.25$ (SE $= 0.11$), with standard errors computed from the numerical Hessian of the concentrated objective function.[1] These parameters are not regression coefficients: they enter only through the construction of $S_{i,t}(\alpha, \beta, \phi)$. An F-test for $H_0 : \alpha = 0$ yields $p > 0.99$, confirming that the restriction is not binding. The structural estimation sample drops to 1,463 (OLS) and 1,414 (PPML) observations because regions must be present in the travel cost matrix and must have positive structural spillover values.

Three aspects of these spillover-shape estimates have direct economic interpretation.

First, $\hat{\alpha} \approx 0$ indicates that the population of the receiving region does not amplify the incoming knowledge spillover. In the context of our model, this means that even small regions could benefit from proximity to innovative hubs, suggesting that geography, rather than market size, determined the ability to absorb external knowledge during this period.

Second, $\hat{\beta} = 1.81$ ($t = 3.52$, $p < 0.01$) is significantly different from zero, confirming that the population of the source region matters for knowledge generation. The point estimate substantially exceeds unity, suggesting increasing returns to scale in the population of the source region, consistent with strong agglomeration economies in innovation (Jaffe, Trajtenberg, and Henderson, 1993).

Third, $\hat{\phi} = 0.25$ ($t = 2.19$, $p < 0.05$) reveals strongly diminishing returns to accumulated knowledge in the spillover process. This finding is consistent with the "fishing out" hypothesis: as the stock of knowledge grows, the marginal contribution of each additional patent to cross-regional spillovers decreases. The first innovations in a region have a higher relative impact on neighboring areas than subsequent ones.

The estimated spillover elasticity $\hat{c}_3$ (coefficient on the structural spillover index) diverges between OLS ($+0.47$, $p < 0.01$) and PPML (-0.11, $p < 0.01$). This sign reversal in the PPML specification may arise because the structural formulation, with its strongly increasing returns to source population ($\hat{\beta} = 1.81$), concentrates spillovers among a few large urban centers; the PPML estimator, which places greater weight on observations with large counts, may then attribute the remaining variation to competitive effects rather than diffusion. The OLS estimator, which weights all observations equally in logs, recovers the positive diffusion channel. This pattern is consistent with a model where spillovers operate at the extensive margin (enabling initial industrialization) rather than the intensive margin (driving further patent growth in already-innovative regions).

1 Block bootstrap standard errors clustered by country (200 replications) yield larger SEs ($SE(\hat{\beta}) = 1.16$, $SE(\hat{\phi}) = 0.63$), as expected when accounting for cross-country heterogeneity.

2.6.3 *Identification and Endogeneity*

A potential concern with gravity models of knowledge diffusion is the endogeneity of the main explanatory variables. In particular, we address three specific threats to identification: (1) reverse causality between population growth and innovation, (2) dynamic panel bias (Nickell bias) arising from the lagged dependent variable, and (3) unobserved path dependence where initial advantages drive long-run outcomes.

	Baseline	No CP lag	Pop 1800	No CP+Pop1800	InitCP (1850)	InitCP+Pop1800
	(1)	(2)	(3)	(4)	(5)	(6)
Lagged Patent Stock (log)	0.785***		0.895***			
	(0.017)		(0.015)			
Population (log)	0.323***	1.275***			0.804***	
	(0.023)	(0.030)			(0.021)	
Spillover Index (log)	0.020**	−0.037**	0.034***	−0.011	−0.031***	−0.006
	(0.009)	(0.018)	(0.009)	(0.023)	(0.011)	(0.012)
Coal Cost (log)	−0.070***	0.138***	−0.119***	−0.008	−0.141***	−0.283***
	(0.022)	(0.038)	(0.023)	(0.046)	(0.026)	(0.028)
Population 1800 (log)			0.192***	1.330***		0.755***
			(0.025)	(0.049)		(0.027)
Initial CP (log)					0.740***	0.963***
					(0.025)	(0.027)
Observations	2023	2023	2023	2023	2023	2023

Table 6: Endogeneity Robustness Checks

Table 6 reports PPML estimates of the baseline gravity patent equation under alternative specifications (all with country and year fixed effects). Column (1) reproduces the baseline model. Column (2) drops the lagged patent stock. Column (3) replaces contemporaneous population with population in 1800. Column (4) combines Columns (2) and (3). Column (5) controls for initial patent stock (cumulated patents up to 1850). Column (6) combines initial patent stock and population in 1800.

First, the positive relationship between population and patenting in our baseline model (Column 1: $\beta_{pop} = 0.323$) could reflect reverse causality: successful industrial clusters attract migration, inflating the population of innovative regions. To gauge this concern, we replace the contemporaneous population with its predetermined 1800 value (Columns 3 and 4). The distribution of European population in 1800 predates the major wave of industrialization and railway construction; it is therefore plausibly free from contamination by late-19th-century innovation shocks and remains highly predictive of subsequent settlement patterns. We note that this approach does not constitute a formal instrumental variable strategy (it replaces a potentially endogenous regressor with a predetermined control) and therefore does not fully resolve reverse causality. Nevertheless, it provides a useful diagnostic. Table 6 (Column 3) shows that when using population in 1800, the spillover coefficient increases to $+0.034$ ($p < 0.01$) and remains highly significant (on the same 2,023 observations). This suggests that our baseline findings are not driven by endogenous migration; if anything, the baseline estimate appears conservative.

Second, we address the concern of Nickell bias. In a fixed-effects model with a lagged dependent variable (knowledge stock), the estimates can be biased in finite samples. Column 2 (R1) drops the lagged stock entirely; this removes the Nickell bias but also omits the region's own absorptive capacity. The results show that without controlling for the region's own history, the spillover coefficient turns negative (Column 2: -0.037, $p < 0.05$). This indicates that the positive spillover effect is conditional on a region's accumulated capabilities: among regions with similar populations, those surrounded by innovation but lacking their own knowledge stock tend to patent *less*, possibly due to competition or agglomeration shadows. The positive diffusion channel ($+0.02$) is only revealed when we account for the region's own dynamic path; the baseline result is therefore not a mechanical artifact of the lagged dependent variable but a conditional economic relationship.

Third, we explore the role of path dependence. A critique of spatial diffusion models is that they may simply capture the persistence of initial advantages: regions that were innovative or well-connected early on continued to be so by the end of the sample. To test this, we explicitly control for the initial patent stock accumulated up to 1850 (Columns 5 and 6). When controlling for initial conditions, the spillover coefficient turns negative (Column 5: -0.031, $p < 0.01$) and becomes small and statistically insignificant once we also use predetermined population (Column 6: -0.006). This reveals that part of the aggregate "diffusion" observed in the baseline model reflects the amplification of early locational advantages. Once we condition on the "starting line," the residual spatial interaction appears weaker and potentially competitive or displacing.

2.6.4 *Robustness: Institutional Controls*

A potential concern with the baseline gravity specification is the omission of institutional quality. If countries with stronger democratic institutions also happen to be located near the British epicenter, the spillover coefficient could be capturing institutional rather than geographic effects.

To address this concern, we augment the gravity model with a measure of institutional quality from the Polity5 dataset (Marshall, Gurr, and Jaggers, 2019). Specifically, we use the *Institutionalized Democracy* score (`democ`), which ranges from 0 (no democratic features) to 10 (full democracy) and captures the competitiveness of political participation, the openness and competitiveness of executive recruitment, and the constraints on executive authority. This composite score is available at the country level for all major European states—including preunification polities such as Prussia, Sardinia, and the Two Sicilies—throughout our period of analysis. We log-transform the variable as log(democ + 1) and merge it into our panel at the decadal level.

Table 7: Institutional Controls

Dependent Variable:		patents	
Model:	(1)	(2)	(3)
Variables			
Lagged Patent Stock (log)	0.7870***	0.7911***	0.8966***
	(0.0086)	(0.0087)	(0.0054)
Population (log)	0.3215***	0.3169***	0.1483***
	(0.0119)	(0.0120)	(0.0068)
Spillover Index (log)	0.0203***	0.0214***	0.0850***
	(0.0046)	(0.0046)	(0.0041)
Coal Cost (log)	-0.0705***	-0.0712***	-0.2222***
	(0.0111)	(0.0111)	(0.0086)
Democracy Score (log)		-0.1709***	-0.0774***
		(0.0343)	(0.0138)
Fixed-effects			
country_id	Yes	Yes	
year	Yes	Yes	Yes
Fit statistics			
Observations	1,837	1,837	1,844
Squared Correlation	0.94605	0.94492	0.92300
Pseudo R^2	0.93028	0.93046	0.90874
BIC	9,518.2	9,500.7	12,221.8

IID standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 7 presents three specifications. The first column reproduces the baseline PPML estimates on the sample with non-missing institutional data ($n = 1{,}837$). The second column adds the democracy score as an additional covariate while retaining country and year fixed effects; the third column removes country fixed effects entirely, so that institutions absorb the cross-country variation previously captured by the fixed effects.

The democracy score enters with a negative and significant coefficient in both specifications: -0.17 ($p < 0.01$) with country fixed effects and −0.08 ($p < 0.01$) without. In the specification with country fixed effects (Column 2), identification comes entirely from within-country variation over time: the coefficient captures whether decades of expanding democratic participation *within* a given country coincided with higher patenting, conditional on knowledge stocks, population, and spillover exposure. The negative estimate indicates that they did not. The major waves of constitutional and franchise reform in 19th-century Europe—the revolutions of 1848, the Reform Acts in Britain, and the unification-era constitutional changes in Germany and Italy—were concentrated in specific decades that do not coincide with the steepest increases in patent intensity; if anything, periods of political upheaval may have temporarily disrupted the stable institutional environment that patenting requires. In the specification without country fixed effects (Column 3), the coefficient shrinks to −0.08, as the negative within-country gradient is partially offset by cross-country variation: some of the least democratic states (Russia, the Ottoman Empire) also had the lowest knowledge stocks and spillover exposure, attenuating the negative relationship in levels.

The coefficients on our variables of interest—lagged patent stock, population, spillover index, and coal cost—remain virtually unchanged across all three specifications. The spillover coefficient moves from +0.020 to +0.021 when institutional controls are added. This stability confirms that the spatial diffusion mechanism captured by our gravity model is robust to the inclusion of institutional quality and is not an artifact of omitted institutional heterogeneity.

2.6.5 *Robustness: Spatial Autocorrelation*

A natural concern with any spatial diffusion model is that the spillover coefficient may reflect residual spatial autocorrelation rather than a genuine causal mechanism. In our setting, the spillover index is by construction a spatially weighted variable; we therefore test whether its significance is inflated by correlated residuals across neighboring regions.

We address this concern through two complementary diagnostics. First, we compute Moran's I statistics on the PPML residuals for each decade, using k-nearest-neighbor spatial weights ($k = 5$). The results reveal modest but significant residual spatial autocorrelation in the decades of most rapid industrial diffusion (1850, 1870, 1880), while the early and late periods show no significant spatial pattern. This pattern is consistent with a correctly specified diffusion model: the spatial structure is largely absorbed by the spillover variable, but some residual clustering persists during the peak of the diffusion wave.

Second, we re-estimate standard errors using Conley's (**conley1999gmm**) spatial HAC estimator (500 km cutoff) and compare them with stan-

	IID SE	Conley SE (500km)	Clustered SE
Lagged Patent Stock (log)	0.7911***	0.7911***	0.7911***
	(0.0087)	(0.0562)	(0.0565)
Population (log)	0.3169***	0.3169***	0.3169***
	(0.0119)	(0.0843)	(0.0813)
Spillover Index (log)	0.0214***	0.0214	0.0214
	(0.0046)	(0.0161)	(0.0133)
Coal Cost (log)	−0.0712***	−0.0712*	−0.0712*
	(0.0111)	(0.0398)	(0.0370)
Democracy Score (log)	−0.1709***	−0.1709	−0.1709
	(0.0343)	(0.2862)	(0.2947)
Num. obs.	1837	1837	1837
Num. groups: country_id	25	25	25
Num. groups: year	7	7	7
Deviance	6317.7827	6317.7827	6317.7827
Log Likelihood	−4615.0683	−4615.0683	−4615.0683
Pseudo R^2	0.9299	0.9299	0.9299

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 8: Gravity Model: Standard Errors Comparison

dard errors clustered by country. Table 8 presents the results: coefficients are stable, while the spillover term remains positive but becomes statistically insignificant once spatial or country-level dependence is accounted for.

This attenuation is expected and we believe it does not invalidate the spillover mechanism. As noted by Head and Mayer (2014), in gravity models of trade and knowledge flows, the spatial structure is the phenomenon under study, not a nuisance parameter. Conley standard errors penalize precisely the spatial correlation that the diffusion variable is designed to capture; they therefore provide a conservative lower bound on significance. The fact that the coefficient remains positive and economically meaningful (+0.021) across all specifications, despite losing statistical significance under the most conservative corrections, suggests that geographic proximity to innovative centers contributed to patenting activity, though the evidence is not robust to spatial clustering adjustments.

2.6.6 *Robustness: Coal Cost Specification*

Our baseline proxies input costs with the minimum travel time from each NUTS2 centroid to the nearest coalfield (equation (22)). An alternative is to use the mean travel time across all coalfields, which weights the cost index by the full spatial distribution of coal deposits rather than by the single nearest source. While the minimum is the more natural economic concept (firms optimise over their sourcing options), the mean provides a useful check on whether results are sensitive to the tails of the coal geography (e.g., a few outlying mines that happen to be very close to some regions).

Table 9 compares two coal-access measures: the baseline minimum travel time to the nearest coalfield (Columns 1–2) versus the mean travel time across all coalfields (Columns 3–4). The coefficients on the lagged patent stock, population, and the spillover index are stable across all four columns. For coal access, the PPML estimates are negative and statistically significant under both measures: -0.070 ($p < 0.01$) with minimum travel time and -0.348 ($p < 0.01$) with mean travel time. The OLS coefficient under minimum travel time is small and statistically insignificant (-0.018), while the OLS coefficient under mean travel time is negative and marginally significant (-0.149, $p < 0.10$). Since PPML is our preferred estimator for count data with many zeros, as recommended by Santos Silva and Tenreyro (2006), the consistent sign and significance in Columns 2 and 4 support the robustness of the input-cost channel. The larger elasticity under the mean measure reflects its broader dispersion: minimum travel time is mechanically compressed toward zero for regions located near a coalfield, while mean travel time retains more cross-sectional variation. Overall, the mean-based specification is a stronger version of the same qualitative result: regions with higher coal access costs tended to patent less, regardless of how coal access is measured.

2.6.7 *Cox Hazard Model: Timing of Industrialization*

While the gravity models capture the determinants of innovation *intensity*, they do not directly address the question of *when* a region crosses the patenting intensity threshold that we interpret as industrial take-off. To study the timing of industrial take-off, we define the industrialization event using the KDE-derived patent density threshold described in Section 2.4.3: a region is considered to have industrialized in the first decade in which its patent density exceeds the threshold value, and regions that never cross the threshold by 1899 are right-censored at 1900. We then estimate a Cox proportional hazards model of the form

$$h(t \mid \mathbf{x}_i) = h_0(t) \exp(\boldsymbol{\gamma}' \mathbf{x}_i) \tag{26}$$

where $h(t \mid \mathbf{x}_i)$ is the hazard rate of industrialization for region i, $h_0(t)$ is the baseline hazard, and $\mathbf{x}_i$ is a vector of time-invariant regional characteristics including distance from London, a circular encoding of bearing from London (sine and cosine), mean population,

Table 9: Robustness: Coal Cost Specification. Columns 1–2 use minimum travel time to the nearest coalfield (baseline); Columns 3–4 replace it with mean travel time across all coalfields.

Dependent Variables:	log_patents	patents	log_patents	patents
Model:	(1)	(2)	(3)	(4)
	OLS	Poisson	OLS	Poisson
Variables				
Lagged patent stock (log)	0.841***	0.785***	0.840***	0.765***
	(0.015)	(0.009)	(0.015)	(0.009)
Population (log)	0.144***	0.323***	0.141***	0.340***
	(0.013)	(0.012)	(0.013)	(0.011)
Spillover index (log)	0.074***	0.020***	0.072***	0.026***
	(0.022)	(0.005)	(0.021)	(0.004)
Coal cost: min distance (log)	-0.018	-0.070***		
	(0.019)	(0.011)		
Coal cost: mean distance (log)			-0.149*	-0.348***
			(0.090)	(0.067)
Fixed-effects				
country_id	Yes	Yes	Yes	Yes
year	Yes	Yes	Yes	Yes
Fit statistics				
Observations	2,023	1,974	2,023	1,974
Squared Correlation	0.85790	0.94597	0.85803	0.94473
Pseudo R^2	0.56545	0.93073	0.56572	0.93065
BIC	3,383.6	9,748.9	3,381.8	9,761.0

IID standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

	Estimate
Distance from London (log)	-2.786^{***}
	(0.187)
Bearing from London (sin)	0.336^{*}
	(0.198)
Bearing from London (cos)	-0.538^{***}
	(0.183)
Mean population (log)	0.110^{*}
	(0.064)
Coal cost (log)	-0.197^{**}
	(0.088)
AIC	1858.99
R^2	0.72
Num. events	228
Num. obs.	249
PH test (global)	0.002

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 10: Cox Hazard Model Results

and coal access cost. We measure distance from London in two ways: straight-line (Euclidean) distance and friction-surface travel time computed from the historical transport network constructed above. The latter measure accounts for the fact that regions at the same Euclidean distance from London could face very different effective communication costs depending on their access to roads and navigable waterways; this distinction is particularly relevant for understanding the south-eastward trajectory of diffusion documented above.

Table 10 presents the results using Euclidean distance as a starting point. Distance from London is the dominant predictor of delayed industrialization: the coefficient of −2.79 ($p < 0.01$) implies that doubling the distance from London reduces the hazard of industrialization by roughly 86%. To respect the circular nature of direction, bearing from London enters through its sine and cosine; these directional terms are jointly significant ($p < 0.01$), consistent with a south-eastward component of diffusion. Mean population is marginally significant ($p < 0.10$), while coal access remains negatively associated with industrialization timing (−0.20, $p < 0.05$).

Crucially, when we replace Euclidean distance with friction-surface travel time, the proximity effect remains the dominant predictor. This is a stronger result: travel time captures the actual communication costs faced by 19th-century agents—shaped by existing roads, navigable rivers, and terrain—rather than mere geographic position. The fact that the south-eastward diffusion pattern holds under this measure suggests that the historical transport network, which connected the British core to northern France, Belgium, and the Rhineland more efficiently than to equidistant regions in other directions, actively channeled the spread of inventive knowledge.

The Cox model relies on the proportional hazards (PH) assumption, i.e., covariate effects are time-invariant and hazard ratios are constant over time. The global proportional hazards test yields $p < 0.01$, suggesting that the PH assumption is violated overall. Covariate-specific tests suggest that non-proportionality is driven primarily by the directional component and coal access, while the distance effect itself satisfies the PH test. Section 2.6.8 reports a targeted sensitivity check that addresses this issue.

2.6.8 *Robustness: Cox Specification*

The baseline Cox model (Table 10) yields a global PH-test p-value of 0.002, indicating some departure from proportional hazards. To verify that our key result is not an artifact of this violation, we re-estimate the Cox model allowing time-varying coefficients for the covariates that fail the covariate-specific PH tests (the directional component and coal access), by interacting them with a centered $\log(t)$ term. Table 11 reports the resulting specification. The proximity effect remains large and precisely estimated, supporting the interpretation that distance from the British epicenter is the most robust predictor of industrialization timing in our data, while some secondary covariates exhibit modest time variation in their effects.

2.7 DISCUSSION AND LIMITATIONS

The gravity model, the Cox hazard model, and the robustness checks all point to geographic proximity as the primary driver of regional innovation intensity and timing. The gravity model confirms that innovation was not an isolated, region-specific phenomenon but was shaped by the interaction between local knowledge stocks and cross-regional spillovers. Structural estimation refines this further: spillovers operate with diminishing returns to accumulated knowledge ($\hat{\varphi} = 0.25$) and strongly increasing returns to source population ($\hat{\beta} = 1.81$), while we find little evidence that the population of the receiving region amplifies incoming spillovers beyond its direct effect on local patenting ($\hat{\alpha} \approx 0$, F-test $p > 0.99$).

In our gravity-like spillover function, a "push" mechanism means that larger origin regions generate disproportionately more outward spillovers ($\beta > 0$), while a "pull" mechanism would mean that larger destination regions attract or absorb more spillovers ($\alpha > 0$). The

	Estimate
Distance from London (log)	-2.918^{***}
	(0.191)
Bearing from London (sin)	0.424^{**}
	(0.206)
Bearing from London (cos)	-0.593^{***}
	(0.185)
Bearing (cos) $\times(\log(t) - \overline{\log(t)})$	1.098^{***}
	(0.325)
Mean population (log)	0.098
	(0.064)
Coal cost (log)	-0.246^{***}
	(0.087)
Coal cost $\times(\log(t) - \overline{\log(t)})$	-0.485^{***}
	(0.177)
AIC	1844.37
R^2	0.74
Num. events	228
Num. obs.	249
PH test (global)	–

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 11: Cox Hazard Model with Time-Varying Effects

strong asymmetry between origin and destination population effects ($\hat{\beta} = 1.81$ vs. $\hat{\alpha} \approx 0$) therefore suggests that diffusion during the Industrial Revolution was driven primarily by large, innovative centers radiating knowledge outward, rather than by large receiving regions pulling knowledge in. The strongly increasing returns to source population imply that a few dominant urban-industrial agglomerations—London, Paris, the Ruhr—were disproportionately influential in spreading knowledge. The diminishing returns to knowledge stocks in the spillover channel ($\hat{\phi} = 0.25$) are consistent with the idea that early innovations had outsized demonstration effects compared to later, more incremental improvements.

The Cox model adds a timing dimension: proximity to London, whether measured by Euclidean distance or friction-surface travel time, is the dominant predictor of when a region crosses the patenting threshold. The proximity effect holds under the travel-time measure as well: the south-eastward diffusion trajectory was not merely a geometric artifact of London's position, but was shaped by the historical transport network, which connected the British core more efficiently to northwestern continental Europe. Coal access is negatively associated with patenting intensity in both models (-0.20, $p < 0.05$), consistent with energy costs acting as a barrier throughout the period. Population is marginally significant in the Cox model ($+0.11$, $p < 0.10$), suggesting that larger regions tended to industrialize earlier, consistent with agglomeration advantages.

The robustness analyses further strengthen our core finding. Institutional controls leave the spillover coefficient virtually unchanged; omitted institutional heterogeneity is therefore not driving the result. Spatial autocorrelation diagnostics confirm that while some residual spatial clustering persists during the peak diffusion decades, the spillover coefficient remains positive and economically meaningful even under the most conservative spatial standard errors. The evidence consistently points to geographic proximity and knowledge flows as the primary drivers, rather than institutional convergence or spurious spatial correlation.

Several limitations deserve mention. First, patents capture only the formal, innovative aspects of the manufacturing transition and miss informal knowledge spillovers and non-patented process improvements. Second, our dataset relies exclusively on patents filed at the British and American patent offices, which may introduce a selection bias toward inventors with international connections. Third, while our friction surface estimates 19th-century transport costs more accurately than Euclidean distance, it assumes constant infrastructure throughout the period, which ignores the endogenous development of railways as industrialization progressed. Fourth, the institutional data is available only at the country level and may mask important within-country variation in legal frameworks and property rights.

Fifth, the structural estimation relies on the assumption that the reduced-form specification correctly captures the data-generating process; misspecification of the functional form could bias the structural parameter estimates. Sixth, our population measure is constructed by aggregating city-level population figures, and therefore captures only the urban component of each region's workforce. For 19th-century Europe, where a large share of inventive activity, including textile cottage industries and artisanal metalworking, originated outside major cities, this measure may systematically understate the innovation potential of predominantly rural regions. To the extent that rural and urban patenting intensity co-varied, the population coefficient and the spillover estimates may be biased toward reflecting urban-industrial agglomeration effects. Future work should address these limitations by incorporating time-varying transport networks, domestic patent records from France and Germany, subnational institutional data, total-population estimates from historical censuses, and alternative measures of technological adoption.

2.8 CONCLUSION

This paper investigates the spatial diffusion of formalized inventive activity across Europe using newly available patent data and a gravity-based framework for knowledge spillovers. Our analysis yields three main findings.

First, we document that the industrialization of continental Europe followed a wave-like diffusion process with a strong spatial component. Using a spatiotemporal kernel density estimation applied to over 120,000 historical patents, we show that the innovation frontier expanded outward from the British epicenter with a distinct south-eastward trajectory.

Second, the gravity model reveals that spatial knowledge spillovers were a significant determinant of regional innovation intensity. The structural estimation further shows that these spillovers operated with diminishing returns to accumulated knowledge ($\hat{\phi} = 0.25$) and were primarily a "push" phenomenon driven by source-region population ($\hat{\beta} = 1.81$), while the receiving region's size played no amplifying role ($\hat{\alpha} \approx 0$). The Cox hazard model confirms that proximity to London, measured both as Euclidean distance and as travel time over the historical transport network, was the dominant predictor of industrialization timing: pre-existing roads and waterways channeled where inventive knowledge spread.

Third, our robustness analysis reveals that the aggregate diffusion pattern is better characterized as the amplification of early locational advantages through spatial proximity rather than as a simple contagion process. When controlling for initial conditions, the residual spatial interaction turns competitive: the geography of the Industrial

Revolution was largely set by early locational endowments, which knowledge flows then reinforced.

These findings bear on debates about the relative importance of geography, institutions, and human capital in long-run development. The results suggest that during industrialization, geographic proximity to the frontier was a first-order determinant of regional outcomes, while institutional quality played a secondary role conditional on location.

REFERENCES

Acemoglu, Daron, Simon Johnson, and James A Robinson (2001). "The colonial origins of comparative development: An empirical investigation." In: *American economic review* 91.5, pp. 1369–1401.

Aghion, Philippe, John Van Reenen, and Luigi Zingales (2013). "Innovation and institutional ownership." In: *American economic review* 103.1, pp. 277–304.

Allen, Robert C (2009). *The British industrial revolution in global perspective*. Cambridge University Press.

Allen, Robert C (2015). "The high wage economy and the industrial revolution: a restatement." In: *The Economic History Review* 68.1, pp. 1–22.

Allen, Treb (2023). *The Topography of Nations*. Working Paper 31795. National Bureau of Economic Research.

Arrowsmith, Aaron (1798). *Map of Europe - Drawn from all the best surveys and rectified by astronomical observation*.

Bergeaud, Antonin and Cyril Verluise (2024). "A new dataset to study a century of innovation in Europe and in the US." In: *Research Policy* 53.1, p. 104903.

Berkes, Enrico, Matthew Lee Chen, and Matteo Tranchero (2026). "Three Centuries of Creative Destruction: The Evolution of British Patenting, 1617–1899." In: *Working Paper*.

Bogart, Dan (2014). "The transport revolution in industrializing Britain." In: *The Cambridge Economic History of Modern Britain* 1, pp. 368–391.

Boldrin, Michele, David K Levine, et al. (2008). *Against intellectual monopoly*. Vol. 62. Cambridge University Press Cambridge.

Bottazzi, Laura and Giovanni Peri (2003). "Innovation and spillovers in regions: Evidence from European patent data." In: *European Economic Review* 47.4, pp. 687–710.

Bottomley, Sean (2014). *The British patent system during the industrial revolution, 1700-1852: From privilege to property*. Cambridge University Press.

Bricker, C. and R.V. Tooley (1989). *Landmarks of Mapmaking: An Illustrated Survey of Maps and Mapmaking*. Dorset Press. ISBN: 9780880293495.

Broadberry, Stephen, Bruce MS Campbell, Alexander Klein, Mark Overton, and Bas Van Leeuwen (2015). *British economic growth, 1270–1870*. Cambridge University Press.

Brunet, Roger (1989). *Les villes européennes*. Paris: La documentation française.

Buringh, Eltjo (2021). "The Population of European Cities from 700 to 2000: Social and Economic History." In: *Research Data Journal for the Humanities and Social Sciences* 6.1, pp. 1–18.

Buzard, Kristy, Gerald A Carlino, Robert M Hunt, Jake K Carr, and Tony E Smith (2020). "Localized knowledge spillovers: Evidence from the spatial clustering of R&D labs and patent citations." In: *Regional Science and Urban Economics* 81, p. 103490.

Comin, Diego A, Mikhail Dmitriev, and Esteban Rossi-Hansberg (2012). *The spatial diffusion of technology*. Tech. rep. National Bureau of Economic Research.

Davey, R., M. Hayes, and J. Norman (1994). "Running uphill: an experimental result and its applications." In: *Journal of the Operational Research Society* 45.1, pp. 25–29.

Fernihough, Alan and Kevin Hjortshøj O'Rourke (2021). "Coal and the European industrial revolution." In: *The Journal of Economic History* 81.2, pp. 372–411.

Griliches, Zvi (1990). "Patent statistics as economic indicators: A survey." In: *Journal of Economic Literature* 28.4, pp. 1661–1707.

Hagerstrand, Torsten (1968). *Innovation diffusion as a spatial process.*

Head, Keith and Thierry Mayer (2014). "Gravity equations: Workhorse, toolkit, and cookbook." In: *Handbook of International Economics* 4, pp. 131–195.

Jaffe, Adam B, Manuel Trajtenberg, and Rebecca Henderson (1993). "Geographic localization of knowledge spillovers as evidenced by patent citations." In: *the Quarterly journal of Economics* 108.3, pp. 577–598.

Keller, Wolfgang (2004). "International technology diffusion." In: *Journal of Economic Literature* 42.3, pp. 752–782.

Kelly, Morgan and Cormac Ó Gráda (2019). "Speed under sail during the early industrial revolution (c. 1750–1830)." In: *The Economic History Review* 72.2, pp. 459–480.

Krugman, Paul (1991). "Increasing returns and economic geography." In: *Journal of Political Economy* 99.3, pp. 483–499.

Licio, Vania (2022). "Coal prices and industrialization in Italy (1861–1911)." In: *Rivista di storia economica* 2, pp. 203–236.

Lucas Jr, Robert E (2009). "Trade and the Diffusion of the Industrial Revolution." In: *American Economic Journal: Macroeconomics* 1.1, pp. 1–25.

MacLeod, Christine (1988). *Inventing the Industrial Revolution: the English patent system, 1660-1800*. Cambridge University Press.

Marshall, Monty G, Ted Robert Gurr, and Keith Jaggers (2019). *Polity5: Political Regime Characteristics and Transitions, 1800–2018*. Center for Systemic Peace.

Mokyr, Joel (1992). *The lever of riches: Technological creativity and economic progress*. Oxford University Press.

Moretti, Enrico (2021). *The geography of innovation: A historical perspective*. Working Paper 28941. National Bureau of Economic Research.

Moser, Petra (2005). "How do patent laws influence innovation? Evidence from nineteenth-century world's fairs." In: *American economic review* 95.4, pp. 1214–1236.

North, Douglass C (1990). *Institutions, institutional change and economic performance*. Cambridge university press.

Özak, Ömer (2018). "Distance to the pre-industrial technological frontier and economic development." In: *Journal of Economic Growth* 23.2, pp. 175–221.

Pollard, Sidney (1981). *Peaceful Conquest: The Industrialization of Europe, 1760–1970*. Oxford: Oxford University Press.

Pomeranz, Kenneth (2000). *The Great Divergence: China, Europe, and the Making of the Modern World Economy*.

Porto Tapiquén, Carlos Efraín (2015). *World Shapefile of Coalfields*. Orogénesis Soluciones Geográficas.

Rodrigue, Jean-Paul (2020). *The Geography of Transport Systems*. 5th. New York: Routledge.

Rogers, Everett M. (2003). *Diffusion of Innovations*. 5th. Free Press.

Santos Silva, J.M.C. and Silvana Tenreyro (2006). "The log of gravity." In: *The Review of Economics and Statistics* 88.4, pp. 641–658.

Sokoloff, Kenneth L (1988). "Inventive activity in early industrial America: evidence from patent records, 1790–1846." In: *The Journal of Economic History* 48.4, pp. 813–850.

Tobler, Waldo (1993). "Three presentations on geographical analysis and modeling." In: *National Center for Geographic Information and Analysis Technical Report* 93.1.

Wrigley, Edward Anthony (2013). "Energy and the English industrial revolution." In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371.1986, p. 20110568.

3

BEYOND THE BEACH: SPATIAL COMPETITION IN TWO DIMENSIONS

ABSTRACT

This paper extends Hotelling's spatial competition framework to two-dimensional markets with linear consumer demand and probabilistic firm choice. Consumer density can be non-uniform across the region, firms compete in prices and locations, and consumers choose according to a logit specification based on the lowest delivered price. We prove existence of price equilibria in the static game and establish existence of subgame perfect Nash equilibria in a dynamic extension with sequential firm entry and fixed entry costs. Depending on parameter values, the model generates both minimum and maximum product differentiation. Numerical simulations illustrate market outcomes under various configurations and consumer distributions.

3.1 INTRODUCTION

The spatial dimension of firm behavior was popularized by Hotelling (1929), who studied location and price setting on a unit interval $[0, 1]$ (a stylized "Main Street"). Both Hotelling and those who followed assumed that consumers choose among firms based on delivered (effective) prices, the defining assumption of spatial competition theory. Related ideas appeared earlier in the work of Launhardt (1885); see Dos Santos Ferreira and Thisse (1996) for discussion.

Despite decades of research building on Hotelling's foundation, three simplifying assumptions have proved persistent. Models are almost always one-dimensional, consumer distributions almost always uniform, and consumer choice almost always deterministic: the nearest firm captures all demand. Each assumption is tractable in isolation; together they leave a substantial gap between theory and practice.

We relax all three at once. The model is set on a compact two-dimensional region with an arbitrary consumer density and probabilistic (logit) choice. In the static case we prove that price equilibria exist. We then extend to sequential firm entry with fixed costs; subgame perfect Nash equilibria exist, and the monopoly entry problem admits a closed-form solution. Numerical simulations show that the model generates both firm clustering and dispersion depending on parameters, and we characterize how market structure responds to entry costs, market size, and competitive intensity.

The literature has addressed each of these problems, though typically one at a time. A well-known modification is the circular model of Salop (1979), which removes the boundary issues present in Hotelling's line. Probabilistic consumer choice was introduced through the conditional logit model of McFadden (1974), later formalized for empirical use by Ben-Akiva and Lerman (1985). These models allow for smoother demand substitution across firms.

Attempts to generalize beyond the line have also continued. Barro (2024), four decades after Salop's initial work, reexamines a circular market, analyzing markups and firm entry. Tsai and Lai, 2005 proposed a triangular market structure. In central place theory, the optimal configuration for a single good is hexagonal, assuming uniform density and equal accessibility, though De Serpa, 1986 argued that the specific polygon shape does not affect equilibrium outcomes. T. Drezner and Z. Drezner (2017) and Z. Drezner and Eiselt (2024) show that when consumers patronize the nearest firm, each firm's market share equals the area of its Voronoi cell. This geometric insight motivates the joint analysis of market shape, firm strategy, and consumer behavior.

A separate question concerns where firms end up relative to each other. Both dispersion and agglomeration are observed in practice. In his model, Hotelling introduced the concept of "minimum differentiation", which describes firms' tendency to cluster around the center of the market. In a spatial market, when the product is homogeneous, the only form of differentiation available to a firm is its location, hence the term minimum differentiation. However, Hotelling's original formulation, as shown by D'Aspremont, Gabszewicz, and Thisse, 1979, is problematic, and a Nash equilibrium does not exist when firms are sufficiently near each other. When transport costs are linear, firms have an incentive to deviate slightly towards the center to capture a larger market share, which makes any possible configuration unstable.

Two-dimensional work goes back at least to the 1970s. Cahan et al. (2021) argues that classic stylized 2-D models often lack realism and that a network-based approach may be more suitable; that work also notes that most generalizations still assume consumers always go to the nearest firm, which is not required. Cox (1987) dropped this assumption, proposing a probabilistic choice model where consumers are more likely to patronize nearby firms but may still buy from more distant ones, captured by a probability vector over firm ranks by distance.[1] Larralde, Stehle, and Jensen (2009) modeled consumer choice with a logit rule in a continuous two-dimensional space, with each consumer choosing based on a generalized delivered price

1 The classical Hotelling model corresponds to the "degenerate" case where the entire probability mass is on the nearest firm.

(mill price plus quadratic transport cost). Both contributions, however, focus more on the location problem than on price.[2]

Wrede, 2015 also introduces probabilistic consumer preferences to Hotelling, adding endogenous residential choice. Firms choose location and price along a linear city, while consumers choose where to live and which firm to patronize, based on a logit utility with random preference heterogeneity. The continuous logit structure guarantees existence and uniqueness of a subgame perfect Nash equilibrium, which resolves the non-existence problem of D'Aspremont, Gabszewicz, and Thisse, 1979 under linear transport costs. This contribution, however, is restricted to 1-D modeling. Most of these models also retain Hotelling's duopoly on a line. Brenner, 2005 relaxes the duopoly restriction, extending the classic Hotelling model to markets with three or more firms competing in location and price, assuming quadratic transportation costs. He shows that, unlike the duopoly case, equilibria with more players exhibit neither maximal nor minimal differentiation. Firms tend to adopt asymmetric and interior positions, and pricing follows a U-shaped pattern, with corner firms often pricing more aggressively. Brenner derives explicit solutions for up to five players and uses numerical simulations for up to nine.

A further common feature of the literature is the assumption of a uniform distribution of consumers. This assumption has been widely explored (Neven, 1986), but more general models, allowing for non-uniform distributions, can encompass several classical frameworks as special cases. For instance, the Bertrand duopoly can be seen as a specific instance of Hotelling competition with a non-uniform consumer distribution, as in Rashid (2016). In his model, the price-setting stage mirrors Hotelling's when consumers are clustered near the two firms located at a given distance. Here, the distance determines the price gap required to induce switching: if prices are in this range, firms share the market; otherwise, the cheaper firm captures all demand. Taken together, this body of work points in a consistent direction: geometry and consumer behavior both matter, and models that fix either one artificially tend to miss things. Hotelling's linear city with uniform consumers and deterministic choice is a useful benchmark, but only that.

The paper proceeds as follows: Section 3.2 presents the core static model and establishes existence results for price equilibria. Section 3.3 extends the framework to dynamic entry with fixed costs, analyzing the monopoly case in detail and characterizing equilibrium market structure. Section 3.4 connects the framework to classical models as limiting cases. Section 3.5 uses numerical simulations to illustrate

2 In Larralde, Stehle, and Jensen (2009), in particular, price is the same for all enterprises, set at a common value, so competition is exclusively spatial; even though price enters consumer utility, it is not subject to strategic optimization.

market outcomes and examine spatial differentiation. Section 3.6 concludes. Proofs and technical details appear in the Appendices.

3.2 THE STATIC MODEL

Consider a compact, convex region $S \subset \mathbb{R}^2$ describing a geographic market. Consumers are located in the space and are immobile. Let $\rho : S \to \mathbb{R}_+$ denote a continuous, bounded consumer density with total mass

$$M = \int_S \rho(s)\, ds.$$

We assume $M > 0$ is finite.

Remark 3.1. Unlike the classical uniform-density assumption, we allow ρ to be non-uniform and potentially multi-modal, reflecting realistic population patterns.

In the market, there is a number N of firms selling a homogeneous good, indexed by $\mathcal{N} = \{1, \dots, N\}$, each located at $x_i \in S$. Firms compete by setting prices $p_i \in [c, \overline{p}]$, where $c > 0$ is the common marginal cost. We assume $\overline{p}$ is low enough to ensure full market coverage: since $P_i^*(s) \in [c, \overline{p} + t\overline{d}]$ for all $s \in S$ and all i, we impose $\overline{p} + t\overline{d} < \alpha/\gamma$, so that $q_i(s) = \alpha - \gamma P_i^*(s) > 0$ throughout the strategy space.[3]

Consumer Behavior and Demand

The effective price paid by a consumer at location s buying from firm i is:

$$P_i^*(s) = p_i + t \cdot d(s, x_i), \tag{27}$$

where $t > 0$ is the transportation cost rate and $d(s, x_i) : S \times S \to [0, \overline{d}]$ represents the transportation cost function, assumed continuous in both arguments and convex in x_i.

Consumers have linear demand. The quantity demanded by a consumer at location s from firm i is:

$$q_i(s) = \alpha - \gamma P_i^*(s), \tag{28}$$

where $\alpha > 0$ and $\gamma > 0$ are demand parameters.

Consumers are more likely to buy from firms offering lower effective prices. We model consumer choice probabilistically with a logit approach similar to Larralde, Stehle, and Jensen, 2009; thus, the probability that a consumer at location s chooses firm i is:

$$\text{Prob}(i \mid s, \mathbf{p}) = \frac{\exp\big(-\beta\, P_i^*(s)\big)}{\sum_{j=1}^{N} \exp\big(-\beta\, P_j^*(s)\big)}, \tag{29}$$

3 Equivalently, one can introduce an outside option or truncate individual demand at zero; we abstract from this to keep the demand system smooth.

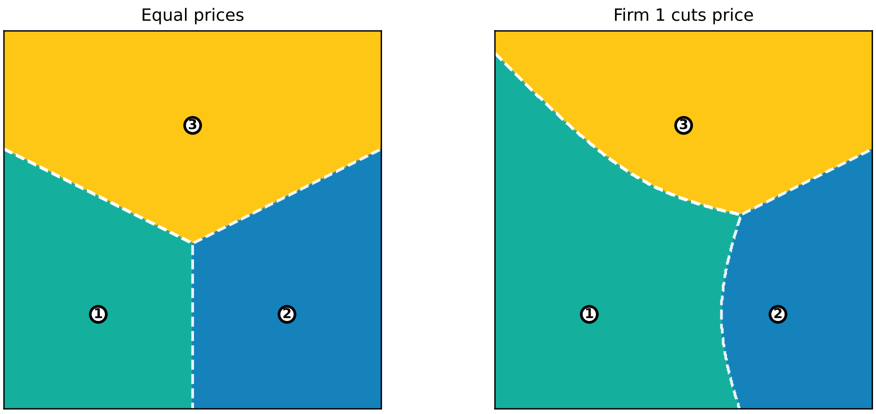


Figure 21: Market areas defined by the lowest effective price (similar to a weighted Voronoi partition). The left panel shows the case when all firms have the same price. The right panel shows the case when firm 1 lowers its price. Dots mark firm locations.

where $\beta > 0$ is the logit parameter reflecting consumer sensitivity to effective price differences. As $\beta \to \infty$, choices become deterministic and consumers select the firm offering the lowest effective price.

Figure 21 illustrates the induced (deterministic) market areas as firm prices vary, showing the case of equal prices (left panel) and the effect of a unilateral price cut (right panel).

The spatial demand for firm i, given the price vector $\mathbf{p} = (p_1, \ldots, p_N)$, is:

$$Q_i(\mathbf{p}) \;=\; \int_S \rho(s)\,(\alpha - \gamma P_i^*(s))\,\mathrm{Prob}\big(i \mid s, \mathbf{p}\big)\,ds. \tag{30}$$

The Optimal Price Problem

In the static scenario—fixed locations, N firms—each firm chooses its price to maximize profit, taking the locations and prices of rivals as given. Firm i solves:

$$\Pi_i(\mathbf{p}) \;=\; (p_i - c)\,Q_i(\mathbf{p}) \tag{31}$$

$$\max_{p_i \in [c, \overline{p}]} \Pi_i(\mathbf{p}) \tag{32}$$

We now establish the existence of price equilibria in this spatial competition framework.

Assumption 3.2 (Quasi-concavity of profits)**.** For each firm i and every fixed $\mathbf{p}_{-i}$, the profit function $\Pi_i(p_i, \mathbf{p}_{-i})$ is quasi-concave in p_i on $[c, \overline{p}]$.

Theorem 3.3 (Existence of Price Equilibrium). *Under the assumptions above (continuous bounded* ρ, $\alpha > 0, \gamma > 0$, $\beta > 0$, $t > 0$, *finite* $\overline{p} > c$, *compact strategy sets* $[c, \overline{p}]$ *where* $\overline{p}$ *ensures* $q_i(s) > 0$*), a logit consumer choice function with linear individual demand, and Assumption 3.2, there exists at least one pure-strategy price equilibrium.*

Proof. (Sketch, following Glicksberg's theorem (Glicksberg, 1952). The complete proof is given in Appendix B)
Step 1: Compact, Convex Strategy Sets. Each firm i chooses p_i from the interval $[c, \overline{p}]$, which is compact and convex.
Step 2: Continuity of Profit Functions. $\Pi_i(p_i, \mathbf{p}_{-i})$ is a continuous function of all prices (see Appendix B).
Step 3: Quasi-Concavity in Own Price. Assumption 3.2 ensures that each firm's objective is quasi-concave in its own price. Appendix B provides a sufficient condition in terms of derivatives of Q_i.
Step 4: Glicksberg's Fixed Point Theorem. With continuous payoffs, compact convex strategy sets, and quasi-concave payoffs, at least one pure-strategy Nash equilibrium exists. □

The existence proof applies Glicksberg's fixed point theorem, using the properties of continuous payoffs, compact convex strategy sets, and quasi-concavity of profit functions (Assumption 3.2).

While the uniqueness of price equilibria can be established under certain conditions (when the direct own-price sensitivity of demand is sufficiently strong relative to competitive cross-price effects), we focus here on existence, which is sufficient for our analysis of the dynamic entry game.

3.3 DYNAMIC ENTRY WITH FIXED COSTS

We now introduce firm entry decisions with fixed costs. Market structure, how many firms enter and where they locate, becomes endogenous.

The Dynamic Entry-Location-Price Game

Consider an infinite-horizon dynamic game with entry cost $F > 0$ and a common discount factor $\delta \in (0, 1)$:

- **Odd periods** $(\tau = 1, 3, \dots)$ **(Entry stage):** A potential entrant (firm $N_{\tau-1} + 1$) observes the locations $\mathbf{x}_{\tau-1}$ of incumbent firms and decides whether to enter. If entry occurs, the entrant chooses location x_E^* to maximize anticipated profit:

$$\max_{x_E \in S} W_E(\mathbf{x}_{\tau-1}, x_E) \tag{33}$$

  where $W_E(\mathbf{x}_{\tau-1}, x_E)$ is the potential entrant's continuation value from entering at location x_E, defined as the discounted net present

value of future price-stage profits minus the fixed cost F under equilibrium continuation play (see Appendix D). Entry occurs if and only if
$\max_{x_E \in S} W_E(\mathbf{x}_{\tau-1}, x_E) > 0$.

- **Even periods** $(\tau = 2, 4, \ldots)$ **(Price competition stage):** Given locations $\mathbf{x}_\tau$, firms simultaneously set prices, leading to the equilibrium price vector $\mathbf{p}^*(\mathbf{x}_\tau)$ as characterized in Section 3.2.

Assumption 3.4 (Regularity of entry values)**.** For every number of incumbents N and every incumbent location profile $\mathbf{x}_N$, the potential entrant's continuation value $W_E(\mathbf{x}_N, \cdot)$ is upper semi-continuous on S.

Initial Entry: The Monopoly Case

The first firm to enter the market, for the first two periods, faces no competition and acts as a monopolist. This special case offers analytical solutions for both optimal pricing and location.

Proposition 3.5 (Monopoly Price Formula)**.** *For a monopolist at location x_1, the profit-maximizing price is:*

$$p_1^*(x_1) = \frac{\alpha}{2\gamma} + \frac{c}{2} - \frac{t \cdot \bar{d}_\rho(x_1)}{2}, \tag{34}$$

where $\bar{d}_\rho(x_1) = \frac{1}{M}\int_S \rho(s) d(s, x_1) ds$ is the weighted average distance to consumers. This solution assumes $\alpha - \gamma c - \gamma t \bar{d}_\rho(x_1^*) \geqslant 0$ to ensure $p_1^*(x_1^*) \geqslant c$, $p_1^*(x_1^*) \leqslant \bar{p}$ to guarantee positive demand quantities, and that all consumers in the market consume from the monopolist.

Proof. For a monopolist at x_1, profit is $\Pi(p_1) = (p_1 - c)\int_S \rho(s)(\alpha - \gamma[p_1 + t \cdot d(s, x_1)])ds$. Differentiating with respect to p_1 and setting to zero:

$$\begin{aligned} \int_S \rho(s)(\alpha - \gamma[p_1 + td(s, x_1)])ds + (p_1 - c)\int_S \rho(s)(-\gamma)ds = 0 \\ M(\alpha - \gamma p_1 - \gamma t \bar{d}_\rho(x_1)) - \gamma M(p_1 - c) = 0 \\ \alpha - \gamma p_1 - \gamma t \bar{d}_\rho(x_1) - \gamma p_1 + \gamma c = 0 \\ \alpha + \gamma c - \gamma t \bar{d}_\rho(x_1) = 2\gamma p_1 \end{aligned}$$

Therefore:

$$p_1^* = \frac{\alpha}{2\gamma} + \frac{c}{2} - \frac{t\bar{d}_\rho(x_1)}{2} \tag{35}$$

The second derivative $\partial^2\Pi/\partial p_1^2 = -2\gamma M < 0$ confirms this is a maximum. □

Proposition 3.6 (Monopolist Location)**.** *The monopolist chooses the location x_1^* that minimizes the weighted average distance to consumers $\bar{d}_\rho(x_1)$.*

Proof. Substituting the optimal price into the profit function:

$$\Pi(x_1) = \frac{M}{4\gamma}(\alpha - \gamma c - \gamma t \bar{d}_\rho(x_1))^2 \tag{36}$$

Assuming $\alpha - \gamma c - \gamma t \bar{d}_\rho(x_1) > 0$ (ensuring positive demand), maximizing profit requires minimizing $\bar{d}_\rho(x_1)$. □

Remark 3.7. For uniform consumer density ($\rho(s) = \text{constant}$) in a convex region S and Euclidean distance, the optimal monopolist location solves the (continuous) Fermat–Weber problem, i.e., it is a geometric median of the uniform distribution on S. More generally, for arbitrary ρ and distance function d, x_1^* is a weighted 1-median (Fréchet median) of the induced distribution.

Remark 3.8. The optimal price decreases with average distance: higher transport costs for distant consumers reduce the quantity they demand, so the monopolist lowers the mill price to retain them.

If the market is initially empty, a necessary condition for the first firm to enter is that the perpetuity value of monopoly profits exceeds the entry cost:

$$\frac{\Pi(x_1^*)}{1 - \delta} > F \tag{37}$$

This condition is exact when no further entry occurs; in general the first entrant's continuation value depends on subsequent entry (see Appendix D). If entry occurs, the first entrant locates at x_1^* and sets price $p_1^*(x_1^*)$ as characterized above in the initial monopoly phase.

Equilibrium Analysis with Multiple Firms

After the first entry, subsequent firms may enter sequentially, each choosing their optimal location given the existing configuration of incumbents.

Theorem 3.9 (Existence of SPNE with Entry)**.** *Under the assumptions of our model, with a positive fixed entry cost* $F > 0$*, a common discount factor* $\delta \in (0, 1)$*, assuming the price equilibrium exists for any configuration of firms, and Assumption 3.4, there exists a subgame perfect Nash equilibrium (SPNE) in the infinite-horizon entry-location-price game.*

Proof. See Appendix D. □

Theorem 3.10 (Non-Uniqueness of SPNE with Entry)**.** *The entry-location-price game generally admits multiple subgame perfect Nash equilibria, distinguished by different numbers of firms, different location configurations, or both.*

Proof. See Appendix D. □

An equilibrium market structure exists, but typically not a unique one. Multiplicity arises from four sources: multiple optimal locations for each entrant, path dependence in sequential entry decisions, strategic entry deterrence by incumbents, and the integer constraint on the number of firms. As a result, comparative statics in N^* are stepwise: parameter changes may leave N^* unchanged until a threshold is crossed.

Equilibrium Properties

Proposition 3.11 (Market Structure Properties)**.** *The equilibrium number of firms N^* in the entry-location-price game with fixed cost F:*

1. *Decreases with F (higher entry costs lead to fewer firms)*
2. *Increases with market size M (larger markets support more firms)*
3. *Decreases with parameters that intensify price competition (β and γ)*
4. *Increases with parameters that soften price competition (larger t)*

Note. *Since N must be an integer, these changes occur in discrete steps (e.g., N^* remains constant for a range of F before jumping down).*

Proof. See Appendix D. □

Higher entry costs and stronger price competition reduce the number of firms that can earn non-negative profits; larger markets and higher transport costs work in the opposite direction by sustaining markups.

Proposition 3.12 (Comparative Statics)**.** *Assuming a price equilibrium exists for any configuration of firms, we obtain the following comparative statics:*

1. *Increasing β (more price-sensitive consumer choice) generally leads to* lower *equilibrium prices p_i^*.*
2. *Increasing γ (higher price sensitivity of demand) generally leads to* lower *equilibrium prices p_i^*.*
3. *Increasing t (higher transportation costs) tends to* increase *equilibrium prices p_i^* by softening competition through greater spatial differentiation.*
4. *Increasing α (higher base demand) generally leads to* higher *equilibrium prices p_i^*.*

Proof. See Appendix C. □

Higher β and γ make consumers more price-responsive, so competition intensifies and prices fall. Higher transportation costs (t) increase spatial differentiation and allow firms to sustain higher prices. Higher base demand (α) raises willingness to pay and pushes prices up.

3.4 RELATIONSHIP TO CLASSICAL MODELS

Our framework connects to classical spatial competition models as limiting cases. In particular, under unit demand and in the deterministic choice limit $\beta \to \infty$, market areas coincide with those in Hotelling's linear city and Salop's circular city.

3.4.1 *Hotelling's Linear City*

Proposition 3.13 (Hotelling Special Case). *Let* $S = [0, 1] \times \{0\}$, $\rho(s) = 1$, $d(s, x) = |s - x|$, *and* $N = 2$. *Assume unit demand and full market coverage. Then as* $\beta \to \infty$, *consumer choices become deterministic and market areas become intervals defined by the lowest effective price* $P_i^*(s) = p_i + t|s - x_i|$, *recovering Hotelling's linear city.*

Proof. As $\beta \to \infty$, $\text{Prob}(i|s, \mathbf{p}) \to 1$ for the firm with the lowest $P_i^*(s)$. With unit demand, $Q_i(\mathbf{p}) \to \text{length}(V_i(\mathbf{p}))$, where $V_i(\mathbf{p})$ is the region where firm i offers the lowest effective price. This recovers Hotelling's model. □

3.4.2 *Salop's Circular City*

Proposition 3.14 (Salop Special Case). *Let* S *be a circle of circumference* L, $\rho(s) = 1/L$, *and let* $d(s, x)$ *be arc distance. Assume unit demand and full market coverage. As* $\beta \to \infty$, *market areas become arc segments, recovering Salop's circular city.*

Proof. Similar to the Hotelling case, $Q_i(\mathbf{p}) \to \text{length}(V_i(\mathbf{p}))/L$ as $\beta \to \infty$, recovering Salop's model. □

Remark 3.15. Both special cases require $\beta \to \infty$ (deterministic consumer choice) and unit demand. Probabilistic demand ($\beta < \infty$) combined with a linear quantity response allows for smoother substitution and a non-trivial quantity margin, both absent from the classical cases. For arbitrary convex distance functions d, our model converges to a generalized Voronoi-Hotelling model as $\beta \to \infty$.

3.5 EQUILIBRIUM ANALYSIS AND SIMULATIONS

While the model guarantees the existence of a price equilibrium, its specific characteristics for $N > 1$ are analytically intractable. We therefore use numerical simulations to illustrate the geometry of market areas, the effect of heterogeneous consumer density, and the sequential entry mechanism of Section 3.3. Unless otherwise stated, the simulations are conducted in a square market $S = [0, 1]^2$ with Euclidean distance. Integrals are approximated on a uniform grid and equilibrium prices are computed by best-response iteration. The full numerical implementation and baseline parameterization are reported in Appendix A.

As a regularity diagnostic connected to Assumption 3.17, for each scenario we compute the curvature ratio $\eta_i(\mathbf{p}^*)$ at the computed equilibrium prices and a conservative own-price scan. The resulting values are reported in Appendix A (Table 12).

We first illustrate how the consumer choice sensitivity β governs the sharpness of market areas. Figure 22 fixes three firm locations and compares the equilibrium outcome for increasing values of β. For low β the market shares are diffuse, while for high β the induced partition becomes nearly deterministic and approaches the Voronoi diagram.

The effect of β on equilibrium prices and profits need not be monotone. Economically, β controls the strength of substitution across firms through the logit term: letting $L_i(s, \mathbf{p}) = \text{Prob}(i \mid s, \mathbf{p})$, we have $\partial L_i / \partial p_i = -\beta \, L_i(1 - L_i)$. When β is small, shares are close to $1/N$ and weakly responsive to prices, so a unilateral price increase does not shift much demand to rivals and equilibrium markups can be high. As β rises to moderate values, substitution becomes stronger (both because of the higher β and because $L_i(1 - L_i)$ is sizeable), intensifying price competition and reducing markups. For very large β, however, choice becomes nearly deterministic: away from the boundaries between firms, L_i is close to 0 or 1 so $L_i(1 - L_i) \approx 0$, and competitive pressure is concentrated on a small set of consumers near indifference loci. In such highly segmented configurations, firms may regain some local market power inside their (almost) Voronoi cells, so equilibrium prices can vary little or even increase slightly with further increases in β.

Consider now the dynamic model. In the first two periods there is only one firm in the economy which will choose location first and then its price. The first firm is unaware whether new firms will enter the market and will behave in this initial phase as a monopolist. We derived in Proposition 3.5 that the monopolist's optimal *mill price* decreases linearly in the transport cost, because higher t raises delivered prices and reduces quantities under linear demand. In particular, at the optimal location x_1^* we have

$$p_1^*(x_1^*) = \frac{\alpha}{2\gamma} + \frac{c}{2} - \frac{t}{2} \, \bar{d}_\rho(x_1^*),$$

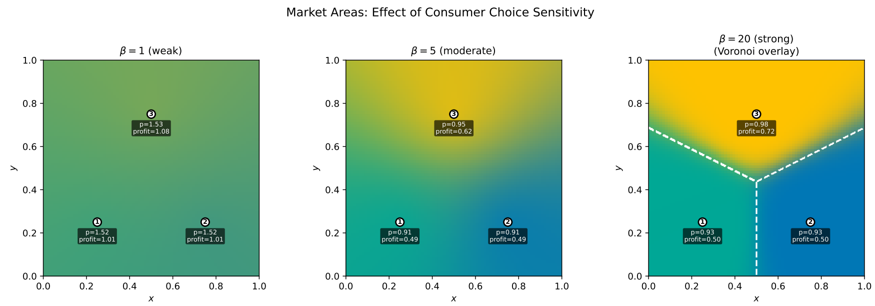


Figure 22: Effect of consumer price sensitivity β on market areas (uniform density, three firms at fixed locations). Colored regions represent market shares; dashed lines overlay the Voronoi partition. As β increases, choice becomes more deterministic and market areas converge to Voronoi cells; equilibrium prices fall sharply from low to moderate β and then vary little at higher β in this configuration. Labels report equilibrium prices and profits.

While the mill price decreases in t, the corresponding average delivered price paid by consumers increases:

$$\bar{P}_1^* \equiv p_1^*(x_1^*) + t\,\bar{d}_\rho(x_1^*).$$

Both objects converge to the standard non-spatial monopoly price with linear demand, $p_m = \frac{\alpha}{2\gamma} + \frac{c}{2}$, as $t \to 0$. Figure 23 plots these comparative statics.

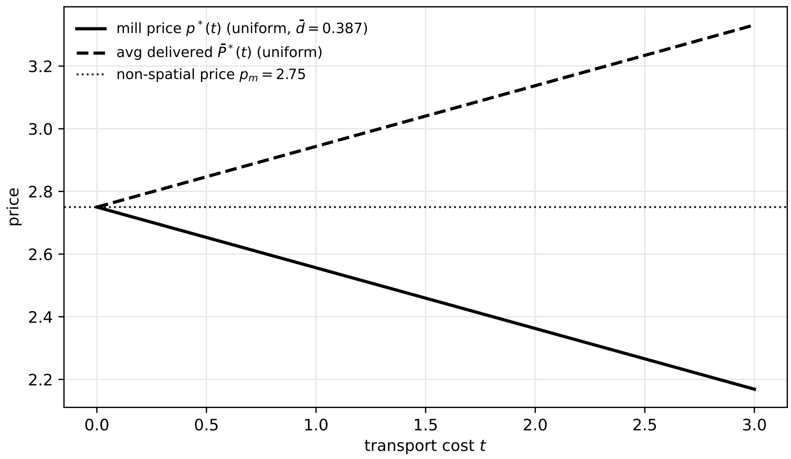


Figure 23: Monopolist case. The optimal mill price $p_1^*(x_1^*)$ (solid) decreases linearly in the transport cost t, while the average delivered price $\bar{P}_1^*$ (dashed) increases linearly. The dotted line marks the non-spatial monopoly price $p_m = \frac{\alpha}{2\gamma} + \frac{c}{2}$, which both objects converge to as $t \to 0$.

After the first and second periods, new firms can choose to enter the market until entry is no longer profitable.

We next illustrate the effect of heterogeneous consumer density on equilibrium prices and market shares. Figure 24 considers a triopoly

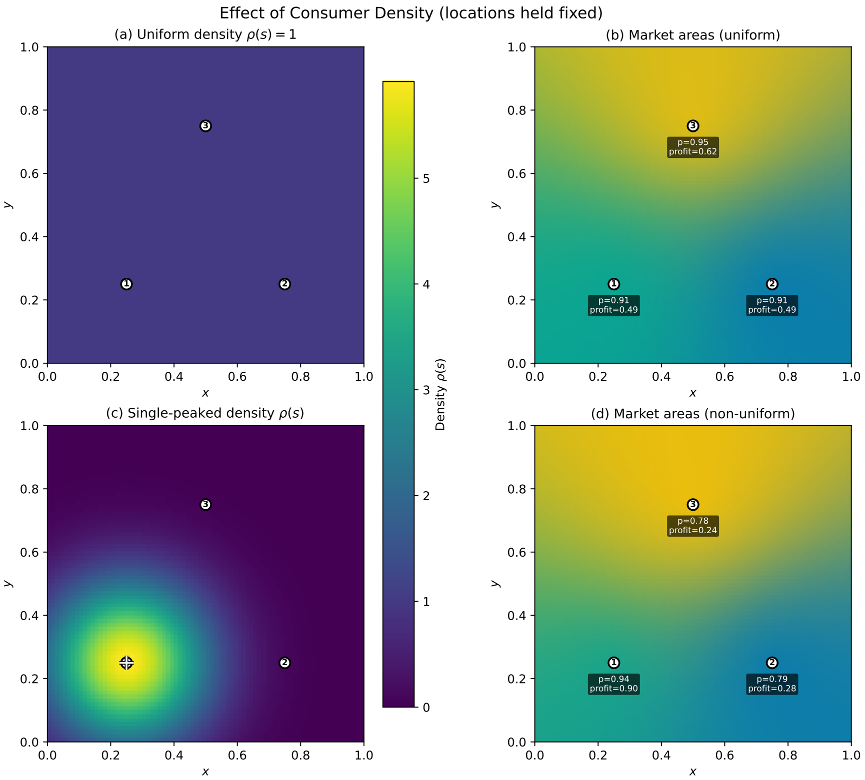


Figure 24: Effect of heterogeneous consumer density holding firm locations fixed (to isolate the role of ρ in price competition). The left column reports the consumer density (uniform versus single-peaked, normalized to integrate to one; peak marked by "+"). The right column reports the corresponding equilibrium market shares. Labels report equilibrium prices and profits.

with fixed locations and compares the equilibrium under uniform density to a market with a single demand peak. The firm located close to the peak captures a larger market share and can profitably set a higher price.

Finally, Figure 25 illustrates a sequential entry path under uniform density and high β. Starting from a monopoly, each entrant chooses a location to maximize its post-entry profit on a finite candidate grid, and prices adjust to the static equilibrium each period. The top panels display selected stages of the entry path; the bottom panel plots the implied equilibrium number of firms $N^*(F)$ as a function of the entry cost F for this realized sequence of entry profits. Additional numerical illustrations are reported in Appendix A.

3.5.1 *Spatial Differentiation: Agglomeration vs. Dispersion*

The simulations also address where firms locate relative to each other, the classic minimum vs. maximum differentiation question (D'Aspremont, Gabszewicz, and Thisse, 1979; De Palma et al., 1985). Two forces com-

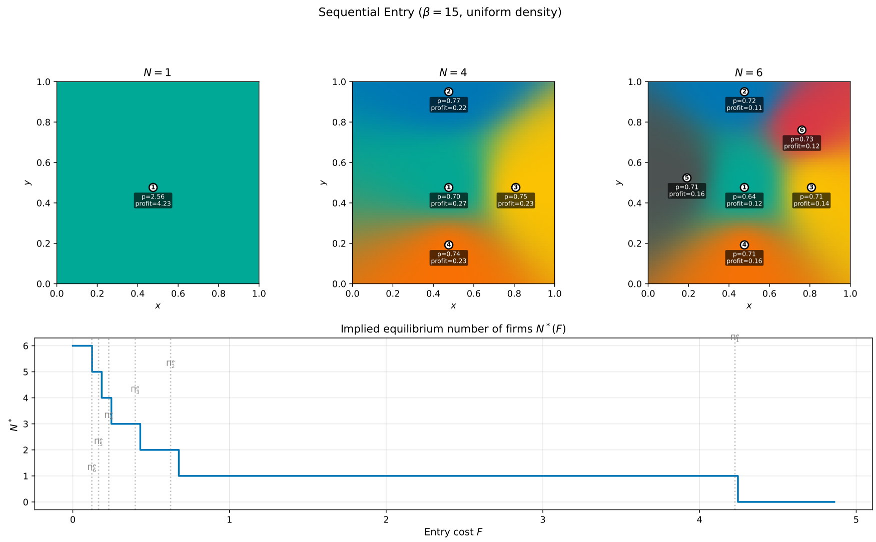


Figure 25: Sequential entry illustration under uniform density ($\beta = 15$). In each period a potential entrant chooses the location that maximizes its post-entry profit on a candidate grid; prices are the static equilibrium prices given locations. The bottom panel plots the entry threshold $N^*(F)$ implied by the sequence of entry profits along this path. Labels report equilibrium prices and profits.

pete: a "demand pull" draws firms toward density peaks, while a "competitive push" drives them apart to reduce price competition. Which dominates depends chiefly on the consumer choice sensitivity β.

Remark 3.16 (Spatial Differentiation Regimes (Numerical)). In our simulations, we observe two regimes:

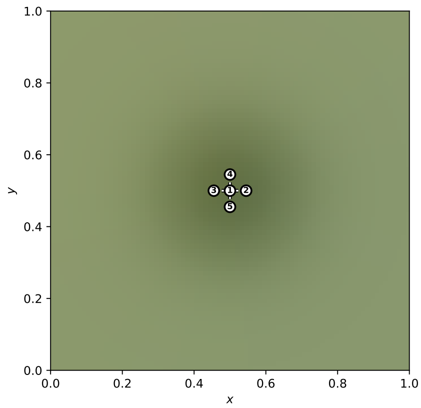
1. **Agglomeration Regime:** For sufficiently low values of β, firms tend to cluster.

2. **Dispersion Regime:** For sufficiently high values of β and transport cost t, firms tend to disperse to soften price competition.

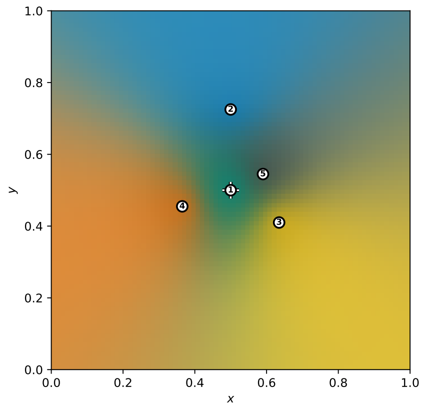
Figure 26 illustrates these regimes computationally in a market with a central density peak. Firms enter sequentially. With low β, the demand-pull dominates and entrants locate close to the density peak (Panel a). With high β and t, the incentive to avoid head-to-head price competition dominates and firms spread out despite the density peak (Panel b).

3.6 CONCLUSION

This paper took Hotelling's framework to two dimensions, allowed for non-uniform consumer density, and introduced probabilistic (logit) choice. We proved that price equilibria exist in the static game and that subgame perfect Nash equilibria exist in the dynamic entry game



(a) Agglomeration Regime (Low β)



(b) Dispersion Regime (High β, high t)

Figure 26: Equilibrium locations for five firms under different differentiation regimes in a market with a central density peak. The background shading indicates consumer density.

with fixed costs, though generally multiple ones, due to location multiplicity and path dependence. The monopoly case admits closed-form solutions: the optimal location minimizes the weighted average distance to consumers, and the optimal price decreases linearly with that distance.

The simulations show that the same model can generate both firm clustering and dispersion. When consumer price sensitivity is low, the pull toward demand peaks dominates and firms agglomerate. When sensitivity is high and transport costs are large, the incentive to avoid head-to-head competition dominates and firms spread out.

The model connects to Hotelling's line and Salop's circle as limiting cases under deterministic choice ($\beta \to \infty$) and unit demand. The two-dimensional, probabilistic version is more general but also harder to analyze: numerical methods fill the gap for multi-firm configurations.

The framework can be brought to data when consumer locations and firm prices are observed; the monopoly case in particular may be tractable with retail scanner data.

APPENDIX

A NUMERICAL IMPLEMENTATION AND ADDITIONAL RESULTS

This appendix describes the numerical implementation used to generate the figures in Section 3.5 and reports additional numerical illustrations.

A.1 *Numerical Setup*

MARKET, GRIDS, AND DENSITIES. We set $S = [0,1]^2$ and approximate integrals on a uniform $n \times n$ quadrature grid (baseline $n = 80$). For non-uniform densities, we normalize ρ to have unit mass, $\int_S \rho(s)\,ds = 1$, so that profit magnitudes are comparable across density scenarios.

PRICE EQUILIBRIUM SOLVER. Given firm locations, equilibrium prices are computed by damped best-response iteration over the compact set $[c, \overline{p}]^N$ (damping 0.5, tolerance 10^{-9}, maximum 500 iterations).

ENTRY ALGORITHM (ILLUSTRATIVE). In sequential-entry figures, an entrant searches over an $m \times m$ candidate grid (baseline $m = 20$) and enters at the location that maximizes its post-entry profit given incumbents; prices then adjust to the static equilibrium. Because the entrant objective can be flat in symmetric configurations, this procedure induces deterministic tie-breaking and can generate path dependence (Figure 29).

BASELINE PARAMETERS. Unless otherwise stated we use $(\alpha, \gamma, t, c, \overline{p}) = (5, 1, 1, 0.5, 3.5)$ and vary β as indicated in each figure.

A.2 *Additional Numerical Illustrations*

Figures 27–29 provide complementary numerical illustrations for the comparative statics and the sequential entry mechanism discussed in the main text.

COMPARATIVE STATICS Figure 27 reports numerical comparative statics of equilibrium prices with respect to $(\beta, \gamma, t, \alpha)$, holding firm locations fixed under uniform density. This complements Proposition 3.12 and the proof sketch in Appendix C.

ENTRY COSTS AND DISCOUNTING To illustrate how discounting rescales the entry threshold, Figure 28 plots $N^*(F)$ for different δ under the heuristic rule $\Pi^e/(1-\delta) > F$, where Π^e denotes the entrant's post-entry profit along a baseline entry path. This is not the equilib-

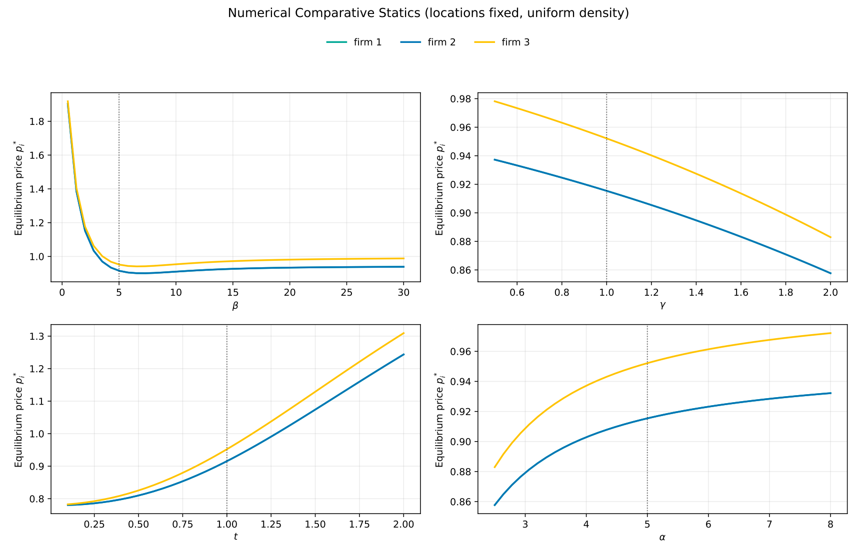


Figure 27: Numerical comparative statics for equilibrium prices under uniform density and fixed locations. Each panel varies one parameter while holding the others at the baseline values.

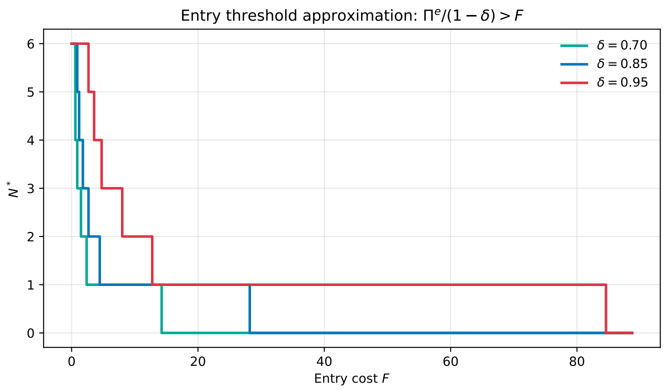


Figure 28: Illustrative effect of discounting on the entry threshold under the heuristic condition $\Pi^e/(1-\delta) > F$.

rium condition of the full dynamic game, but it helps interpret the role of δ.

MULTIPLICITY AND PATH DEPENDENCE Figure 29 illustrates how multiple optimal entry locations can arise in symmetric configurations. Starting from the same set of incumbents, different deterministic tie-breaking in the candidate search can lead to distinct final configurations.

A.3 *Regularity Diagnostics for Assumption (R)*

Table 12 reports $\max_i \eta_i(\mathbf{p}^*)$ at the computed equilibrium prices, together with a conservative own-price scan η^{scan} on a one-dimensional

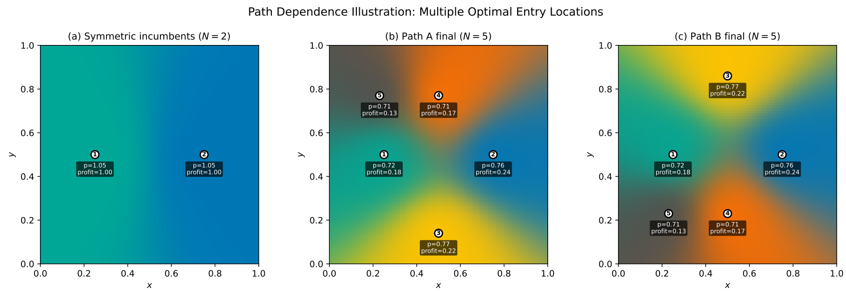


Figure 29: Path dependence illustration (uniform density). Panels (b) and (c) show two distinct final configurations obtained from the same incumbents under different deterministic tie-breaking in the entrant's candidate search.

Table 12: Regularity diagnostics for Assumption (R).

Scenario	N	$\max_i \eta_i(\mathbf{p}^*)$	$\max_i \eta_i^{\text{scan}}$
Duopoly, uniform, $\beta = 5$	2	0.2513	0.9884
Triopoly, uniform, $\beta = 5$	3	0.5757	0.9922
Entry eq., uniform, $N = 6$, $\beta = 15$	6	0.8097	1.1255
Triopoly, uniform, $\beta = 50$ (stress test)	3	0.7568	2.4116
Triopoly close, $\beta = 50$ (stress test)	3	0.4781	1.0000

Notes: $\eta_i(\mathbf{p}) = \frac{Q_i(\mathbf{p})\,Q_i''(\mathbf{p})}{(Q_i'(\mathbf{p}))^2}$ as defined in Appendix B. We define $\eta_i^{\text{scan}} \equiv \max_{p_i \in [c,\overline{p}]} \eta_i(p_i, \mathbf{p}^*_{-i})$ (computed on a one-dimensional grid), holding rivals fixed at $\mathbf{p}^*_{-i}$. This is a conservative diagnostic and is not a test of Assumption 3.17 (which conditions on interior critical points and must hold for all $\mathbf{p}_{-i}$). Hence $\eta_i^{\text{scan}} > 2$ does not by itself imply a violation of (R); it indicates that the curvature ratio can exceed 2 away from critical points in extreme regimes (high β).

grid $p_i \in [c, \overline{p}]$ holding $\mathbf{p}_{-i} = \mathbf{p}^*_{-i}$ fixed. The scan is a diagnostic for numerical regularity and should not be interpreted as a formal test of Assumption 3.17, which conditions on interior critical points.

B EXTENDED PROOF OF THEOREM 1: EXISTENCE OF PRICE EQUILIBRIUM

We apply Glicksberg's fixed point theorem (1952) to establish existence of a pure-strategy Nash equilibrium.

COMPACT, CONVEX STRATEGY SETS Each firm $i \in \mathcal{N}$ chooses $p_i \in [c, \overline{p}]$. This interval is compact and convex. The assumption on $\overline{p}$ is that $\overline{p} < \min_{i,s}(\alpha/\gamma - t \cdot d(s, x_i))$ and $c < \overline{p}$, ensuring $q_i(s) = \alpha - \gamma P_i^*(s) > 0$ for all $s \in S$ and $p_i \in [c, \overline{p}]$.

CONTINUITY OF PROFIT FUNCTIONS $\Pi_i(\mathbf{p}) = (p_i - c)Q_i(\mathbf{p})$. The firm's i demand is $Q_i(\mathbf{p}) = \int_S \rho(s)\,(\alpha - \gamma P_i^*(s))\,\text{Prob}\big(i \mid s, \mathbf{p}\big)\,ds$. Notice that:

1. $P_j^*(s) = p_j + t \cdot d(s, x_j)$ is continuous in p_j.
2. $q_i(s) = \alpha - \gamma P_i^*(s)$ is continuous in p_i (and positive by assumption on $\overline{p}$).
3. $\text{Prob}(i|s, \mathbf{p}) = \exp(-\beta P_i^*(s)) / \sum_k \exp(-\beta P_k^*(s))$ is continuous in all p_k. Denominator is strictly positive.
4. The integrand $\rho(s) q_i(s) \text{Prob}(i|s, \mathbf{p})$ is continuous in $\mathbf{p}$ and bounded on $S \times [c, \overline{p}]^N$ (since ρ bounded, $q_i(s) \in (0, \alpha]$, $\text{Prob} \in [0, 1]$).
5. By dominated convergence, $Q_i(\mathbf{p})$ is continuous in $\mathbf{p}$.
6. $\Pi_i(\mathbf{p})$ is continuous as a product of continuous functions.

QUASI-CONCAVITY IN OWN PRICE Glicksberg's theorem requires each $\Pi_i(\mathbf{p})$ to be quasi-concave in p_i. This is imposed in Assumption 3.2. A sufficient condition can be stated in terms of derivatives of Q_i.

Fix firm i and rivals' prices $\mathbf{p}_{-i}$. Write $\Pi_i(p_i, \mathbf{p}_{-i}) = (p_i - c)Q_i(p_i, \mathbf{p}_{-i})$, where

$$Q_i(\mathbf{p}) = \int_S \rho(s)\, q_i(s, \mathbf{p})\, L_i(s, \mathbf{p})\, ds, \qquad q_i(s, \mathbf{p}) = \alpha - \gamma P_i^*(s), \qquad L_i(s, \mathbf{p}) = \text{Prob}(i \mid s, \mathbf{p}).$$

Under our assumptions, differentiation under the integral sign is justified. Using $\partial L_i / \partial p_i = -\beta L_i (1 - L_i)$ we obtain:

$$Q_i'(\mathbf{p}) = -\int_S \rho(s) \Big[\gamma L_i(s, \mathbf{p}) + \beta q_i(s, \mathbf{p}) L_i(s, \mathbf{p})(1 - L_i(s, \mathbf{p}))\Big]\, ds \;<\; 0,$$
$$Q_i''(\mathbf{p}) = \beta \int_S \rho(s)\, L_i(s, \mathbf{p})(1 - L_i(s, \mathbf{p})) \Big[2\gamma + \beta q_i(s, \mathbf{p})(1 - 2L_i(s, \mathbf{p}))\Big]\, ds.$$

At an interior critical point $\partial \Pi_i(p_i, \mathbf{p}_{-i}) / \partial p_i = 0$, we have $p_i - c = -Q_i / Q_i'$, and

$$\Pi_i''(p_i, \mathbf{p}_{-i}) = \frac{2(Q_i')^2 - Q_i Q_i''}{Q_i'}.$$

Since $Q_i' < 0$, this critical point is a strict local maximum whenever $2(Q_i')^2 > Q_i Q_i''$. Define the curvature ratio

$$\eta_i(\mathbf{p}) \;\equiv\; \frac{Q_i(\mathbf{p})\, Q_i''(\mathbf{p})}{(Q_i'(\mathbf{p}))^2}.$$

Then the strict second-order condition at any interior critical point is equivalent to $\eta_i(\mathbf{p}) < 2$.

Assumption 3.17 (Regularity condition (R))**.** For every firm i, every $\mathbf{p}_{-i}$, and every interior $p_i \in (c, \overline{p})$ such that $\partial \Pi_i(p_i, \mathbf{p}_{-i}) / \partial p_i = 0$, we have $\eta_i(p_i, \mathbf{p}_{-i}) < 2$.

Under Assumption 3.17, every interior critical point of $\Pi_i(\cdot, \mathbf{p}_{-i})$ is a strict local maximum, so $\Pi_i(\cdot, \mathbf{p}_{-i})$ is quasi-concave on $[c, \overline{p}]$. In particular, Assumption 3.2 holds.

Remark 3.18. In the monopoly case $N = 1$, we have $L_1(\cdot, \mathbf{p}) \equiv 1$, hence $Q_1'' \equiv 0$ and $\eta_1 \equiv 0$, so Assumption 3.17 is automatically satisfied.

APPLICATION OF GLICKSBERG'S THEOREM All conditions are met, so at least one pure-strategy Nash equilibrium exists.

C COMPARATIVE STATICS (PROOF SKETCH)

This appendix sketches the proofs for Proposition 3.12 and explains the economics behind each result. A complete sign characterization in full generality requires additional restrictions on primitives (or symmetry): the Jacobian and cross-effects are hard to sign in a general two-dimensional configuration.

C.1 *Preliminaries*

Let $\mathbf{p}^*$ denote the equilibrium price vector. At equilibrium, each firm's first-order condition must be satisfied:

$$F_i(\mathbf{p}^*) = \frac{\partial \Pi_i(\mathbf{p}^*)}{\partial p_i} = Q_i(\mathbf{p}^*) + (p_i^* - c)\frac{\partial Q_i(\mathbf{p}^*)}{\partial p_i} = 0 \quad \forall i \in \mathcal{N} \tag{38}$$

To analyze how equilibrium prices change with respect to a parameter $\theta \in \{\beta, \gamma, t, \alpha\}$, we apply the implicit function theorem to the system of equations $F_i(\mathbf{p}^*, \theta) = 0$. Define the Jacobian matrix $J = [\frac{\partial F_i}{\partial p_j}]_{i,j}$. The changes in equilibrium prices with respect to θ are given by:

$$\frac{d\mathbf{p}^*}{d\theta} = -J^{-1} \cdot \frac{\partial \mathbf{F}}{\partial \theta} \tag{39}$$

where $\frac{\partial \mathbf{F}}{\partial \theta} = [\frac{\partial F_i}{\partial \theta}]_i$ is the vector of partial derivatives with respect to θ.

Under conditions where the Jacobian J is invertible, and if all elements of $\frac{\partial \mathbf{F}}{\partial \theta}$ have the same sign, and the off-diagonal elements of $-J^{-1}$ are non-negative, the corresponding elements of $\frac{d\mathbf{p}^*}{d\theta}$ will have the same sign as $-\frac{\partial F_i}{\partial \theta}$. This allows us to determine the direction of price changes.

C.2 *Effects of Changing β (Choice Intensity)*

For the choice intensity parameter β, we have:

$$\frac{\partial F_i}{\partial \beta} = \frac{\partial Q_i}{\partial \beta} + (p_i^* - c)\frac{\partial^2 Q_i}{\partial p_i \partial \beta} \tag{40}$$

As β increases, consumers become more price-sensitive, making $\frac{\partial Q_i}{\partial p_i}$ more negative. This increased elasticity pushes firms to reduce prices.

At equilibrium, this effect typically dominates:

$$\frac{\partial F_i}{\partial \beta} < 0 \tag{41}$$

If so for all i:

$$\frac{dp_i^*}{d\beta} < 0 \tag{42}$$

C.3 *Effects of Changing γ (Price Sensitivity)*

For the demand price sensitivity parameter γ, we have:

$$\frac{\partial F_i}{\partial \gamma} = \frac{\partial Q_i}{\partial \gamma} + (p_i^* - c)\frac{\partial^2 Q_i}{\partial p_i \partial \gamma} \tag{43}$$

As γ increases, each consumer's quantity demanded falls for a given effective price; demand also becomes more price-responsive.

Both effects push in the same direction:

$$\frac{\partial F_i}{\partial \gamma} < 0 \tag{44}$$

Under the same condition:

$$\frac{dp_i^*}{d\gamma} < 0 \tag{45}$$

C.4 *Effects of Changing t (Transportation Cost)*

As t increases, spatial differentiation becomes more important relative to price differences, so demand becomes less price-elastic. This is the dominant effect:

$$\frac{\partial F_i}{\partial t} > 0 \tag{46}$$

Then:

$$\frac{dp_i^*}{dt} > 0 \tag{47}$$

C.5 *Effects of Changing α (Base Demand)*

Higher α raises quantity demanded proportionally across all firms:

$$\frac{\partial F_i}{\partial \alpha} > 0 \tag{48}$$

Hence:

$$\frac{dp_i^*}{d\alpha} > 0 \tag{49}$$

D ANALYSIS OF DYNAMIC ENTRY GAME

This appendix proves the existence results for the dynamic entry game stated in Section 3.3 and explains why multiple equilibria typically arise.

D.1 *Proof of Theorem 3.9: Existence of SPNE*

The proof proceeds in three steps:

Step 1: Bounded Number of Firms

We first establish that the entry process eventually terminates at some finite number of firms.

Lemma 3.19 (Bounded Entry)**.** *There exists a finite upper bound $\bar{N}$ such that if $N_\tau \geqslant \bar{N}$, no further entry will occur in any SPNE.*

Proof. Fix any number of firms $N \geqslant 1$, any location configuration $\mathbf{x}_N$, and any price vector $\mathbf{p} \in [c, \overline{p}]^N$. For each $s \in S$ and each $i \in \{1, \dots, N\}$, the effective price satisfies $c \leqslant P_i^*(s) \leqslant \overline{p} + t\overline{d}$, where $\overline{d} = \max_{s,x \in S} d(s, x)$. Hence the logit share obeys the uniform bound

$$L_i(s, \mathbf{p}) \;=\; \frac{\exp(-\beta P_i^*(s))}{\sum_{j=1}^{N} \exp(-\beta P_j^*(s))} \;\leqslant\; \frac{\exp(-\beta c)}{N \,\exp(-\beta(\overline{p} + t\overline{d}))} \;=\; \frac{K}{N},$$

where $K \equiv \exp(\beta(\overline{p} + t\overline{d} - c))$ is finite and independent of N and $\mathbf{x}_N$.

Moreover, individual demand satisfies $q_i(s, \mathbf{p}) = \alpha - \gamma P_i^*(s) \leqslant \alpha - \gamma c$. Therefore,

$$Q_i(\mathbf{p}) \;=\; \int_S \rho(s)\, q_i(s, \mathbf{p})\, L_i(s, \mathbf{p})\, ds \;\leqslant\; (\alpha - \gamma c)\, M\, \frac{K}{N},$$

and for each firm i,

$$\Pi_i(\mathbf{p}) \;=\; (p_i - c)\, Q_i(\mathbf{p}) \;\leqslant\; (\overline{p} - c)\, (\alpha - \gamma c)\, M\, \frac{K}{N} \;=\; \frac{C}{N},$$

where $C \equiv (\overline{p} - c)(\alpha - \gamma c)MK$.

Consider now an entry stage with N incumbents. If a new firm enters, the subsequent price subgame has $N + 1$ firms, so its one-period equilibrium profit is bounded above by $C/(N + 1)$. Since N can only increase over time, its continuation value from entry is bounded by the perpetuity $C/((N + 1)(1 - \delta))$. Choose

$$\bar{N} \;\equiv\; \left\lceil \frac{C}{F(1 - \delta)} \right\rceil + 1.$$

Then for every $N \geqslant \bar{N}$, any entrant's continuation value is strictly below F, so entry is unprofitable. Therefore, no further entry occurs once $N_\tau \geqslant \bar{N}$. □

Step 2: Existence of Optimal Strategies at Each Stage

Lemma 3.20 (Price Equilibrium Existence)**.** *For any number of firms* N *and location configuration* $\mathbf{x}_N$*, a pure-strategy Nash equilibrium of the price-setting game exists.*

Proof. This follows directly from Theorem 3.3. The price-setting game satisfies the conditions of Glicksberg's theorem. □

Lemma 3.21 (Optimal Location Choice)**.** *For any configuration of incumbent firms* $\mathbf{x}_{inc}$*, if entry is profitable, there exists an optimal location choice for the entrant.*

Proof. Fix incumbents' locations $\mathbf{x}_{inc}$ with N firms. For each prospective entrant location $x_E \in S$, consider the $(N+1)$-firm price subgame with location vector $(\mathbf{x}_{inc}, x_E)$, and let $\mathcal{E}(\mathbf{x}_{inc}, x_E) \subset [c, \overline{p}]^{N+1}$ denote its set of Nash equilibria. By Theorem 3.3, $\mathcal{E}(\mathbf{x}_{inc}, x_E)$ is non-empty for every x_E; it is compact because it is a closed subset of the compact set $[c, \overline{p}]^{N+1}$. Moreover, $\mathcal{E}(\mathbf{x}_{inc}, \cdot)$ has a closed graph (equilibrium conditions are preserved under limits by continuity of payoffs), so it is upper hemi-continuous.

Define the entrant's equilibrium-profit value function as the maximum profit over price equilibria,

$$V_E(\mathbf{x}_{inc}, x_E) \equiv \max_{\mathbf{p} \in \mathcal{E}(\mathbf{x}_{inc}, x_E)} \Pi_{N+1}(\mathbf{p}; \mathbf{x}_{inc}, x_E).$$

The profit function $\Pi_{N+1}(\mathbf{p}; \mathbf{x}_{inc}, x_E)$ is continuous in $(\mathbf{p}, x_E)$ by dominated convergence (the integrand is continuous in distances and uniformly bounded on $S \times [c, \overline{p}]^{N+1}$). Berge's maximum theorem then implies that $V_E(\mathbf{x}_{inc}, \cdot)$ is upper semi-continuous and attains a maximum on the compact set S. Hence there exists an optimal entrant location $x_E^* \in \arg\max_{x_E \in S} V_E(\mathbf{x}_{inc}, x_E)$. □

Step 3: Construction of SPNE

We construct an SPNE by backward induction on the number of incumbents, using the upper bound $\bar{N}$ from Lemma 1.

TERMINAL CASE ($N \geqslant \bar{N}$). By Lemma 1, when $N \geqslant \bar{N}$ no further entrant can recoup the fixed cost, so "do not enter" is optimal in every entry subgame. In every price subgame, a Nash equilibrium exists by Theorem 3.3; playing a Nash equilibrium in each price subgame is sequentially rational. Therefore an SPNE exists for every subgame with $N \geqslant \bar{N}$ incumbents.

INDUCTIVE STEP. Fix $N < \bar{N}$ and suppose that for every location configuration $\mathbf{x}_{N+1}$ there exists an SPNE of the continuation game starting from the next entry stage with $N+1$ incumbents. Consider any subgame that begins at an entry stage with N incumbents located at $\mathbf{x}_N$. If the potential entrant stays out, its payoff is 0. If it enters and chooses $x_E \in S$, the subsequent price subgame has $N+1$ firms and admits a Nash equilibrium by Theorem 3.3; thereafter, play follows an SPNE of the continuation game by the inductive hypothesis. Let $W_E(\mathbf{x}_N, x_E)$ denote the entrant's continuation value from entering at x_E (net of the fixed cost F) under this continuation play. By Assumption 3.4 and compactness of S, the set of optimal entry locations $X_E^*(\mathbf{x}_N) \equiv \arg\max_{x_E \in S} W_E(\mathbf{x}_N, x_E)$ is non-empty. To complete the strategy definition, we employ a deterministic tie-breaking rule (e.g., lexicographic selection) to select a unique $x_E^* \in X_E^*(\mathbf{x}_N)$. The strategy profile is then: the entrant enters at x_E^* if $W_E(\mathbf{x}_N, x_E^*) > 0$ (and stays out otherwise); if entry occurs, play proceeds according to the inductively defined SPNE. This construction satisfies sequential rationality at the entry stage and, by hypothesis, in all subsequent subgames.

Combining these strategies yields an SPNE for the subgame with N incumbents. By induction, an SPNE exists starting from the initial empty market.

D.2 *Proof of Theorem 3.10: Multiplicity of Equilibria*

We now explain why multiple SPNE typically exist. The argument is constructive: we exhibit three independent sources of multiplicity.

Source 1: Multiple Optimal Locations

Lemma 3.22 (Location Multiplicity). *There exist market configurations where an entrant has multiple optimal locations that yield equivalent profits.*

Proof. Consider a symmetric case with $\rho(s)$ uniform on a disc S, and two incumbents located at opposite sides of the disc's circumference. By symmetry, a third entrant has at least two optimal locations that yield identical profits. Each of these location choices leads to a different market configuration and affects the entry decisions and optimal locations for subsequent entrants. □

Source 2: Path Dependence

Lemma 3.23 (Path Dependence). *Different sequences of entry decisions can lead to different equilibrium market structures.*

Proof. Let $\mathbf{x}_A = (x_1^A, x_2^A, \dots, x_{N_A}^A)$ and $\mathbf{x}_B = (x_1^B, x_2^B, \dots, x_{N_B}^B)$ be two different location configurations reached through different entry sequences. Because locations affect equilibrium profits, the continuation value of entry $\max_{x_E \in S} W_E(\mathbf{x}, x_E)$ can differ across configurations. It is therefore possible that $\max_{x_E \in S} W_E(\mathbf{x}_A, x_E) > 0$ while $\max_{x_E \in S} W_E(\mathbf{x}_B, x_E) < 0$, meaning entry continues in path A but stops in path B. This leads to equilibria with different numbers of firms. □

These sources of multiplicity establish that the entry-location-price game typically admits multiple SPNE, completing the proof of Theorem 3.10.

D.3 Market Structure Properties

We now provide the reasoning for the comparative statics on market structure presented in Proposition 3.11.

Proposition 3.24 (Entry Cost Effect)**.** *The equilibrium number of firms* N^* *is non-increasing in the entry cost* F*.*

Proof. Entry occurs in an entry subgame if and only if $\max_{x_E \in S} W_E(\mathbf{x}_{N-1}, x_E) > 0$, where the fixed cost enters the continuation value subtractively. Increasing F weakly decreases W_E in every subgame, which cannot increase the number of firms that enter in equilibrium. □

Proposition 3.25 (Market Size Effect)**.** *The equilibrium number of firms* N^* *is non-decreasing in the market size* $M = \int_S \rho(s)ds$*.*

Proof. A proportional increase in $\rho(s)$ across all locations increases per-period demands and profits proportionally. This raises entrants' continuation values W_E, so entry becomes more likely for a given number of incumbents and equilibrium market size may increase. □

Proposition 3.26 (Competition Parameters)**.** *The equilibrium number of firms* N^* *is:*

1. *Non-increasing in* β *(choice intensity parameter)*
2. *Non-increasing in* γ *(demand price sensitivity)*
3. *Non-decreasing in* t *(transportation cost parameter)*

Proof. From Section C, the comparative statics discussion suggests that:

- Higher β and γ intensify price competition, reducing equilibrium prices and profits
- Higher t softens price competition, increasing equilibrium prices and profits

Since the entry decision is based on the sign of $\max_{x_E \in S} W_E(\mathbf{x}_{N-1}, x_E)$, lower profits make entry harder to justify, while higher profits make it easier. The effects on N^* follow accordingly. □

REFERENCES

Barro, Robert J (2024). *Markups and entry in a circular hotelling model*. Tech. rep. National Bureau of Economic Research.

Ben-Akiva, M. and S. R. Lerman (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.

Brenner, Steffen (2005). "Hotelling Games with Three, Four, and More Players." In: *Journal of Regional Science* 45.4, pp. 851–864.

Cahan, David, Hsiao-Han Chen, Logan Christie, and Arkadii Slinko (2021). "Spatial Competition on 2-Dimensional Markets and Networks When Consumers Don't Always Go to the Closest Firm." In: *International Journal of Game Theory* 50.4, pp. 945–970.

Cox, Gary W. (1987). "Electoral Equilibrium under Alternative Voting Institutions." In: *American Journal of Political Science* 31.1, pp. 82–108.

D'Aspremont, Claude, Jean Jaskold Gabszewicz, and Jacques-François Thisse (1979). "On Hotelling's "Stability in competition"." In: *Econometrica: Journal of the Econometric Society* 47.5, pp. 1145–1150.

De Palma, Andre, V. Ginsburgh, Y. Y. Papageorgiou, and J. F. Thisse (1985). "The Principle of Minimum Differentiation Holds under Sufficient Heterogeneity." In: *Econometrica* 53.4, pp. 767–781.

De Serpa, Allan C. (1986). "Equilibrium Location in a Uniform Market: A Geometrical Approach." In: *Regional Science and Urban Economics* 16.4, pp. 535–542.

Dos Santos Ferreira, Rodolphe and Jacques-François Thisse (1996). "Horizontal and vertical differentiation: The Launhardt model." In: *International Journal of Industrial Organization* 14.4, pp. 485–506. ISSN: 0167-7187.

Drezner, Tammy and Zvi Drezner (2017). "Leader-Follower Models in Facility Location." In: *Spatial Interaction Models: Facility Location Using Game Theory*. Ed. by Z. Drezner and H. A. Eiselt. Springer, pp. 73–104.

Drezner, Zvi and Horst A. Eiselt (2024). "Competitive Location Models: A Review." In: *European Journal of Operational Research* 316.1, pp. 5–18.

Glicksberg, I. L. (1952). "A Further Generalization of the Kakutani Fixed Point Theorem, with Application to Nash Equilibrium Points." In: *Proceedings of the American Mathematical Society* 3.1, pp. 170–174.

Hotelling, Harold (1929). "Stability in Competition." In: *The Economic Journal* 39.153, pp. 41–57. ISSN: 00130133, 14680297.

Larralde, Hernan, Julien Stehle, and Pablo Jensen (2009). "Analytical Solution of a Multi-Dimensional Hotelling Model with Quadratic Transportation Costs." In: *Regional Science and Urban Economics* 39.3, pp. 343–349.

Launhardt, Wilhelm (1885). *Mathematische Begründung der Volkswirthschaftslehre*. W. Engelmann.

McFadden, Daniel (1974). “Conditional Logit Analysis of Qualitative Choice Behavior.” In: *Frontiers of Econometrics*. Ed. by Peter Zarembka. New York: Academic Press, pp. 105–142.

Neven, Damien J. (1986). “On Hotelling’s competition with non-uniform customer distributions.” In: *Economics Letters* 21.2, pp. 121–126.

Rashid, M. (2016). “Generalising Bertrand competition: a special case of the Hotelling model.” In: *Economic Theory Bulletin* 4.2, pp. 187–198.

Salop, Steven C. (1979). “Monopolistic Competition with Outside Goods.” In: *Bell Journal of Economics* 10.1, pp. 141–156.

Tsai, Pei-Lin and Chin-Diew Lai (2005). “Location Choice in a Triangular Market.” In: *Annals of Regional Science* 39.1, pp. 137–151.

Wrede, Matthias (2015). “A continuous logit Hotelling model with endogenous locations of consumers.” In: *Economics Letters* 126, pp. 81–83.