FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

The Diffusion of Institutional Innovation

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Business Informatics

by

Alexandre Tanner, BSc
Registration Number 1425161

to the Faculty of Informatics

at the TU Wien

Advisor: Ass.-Prof. Mag. Michael Filzmoser, PhD

Vienna, 22nd August, 2017

Alexandre Tanner | Michael Filzmoser

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.ac.at

Declaration of Authorship

Alexandre Tanner, BSc
1060 Vienna

I hereby declare that I have written this Thesis independently, that I have completely specified the utilized sources and resources and that I have definitely marked all parts of the work - including tables, maps and figures - which belong to other works or to the internet, literally or extracted, by referencing the source as borrowed.

Vienna, 22nd August, 2017

Alexandre Tanner

Acknowledgements

The day has come, the end of a seemingly never-ending diploma thesis and formal schooling. Thank you to Eva Tanner for proof reading this thesis, and to Thomas Auzinger for providing the LaTeX template. Last but not least, my gratitude goes to Michael Filzmoser for supervising this work over the past year.

Abstract

Institutions are society's *rules of the game* and are increasingly recognized as playing a crucial role in providing the background conditions for economic growth. An institutional innovation is a perceived improvement to these rules of the game. The literature on the diffusion of institutional innovation remains poorly explored and is often not differentiated from social or cultural innovations.

In this exploratory study, the diffusion duration, pattern, and process was analyzed and cross-compared for two major cases of institutional innovation — the international gold standard and state-driven compulsory education — from a historical perspective on temporal and spatial dimensions. We used standardized quantitative methods commonly employed in diffusion studies that allow cross-comparison with previously studied diffusion processes, by curve fitting time series data with a three-parameter logistic function and its linear transformation known as the Fisher-Pry transform, and then deriving the diffusion rate that can be compared.

Our results suggest that institutional innovation diffuses in the same pattern that technological innovation was shown to follow in past empirical studies: An s-shaped curve best modeled by a three-parameter logistic function. Institutional innovation is strongly susceptible to social barriers, and its diffusion is a function of the degree of influence of those opposing it. This work can help modelers by bounding the range of the diffusion rate of institutional innovation.

Kurzfassung

Institutionen bestimmen die *Spielregeln* der Gesellschaft und es wird allgemein anerkannt, dass sie zunehmend eine entscheidende Rolle bei der Schaffung von geeigneten Grundlagen für das Wirtschaftswachstum einnehmen. Eine institutionelle Innovation ist eine augenscheinliche Verbesserung dieser Spielregeln. In der Literatur bleibt die Verbreitung institutioneller Innovationen jedoch ungenügend erforscht und oft fehlt eine Abgrenzung zu sozialen oder kulturellen Innovationen.

In dieser Studie wurden Diffusionsdauer, Diffusionsmuster und der Diffusionsprozess analysiert und für zwei wesentliche Fälle institutioneller Innovation aus einer historischen Perspektive bezüglich Zeit- und Raumdimension untersucht: Der internationale Goldstandard und die staatliche Pflichtschuldbildung. Wir verwendeten standardisierte quantitative Methoden, die üblicherweise in Diffusionsstudien eingesetzt werden. Diese Methoden ermöglichen einen Vergleich mit bereits untersuchten Diffusionsprozessen. Dies kann durch Kurvenanpassung von Zeitreihendaten mit einer dreiparametrigen logistischen Funktion und ihrer linearen Transformation erstellt werden, die als Fisher-Pry-Transform bekannt ist. Daraus kann die Diffusionsrate abgeleitet und verglichen werden.

Die Resultate zeigen, dass institutionelle Innovation eine Diffusion in demgleichen Muster aufzeigt, wie dies auch in vergangenen empirischen Studien ermittelt wurde: Eine s-förmige Kurve, die am besten durch eine dreiparametrige logistische Funktion dargestellt werden kann. Institutionelle Innovationen sind sehr anfällig gegenüber sozialen Hindernissen. Ihre Diffusion ist abhängig von Macht und Einfluss ihrer Opposition. Die vorliegende Arbeit leistet einen Beitrag zur quantitativen Innovationsforschung, indem sie dem Spektrum der Diffusionsrate der institutionellen Innovation einen Rahmen setzt.

Contents

CHAPTER 1

Introduction

> *"Commerce and manufactures can seldom flourish long in any state which does not enjoy a regular administration of justice, in which the people do not feel themselves secure in the possession of their property, in which the faith of contracts is not supported by law, and in which the authority of the state is not supposed to be regularly employed in enforcing the payment of debts from all those who are able to pay. Commerce and manufactures, in short, can seldom flourish in any state in which there is not a certain degree of confidence in the justice of government."* [1, p.387]

- Adam Smith, *The Wealth of Nations*

Throughout history, why have some countries experienced prosperous economic development while others stalled? Part of the answer lies in *institutions* according to North[1], "Institutions have been devised by human beings to create order and reduce uncertainty." [2, p.97]. North argued that certain countries had little economic development because they lacked the institutional innovations which were adopted in richer countries, such as a legal system, with the power to enforce property rights and contracts [2]. These institutions shape incentives for economic growth and reduce market uncertainties. He compared institutions to the rules of a game, and organizations to the players [3]. Already back in the 18th century, Smith, in his magnum opus "The Wealth of Nations" [1] described the primordial importance of institutions in providing the conditions that allow economies to flourish.

Therefore, in the grand scheme of human activity, innovation and institutional evolutions are not optional pursuits. They are indispensable for the continued existence of our economic and social system as we constantly require readjusting our changing environment,

[1] Winner of the 1993 Nobel Prize in Economics.

while addressing the global challenges from our growing population demands. Unlike natural evolution, where beneficial mutations for survival are purely random, human innovations have the benefit of being targeted non-random adaptations.

From a business point of view, innovation can be defined as the process of transforming a new idea or invention into a service or product, which fulfills a market need [4]. However, according to Schumpeter [5], the radical innovations, which do not yet have a market demand, are even more important and an essential part of economic development. The acceptance of an innovation can be measured by its spread, and the percentage of the population who have adopted the innovation; this is also termed *diffusion* [6]. Diffusion is defined by Rogers [7] as the process in which an innovation spreads over time within a social system [7]. The diffusion pattern has been observed to follow logistic growth for technological innovations, which can be modeled by a logistic function; in the literature this logistic pattern is referred to as an s-shaped curve or simply *s-curve* [8, 9].

Innovations are not limited to technological changes; an innovation which affects the norms of society would be an institutional innovation, which is the type of innovation that we will study in this thesis. When combining Rogers' definition of innovation namely "an idea, practice, or object that is perceived as new by an individual or another unit of adoption." [7, p.27] and North's definition where "Institutions are the rules of the game in a society or, more formally, are the humanly devised constraints that shape human interaction." [10, p.3]. We define institutional innovation as a perceived novel change in the *rules of the game* by a unit of adoption. This can be a change in existing rules or the creation of new rules.

According to Grübler [11], institutional and organizational innovations may have been the driving forces for change because they provided the environment for technological changes and transformed the socio-economic situation of the general population. Throughout the last couple of centuries, the Western world has gone through successive waves of extensive institutional changes, with constant readjustments to the rules of the game, sometimes disruptive, and adapting to the altering environment stemming from the interrelated technological and organizational changes. Much of these changes, occurred only after the Industrial Revolution which first emerged in Britain. Britain was the earliest to adhere to strict rules of law and institutions that promoted economic growth [2] that may explain why the Industrial Revolution first arose there and not elsewhere.

An example of an institutional innovation is capitalism, which has spread throughout the world — in its modern form after the Industrial Revolution — and has substituted over its duration other types of economic systems. Capitalism was an institutional innovation because it was a new way of organizing society. Another example of a diffusing institutional innovation is the spread of democracy in the 19th century that has been progressively substituting autocracies: Namely, it came about only after an increased threat of an organized revolt against the elite, after the French Revolution, urbanization and industrialization of the West that increased the power and coordination of the masses [12]. After the wave of democratization in Europe, the 1800s saw a cluster of major institutional innovations diffuse coinciding with the strong growth of per capita incomes

[13]. Universal state-driven compulsory education systems diffused across Europe and the USA and, as a result of increased trade flows from the Industrial Revolution, a new international monetary regime was adopted by all major economies named the international gold standard. The gold standard helped finance the Western transition from an agrarian-based economy to an industrialized economy [14]. The compulsory education and gold standard cases will be the subject of our within-case analysis in Chapter 4. The 1800s also saw a major change in the *rules of the game* with the abolition of slavery in the British Empire, who was the dominant international economic power at the time.

1.1 Problem statement

Up to now, there has been little research on the diffusion of institutional innovations, with the majority of diffusion research confined to the diffusion of technological innovations. Institutional innovations were often not differentiated from *social* or *cultural* innovations. The reason we study historical cases is that natural and social processes have a tendency to follow a pattern, and modeling the process' pattern gives one the power of prediction. In the process of analyzing the diffusion process of historical cases, we intend to not only understand in what manner the diffusion proceeded, but also *why* the diffusion happened. To our knowledge, there is no literature on the diffusion pattern of institutional innovation, and no approximation of the diffusion duration (the number of years it takes to spread) of institutional innovations. With regards to future developments that require institutional changes, It would be of interest to have an approximation of the diffusion rate (speed) of institutional innovations in response to technological and socioeconomic changes.

In this thesis, in a pursuit to understand the dynamics and process of institutional innovation, we attempt to answer the following questions:

- *Is there a pattern in the diffusion of institutional innovation?*
- *If so, then can the diffusion pattern be closely modeled by the logistic function?*

It is not the objective of this work to develop a new theoretical framework on the diffusion of institutional innovation, but instead, our aim is an empirical research of major cases of institutional changes that integrates the diffusion theory literature, the historical context of the diffusion, and the quantitative framework developed by diffusion theory scholars at the International Institute for Applied Systems Analysis (IIASA)[2].

1.2 Outline

This thesis is divided into four chapters, including the introduction. In Chapter 2, we present the state of the art of diffusion theory, regarding temporal and spatial dimensions.

[2]IIASA is an interdisciplinary research institute in Laxenburg, Austria.

We also introduce in the same chapter, the descriptive models that we use to test our second research question, notably the three-parameter logistic function. In Chapter 3 we introduce our methodology for evaluating the research questions, which is a case study approach integrated with the quantitative framework used by IIASA scholars to study diffusion patterns. In Chapter 4 we present the results of our case study approach, through a within-case analysis and a cross-case analysis. Finally, we conclude our work in Chapter 5 with a summary of the results and a discussion on future work.

CHAPTER 2

Diffusion Theory

"Society is imitation, and imitation is a kind of somnambulism." [15, p.87]

- Gabriel Tarde, *The Laws of Imitation*

In this chapter, we present the necessary theoretical background on diffusion theory to allow the reader to become familiar with the necessary concepts of diffusion theory. This theoretical context includes the models used by diffusion researchers at IIASA in their quantitative framework of diffusion theory.

Diffusion research is a key field in innovation studies, Grübler defines it as the study of "the spread, adoption and effects of innovations within a social system" [16, p.26]. The main focus of this field was for a long time, concentrated on understanding the diffusion of technological innovations, but this is now being expanded to social changes. Due to the large volume of works and heterogeneity of the field, we only cover the literature that we deem most relevant to our research.

In Section 2.1, we present the state of the art of diffusion theory, starting with the common concepts, and then expanding to more detailed cases. In Section 2.2, we present the descriptive models that are used to measure the diffusion of innovation in this work, notably the s-shaped logistic function.

2.1 State of the art

In 1890, Tarde published "The laws of imitation". In it, he attempted to explain the universal importance of *imitation*. In present-day terms, *adoption* of an innovation is employed instead of imitation [7]. Tarde claimed everything in society is either invention or imitation, the latter being the main drive for ideas, beliefs, knowledge, and innovations to spread [15]. He observed a decimation of most inventions, with only a tenth of them being imitated. In Tarde's view, humans are naturally imitative; even enslaved by an imitative nature, to the point where they are not self-aware of their imitative actions. "Both the somnambulist and the social man are possessed by the illusion that their ideas, all of which have been suggested to them, are spontaneous." [15, p.77]. He describes the 'super-social' innovators, as those who momentarily awake from somnambulism. Crucially, he stated that imitation, which is like a self-spreading contagion, spreads in the same pattern as populations as defined in models by Malthus and Darwin, termed *geometric progression*. "A slight incline, a relatively sharp rise, and then a fresh modification of the slope until the plateau is reached." [15, p.127]. Since Tarde's writings, this pattern is commonly referred to, in diffusion theory, as the s-curve [7]. The act of imitation also plays a role in defining "intelligence" in the field of artificial intelligence. Alan Turing proposed "The imitation game" in [17] — if a machine is indistinguishable from a human in a text-based conversation, then it can be considered intelligent. Thus, the ability to imitate is a requirement to being regarded as an intelligent being.

The most influential scholar, in the field of diffusion studies, is Rogers. His 1962 book *Diffusion of innovations* constructed the basis of the theory of the diffusion of innovations, as a qualitative theoretical framework [7]. The book is now in its fifth edition, taking into account modern developments, such as the spread of the internet. Scholars at IIASA have been studying the diffusion of technological innovations for several decades.They have significantly contributed to the quantitative and qualitative framework of describing diffusion processes, with a focus on transport and energy-related innovations [16, 18, 19, 20].

According to Rogers and other scholars, the diffusion of technological innovation tends to follow an s-shaped curve, based on empirical evidence, when the number of adopters is plotted cumulatively, or as a bell shaped curve when plotted by frequency [7, 9, 21, 22, 23, 24]. The variation in the s-shape among the different processes is due to the differences in their rate of adoption. Grübler, who is among the leading scholars at IIASA − wrote: "The basic patterns of technological substitution and diffusion are largely invariant across a large and diverse set of historical examples." [19, p.260]. Rogers identified four key elements as reoccurring in every case of diffusion research:

1. *The innovation*:
   Defined as: "an idea, practice, or object that is perceived as new by an individual or other unit of adoption" [7, p.27]. The definition makes no mention of a positive effect, as an innovation can also have detrimental effects on the adopters or society.

   The bulk of diffusion studies has looked at the diffusion of technological innovation. In diffusion research, the term 'technological innovation' is used broadly to include not only material innovations but also knowledge, processes [19] and information innovations [7].

   Innovations sometimes diffuse as part of a group of interdependent innovations, these are called *technology clusters* and are defined by Rogers as "one or more distinguishable elements of technology that are perceived as being closely related." [7, p.28]. An example of a technology cluster is one that comprises oil, cars and petrochemicals [16]. An innovation is not necessarily used the same way by all adopters; there is a concept in diffusion studies called *re-invention* [7], where the innovation is modified from its original form during the adoption process to suit the adopter's needs. Tarde pointed out that imitations are modified when spreading; he gave the example of religious myths and languages, transforming over time, from nation to nation [15]. Grübler et al. talk about incremental changes to innovations after they have been adopted by the market [19], these incremental changes lead to better performance and lower costs over time. An innovation rarely remains the same during the whole duration of its diffusion. An example is the automobile, today they are very different from the initial ones that replaced horses, FM radio, seat belts and airbags are now standard options, and they will keep changing with the advent of the self-driving electric automobile. The re-invention can even occur when the innovation has reached a mature stage, such as is the case for electronic appliances (e.g. smart thermostats, connected washing machines and lights) that are currently being re-invented for the Internet of things (IoT).

2. *Communication channels*:
"Is the means by which messages get from one individual to another" [7, p.32]. Examples of communication channels are an individual's social network, mass media, or face-to-face. As explained in Tarde's theory, humans are imitative by nature, and exposition to new ideas through any communication channel is how those ideas will spread. Similar actors, regarding socioeconomic status and education, will adopt around the same time, due to fewer communication barriers between them [7]. A number of similar countries, in terms of culture and low communication barriers can turn them into a *country cluster*, countries with "similar diffusion patterns" [16, p.262].

3. *Time*:
It is the key dimension in the diffusion process through which we measure the number of adopters of an innovation, and with which we determine whether an individual (or another unit of adoption) is an early, or late adopter (see Figure 2.1). The rate of adoption is a *temporal* measure and will be further explained in section 2.1.1. Also relating to time, *timing* of the innovation is crucial [25]. The light bulb cannot successfully diffuse before the diffusion of electricity.

4. *Social system*:
Defined as "a set of interrelated units that are engaged in joint problem solving to accomplish a common goal" [7, p.38]. The unit's social system has an impact on the diffusion decision and rate [7], due to factors such as the system's *norms*, *opinion leaders* (those who persuade others to adopt), and the *type of innovation-decision*. The three types of innovation-decisions defined by Rogers [7] are: *optional innovation-decision*, in which the choice of adopting an innovation, is made by

> an individual independently of the other members of the system, some adopt and others do not; *collective innovation-decision*, where the decision to adopt is by agreement among all members of the system, either all or none adopt; and finally *authority innovation-decision*, where the decision to adopt an innovation is taken by a select elite of individuals, all must follow the decision taken. Technological innovations adopted by consumers tend to be optional decisions; organizational and institutional innovations tend to be collective or authority-decisions.

The recognition of a basic pattern facilitates prediction, which is a key topic in diffusion studies. Tarde speculated about the predictive power of the geometric progression of imitation. This would allow producers to predict the demand for their goods in future years, based on the sales in the current year [15]. According to Christensen [24], The s-curve can be used as a descriptive or predictive model for technological innovations at an industry level; therefore, it is a valuable tool in the field of strategic management for high-level analysis. In an empirical study, Debecker and Modis concluded that once the inflection point has been reached, the logistic function can be used to predict the upper asymptote of a growth process, within an accuracy of 20% with a 95% confidence interval [26]. Scholars at IIASA remain grounded on the forecasting power of their models: "Long-term technological forecasting still remains elusive; however, historical analysis and improved numerical modeling, together, can sharply increase the ability to anticipate technological changes and their environmental impacts." [19, p.248]. Grübler categorizes two types of processes for the diffusion of an innovation [16]: (i) *diffusion/growth* processes and *substitution/replacement* processes.

Growth processes are cases in which an innovation diffuses in a new market. Though Grübler admits that, in reality, no innovation diffuses in a vacuum, market dynamics imply that there will inevitably be competing forces, and for each case, the market environment should be analyzed [16]. Substitution processes involve the replacement of an existing technique by another. A common historical example is the replacement of horses by the automobile in the early 20th century. Once a technology has finished diffusing, it tends to be slowly substituted by newly emerging technologies [23, 19]. It is often the case that the distinction between diffusion and substitution processes is not clear cut, as there are often competing innovations over long periods. Grübler gives the example of the railroad, which starts out as a pure diffusion process in a new market, but over a period, also substituted stagecoaches [19].

2.1.1 Rate of adoption

Diffusion is not instant, and an innovation can only spread as fast as information can travel. Rogers defines the rate of adoption as the number of adopters over a specified time period. This is the measure of the speed at which, an innovation spreads. There is no widely used standardized measure among scholars for the diffusion process according to [8]. "It appears that relatively little systematic work has been carried out in the various diffusion research disciplines to analyze a diffusion process based on a systematic

Table 2.1: Examples of diffusion with corresponding diffusion rate Δt. Source: table adapted from [19, p.264].

Case	Δt (years)[1]	Diffusion (D) Substitution (S)
Growth of railways, World [16]	60	D
Growth of railways, France [16]	47	D
Stock options compensation for executives[2] [30]	42	D
Air conditioners in homes, Japan [31]	19	D
Cellphone subscriptions, Scandinavia and Japan [29]	16	D
Cars vs horses, USA [32]	12	D
Washing detergent vs. soap, USA [27]	9	S

[1] Δt rounded to closest integer.
[2] The Δt of using stock options as a top executive compensation is estimated from Figure 3 in [30, p.48].

and comparable set of measures of the temporal and spatial spread of an innovation." [8, p.457].

In an attempt to homogenize the field, Grübler et al. at IIASA have taken a reoccurring quantitative approach to measuring the speed of the different diffusion processes, by using the Δt measure, referred to by various names, commonly as the *diffusion rate*, *takeover time* by Fisher and Pry [27], or as *characteristic duration* by [28], which is the length in years, that an innovation takes to go from 10% to 90% of its duration [9, 29, 19, 8, 23, 16, 9]. The use of Δt facilitates a quick comparison of the diffusion rate between the processes studied; IIASA has case studies with Δt's as short as 13 days, for the diffusion process of a resistance to technology movement in 1830 [9], to a diffusion rate of 110 years for the replacement of traditional energy (fuelwood) by coal [19]. The Δt for the automobile to replace horses was a mere 12 years [19]. A Δt is categorized as long in [19], if it is over 40 years, as medium if it is between 40 and 20 years, and fast if it is under 20 years. Table 2.1, partially adopted from [19, p.264], with the addition of some more recent cases, lists the Δt of over a half dozen diffusion processes.

In [19], it is observed that substitution processes tend to be faster than growth processes because the replacement innovations are often compatible with the existing environment (network externalities) which was required for the initial technique to diffuse; such as the automobile benefiting from existing road networks which were built for horse carriages, "Extensive road networks existed long before the advent of the automobile." [16, p.127].

Tarde suggested that modern society and a larger human population leads to faster diffusion [15]. According to Rogers' theory [7], the rate of adoption depends on five determinants:

1. *The perceived attributes of an innovation.*
2. *Types of innovation-decision* (authoritative, collective or optional).

3. *Communication channels* (mass media, social networks, etc.).
4. *Nature of the social system* (norms).
5. *Extent of change agents' promotion efforts* (Opinion leaders).

The perceived attributes of an innovation are the most significant explanatory variables for the rate of adoption [7, 16]:

1. *Relative advantage*: The perceived advantage that the adopter gains from the innovation can be of economic nature, or social nature, such as prestige. The relative advantage is considered the most important attribute in predicting the rate of adoption. According to Mohr: "The quest for prestige rather than the quest for organizational effective or corporate profit motivates the adoption of most new programs and technologies" [33, p.126]. Profitable innovations and those requiring only small investments were adopted faster by firms according to Mansfield [22]. Rogers considers the lack of an immediate perceived advantage from *preventive innovations*, a reason for their observed slow adoption rate. The goal of a preventive innovation is to prevent a future negative event, an example of such is the automobile seat belt. Due to their regulatory nature, certain institutional innovations can be preventive innovations, such as the carbon tax. Finally, policies to increase the relative advantage, such as incentives or subsidies increase the number of adopters, but not sustainably [7].
2. *Compatibility*: The second most important quality, is the degree to which an innovation is perceived to be compatible regarding culture, values, previous experience, technology, and needs. The network effect is a fundamental element for innovations that increase in value with the number of adopters, such as the telephone [18]. In [34], the author considers the late adoption of compulsory education of southern US states as a cultural incompatibility of the southern legislators with the idea of providing African Americans with an education (see Section 2.1.3).
3. *Complexity*: Innovations that are complicated to understand or use, are adopted slower.
4. *Trialability*: Innovations that can be trialed before their adoption, are adopted faster, due to the lower degree of uncertainty for the adopter. The unit of adoption learns and makes a decision based on their own experience.
5. *Observability*: Innovations that are visible to non-adopters (such as learning from their neighbor) with observable results, are adopted faster. Roof solar-panels or wind turbines are examples of innovations with high observability.

One of Rogers' most influential contributions to the field, is his adopter stage categorization. He labeled the different adopter stages into five categories based on innovativeness [7]:

1. *Innovators*: first stage, 2.5% of adopters.
2. *Early adopters*: second stage, 13.5% of adopters.
3. *Early majority*: third stage, 34% of adopters.
4. *Late majority*: fourth stage, 34% of adopters.
5. *Laggards*: last stage, 16% of adopters.

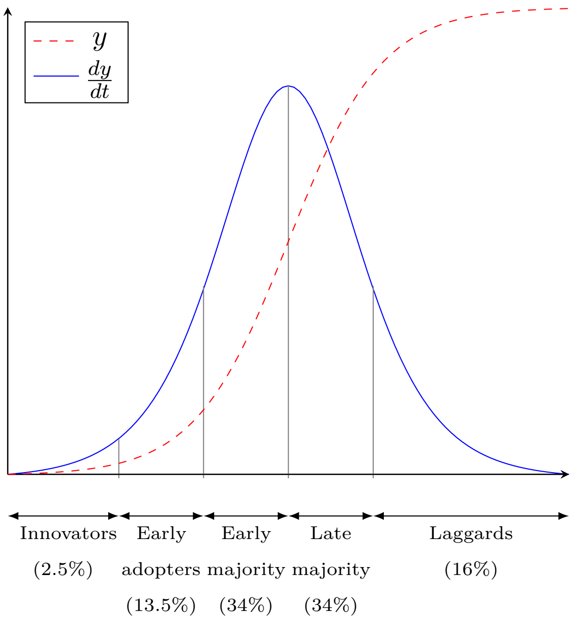


Figure 2.1: Rogers' adopter categorization [7] overlaid onto the sigmoidal s-curve (red) with its derivative $\frac{dy}{dt}$ (blue).

In Figure 2.1, we displayed the Rogers' categories under the simple logistic function and its derivative function which is a bell-shaped curve. In diffusion theory, this may be referred to as *Roger's curve*. The inflection point of the s-curve is the point in time where the bell curve (representing the frequency) is at its peak (see Section 2.2.2) and the point where the *early majority* turns becomes the *late majority*.

Looking at the stages from the technological side, IIASA scholars have categorized six stages of technological innovation, with the notable presence of the *invention* stage, which is distinct from the innovation stage, citing Schumpeter as an influence [19, p.3-4]:

1. *Invention*: 0% of market share.

2. *Innovation*: 0% of market share.
3. *Niche market commercialization*: 0-5% of market share.
4. *Pervasive diffusion*: 5-50% of market share.
5. *Saturation*: up to 100% of market share.
6. *Senescence*: declining market share.

How about factors that can negatively affect the rate of adoption? Grübler identified that *opposition to change* was a reoccurring factor throughout history which influences the rate of diffusion [9]. This resistance can effectively oppose innovations, by either rejecting those that are not socially sustainable or by forcing innovations to meet certain societal concerns. Glaziev suggests organizations as having a natural tendency to internally oppose innovation, due to their strive for stability [35]. Innovations add a degree of uncertainty, and therefore instability.

According to Grübler and Nakicenovic, the rate of adoption is slower but with a higher adoption rate among the initial adopting countries (first mover), while a faster rate, but smaller total adoption rate is observed in subsequent markets (follower). This can be termed the *catch-up effect*. Glaziev views this as an advantage, and called it, the *advantage of the backward* [35]: the laggards can benefit from the learning experience of earlier adopters, to efficiently use their resources to close the gap.

2.1.2 Spatial diffusion

There is a *spatial* dimension to diffusion: innovations spreads out geographically from their innovation center. Grübler states that the innovation tends to grow slower but with a higher adoption rate in the initial market of the innovation, then subsequently, having faster growth but a smaller adoptive rate in subsequent markets [9]. Torsten Hägerstrand – who conceptualized much of the spatial diffusion theory [36, 37] – and Gould – in a seminal report on spatial diffusion – used the wave analogy: "Like all waves, innovation pulses across a landscape tend to lose their strength with distance from the source of the disturbance." [38, p.17].

There are four categories for spatial diffusion [38]:

1. *Expansion diffusion*:
   This common type of process describes the spread of an idea through a population, the total number of knowers increases over time.
2. *Relocation diffusion*:
   This process concerns the type of process, in which the carriers move over time to a new location; an example is the process of urbanization.

	Contagious	**Hierarchical**
Expansion	Ideas and innovations at local level; diseases; cooperative societies.	Ideas, innovations, fads through urban and central place structures.
Relocation	Migration waves; hollow frontiers.	Movements of scholars in universities; transfers of students; “stepping stone” migration.

Figure 2.2: Cross-classification of the types of spatial diffusion processes. Source: [38, p.15]

3. *Contagious diffusion*:
   Diffusion processes, in which spatial proximity is a key factor. The spread starts from one or multiple innovation centers and diffuses like a wave. The wave analogy also conveys the concept of *neighborhood effect*, affecting those close-by, first.

4. *Hierarchical diffusion*:
   The processes, during which certain geographical areas or groups of people are leapfrogged. Those at the top of a social hierarchy are the early adopters, such as the wealthy.

A diffusion process is not strictly restricted to one category. A process can be a hierarchical and expansion process, such as the spread of the internet, which at first, was confined to a community of scholars, before trickling down to all echelons of society. Gould established a two-dimensional classification for the processes in Figure 2.2. On one dimension, it differentiates *whether* the process increases the number of adopters (expansion) or not (relocation); and on the second dimension, it differentiates *how* the process diffuses spatially, whether spatial proximity (contagious) or hierarchical status (hierarchical) is the key factor.

In addition to the types of spatial diffusion processes, there are different types of *barriers* to spatial diffusion. These are in most cases permeable; they slow down the diffusion rate but do not block it completely. Gould listed two preeminent types of barriers [38]: first is the *physical* barrier, which can be *permeable*, *absorbing* or *reflecting*. Examples are mountains, deserts, or oceans: these barriers became more permeable with time by virtue of technological innovations; and secondly, *cultural* barriers, which comprises *linguistic*, *religious*, *political*, and *psychological* barriers.

These spatial diffusion processes can be analyzed on different geographical scales [38]:

- *Micro-level*:
  At this scale, the spatial diffusion process is considered at a person to person level.

- *Urban-level*:
  Individuals are aggregated into small areas, such as districts or towns.

- *Regional-level*:
  Aggregation of individuals up to a regional scale.

- *Macro-level*:
  The final scale considers the process at the national and international level. Institutional innovations are inclined to be analyzed on a macro scale, as we analyze the adoption process on a country basis.

Gould described the spatial diffusion of a major institutional innovation: the diffusion of *banking* in the undeveloped early America (1800s) [38], permitting the spread of local credit lines to fuel the economic development of a country, where a culture of entrepreneurial spirit reigned. The country's innovation center for banking was on the north east coast in 1810, spreading southward down the east-coast and westward from Philadelphia along the Ohio river. The spatial spread tended to follow in the same wave as the settlements, with banks flowing up to the frontier, in industrial centers and areas where agricultural plantations were based. The pattern was not a simple expansion west and southward, but also progressively filling in the leapfrogged areas, between the innovation center and frontier. "The overall patterns of banking adoption are not always sharp" [38, p.66]. The growth pattern was also not unidirectional, with many banks failing during economic hardships, due to an over-expansionary wave of banks during the growth years.

2.1.3 Diffusion of institutional innovation

Young in 2011 stated that: "in contrast with the literature on technological progress, relatively little is known about the ways in which new institutions are created and how they become established within a given social framework." [39, p.21285]. In his paper's evolutionary game theory model [39], he considers the diffusion of institutional (social) innovation as initially slow because of three main reasons:

1. *Lack of information* about the superiority of the innovation because of low observability.

2. The innovation requires *re-invention* because the original form does not work well in practice.

3. Institutional (social) innovations requires *coordination* with others and benefits from increasing returns, similar to the network externalities for technological innovations. "an individual who invents a new form of legal contract cannot simply institute it on his own: First, the other parties to the contract must enter into it, and second, the ability to enforce the contract will depend on its acceptance in society more generally" [39, p.21285].

The diffusion rate in his mathematical model, depends on three factors: the *network topology*, the *relative advantage* of the innovation, and the amount of *noise* (which impacts the *coordination*).

Young also proposed a complex *learning model* [40], to model the adoption of innovations by taking into account the *resistance level* of each agent. The idea of the learning model is that adoption by an individual depends on their resistance level and the experience of earlier adopters; some agents are more risk-adverse than others and will require much information from previous adopters before deciding to adopt. In contrast to the logistic function, this model allows super-exponential growth rates to be modeled in early stages. Young demonstrated his learning model in [41] to curve fit the s-shaped pattern in the diffusion of stock options as a form of compensation for executives. The curve and data exhibited a striking jump-start in the early stage due to a change in tax laws.

There is evidence among empirical studies for the s-curve pattern in the diffusion of institutional innovations. Pemberton concluded in a 1936 study that *cultural* diffusion followed an s-shaped diffusion process. He used the integral of the normal distribution (which has an s-shaped pattern, see Section 2.2.2) to model it. His cases of cultural diffusion included the diffusion of postage stamps, compulsory education laws in the USA, and the adoption of tax limitation laws [34]. Postage stamps can also be considered a technological innovation. Pemberton defines *cultural* diffusion, as encompassing the adoption process of technical and social innovations [42]. Culture has overlapping concepts in common with the definition of institution, which are *norms* or *customs.* Concerning the *normality* in the adoption pattern, he stated that the complex factors for an early adoption are counterbalanced by the factors causing a late adoption, resulting in an s-shape when plotted cumulatively. “The most probable time of trait acceptance is the middle or average time and the probabilities of frequencies in other time periods follow a normal distribution.” [34, p. 550].

Pemberton initially observed that when plotting the number US states who adopted compulsory education, the plot did not resemble the typical s-curve. It was only when separating the states into: northern and western; and southern states, that the clear s-shape pattern appeared. The author concluded that “the south represented an entirely different culture area.” [34, p.554]; and, according to him, the *cultural* difference was that the two populations were heterogeneous, due to the high number of African Americans living in the southern states. “Compulsory school laws to the southern legislator meant providing the same school facilities for Negro children as for whites. Such laws meant, too, the adoption of the principle that Negro children should have an education.” [34, p.554]. The factor of heterogeneous populations is explained in detail in Rogers’ diffusion theory [7]. The time gap, between the average adoption date of compulsory education between northern and southern states, is *30 years.* According to Bandiera et al. [43], this time gap can be quantitatively explained by the difference in the number of migrants between the different states. The states with a significant share of European migrants from countries without compulsory education in their countries of origin, had to adopt compulsory schooling laws earlier, to educate and culturally homogenize the migrants;

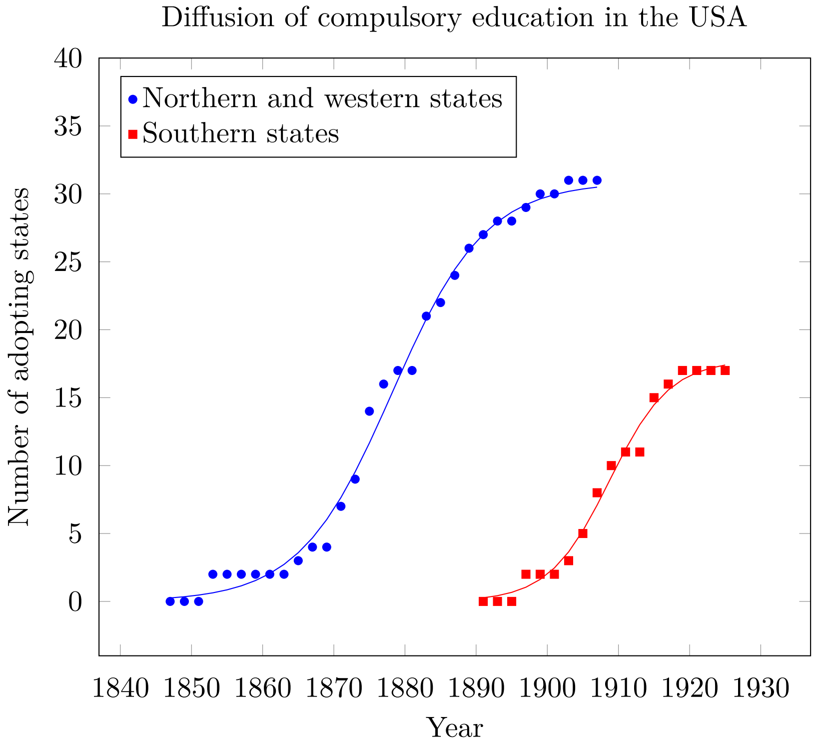


Figure 2.3: Diffusion of compulsory education among states in the USA. Time series from 1847 to 1925; two year intervals. Fitted 3 parameter logistic curve (Northern: $k = 30.871$, $b = 0.153$, $\Delta t = 28.66$ years and $t_m = 1878.261$; southern $k = 17.776$, $b = 0.236$, $\Delta t = 18.6$ years and $t_m = 1908.739$). Source: adapted from [34, p.555]; data source: [44, p.514-515]

these compulsory education laws were, therefore, a method of *nation-building*.

In Figure 2.3, we replotted the diffusion with the original data. For simplicity, we opted to plot, both northern and southern states on the same graph, and curve fit with the logistic function, and not with the integral of the normal distribution as it was originally done in [34]. Due to the typical s-shape of this case, the differences in the goodness-of-fit between the cumulative normal curve and logistic curve fittings would be trivial.

In a 1991 study of the diffusion of democracy[1] [45], the authors studied whether democratization follows the same pattern as technological innovation, specifically the technological substitution pattern as described by the Fisher-Pry model (see Section 2.2.3) [27]. “The evolution of new types of community, too, is a form of innovation.” [45, p.20]. They concluded the following: “Our analysis shows that the cumulative growth in the fraction of the world’s population that comprises its democratic communities conforms to the

[1]Criteria to be considered a democracy by [45], is an ’institutional democracy’ score of 6 or over (maximum is 10) in the POLITY II data set (`http://doi.org/10.3886/ICPSR09263.v1`).

Fisher-Pry model" [45, p.30]. The Fisher-Pry model (further explained in subsection 2.2.3) is a linearization of the s-curve, focusing on the substitution of one innovation (autocracy) by another (democracy). Therefore, if an innovation conforms to the Fisher-Pry model, then it will have an s-curve when plotted cumulatively over time. The data estimates that democratization remains in the early majority stage in 1987.

In (a) of Figure 2.4, we plotted the model, with the original data from the study of the diffusion of democracy by Modelski [45], and added the simple logistic model in (b), for a more familiar view of the diffusion. More evidence is required, before confirming the long-term conclusion derived in [45], as the Fisher-Pry model assumes that the substitution proceeds unilaterally, to completion, which is an optimistic view for the diffusion of democracy. The data shown in the study, up to 1987 does indeed conform to a Fisher-Pry model, which shows exponential growth during the early stage, ignoring the outlier, which is the growth of autocratic rule during World War II.

Their explanation for democracy, following the same pattern as technological innovations, was that one of the most important instruments of democracy, *collective choice* can be seen as a technological innovation which improves the well-being of its adopters, and its diffusion process will, therefore, follow in the same manner. They consider the process of democratization as long ($\Delta t = 176$ years), due to the slow nature of community change.

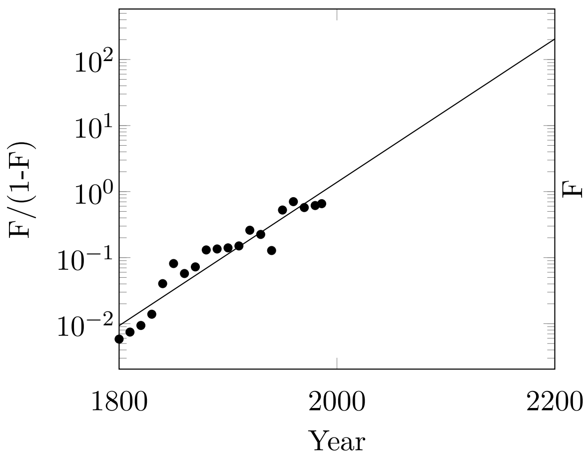
In their quantitative analysis, in 1986, only 40 percent of the world population lived under a democracy. They extrapolated the curve beyond 1986 to predict that 90% of the global population would be living under a democracy by 2075. Any long-term extrapolations before 60% has been reached remains speculative with the logistic function according to [23]. An interesting study would be to conduct a follow-up analysis of the updated data set from POLITY IV[2], which has data up to 2015. For a better curve fit, one could also split the curve into several successive waves, with the first wave ending after World War II, and the second wave peaking at the end of the Soviet Union.

2.2 The logistic function

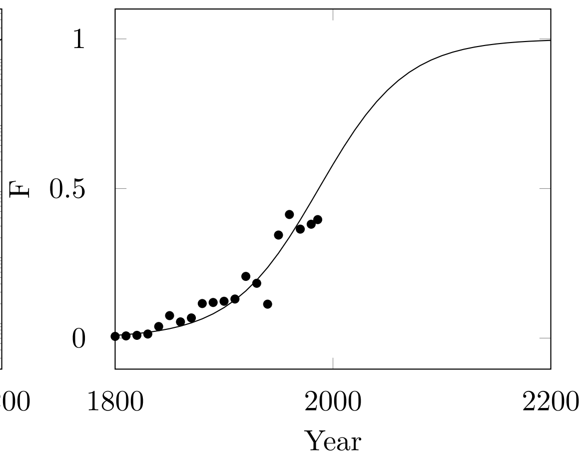
In this section we discuss in detail, the commonly used models for modeling the diffusion of innovations, they are also the ones that we used in this thesis to measure the diffusion of our cases of institutional innovation: the logistic function, and its variants, notably the Fisher-Pry transform.

The logistic function is a simple yet powerful nonlinear function with the following pattern: A slowly marginally accelerating rate of growth until reaching an inflection point, at this point the function goes from convex to concave in shape, and its rate of growth decreases until the function reaches an upper asymptote. For curve fitting s-shaped diffusion patterns, the logistic function is our model of choice.

[2] http://www.systemicpeace.org/inscrdata.html



(a) Fisher-Pry substitution model [45, p.29].



(b) Plotted as an s-shaped logistic curve.

Figure 2.4: Diffusion of democracy with data from 1800-1986, fraction of world population living in a democratic country in proportion to total world population, $F = y/K$, curve parameters: $b = 0.025$, $t_m = 1987$, $\Delta t = 176$ years, $K = 1$. Source: adapted from [45, p.29]

2.2.1 Background and application

The s-shaped curve function is used to model systems in a variety of fields. In this thesis, we will call it the *logistic function*, in reference to the original term, proposed by Belgian Mathematician Verhulst, who introduced it in [46] and [47] in his study of human population growth. Verhulst built on the writings of the Economist Malthus, who within the context of the French Revolution hypothesized that population growth would go ahead of the resources required to sustain that population. The British Economist was the first to come up with a model for understanding population dynamics and observed that “Population, when unchecked, goes on doubling itself every twenty-five years, or increases in a geometrical ratio.” [48, p.8].

The logistic function is referred to by different terms depending on the application and field, and some refer to a slight variation or a special case of the logistic function. The common names include S-Shaped curve or s-curve, sigmoidal function or sigmoid curve, growth function, generalized logistic function, hyperbolic tangent function (tanh) and autocatalytic growth-curve.

Its application encompasses many fields, well beyond the diffusion of innovation, and population growth. The inverse of the sigmoid function is the *logit* function which is commonly used in econometrics. In [49], botanists apply a variation of the sigmoid Function to model the growth of vegetation. In the field of biostatistics, Berkson [50], who is credited with introducing the logit model, suggested that the logistic curve was a suitable replacement for the integral of the normal probability function in modeling the lethal dosage of drugs, due to it being easier to compute. In chemistry, it is referred

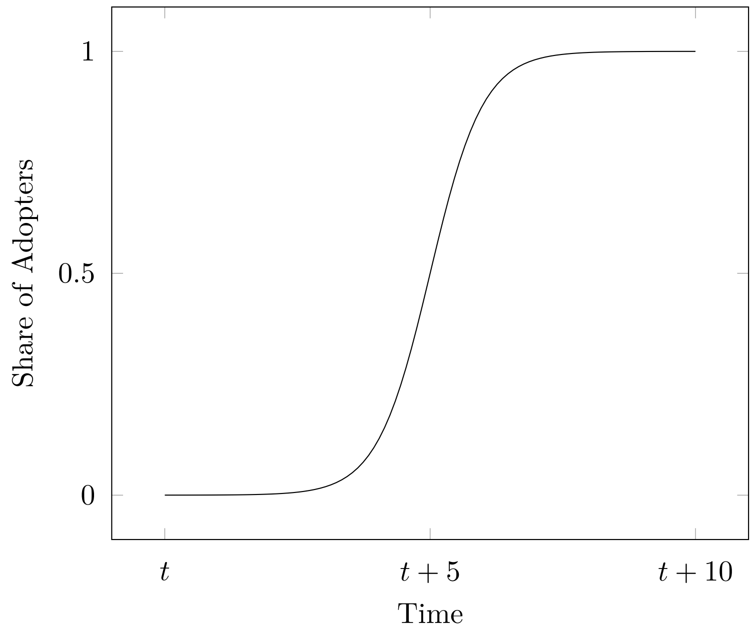


Figure 2.5: The logistic curve.

to as the autocatalytic growth-curve, because it is used to model the reaction time of autocatalyzed reactions. In the field of machine learning, the sigmoid function is used for input-output mapping of neurons [51]. A commonality of these processes is the rate at which information spreads [52] and the natural ceiling to their growth, often referred to as the *carrying capacity.* Exponential and linear models do not have this upper limit to growth, and therefore, they will poorly describe any biological process, in an environment of finite resources. The logistic function has a universal applicability to describe any natural growth process with the S pattern; it is also popular because special cases, like the standard logistic function, can be used as a probability distribution function.

2.2.2 Functions and parameters

The most basic form of the logistic function, is commonly referred to as the *sigmoid function* or *standard logistic function* (2.1). It has a value range of $[0, 1]$ (just like probability distributions). Therefore, it can be interpreted as a cumulative density function (CDF).

$$y(t) = \frac{1}{1 + e^{-t}} \tag{2.1}$$

Graphically, the sigmoid looks like the integral of the normal distribution, thus the derivative of a logistic function will resemble a Bell curve because the fundamental theorem of calculus states that the derivative of a CDF is a probability density function (PDF). The derivative allows us to analyze the rate of growth of a function at point t.

The derivative of a function, remains a function, the following is the sigmoid's first derivative function:

$$\frac{dy}{dt} = \frac{e^{-t}}{(1+e^{-t})^2} \tag{2.2}$$

A convenient property about the sigmoid's derivative function, is that it can be rewritten as simply:

$$\frac{dy}{dt} = y(t)(1 - y(t)) \tag{2.3}$$

This rewritten derivative function is referred to as the *hazard rate* in [40], it is the speed at which non-adopters become adopters. The ease of calculation of its derivative makes the logistic function a suitable choice for feed forward neural networks [53]. We will show how the simplified derivative equation from 2.3 can be derived.

Let $y(t) = \frac{1}{1+e^{-t}}$. So,

$$\begin{aligned}
\frac{dy}{dt} &= \frac{e^{-t}}{(1+e^{-t})^2} \\
&= \frac{1+e^{-t}-1}{(1+e^{-t})^2} && \text{(+1 and -1 to numerator)} \\
&= \frac{1+e^{-t}}{(1+e^{-t})^2} - \frac{1}{(1+e^{-t})^2} \\
&= \frac{1}{1+e^{-t}} - \frac{1}{(1+e^{-t})^2} \\
&= \frac{1}{1+e^{-t}} - \frac{1}{1+e^{-t}}\frac{1}{1+e^{-t}} \\
&= \frac{1}{1+e^{-t}}(1 - \frac{1}{1+e^{-t}}) && \text{(factorization)} \\
&= y(t)(1-y(t)) && (\text{since } y(t) = \frac{1}{1+e^{-t}})
\end{aligned}$$

The derivative of the logistic function is defined because the curve is continuous, without jumps, and doesn't have any sudden edges which would make the derivative undefined. It is non-negative for all t because the s-curve never decreases. In Figure 2.6, the global maximum is the point with the highest rate of growth, we will denote it with t_m, and we will also refer to it as the inflection point. To know whether the function is concave upwards or concave downwards, computing the second derivative is necessary.

The second derivative gives us information on how the rate of growth of our logistic function changes over time. A positive value means that the rate of growth is continuously

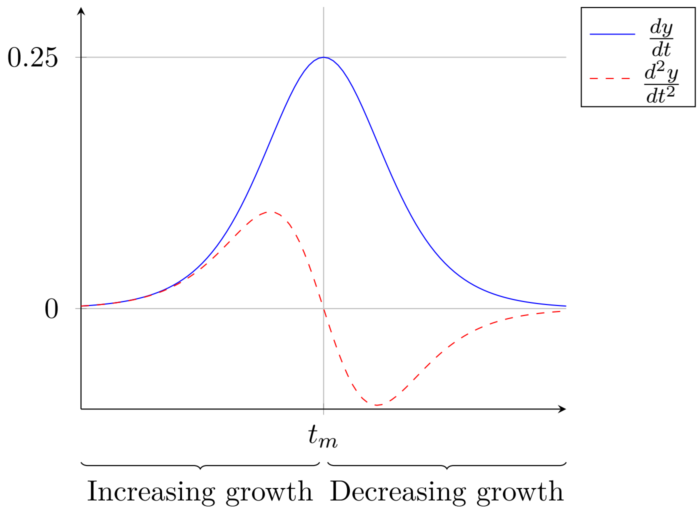


Figure 2.6: First derivative and second derivative of the standard logistic curve.

increasing (As seen in Figure 2.6, up to t_m, this portion is concave upwards). A value of zero signals the inflection point (At t_m). A negative value means that the rate of growth is continuously decreasing, concave downwards (After t_m, this portion is concave downwards). The derivatives of all s-shaped functions will follow this pattern.

Second derivative of the sigmoid function:

$$\frac{d^2y}{dt^2} = -\frac{2}{(1+e^t)^3} + \frac{3}{(1+e^t)^2} - \frac{1}{(1+e^t)} \tag{2.4}$$

As briefly mentioned earlier, the first derivative graph resembles the probability density function of a normal distribution and Roger's adoption curve. If our data is plotted in terms of frequency instead of cumulative, then the s-curve would turn into a bell shaped curve (see Figure 2.1). Inversely, the integral of any bell curve function would be s-shaped. This is why many cumulative distribution functions are sigmoidal in shape.

Up to this point, we have not discussed any parameters; this is because, in the sigmoidal logistic function, the common parameters, known as b and K are equal to 1, and there is no time constant ($t_m = 0$). In this thesis, we use the three-parameter logistic function (2.5) for curve-fitting; the three parameters allow more flexibility for curve-fitting compared to the standard logistic function. The data is plotted in terms of cumulative adopters, and not frequency because the bell shape of the frequency distributions is more sensitive to short-term *noise* [54]. Scholars in [28] describe the three parameter logistic function as "attractive for modeling s-shaped growth because it is a parsimonious model where the three parameters have clear, physical interpretations." [28, p.2]; it is commonly used in diffusion studies [9, 23, 29] and has the following equation:

$$y(t) = \frac{K}{1 + e^{-b(t-t_m)}} \tag{2.5}$$

Where:

K: is the upper asymptote of the curve.

b: is the growth rate of the curve.

t_m: is the midpoint of the curve, inflection point, it's a time constant.

e: is the base of the natural logarithm, it has a value of ≈ 2.718.

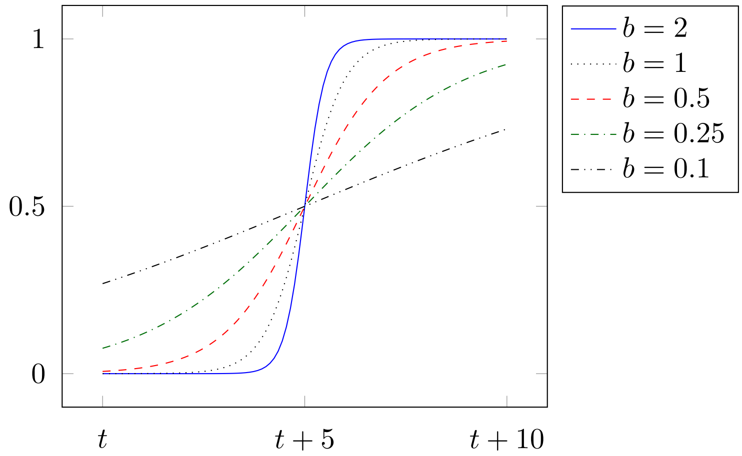
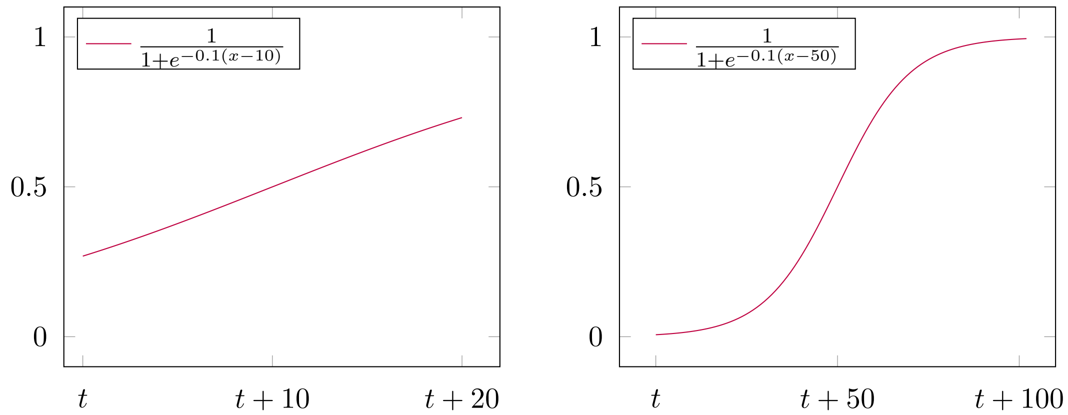


Figure 2.7: 3 parameter logistic function with different b parameters. Holding $k = 1$ and $t_m = 5$.

The diffusion time can be represented by the time constant Δt; it is the number of years an innovation takes to go from 10% to 90% of its duration (K). The full diffusion time from 1% to 99% is equal to $2\Delta t$. This metric has been used to measure the *duration* of the diffusion in previous studies by [27] and academics at IIASA [8, 9, 23, 19, 29]. The advantage of Δt is that it uses time (years) as a metric, which is much easier to compare, in contrast to comparing the slopes b.

K, b, and t_m determine the shape of the logistic function. Parameter b is the steepness of the curve at point t_m. When holding K and t_m as constants, the following holds true:

- A smaller b value means the curve has a *shallower* slope, and the time of diffusion Δt will be longer.
- A larger b value indicates the curve has a *steeper* slope, and the time of diffusion Δt will be shorter.



(a) $b = 0.1$, Time range of 20 (years)

(b) $b = 0.1$, Time range of 100 (years)

Figure 2.8

In [9], Grübler noted that Δt can be estimated by $\frac{ln(81)}{b}$. This implies that halving the b value from 1 to 0.5 means doubling the time of diffusion Δt, and also that Equation 2.5 car be rewritten as:

$$y(t) = \frac{K}{1 + e^{-\frac{ln(81)}{\Delta t}(t-t_m)}} \tag{2.6}$$

In Figure 2.7, five logistic curves are plotted, with different b values. They all intersect at t_m, the point where their slope b varies. The lowest b value of 0.1 starts to resemble a linear function. However, this is solely due to the restricted domain range of the graph. In Figure 2.8, we compare two functions with the same shallow slope (b) parameter, side-by-side, but over different time ranges, one over a domain of 20, the other over a domain of 100. The function that resembled a linear function turns into an s-shaped curve with a longer domain. This tells us that our domain greatly impacts the shape of our curve fitting functions, and it is not possible to correctly curve fit long Δt diffusions with a short range of data.

Parameter K is the upper limit of the asymptote. It is the maximum y value of our actual data. Depending on the case, it may represent the maximum of the cumulative GDP per capita of adopters, or simply the aggregate number of adopters.

$$\lim_{t \to \infty} y(t) = K \tag{2.7}$$

t_m is the point in time when the growth of the curve is at its maximum and the inflection point. In simple terms, t_m is the middle of the curve, around which the curve is symmetrical. It is equal to the point where the curve reaches $\frac{K}{2}$ when the first derivative is at its global maximum, and when the second derivative equals zero (see Figure 2.6).

$$\frac{d^2y}{dt^2}(t_m) = 0 \tag{2.8}$$

Another three parameter function which outputs an s-curve is the *Gompertz function*. This curve features alongside the logistic function, in the quantitative analysis of growth processes in IIASA's diffusion research. The Gompertz function is useful to use alongside the logistic function because contrary to the latter it is an asymmetric function. It is defined as the following [16]:

$$y(t) = Ke^{-e^{-b(t-t_m)}} \tag{2.9}$$

Where:

K: is the upper asymptote of the curve.

b: is the growth rate of the curve.

t_m: is the midpoint of the curve, inflection point, it's a time constant.

e: is the base of the natural logarithm with value of ≈ 2.718

The Gompertz function has a $\Delta t = \frac{1}{b} log \frac{log 10}{log(10/9)} \approx \frac{1}{b} 3.0844$, and the noteworthy property of $y(t_m) = \frac{K}{e}$.

There exist more complex variants of the logistic curve, with more than three parameters, which can give a better goodness of fit, but the three parameter logistic function remains the best option in our case, due to its simplicity, the ease of estimating its parameters and resulting flexible curve-fit over s-shaped data. It is also the function used by IIASA researchers and allows cross-case comparisons of diffusion rates. A more elaborate model with more parameters does not make it a better model, in the words of the British Statistician George Box:

> “Since all models are wrong the scientist cannot obtain a "correct" one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena. Just as the ability to devise simple but evocative models is the signature of the great scientist so overelaboration and overparameterization is often the mark of mediocrity” [55, p.792]

2.2.3 Fisher-Pry transform (substitution model)

Growth models such as the logistic function do not take into account any previous innovation that is being substituted, *substitution models* do. Though substitution models are much more relevant for cases of technological innovation – due to the never-ending wave of substituting technologies – there are cases where they remain relevant for cases of institutional innovation, such as the diffusion of democracy (in replacement of autocracy) [45]. We will describe the Fisher-Pry transform model which is a linearization of the three parameter logistic function [27, 28]. The original binary model was designed as a substitution model for technological change in 1971 by Fisher & Pry, modeling only *two* competing innovations with any market share gain from one innovation equivalent to the market share loss from the other, as in any zero-sum game with two players. The model makes three assumptions:

1. "Many technological advances can be considered as competitive substitutions of one method of satisfying a need for another.
2. If a substitution has progressed as far as a few percent, it will proceed to completion.
3. The fractional rate of fractional substitution of new for old is proportional to the remaining amount of the old left to be substituted." [27, p.1]

The variant we use, is called the *Fisher-Pry transform* [28] or *logit transform* [18], because it transforms a nonlinear function (the three parameter logistic function) into a linear one. The equation is simply the relative market share ($K = 1$) of the new innovation, denoted by $F(t)$, in proportion to the remaining market share, which still belongs to the old innovation, $1 - F(t)$.

$$FP(t) = \frac{F(t)}{1 - F(t)} \tag{2.10}$$

Where:

$$F(t) = \frac{y(t)}{K} = \frac{1}{K}\frac{K}{1 + e^{-b(t-t_m)}} = \frac{1}{1 + e^{-b(t-t_m)}}$$

Though in this form, it is not yet linearized. To transform a function which has an exponent in it – such as the logistic function – into a linear one, we must take the natural logarithm on both sides of the equation to eliminate the exponent (the natural logarithm is the inverse of the exponent):

Let $ln(FP(t)) = ln(\frac{\frac{1}{1+e^{-b(t-t_m)}}}{1-\frac{1}{1+e^{-b(t-t_m)}}})$. So,

$$= ln(\frac{1}{(1+e^{-b(t-t_m)})(1-\frac{1}{1+e^{-b(t-t_m)}})}) \quad \text{(fraction rule)}$$

$$= ln(\frac{1}{(1+e^{-b(t-t_m)})(\frac{e^{-b(t-t_m)}}{1+e^{-b(t-t_m)}})}) \quad \text{(Given } 1-\frac{1}{1+e^{-b(t-t_m)}} = \frac{e^{-b(t-t_m)}}{1+e^{-b(t-t_m)}}\text{)}$$

$$= ln(\frac{1}{\frac{(e^{-b(t-t_m)}(1+e^{-b(t-t_m)}))}{1+e^{-b(t-t_m)}}})) \quad \text{(Multiply denominator fractions)}$$

$$= ln(\frac{1}{e^{-b(t-t_m)}}) \quad \text{(Cancel out common factor } (1+e^{-b(t-t_m)})\text{))}$$

$$= ln(1) - ln(e^{-b(t-t_m)}) \quad \text{(Log rule } ln(\frac{a}{b}) = ln(a) - ln(b)\text{)}$$

$$= -ln(e^{-b(t-t_m)}) \quad (ln(1) = 0)$$

$$= -(-b(t-t_m)) \quad \text{(log rule } ln(e^x) = x\text{)}$$

$$= b(t-t_m))$$

As we have shown, the function can be linearized by using the natural logarithm, with Equation 2.11 as given by [28, p.5], showing the linear form with slope b.

$$ln(FP(t)) = b(t-t_m) = \frac{ln(81)}{\Delta t}(t-t_m) \tag{2.11}$$

Our data composed of d_t points of data must also be transformed similarly, for the data to be correctly displayed:

$$FP(d_t) = \frac{\frac{d_t}{K}}{1-\frac{d_t}{K}} \tag{2.12}$$

With the y-axis ($\frac{y(t)}{1-y(t)}$) plotted on a logarithmic scale, and the x-axis (time t) remaining on a linear scale, our resulting plot is semi-logarithmic. A steep linear slope (near vertical) implies a fast substitution process, a shallow slope (near horizontal) implies a slow substitution process. If the data points follow a s-curve, the resulting data points will resemble a linear line when transformed and plotted on a semi-log graph, and they will be curve fit by the linear function given by 2.11. There are only two parameters (b and t_m) because K is normalized to 1 ($K = 1$). As indicated in [27, 28], t_m is equal to the time where $FP(t) = 10^0 = 1$, and Δt is equal to to the time between 10^{-1} and 10^1. In order to make the Fisher-Pry transform graphs more intuitive to read, IIASA scholars have been labeling the right side of their plots with the percentage of saturation, as in [16].

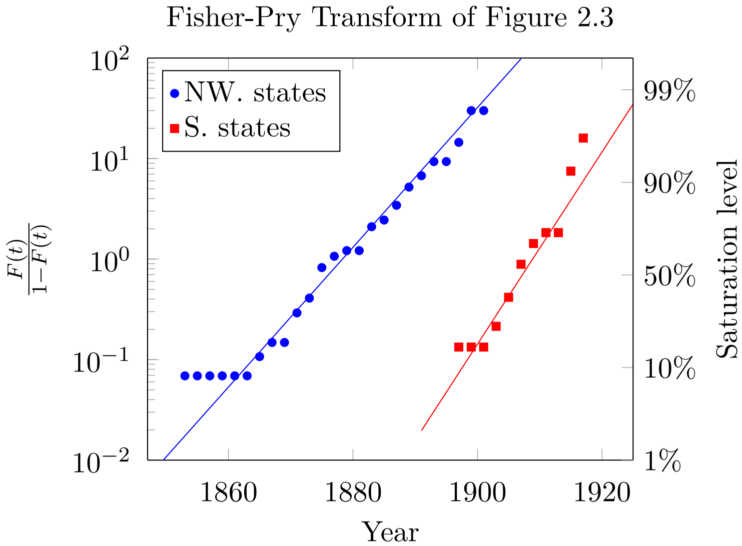


Figure 2.9: Diffusion of compulsory education among states in the USA. Time series from 1847 to 1925; two year intervals. Fisher-Pry transform (northern and western states: $b = 0.16$, $\Delta t = 27$ years and $t_m = 1878.31$; southern states $b = 0.22$, $\Delta t = 20$ years and $t_m = 1908.86$). Source: adapted from [34, p.555]; data source: [44, p.514-515]

In Figure 2.9, we plotted the Fisher-Pry transform of the diffusion of compulsory education in the USA, from the Pemberton's study [34] described in Section 2.1.3 and Figure 2.3. The Y-axis represents the fraction of states that have adopted a compulsory education law, in proportion to the remaining states that haven't. There was no ongoing substitution process of one innovation by another in this case, but this model can be used as an alternative visualization of diffusion case studies. A difference in slope between different cases is more apparent with the Fisher-Pry transform compared to the s-curve view, but early data points which don't follow an exponential distribution are also more apparent.

According to [56], the limitations of this model is that it assumes the substitution will proceed unilaterally and never reverse itself, once an agent switches, they do not switch back, within the time frame of analysis. Also, there is often low data availability of the early stages of the substitution processes. Fisher & Pry recommend not applying the model to a process until the new innovation has reached a few percentage of the total market [27]. Finally, it only considers a binary substitution, one for another, when in reality there are often more than two competing innovations at a certain time. Marchetti-Nakicenovic adapted the Fisher-Pry substitution model, to allow it to model more than two competing technologies [20]. Just like the logistic function, it assumes symmetry, but it is not necessarily a sign of weakness of the model: "In a majority of cases asymmetric substitution patterns are an indication that additional technologies compete on the market." [16, p. 21].

CHAPTER 3

Methodology

"Essentially, all models are wrong, but some are useful." [57, p.424]

- George E. P. Box, *Empirical model-building and response surfaces*

In this chapter, we present the methodology used to derive our results. In particular, we explain our case study process and a detailed overview of the quantitative framework applied to analyze the case studies.

3.1 Case study research

It is only possible to develop an understanding of the diffusion pattern of institutional innovation, its factors, its rate of diffusion, and its barriers, by studying specific historical cases. Thus, we have decided to adopt a case study approach to this work. By focusing on several exemplary cases of institutional innovation, we intend to understand the dynamics and the general pattern of this type of innovation. An advantage in studying institutional innovation is that in most cases, the required quantitative data can be collected from publicly available historical archives or institutional data sets, and we can, therefore, avoid the possible unreliability of interview data or third-party consulting research data.

Our aim is not to provide a novel understanding of our case study topics, but to integrate the existing economic history literature and longitudinal data of the diffusion with IIASA's quantitative framework described in Section 3.2. Despite a quantitative nature to our analysis, which considers the multiple dimensions of the diffusion, we ground it within its historical context. The qualitative analysis of our cases allows us to explain the dynamics that triggered or hindered the diffusion process, such as interdependent adopters and social diffusion barriers. In each case study, we describe the historical development and context of the diffusion. A simple description of the temporal diffusion pattern is not meaningful without understanding the underlying determinants and dynamics that caused the process to diffuse in the manner that it did.

Within the case studies, we analyze the institutional innovation on a country by country basis; we compare the adoption process and context between the units of adoption (countries). Thus, each case study is composed of smaller subcases, and we already form an initial comparative analysis in the within-case analysis.

Our case study methodology follows the theory development process as suggested by K. Eisenhardt [58], which is a method to build theory from reviewing and analyzing case studies. Our resulting theory is compared to the existing literature conducted by IIASA researchers, in particular [16, 9, 8, 19]. The following eight-step process is from [59] and is based on Eisenhardt's process.

1. **Literature research**: Review of the literature about the diffusion of (institutional) innovation, including spatial diffusion, and a literature review of the economic history of the case studies in question.
2. **Selection of case studies**: Identification of relevant case study topics that are not constrained by data availability. Identify how the diffusion of the case study subject can be analyzed, which elements are to consider.
3. **Instrument determination**: Establish data sources for the case studies. Longitudinal data is necessary for the study of diffusion over time. Identify the tools to be used for the temporal and spatial diffusion analysis.
4. **Data collection**: Iterative data collection.
5. **Within case analysis**: Study the data and validate it from different sources. Observe the data from different dimensions. Analyze the historical diffusion, derive the determinants of the adoption decision, and rate of diffusion.
6. **Cross-case analysis**: Comparison of diffusion patterns and processes among case studies. A cross-case analysis allows identifying consistent patterns.
7. **Theory building**: Construct theory on the patterns of the diffusion of institutional innovation. The resulting theory and findings are incorporated into the cross-case analysis.
8. **Literature incorporation**: Link resulting theory with existing literature covered in the state-of-the-art. The existing research is integrated into the within and cross-case analysis sections.

During the case studies selection, we have to abide by certain restrictions. Since this is an exploratory study into the pattern of institutional innovation, the cases must be illustrative cases of institutional innovation, and not peripheral ones. We excluded any case study topics without long-term longitudinal data, thus restricting us to past cases which have finished diffusing or are near the end of their diffusion duration. For a broader analytic power, we want one study of each of the two diffusion processes (see Section 2.1), one that is a *growth* process and one that is a *substitution* (replacement) process.

The specific weakness of the case study approach for analyzing the diffusion of institutional innovation is first, that of selection bias when selecting our case studies and secondly, a tendency for pro-innovation bias during the selection. As explained by Rogers in [7], we tend to pick the "successful" innovation cases, while ignoring the innovations that have failed to diffuse. The innovations that failed to diffuse will also leave less of a trace, and be harder to study due to a lack of data availability, which orients the research towards the successful ones. Rogers prominently discusses the pro-innovation bias in diffusion theory research in [7]. It is the tendency to assume and imply that the innovation being studied, should diffuse and be adopted by everyone. The decision to adopt an innovation should be regarded objectively to avoid this bias, with no predefined right or wrong decision because, in the adopter's eyes, the decision to adopt is always the correct one. Another limit in our case study approach is the generalization of the resulting theory: it is somewhat confined to the specific context of the case. We can not assume that institutional innovations diffuse at the same rate in the 19th century as in the 21st century.

The structure of our case studies follow a standardized format to facilitate comparison: after an introduction to the topic, we present an overview (and any limitations) of the data that is used to estimate the temporal and spatial diffusion. Then, the historical context of the diffusion is detailed, followed by a quantitative description of the temporal and spatial dimensions of the diffusion. Finally, a section on the determinants of the diffusion summarizes the qualitative aspects of the diffusion. In the analysis of the determinants of the diffusion, we integrate Rogers' theory on the perceived attributes of an innovation [7]. The attributes (presented in detail in Section 2.1.1) are the perceived relative advantages of the innovation, the compatibility, the complexity, the trialability, and the observability of an innovation.

3.2 Quantitative framework

In the early-day diffusion studies in the 1930s, diffusion academics attempted to curve fit their data with the normal distribution [34, 60], but it lacked in flexibility and goodness of fit. In contemporary studies, notably at IIASA, the logistic function, or a variation of it, is used to curve fit cases of diffusion of innovation with success. According to the authors of [61], the simple logistic function provides the best fit to the diffusion data compared to other variants.

For the within and cross-case analysis, we follow the quantitative framework used by IIASA researchers [16, 19] and the previously explored literature in Chapter 2 to analyze the diffusion patterns with the methods described. Our methodology is primarily derived from the enormous work of Grübler on the diffusion of different infrastructures in [16].

The central quantitative analysis consists of curve fitting the three-parameter logistic function or the Gompertz function (whichever fits best) to a time series (actual data) of a growth process to analyze the long-term diffusion pattern, from which we can extract the rate of diffusion Δt. This rate can then be compared to the Δt of other cases, and also with the existing diffusion theory literature published by IIASA, but it restricts us to use Equation 2.5 as our s-curve. In [8], Grübler presented a meta-analysis of the Δt of 265 diffusion processes, which had a mean Δt of 41 years across all the processes. Even though the Fisher-Pry transform (see Section 2.2.3) is a substitution model, it is a transformation of Equation 2.5 and therefore will still have the same Δt. We do not use the Shafir-Kabir model in this work because it is better suited to model substitution processes which are more common in technological diffusion processes.

On what measure we decide to analyze the diffusion depends on the case, and it will differ between cases. For the diffusion of railways, Grübler used the growing length of the railways as a measure of the diffusion [16]; for an institutional innovation like compulsory education, the diffusion can be measured through the proportion of the population living under compulsory education, or through the enrollment ratio.

The key variable in diffusion studies, is *time*, the temporal dimension; we investigate the process of adoption over a period of time. As stated in Section 2.1.3, only cases where the actual data has reached 60% of the estimated upper asymptote, can the pattern be reliably described by a logistic function [23]. Until reaching t_m, the s-curve remains indistinguishable from an exponential curve. To avoid speculative results, we ruled out recent cases of institutional innovation from our study which were still in the phase of exponential growth before t_m.

As mentioned in the state of the art of diffusion theory, there exists an often ignored, secondary dimension, *spatial* diffusion. Our method consists of using the theoretical framework of [38] as defined in Section 2.1.2. Regarding geographical scales, due to the nature of institutional innovations, we analyze the diffusion processes on a *macro-level*, which considers processes on a national and international level. The plotting of the spatial dimension diffusion is done with open-source historical maps in the SVG format[1] of the geographical regions in question. The national borders of European countries changed during the 19th century. Thus different maps are used depending on the year.

3.2.1 Parameter estimation and normalization

To evaluate whether a time series of data can be fitted by a function, we use a process called *curve fitting*; this process is similar to constructing a regression, but they are not to be confused with each other. We do not attempt to explain the relationship between independent variables and dependent variables, but to simply try to investigate whether a nonlinear s-shaped function can closely describe a time series.

[1] Our SVG maps are sources are from Wikimedia: `https://commons.wikimedia.org/wiki/File:Blank_US_Map_(states_only).svg` and `https://commons.wikimedia.org/wiki/Category:Blank_SVG_maps_of_Europe_for_historical_use`

To curve fit a logistical s-shape to our data, we must estimate the parameters of the function which deliver a curve with the closest possible fit to the data. The parameter estimation is done without *eyeballing*, by using software applications that can find the optimal estimated parameters, with methods such as the nonlinear least squares, which iteratively adjusts the parameters until the residuals sum of squares is minimized. In [28], the authors recommended corroborating the logistic function fit with the Fisher-Pry transform because it weighs the data points more strongly in the formative and late stages. There exist two software applications based on the IIASA research, that can estimate growth and substitution curves to curve fit longitudinal data, both are freeware:

1. *Logistic Substitution Model II*[2] (LSM2) by IIASA.
2. *Loglet Lab*[3] by Rockefeller University (version 4.0 was released in 2016).

In addition to the parameter estimation method of minimizing the least squares of the residuals, Loglet Lab 4.0 added an option of using a Monte-Carlo method based on a genetic algorithm but we observed no significant improvement in the standard error when testing with it. Table 3.1 is a non-exhaustive list of the different fits offered by the two applications and the nls package in the free statistical software *R*. Out of these fits, during our within-case and cross-case analysis, we use: the three-parameter logistic and Gompertz fit as our primary models; complemented by the Fisher-Pry transform when the diffusion process is deemed a substitution process, and finally the standardized residuals plot for an in-depth review of the goodness of fit.

Table 3.1: Comparison of the available (non-exhaustive) fits between Loglet Lab 4.0, LSM2 and R (nls package)

Functionalities	**Loglet Lab**	**LSM2**	**R (nls)**
Logistic fit	✓	✓	✓
Logistic fit with intercept			✓
Fisher-Pry transform	✓	✓	✓
Gompertz fit	✓	✓	✓
Linear fit		✓	✓
Exponential fit		✓	✓
Standardized residuals	✓[4]		✓

We use LSM2 to calculate the Δt and R^2 values of our curve fits. For the parameter estimation, we use LSM2 and the nls package in R. The *nls*[5] package can estimate our logistic function parameters, using the Gauss-Newton algorithm (by default), when given

[2]http://www.iiasa.ac.at/web/home/research/researchPrograms/TransitionstoNewTechnologies/LSM2.en.html
[3]https://phe.rockefeller.edu/LogletLab/
[4]Only in Loglet Lab 3.0 Javascript Edition: http://lizardinthesun.com/loglet/
[5]https://stat.ethz.ch/R-manual/R-devel/library/stats/html/nls.html

the formula and sensible initial parameter values. The nls package in R is overall superior as a parameter estimation tool because it grants the flexibility (and transparency) to write the exact functions to be used. Loglet Lab and LSM2 do not allow the user to modify the logistic function formula. Thus, a three-parameter logistic fit with intercept is only possible with the nls package in R.

The logistic function with intercept is useful in situations when the initial values of the data are above zero [16]. This can be the case in institutional innovation if a large country is the first adopter far before any other adopters, the high initial value is better modeled by a logistic fit with a constant. The larger units of adoption and perhaps the low number of units in the analysis of institutional innovation also justifies the possible addition of the constant, though the Δt and general shape of the function is completely unaffected by its addition. In Equation 3.1, we display the logistic function with intercept. The sole mathematical difference to the normal logistic function is that adding a constant results in $y(0) = c$ instead of $y(0) \approx 0$ as previously with Equation 2.5.

$$y(t) = \frac{K}{1 + e^{-b(t-t_m)}} + c \tag{3.1}$$

Where:

K: is the upper asymptote of the curve.

b: is the growth rate of the curve.

t_m: is the midpoint of the curve, inflection point, it is a time constant.

e: is the base of the natural logarithm, it has a value of ≈ 2.718.

c: is the constant (y-intercept).

Using R to curve fit logistic functions is also a future-proof method due to the popularity of R in the domain of statistics and the R project will most probably continue to be supported for many decades. The earlier, Java-based versions of Loglet Lab are already unsupported on modern operating systems, and there is no guarantee of continued support or availability of the recent release version of Loglet Lab. The latest version of LSM (LSM2) was released back in 2006. The future availability of Loglet Lab only came into question during the writing of this work, because of multiple periods of non-availability of Loglet Lab 4.0, which is solely offered as a SaaS (web-based software).

When comparing the resulting estimated parameters in the three applications, they result in the same values when using the default least squares method. Solely the Loglet Lab Monte-Carlo parameter estimation method can output different results (it is the default method of Loglet Lab 4), but we do not use it in this work. The precision of the estimated slope value b is the most important parameter because it determines Δt.

Concerning the plotting of the curves with the estimated parameters, we choose not to use any plot outputs of the three mentioned applications. Instead, we use the LaTeX *PGFPlots*[6] package, which offers complete manual control of the outputted two-dimensional plots. Both Loglet Lab and LSM2 have insufficient options for customizing the outputted plots.

The advantage of LSM2 and Loglet Lab compared to R is that they are user-friendly to use and have a gentle learning curve. They are specifically developed to facilitate the task of researchers who do not have the mathematical or R programming skills to conduct parameter estimations of growth curves. Also, both applications can automatically "guess" the initial parameter values. Unfortunately, with the nls package in R, it is necessary to estimate realistic initial parameter values manually, this is a universal problem when using nonlinear least squares methods. The Gauss-Newton algorithm will not converge the parameter values towards the resulting estimated parameter values if the initial values are outside a reasonable range. To determine appropriate starting values, we use the following ranges:

- $K = 1$ if the time series data set is normalized, otherwise K is equal to the maximum of what we are observing (e.g., if we are counting the cumulative number of adopters, K is equal to the upper limit of the cumulative number of adopters).
- $b = [0.01; 0.5]$, lengthy diffusion processes have a smaller b value, conversely short diffusion processes have a larger b value (see Section 2.2.2).
- t_m is around the middle of the time series' period (e.g., if the time series ranges from 1750 to 1850, t_m should have an estimated value of 1800).
- c is approximately equal to the initial value in y (the c parameter is only added when using the logistic fit with constant).

Listing 3.1 displays the commands we use to estimate the parameters of a three-parameter logistic function with the nls package in R. Lines 8 and 13 can be adapted, for other models. The Fisher-Pry transforms don't require any work in R. We can convert our logistic function (with its estimated parameters from nls) directly into the Fisher-Pry transform function in our PGFPlots LaTeX code, using the formula $\frac{F(t)}{1-F(t)}$ and plotting it on a semi-logarithmic graph (see Section 2.2.3 for details on the Fisher-Pry transform). Loglet Lab does not show the early, and late actual data points in its Fisher-Pry transform plot outputs as recommended by the model's authors in [27], due to the model's weakness in assuming that the process begins in the infinite past. We followed the same method of ignoring the initial and last 10% (or outliers) of actual data points for our Fisher-Pry plots.

A technique used by IIASA scholars [16, 23], is to normalize K to $K = 1$, by dividing each value in the adopter's column by K, thus turning it into a fractional adoption rate. It is

[6] http://pgfplots.sourceforge.net/

```
# After importing csv file with data. Set x to 'time' column of csv, and y to 'adoption' column.

x = csvfile$year
y = csvfile$adopters

# Define the three-parameter logistic function

logistic = function(params, x) {params[1] / (1 + exp(-params[2] * (x - params[3])))}
# with constant: logistic = function(params, x) {(params[1] / (1 + exp(-params[2] * (x - params[3])))) + params[4])

# The starting values of k, b, and tm must be estimated below by the user, and they must be reasonable values, or an error will occur.

curvefit <- nls(y~k/(1 + exp(-b * (x-tm))), start=list(k=42300,b=.109,tm=1870))
# with constant: nls(y~(k/(1 + exp(-b * (x-tm))))+c, start=list(k=0.8,b=.5,tm=1875, c=0.04))

# save estimated parameters to params

params=coef(curvefit)

# Plot the estimated fitted logistic curve
yhat <- logistic(params,x)
plot(yhat,type="l")

# Show actual data points on plot
points(y)

# Display estimated parameters
params
```

Listing 3.1: R code to estimate the parameters of the curve fitted 3 parameter logistic function.

the same process as in the Fisher-Pry transform. This normalization is necessary when displaying several different cases of diffusion of institutional innovation on the same graph. It scales the same upper limit of the plot to the maximum adoption. Normalized curves have their y-axis labeled with 'Index ($K = 1.00$). Without normalization, the comparison is more difficult due to different units, sometimes y is in the fractional GDP per capita, and sometimes y is in the cumulative number of adopters. The cross comparison remains limited when the y is calculated on different scales.

3.2.2 Goodness of fit

A question mark comes to analyzing whether our fitted s-shaped curve fits the actual data. In linear models, it is standard to use the R^2 and a residuals plot to judge the goodness of fit. The R^2 is far from an optimal measure of fit for nonlinear models according to [62], who analyzed the R^2 in nonlinear curve fitting. The author recommends abstaining from using the R^2 as a measure of the goodness of fit: "The use of highly inferior nonlinear models is reflected only in the third or fourth decimal place of R^2 and thus the description of single models when using R^2 is not meaningful, as this measure tends to be uniformly high when a set of models is inspected." [62, p.8]. However, despite this deficiency in nonlinear models, in IIASA's diffusion research, the R^2 still plays a key role in the evaluation of the goodness of fit of the curves [16, 29, 23]. The R^2 is also calculated inside of IIASA's LSM2 software. In an IIASA meta-analysis study [23], Wilson defined two criteria to establish the acceptability threshold of a logistic fit, a high minimum value of R^2 should somewhat compensate for the weakness of the R^2 in nonlinear models:

1. a minimum R^2 of 95% is necessary to verify that the model fits the data.
2. the actual K must have surpassed 60% of the estimated K parameter.

In addition to these two criteria, to verify if the logistic or Gompertz curve fits our data, we plot and analyze the standardized residuals of the better performing curve. Though, as mentioned by Grübler, just from plotting the data point and the fitted curve, one should already be able to tell whether the fit is good: "the human eye is an excellent guide for judging whether a particular technological diffusion or substitution path follows an s-shaped, e.g., logistic, pattern" [18, p.61]. The first 5% and the last 5% of the curve are not expected to fit well [26]. We go above the minimum 95% (0.95) R^2 criteria to judge whether the estimate of the logistic or Gompertz closely fits the data by judging any R^2 above 99% (0.990) to be an indication of a good fit. However, it remains indispensable to look at the outliers and any patterns in the residuals. An R^2 of 99% is much easier to attain in nonlinear functions compared to linear functions as explained in [62].

Residuals are the difference between the actual data (d_t) and the estimated values of our logistic curve $y(t)$. Therefore, the smaller the residuals, the better. A non-uniform distribution of positive and negative residuals around the zero axis, which looks like random noise, signifies a 'good' goodness of fit of the fitted logistic curve [28]. The presence of many outliers and a clear pattern among the residuals graph can mean a bad fit, and that our data does not follow an s-curve, though it can be argued that a pattern in the residuals be expected with nonlinear models even when the fit is good. We plot the *standardized* residuals to detect and better understand any outliers, and to measure the spread of our residuals. Assuming the standard residuals $sr \sim N(0, 1)$, then no more than 5% of residuals should fall outside of the value of $|2|$. Standardized residuals are superior to raw residuals in cross-comparative case studies because the values can be cross compared.

The residual standard deviation σ given by Equation 3.4 is a critical value, it helps us understand how far our fitted model differs from the actual data points on average. If comparing several fits, the one with the lowest residual standard deviation, or standard error, is best. The equation for the residuals, of a data set of n data points is given by 3.2[7]:

[7] In econometric notation, the formula would be written as: $\hat{e}_t = Y_t - \hat{Y}_t$, with $\hat{Y}_t$ the estimator of Y_t. The notation of $\hat{u}$ is also sometimes used, to refer to residuals, to avoid confusion with the errors e.

$$r_t = d_t - y(t) \tag{3.2}$$

Where:

$t = \{1, .., n\}$.

d_t: Actual data point, at time t.

$y(t)$: Estimated curve fitted logistic function at time t.

The standardized residuals sr_t can then be calculated similarly to calculating a z-score, by Equation 3.5, given the standard deviation and mean of the residuals. The value of sr_t is positive when the residual is larger than the average, and negative if it's smaller.

$$\bar{r} = \frac{1}{n}\sum_{t=1}^{n} r_t \tag{3.3}$$

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(r_t - \bar{r})^2} \tag{3.4}$$

$$sr_t = \frac{r_t - \bar{r}}{\sigma} \tag{3.5}$$

Finally, the K value (the carrying capacity of the curve) gives us information about the goodness of fit. The difference between the actual and estimated K value is a method to verify if a model fits. If the actual K is equal to 1, then we can expect an estimated K value between 0.95 and 1.05. If the estimated K value differs widely, then the model is not a good fit. If a model does not fit a set of data at all, then an estimated K parameter will not even be found by the software applications.

CHAPTER 4

Results

"Progress, then, is a kind of collective thinking, which lacks a brain of its own, but which is made possible, thanks to imitation, by the solidarity of the brains of numerous scholars and inventors who interchange their successive discoveries." [15, p.148-149]

- Gabriel Tarde, *The Laws of Imitation*

In this chapter, in an attempt to answer our research question on the diffusion pattern of institutional innovation, we present the results of our case study analysis based on the methodologies as introduced in Section 3 — starting by within-case analyses of two major cases of institutional innovation, and then a cross-case analysis. Each case study is structured as a self-contained chapter.

Our focus in this section is on analyzing the diffusion patterns of several cases of institutional innovation that were a result of the industrial revolution on a macro scale with the comparison of the dynamics in numerous countries. The study of cases relating to the industrial revolution and 19th century is a natural fit for institutional innovations because of the dramatic social change brought about by extensive technological changes (and vice versa). Concerning the industrial revolution, Grübler stated: "Perhaps the intellectual and institutional/organizational changes were the most fundamental, in that they provided an environment favorable for systematic experimentation (creation) and commercial application (diffusion) of innovations." [11, p.7]. One of the major institutional changes in the 1800s was the wave of democratization of European countries in a substitution process of autocracies; its diffusion process was studied in [45]. In [12], Acemoglu et al. describe the diffusion process of democracy as an inevitable adoption in Europe after the French Revolution, industrialization and urbanization of society increased the chances of an organized revolt against the elite: "Democracy did not emerge from the voluntary acts of an enlightened elite. Democracy was, in many ways, forced on the elite, because of the threat of revolution." [12, p.461]. The wave of democratization brought about other institutional changes, such as the rise of trade unions and universal compulsory education [12].

In our first case study in Section 4.1, we analyze the diffusion of universal compulsory education in Europe and the US, which only happened on a state-level across whole nations after the industrial revolution began. Before the spread of compulsory schooling, education was limited to a privileged elite and the clergy. The extent of the importance of schooling towards economic growth during the Industrial Revolution is not known, because it is hard to reliably measure how much of a contribution it made [63]. The diffusion of compulsory education can be measured by two metrics: through the enactment of compulsory schooling laws or the growth of the compulsory level enrollment ratios; in our case study, we analyze the diffusion through both methods.

Our second case study in Section 4.2 concerns the diffusion of the gold standard. As a result of the increased trade and capital flows from the industrial revolution: nations, central banks, and economists battled to come up with a better monetary system; the gold standard became the international monetary regime in the late 19th century. The growth and diffusion of the banking system — developed by necessary optimizations under institutional innovations such as the gold standard — helped finance the transition in the West from agrarian to industrialized societies [14].

4.1 Compulsory education

By the beginning of the twentieth century, the majority of Western European countries had adopted compulsory education legislation — compulsory education laws oblige children to attend school until they reach a specified age; this age differs by country and has been adjusted upwards over time. The diffusion process of compulsory education was highly heterogeneous among countries as a result of political, religious and socioeconomic factors. We consider universal compulsory education to be an exemplary case of institutional innovation because its introduction had the effect of creating state-driven mass education institutions which are now ubiquitous and indispensable in nations worldwide. This new model replaced the existing educational institution, which up to that point was private and religious-driven schooling reserved for a small share of the population [64]. The institution of mass compulsory education was described by [65] as: "Institutionally chartered to be universal, standardized, and rationalized. (...) Education is a mass institution in the sense that it incorporates everyone, cutting across such lines of differentiation as ethnicity, region, class, and gender." [65, p.147].

Like most innovations which threaten the status quo, the diffusion of compulsory education in Western Europe was not without opposition: the bourgeoisie had a stake in maintaining their children's educational advantage over the masses [64]; the Church — in conflict with rising secular states after the Enlightenment and French Revolution — wanted to remain the utmost authority over schooling, due to the fundamental importance of schooling in maintaining and increasing their followers; the aristocrats and clergy also feared that an educated peasantry could spark a rebellion against their rule [65], this was mirrored in the United States, where slave owners in the southern states feared dissatisfaction and rebellion from their slaves if they became literate — this opposition in the United States even resulted in, what is indeed the opposite of a compulsory education law: ratified legislation making it a crime to teach slaves to read or write [66].

In Sections 4.1.2 and 4.1.3 we analyze the diffusion of compulsory education enactments at the national level, in Europe, and in the US. In Section 4.1.5 we examine the diffusion of the enrollment ratios, which is the ratio of the number of individuals enrolled in schools in proportion to the total number of individuals who are eligible to be enrolled. If we regard compulsory education as an invention, then the enrollments ratio can be interpreted as a measure of the actual diffusion of the innovation. Finally, in Section 4.1.6 we describe the determinants of the diffusion.

4.1.1 The data

It is important to differentiate the two elements of analysis: the diffusion of *compulsory education* legislature only focuses on the laws that are enacted, and not on the actual intended result, which is to increase the schooling *enrollments ratio.* The analysis of both elements also allows one to calculate the *time gap*, which is the duration of time between the adoption of the institutional innovation, and the intended result. Our data source for the adoption dates of compulsory education laws is from [44, p.514] for the US states, and from [64, 65, 43] for Europe. We grouped the data over a two-year interval for the US, and a five-year interval for Europe. When we mention Europe, we refer to the 16 Western European countries in Table 4.1 excluding Russia.

In our study of the diffusion of *enrollment ratios* — which can be considered a better measure of the actual growth of compulsory schooling compared to the law enactments — the analysis will be on *secondary* level enrollments due to better data availability over the whole duration of the diffusion compared to primary level enrollments. It is common practice for education data to be split into three levels: *primary* (primary schools), *secondary* (middle schools and high schools), and *tertiary* (universities and other post-secondary schools). We discuss the available primary enrollment ratios in Section 4.1.2 and in Section 4.1.3 within the limitations of the available data, and then we analyze the secondary enrollment ratios in Section 4.1.5. The ‘time gap’ measure in Table 4.1 is the number of years between the adoption of a compulsory education law and reaching a 90% primary enrollment ratio, 90% is used instead of 100% because certain countries will experience a saturation level that never reaches the full 100% enrollment ratio. There are two principal types of enrollment ratios: the gross enrollment ratio, which can go over 100%, this is often the case in developed countries, e.g., Sweden had a gross enrollment ratio of 132.91%[1] in 2014 because older students who are repeating a school year are counted; and the adjusted enrollment ratio — this is the ratio that is used in our data

[1] Gross enrollment ratio, secondary, both sexes, in percentage. Source: http://data.uis.unesco.org/

— which controls for repeating students. The adjusted ratio is superior because it gives a better representation of the actual percentage of youth who are enrolled [67]. We consider the enrollment ratio, which is a percentage, to be a better representation of the institutional investment and adoption of compulsory education, than the cumulative number of enrollments, which does not control for periods of enrollment surges after a high birth period.

Our data source for the enrollments ratio is the data set provided in [67] which includes time series data of enrollment ratios in individual European countries and the US. The secondary level data in the data set is an adjusted secondary level school enrollment ratio, which allows cross-country comparison and accounts for: (i) differences in categorizing primary and secondary between countries (i.e, standardized cutoff age to differentiate between primary and secondary schooling); (ii) territorial changes (such as the separation of Austria-Hungary) [67]. The secondary level data does not take into account homeschooling or other informal schools, which could slightly skew the enrollment ratio downwards for certain countries with a culture of homeschooling such as the US.

4.1.2 The historical background of the diffusion

In this section, we will describe the historical context around the diffusion of compulsory education of the most relevant countries in Western Europe along with references to the legislative adoption dates and primary enrollment ratios. The context is crucial to understand the underlying reasons behind the initial growth and eventual expansion of state-driven compulsory education around Europe. A much more detailed overview of the historical construction of compulsory education systems in Western Europe can be found in the work of [65].

Wide-scale compulsory education has its innovation origin in central Europe, in Prussia under Frederick William I in 1716 [65]. He made village schools compulsory for working class children — the children of the elite could already afford some form of private schooling. This compulsory law was reiterated by his monarchic successor: Frederick the Great, in 1763 at the end of the Seven Years' War. A move described by [65], as a technique of reconstruction after the human and capital losses of the war. [68] claims this reiteration for education was a military technique to condition the children from the different regions of a fragmented Prussia, to unify ideologically with the state.

The relationship between education and war is not coincidental. Firstly, [68] describes a historical union between Prussia's military and education, with one side providing science skills and the other discipline; Aristotle once said: "The system of education and the greater part of the laws are framed with a view to war" [69, Book VII, p.4]. Secondly, the adoption and reiteration of mass education laws were used in Prussia (and later the German Empire), and the Austro-Hungarian Empire as an attempt to increase national competitiveness after military wars [65].

But it was not until the early 1800s that compulsory education in Prussia became truly institutionalized resulting in greatly increased enrollments [64]. The growing enrollments

Table 4.1: Dates of adoption of a compulsory education law and primary enrollment ratios[1].

Country	Adoption of compulsory education law (year)	Primary level enrollment ratio in year 1875 (%)	Estimated 90% primary level enrollment ratio reached (year)	Time gap between law adoption and 90% primary level enrollment ratio (years)
Austria[2]	1869	54	1915	46
Belgium	1914	76	1918	4
Denmark	1814	98	1840	26
France	1882	100	1870	-12
Finland	1921	11	1918	-3
Germany[3]	1871	53	1913	42
Great Britain[4]	1880	50	1887	7
Greece	1834	33	1959	125
Ireland	1892	36	1947	55
Italy	1877	57	1927	50
Netherlands	1900	81	1927	27
Norway	1848	76	1890	42
Portugal	1844	32	1955	111
Spain	1838	56	1967	129
Sweden	1842	100	1855	13
Switzerland	1874	90	1875	1
Russia	1918	11	1927	9
US	1890[5]	100	1857	-33

[1] Data sources: [67, 64, 65, 43, 44]
[2] We chose the more conservative 1869 law adoption date as the relevant date compared to an earlier date, due to the state-controlled and non-transitory nature of the later law [65].
[3] We decided to use the conservative date of 1871, when the compulsory education law started applying to all states after the German unification at the end of the Prussian-Franco war [43, 65].
[4] Scotland already adopted a compulsory education law in 1872 [43].
[5] Compulsory education laws are imposed at a state level, 1890 is an estimated non-weighted average (t_m) for all states.

were sprung on by educational reforms under the guidance of Wilhelm von Humboldt, influenced by the Swiss educational reformer Pestalozzi and the German philosopher Fichte. The latter preached nationalist ideals declaring that every citizen must devote oneself to maintaining the *ethos* of the nation. This nationalist ethos could only be maintained spatially within the whole of the nation and over time through a national education system, which like the military, is compulsory and state-funded [68].

Austria's Joseph II mimicked his Prussian counterpart by adopting a compulsory education law in 1774, coinciding with the end of the Seven Years' War and a temporary end to the Jesuit control over the Austrian education system [65]. Though enrollments did not expand greatly until a century later, when secularization of the educational system took place after the adoption of a new compulsory education law in 1869; before then, the expansion of the state school enrollments were slowed due to the reactionary movement that opposed the social and political changes of the French Revolution, and the control of the Catholic Church over the educational system until 1868 [65]. Austria (Austria-Hungary at the time) and Germany have a similar primary enrollment pattern in response to their definitive compulsory education laws adopted near 1870, with their time gap in Table 4.1 approximating forty-five years, so despite very early attempts in the 18th century to impose compulsory education, it was not until a couple of centuries later, near the beginning of World War I, that both countries reached a 90% primary level enrollment ratio.

Scandinavian countries were successful early adopters of compulsory education in law and in execution with high primary level enrollment rates very early on. Denmark adopted a compulsory education law in 1814 after establishing an educational commission 25 years prior, in an attempt to strengthen the country after conceding territorial losses (in the Treaty of Kiel which was signed in January 1814) at the end of the Napoleonic Wars [70, 65]. Sweden modeled a system based on Denmark's success, pushed by liberal reform which was influenced by the French Revolution, resulting in the adoption of a compulsory state schooling system in 1842. As indicated in Table 4.1, both countries were the earliest in Western Europe to reach a 90% primary level enrollment ratio; Denmark reached 90% by 1840 and Sweden by 1855. Norway lagged slightly behind its Scandinavian neighbors with compulsory legislature coming six years after Sweden, but with a larger time gap (42 years) until it reached a 90% primary enrollment ratio (in 1890). If we consider Finland as part of Scandinavia, then we can note that its process was very different from other Scandinavian countries, and rather resembled the laggard adoption process of Russia, of which it was still a part of, until the end of World War I. Finland is the country in our data set with the most recent date (1921) to adopt compulsory education legislature; but the Finns reached a 90% primary enrollment rate at a similar date as the late majority of European countries, in 1918, three years before they even introduced the compulsory education law.

In France, the educational system was subject to strong competition between the state and Catholic Church. Early attempts by the state to impose universal compulsory education during a short period of separation of Church and state failed at the end of the

18th century, with the return of the Church's authority over educational matters during Napoleon Bonaparte's reign. The French Emperor's educational focus was on educating an elite, that would run the governmental administration and military of the country [65], thus a universal education law did not happen under his reign despite being quoted in the memoirs of a close adviser as proclaiming the primordial importance of a universal education for all French citizens:

> "Of all our institutions public education is the most important. Everything depends on it, the present and the future. It is essential that the morals and political ideas of the generation which is now growing up should no longer be dependent upon the news of the day or the circumstances of the moment. Above all we must secure unity : we must be able to cast a whole generation in the same mould." [71, p.60-61].

Napoleon recognized the potential and utility of public education for instilling loyalty in a whole new generation. But, it was only a decade after the formation of the Third Republic, after the defeat of his nephew Napoleon III and the removal of the Church's influence over state affairs, that France managed to establish a state educational system in 1882 [65]. As shown in Table 4.1, a high primary enrollment ratio was already reached by the time of the law adoption, this can be explained by the explosive growth in primary enrollment rates [67] after the reforms of the new government from 1870 onwards; and by the Church's actions, by firstly delaying the law adoption to nearly a decade after the European average of 1875 (which is illustrated by the inflection point in Figure 4.1) and secondly, by expanding their own educational system after the earlier secularistic threats, which resulted in the growth of primary level enrollment rate before compulsory schooling came into law [64, 65].

England's educational system prior to the adoption of compulsory education in 1880 was targeted towards educating the elite in private, mostly Anglican schools [65]. The turning point according to [65], was firstly the reforms in the 1860s driven by the rise of liberal and democratic movements to educate the working class; and secondly the rising threats towards the British industrial world dominance by the German unification in 1871 and the rapid economic expansion of the US after its civil war. The result was an explosive growth in primary enrollments during the late 1870s and 1880s, having stood at only a 50% primary enrollment ratio in 1875 as shown in Table 4.1; they approached primary enrollment ratio saturation within 7 years of adoption. The push towards a reformed educational system to improve British competitiveness on the international arena was therefore realized, but educational reforms in 1880, 1902 and 1944 did not stop them from eventually losing their superpower status [65].

Italy, after a turbulent period with a lack of national identity post-unification, adopted a compulsory education system in 1877 [65], but a weak and fragmented state resulted in a fifty-year gap between adoption and a 90% primary enrollment ratio. Other southern European countries — Spain, Greece, and Portugal — had a similar problem with weak states that failed to implement and enforce the compulsory education laws that were legislated early in the 19th century, ahead of most Western European countries (t_m = 1875). Spain, Greece, and Portugal have the largest time gap (see Table 4.1) between law adoption and reaching 90% primary enrollments: 129, 125 and 111 years respectively, which can be attributed to ineffective and fragmented states [65].

In the 20th century, after the diffusion of compulsory education laws and primary enrollment ratios reached saturation, the compulsory law started going through iterative processes of *re-invention* which is still ongoing today. The re-invention comprises progressive increases in the duration of compulsory schooling. The compulsory period has now started to fully encompass the secondary level, reaching an ending age of up to 16, 17 or even 18 years old [72]. In the Netherlands until 2007, students had to attend secondary schooling until the age of 17; after 2007, this was increased to 18 years old [73]. A similar re-invention process took place in Austria in 2017, the ending age of compulsory education age was increased to 18 from 15 in order to tackle youth unemployment [74, 72].

4.1.3 The temporal diffusion of compulsory education laws

In Europe

From Figure 4.1, we can conclude that the spread of compulsory education legislature adoption in Europe — in terms of the percentage of the total population[2] — can be closely modeled ex post by a logistic function. The peak period of adoption occurred in the 1870s and 1880s with the large powers: Germany, Italy, France and the United Kingdom, all legally adopting compulsory education laws in quick succession (see Table 4.1). Adoption approached saturation in the 1880s, with over 90% of Western European already living under compulsory education laws (see Figure 4.1). Our resulting estimated curve fit has a saturation level of 100% (fixed $K = 1$), an inflection point which is reached in year 1875 (t_m), a Δt of 25.77 years (the amount of years it takes for the process to go from 10% to 90% of its duration) and an R^2 of 99.1%. The logistic performs slightly better in terms of the R^2 of the estimate, compared to the Gompertz fit ($R^2 = 98.6\%$); this is because it does not underestimate the growth as strongly as the Gompertz in the early stage (1800-1870). After 1870, both functions perform near identically. If we do not fix the K parameter to 1, then the resulting estimated parameters do not differ by much, with negligible changes in the estimated parameters of both curves and an insignificant impact on the goodness of fit. The estimated Gompertz K parameter resolves to 0.996, which is very close to the actual K of 1, a sign of a good fit overall, this results in a marginally superior R^2 of 98.7%. Nevertheless, for the logistic, not fixing the K parameter does result in a shorter Δt of 24.7 years.

[2]We calculate the data in terms of percentage of the total European population (instead of total population) to control for the explosive population growth during the period, though in both cases, the shape is logistic. The data source for our population data is Maddison's data set [13]

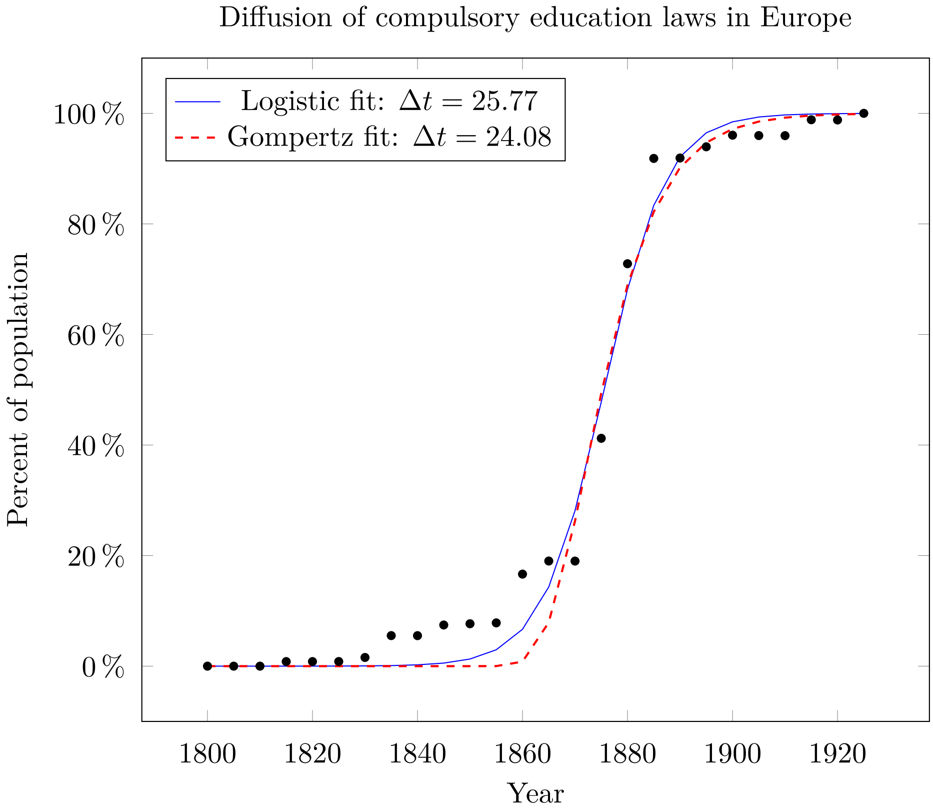


Figure 4.1: Diffusion of compulsory education laws in Western Europe (in percent of population living under compulsory education laws). Time series from 1800 to 1925. Fitted 3 parameter logistic function in red ($K = 1$, $b = 0.17$, $t_m = 1875.52$ and $\Delta t = 25.77$) and a fitted Gompertz function in blue ($K = 1$, $b = 0.128$, $t_m = 1872.275$ and $\Delta t = 24.08$. Data sources: [13, 43, 67]

In Figure 4.2, we plot the standardized residuals of the logistic fit with K fixed to 1. The standard error of the residuals (σ) is 0.0456. The plot demonstrates a good fit of the logistic curve because only one standardized residual is less than -2 and none are larger than 2, out of a total of 26 data points (the number of outliers should not surpass 5%). A non-random pattern in the residuals is expected for a logistic fit because it is a nonlinear function.

Concerning fits that are not s-shaped, the modified exponential yields no estimated model fit (i.e., the actual data can not be curve fit by the modified exponential because it differs too much). When we look at how the linear function performs, we find that despite a R^2 of 84.9%, the residuals of the linear fit has a clear pattern of overestimating prior to the inflection point of 1875, and underestimating after, thus indicating that this diffusion process is significantly better described by a logistic s-shape function.

The slight deviation of the actual data with the estimated logistic curve during a twenty-

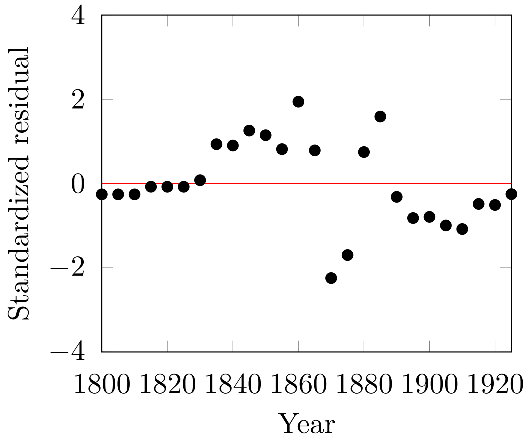


Figure 4.2: Standardized residuals of the compulsory education laws diffusion in Europe, with the three-parameter logistic function from Figure 4.1.

five year period between 1835 and 1860 is due to the bandwagon of southern countries (each had a population size multiple times larger than the Scandinavian countries who adopted early) adopting compulsory education legislation. This strongly expanded the early growth of the actual data upwards; but, as described in Section 4.1.2, these adoptions were only for political show as these countries had extremely low enrollments until a century later despite, perhaps, a legitimate intent. In corroboration with Grübler's analysis [16] — in his case, of technological innovations (infrastructure) — we notice in Table 4.1, without going in-depth with a statistical analysis, that the later adopters experienced a slight *catch-up* effect in their enrollment rates, with lower time gaps between adoption and reaching a saturated enrollment ratio, though this was sometimes due to the fact that the primary enrollments were already high before the late law adoption. By 1927, the majority of countries (with the exception of the southern European countries) had a primary level enrollment ratio of over 90%.

We can conclude from the last column of Table 4.1, that there is no overall pattern in the time gap between compulsory education law adoption, and reaching a high enrollment ratio for Western European countries; the time gaps among countries are wide-ranging: from −12 years (excluding the US) to 129 years. Surprisingly, it implies that adopting a compulsory education law had little impact on primary enrollment ratios, and this weak relationship was already noted by [64].

In the US In Section 2.1.3 we briefly oversaw the 1936 study [34] by Pemberton on the diffusion of compulsory education in the US. The paper did not carry out a comprehensive analysis of the diffusion process (barely a couple of pages were dedicated to the case), though the author noted an s-curved diffusion shape when dividing the states into north and south — this separation is necessary because the two groups were (maybe still are) culturally heterogeneous and adopted compulsory education on two different time lines. For a lack of space, we do not provide a historical context in the same way as we did for Western Europe, but we will describe the *temporal* and *spatial* dimensions of the diffusion process. The laws on compulsory education in the US are not decided at the federal level (i.e., there is no federal law imposing compulsory education), hence each state must individually decide whether to adopt the law or not. Thus, we can analyze how the diffusion rate proceeded among states and it becomes a good case study to analyze how an institutional innovation diffuses.

When comparing to Western Europe, the United States of America was ahead of the curve in primary enrollments, reaching an 80% primary enrollment ratio nationwide as early as 1850 [67] — this rate is slightly behind Sweden and Denmark but ahead of other European countries, nevertheless it is multiple times more remarkable considering the population of the US was 23 million in 1850 [75] compared to Sweden's 3.5 million and Denmark's 1.5 million [13].

A possible direct comparison of the estimated curve parameters (US and EU curves are shown in Figures 2.3 and 4.1) of the two continents remains limited. For the US, we calculate in terms of the cumulative number of states that adopted compulsory education, whereas with Europe we calculate in terms of the percentage of Europeans living under compulsory education laws. This limitation is due to the lack of the required yearly census data on the population size of US states (US Census data only provides updated data per decade), for the years in question (1853-1925). In this instance, we consequently attribute equal weight to all 48 states (excluding Alaska and Hawaii) — the analytic value does not suffer as both methods are relevant to model the spread of an institutional innovation.

With the condition of clustering the states into two groups: northern and southern, as originally done by [34], we observe that both groups are best described by the three-parameter logistic curve as illustrated in Figure 2.3. The northern states — who started their adoption much earlier, in 1853 with Massachusetts and New York — have a Δt of 28.655 years; the southern states have a Δt of 18.68 years. The main conclusion we can derive is the South has a t_m (the year 1908) 30 years behind the t_m (the year 1878) of the northern states, but 10 years shorter diffusion duration (Δt). This corroborates one of Grübler's findings [16] — for the diffusion of technological innovation — that the late majority and laggard adopters benefit from a *catch-up* effect and tend to have a shorter diffusion duration.

Concerning the performance of the other curves, in both northern and southern cases, the Gompertz curve is inferior to the logistic curve at modeling the early growth stage, but of similar performance after a third of the duration has been reached; a similar trend was observed in our analysis of Western Europe. The R^2 is at a comparable percentage for both curves, with an identical 99.3% for the northern states, and between 98.7% and 98.2% for the southern states. The modified exponential curve could not be curve fit to the actual data, in both northern and southern states. The performance of the linear curve does seem as first hand, relatively close to the performance of the logistic, with an R^2 of 94.5% for the northern states and 95.3% for the southern states, but when observing the residuals plot, the biased homoscedastic pattern of its residuals does show that the logistic or other nonlinear s-curve functions are more appropriate fits.

When holding the upper asymptote parameter constant ($K = 31$ and $K = 17$, for northern and southern states respectively), we notice a decrease in the slope for the southern states' curve, which results in a slightly shorter Δt, but no significant difference for northern states. Thus, in addition to the better R^2 measure, we can say that the logistic curve is slightly better fitted for describing the diffusion of compulsory education law adoptions in northern states compared to southern states.

4.1.4 The spatial diffusion of compulsory education laws

In Europe

With regards to the *spatial* dimension of the diffusion of compulsory education laws in Europe, we refer to Figure 4.3. The innovation center started in central and northern Europe, in predominantly Protestant countries (Austria-Hungary being the counterexample as a Catholic nation). We refer to transitory compulsory education when an earlier law adopted by Prussia and Austria-Hungary, that did not lead to state-controlled educational systems, were not permanent in nature compared to the subsequent law enactments near 1870. As the innovation spread throughout Scandinavia, the expansion reached a cultural barrier at the Swedish border with Finland, which was still part of the Russian Empire. Thus, the spatial diffusion pattern did not follow a contagious diffusion process (see Section 2.1.2), but a *hierarchical* and *expansion* diffusion process with cultural (political) barriers. By 1860, the innovation leapfrogged into much of Central Europe and was adopted in the peripheral European countries (Spain, Kingdom of Sardinia and Greece). Up to 1879, the innovation spread centrally to over half of Western Europe, ranging all the way eastwards to Russia from the vertical line of borders comprising the eastern borders of France, Belgium, Luxembourg and the Netherlands. In the next decade, the expansion continued westwards to the two dominant European countries of the early 19th century [65]: the United Kingdom and France. The spread reached completion after the turn of the century, with Belgium (1914) and Finland (1921), the last two European countries to adopt universal compulsory education laws.

In the US

The spatial dimension of the diffusion of compulsory education had a similar pattern in the US — however, this time not only does it follow an *expansion* diffusion process, but also a *contagious* diffusion process as seen on Figure 4.4: where local proximity is key and the innovation spreads geographically to those nearby like a wave or virus.

In the year 1860, the innovation center was situated on the Eastern Seaboard, with Massachusetts and New York — the first states to adopt compulsory education in the 1850s. From then on, in the 1860s it spread to nearby states such as the District of Columbia and Vermont. A sub-innovation center appeared on the West Coast in 1871, with the adoption of compulsory education by the state of Washington, and a few years

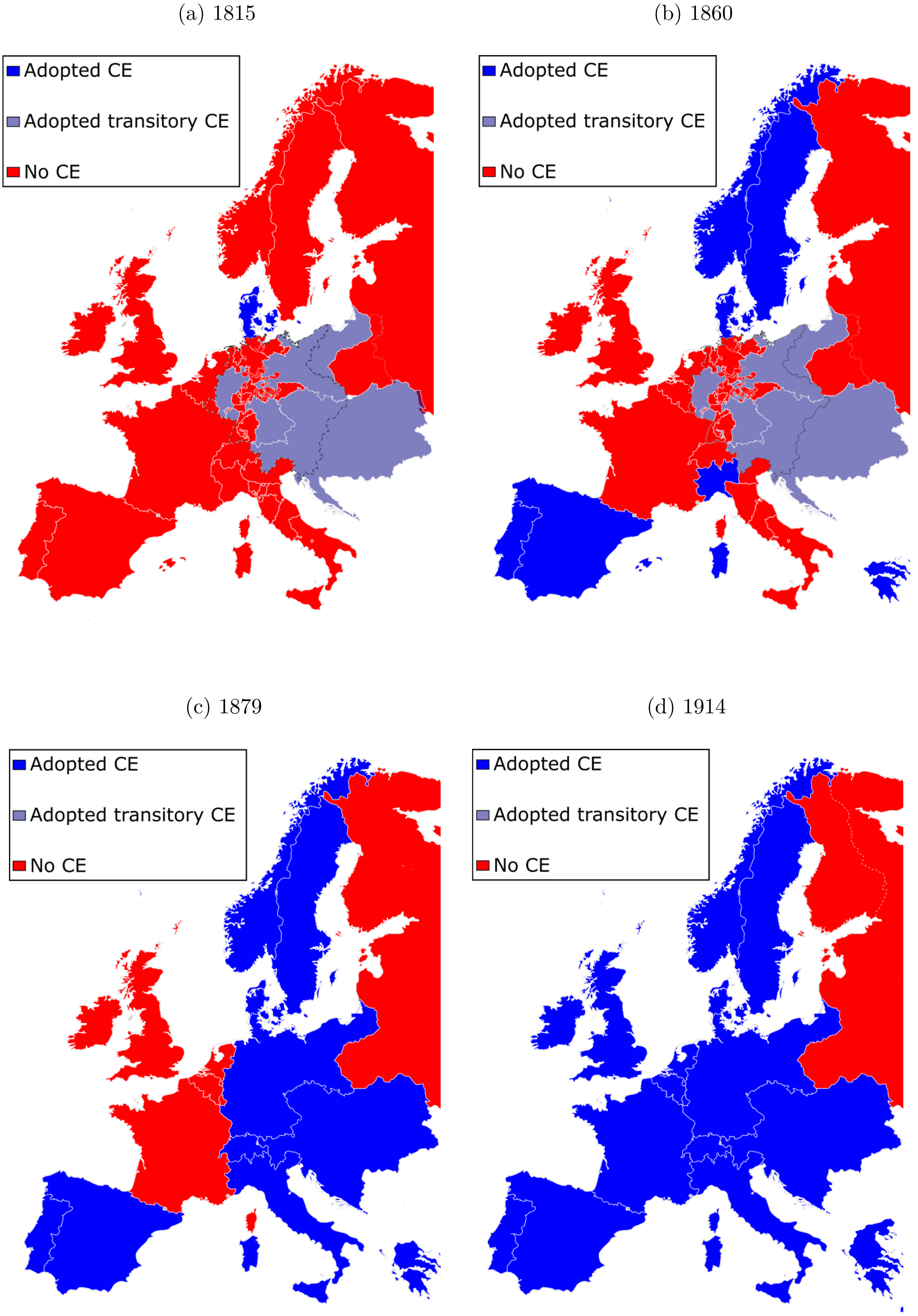


Figure 4.3: Spatial diffusion of compulsory education laws in Europe between 1815 and 1914. Data source: see Table 4.1.

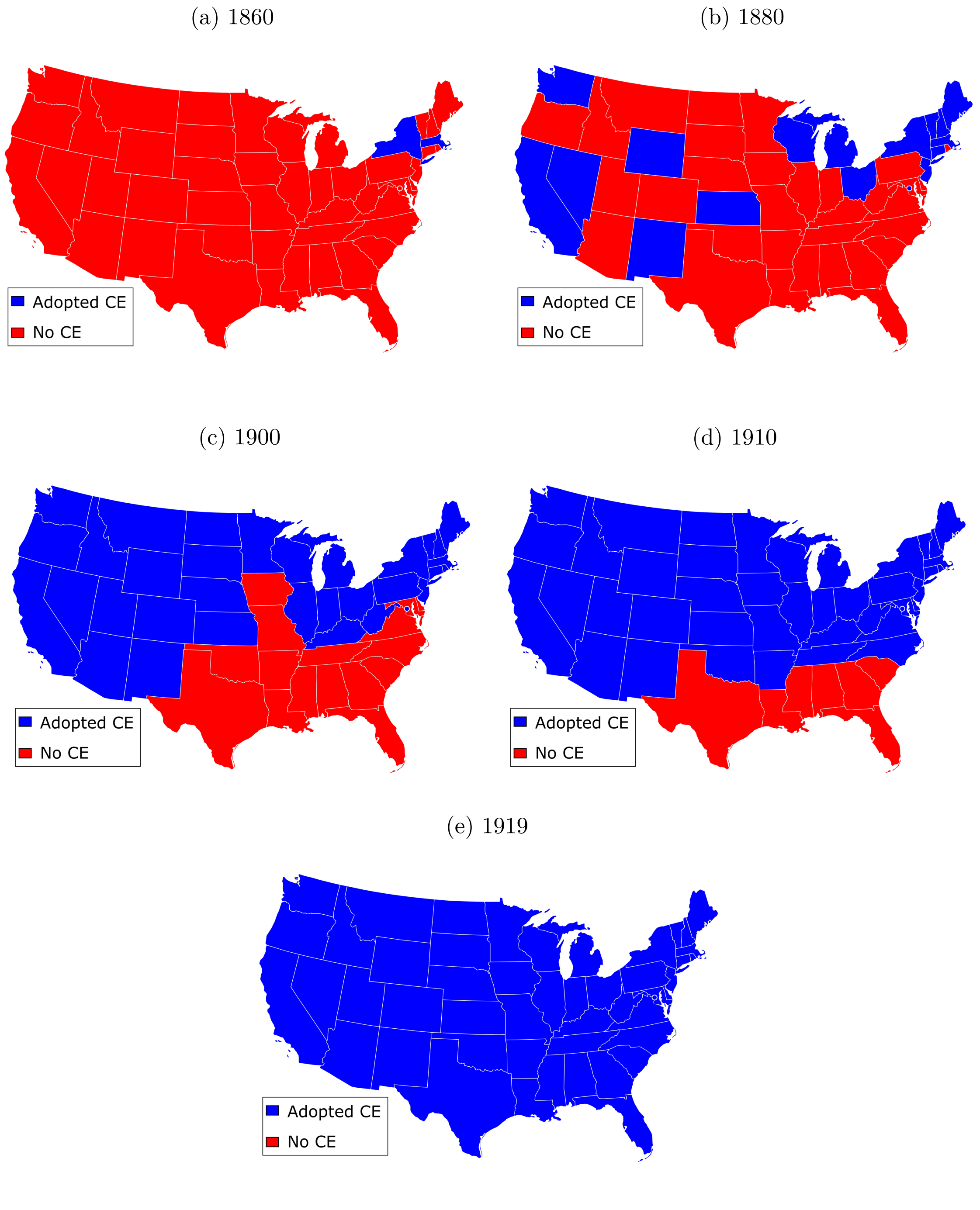


Figure 4.4: Spatial diffusion of compulsory education laws in the US between 1815 and 1919. Data source: [44].

later the sub-innovation center spread down to California. The general trend from this point on was a wave of adoption from north to south, and from the coastal regions inwards. By 1910, the adoption wave reached the outskirts of its final region of resistance, a region informally referred to as the “Bible Belt” and more importantly, it was also the Confederate States of America during the period in the 1860s and the American Civil War (a region with an economic reliance on the institution of slavery which was a root cause of the American Civil War). The final wave of adoption spread throughout the south central and southeastern states through the 1910s, and by the end of World War I, all 48 US states had adopted compulsory education legislature.

We recognize the same situation as in Western Europe concerning the diffusion barrier, with the main barrier to the spatial diffusion in the US, being a *cultural* barrier, notably conservatism which by definition is adverse to change. The southern states form part of a separate cultural area within the US, as noted in [34]. In terms of demographics, the northern and southern states were vastly different at the time, with the northern states having more European migrants [43] and southern states having a larger number of African-Americans [34]; the latter used to make up a large percentage of the total population of the south central and southeastern states before the Second Great Migration. This demographic difference was the critical cultural difference between the North and South according to [34], and the southern state lawmakers were reluctant to adopt a universal compulsory education that would incorporate everyone. It was illegal to educate slaves in many southern states prior to the abolition of slavery in 1865 [66, 43] with slave owners wary of educating their slaves, stemming from fears of a slave rebellion — over 84% of African-Americans born in the 1840s in the south were illiterate [66]. And the social resistance from policymakers continued for decades after, with legal caveats included in the compulsory legislature which restricted the African-American population in the South from enrolling in schools [66]. On a micro-scale, the spatial diffusion process was a *hierarchical* and *expansion* process; hierarchical because certain groups of people were leapfrogged (African-Americans) despite the universal nature of compulsory education.

Another fundamental facet which impacted the decision and timing of the adoption of compulsory education was put forth in an empirical study by [43], which claims that the states who adopted earlier (northern and specifically northeastern states) were the ones with the highest percentage of European migrants that originated from countries with no compulsory education. Indeed, the intention of the policymakers was to instill civic values in the migrants who come from countries with no compulsory schooling; the intended impact was to reach all migrants, not just the migrant children, the educated migrant children could transmit these values to their parents at home. As shown in Subfigure 4.4a in 1860, one of the first adopters was New York, a main entry port to the US, and first point of arrival for European migrants who came on ships from across the Atlantic. Even though there was a movement towards instilling civic duties in European migrants through compulsory education, there was a lack of movement towards providing the same for African-Americans in the southern states, thus displaying the cultural difference between the northern and southern states. In summary, the adoption of compulsory education in the United States was a method of nation-building and "Americanization" during a period of mass migration but the diffusion also encountered a social barrier from (southern) states with a cultural resistance towards adopting a universal compulsory education for all citizens.

4.1.5 The temporal diffusion of secondary level enrollment ratios

In this section, we analyze the temporal diffusion of *secondary* enrollments rates in Western Europe, the US and Russia[3]. After saturation (100%) of the primary enrollment ratios were reached, countries started investing in a second wave of compulsory education: at the secondary level. This often meant increasing the school-leaving age and expanding the state's educational system. The grand majority of the growth of secondary education in Europe happened post-World War II.

We are analyzing the estimated curve fittings of *secondary* enrollment rates due to a lack of data for the full duration of primary enrollment rates before 1820, at which point, most developed countries already had primary enrollment rates well above 10% but still very low secondary enrollment rates (e.g., Sweden had an estimated 80% primary enrollment ratio in 1820 but approximately 0.11% for its secondary enrollment ratio [67]), thus the necessary data to analyze the whole duration of secondary enrollment rates is available and more accurate, even for the earliest adopters.

Our analysis on secondary enrollments begins in 1850, at this point in time the US was the most educated nation in the world if measured in terms of secondary education with around a 10% enrollment ratio; most European countries, with the exception of what is today's Germany (who had around a four percent enrollment ratio), had less than one percent secondary enrollments. As shown in Table 4.2, the secondary level diffusion process only reached the takeoff point of 10% post World War I in all countries except the US. The early European countries to experience growth in secondary enrollments were the countries on the losing side of World War I: the German Empire and Austria. The slowest to reach 10% enrollments were the southern countries: Spain (1957) and Portugal (1958), who were over a decade behind the third slowest, Italy (1947). By looking at the inflection point (t_m) of each country's estimated logistic[4] curve in Table 4.2, we can observe the countries who's overall diffusion process happened later in time. The countries who were slowest to reach 10% secondary enrollment ratio, were also the slowest overall, with Portugal reaching its inflection only as late as 1986, Spain in 1972, and Italy in 1971; the rest of Europe had their inflection point (t_m) predominantly in the 1960s. All Western European countries in Table 4.2 are currently close to saturation, with over 90% in secondary enrollment ratios in 2010 — the majority at 100%.

Regarding the diffusion duration of secondary enrollments, firstly, the Δt spans from 34 years to 91 years in Western Europe (see Table 4.2). The diffusion process of secondary

[3] We only focus on Western Europe, and other developed countries, due to the lack of available data for enrollment rates in developing countries from 1820 to 1945.

[4] Our parameter estimates in the table are calculated with the logistic function because of its superior fit in this case over the Gompertz overall.

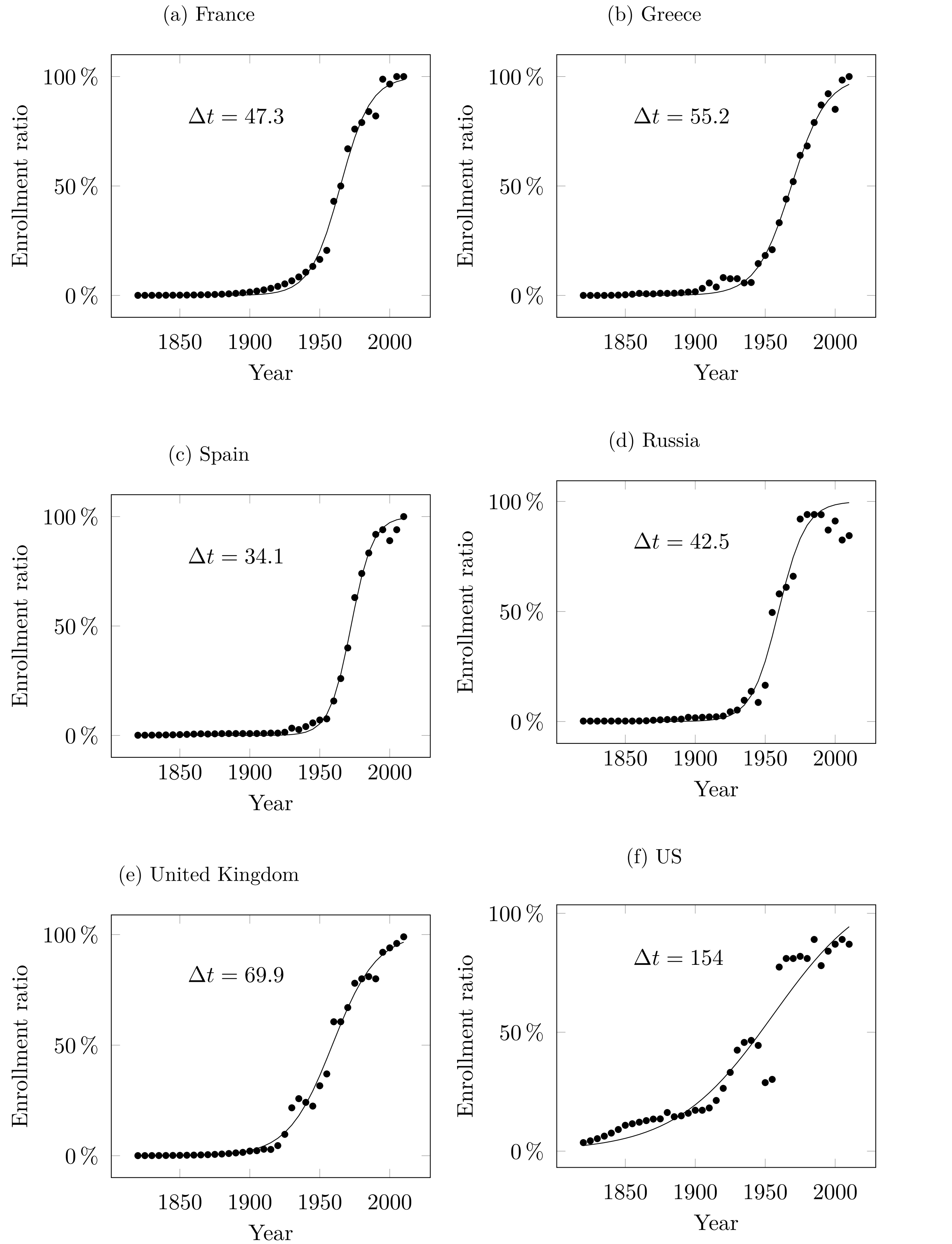


Figure 4.5: Diffusion of secondary school enrollments in six countries between 1820 and 2010. For parameters and data source, see Table 4.2.

Table 4.2: Diffusion of secondary level enrollments[1] from 1820 to 2010.

Country	Estimated 10% secondary enrollment ratio reached (year)	t_m (year)	Δt (years)	Secondary level enrollment ratio in 2010 (percentage)	R^2 of estimate
Austria	1920	1965	59.0	96.4	.979
Belgium	1946	1954	45.3	96.89	.987
Denmark	1944	1962	54.1	100	.991
France	1940	1964	47.3	100	.994
Finland	1940	1964	47.4	100	.992
Germany	1918	1962	91.1	99.7	.910
Greece	1942	1969	55.2	100	.995
Ireland	1944	1966	43.5	100	.994
Italy	1947	1971	58.3	95.8	.988
Netherlands	1925	1960	62.1	100	.993
Norway	1933	1964	47.8	96	.991
Portugal	1958	1986	42.2	100	.946
Spain	1957	1972	34.1	100	.997
Sweden	1943	1962	46.7	93	.987
Switzerland	1925	1970	59.8	93	.984
UK	1926	1959	69.9	99	.991
Russia	1937	1960	42.5	84.4	.980
US	1847	1945	131.7	87	.928

[1] Using the logistic curve, K fixed to 100%, data source: [67]

enrollments can be labeled as *slow* according to Grübler's categorization (above 40 years) [19]. The only country with a diffusion time under 40, is Spain — who experienced a late, but explosive growth in enrollments during the 1960s and 1970s — with a Δt of 34 years. Secondly, we observe that the first countries to expand their secondary enrollment ratios (i.e., the early adopters of secondary schooling) are the countries with the longest Δt's; all countries who reached an estimated 10% enrollment ratio before 1925, have a Δt of 59 years or above. Thus, this means that the later "adopters" experienced a *catch-up* effect. This reoccurring effect in diffusion theory has already been observed in our analyses in the previous section.

The diffusion pattern of the secondary enrollment ratios across all countries in Table 4.2 are remarkably *logistic* in shape. Specifically certain countries, such as Spain and Greece (see Subfigures 4.5c and 4.5b) — who experienced the majority of their growth later than the average (i.e., they are the laggards) — can have their diffusion pattern near perfectly modeled by the logistic function. We deem that this is due to the catch-up effect and possibly due to the fact that nearly the entirety of their growth happened after World

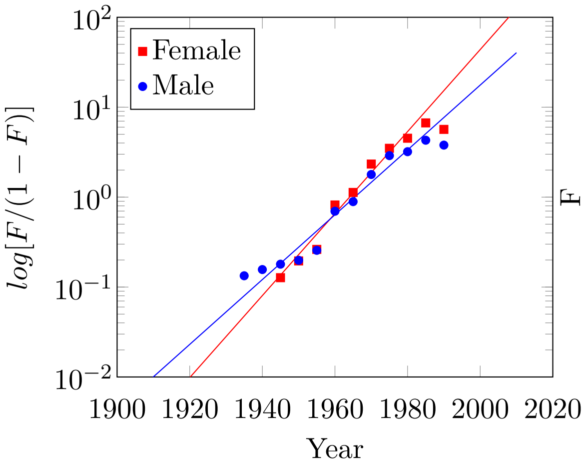
War II, minimizing any possible discontinuities which could have been caused by the wars which ravaged European countries for the first half of the 20th century. The R^2 of the logistic estimates in Table 4.2 strengthens the argument of a logistic pattern — we judge any R^2 for the estimate of the logistic above 0.990 to be an indication of an excellent fit but it remains indispensable to look at the outliers and any patterns of the residuals.

Both the UK and France have a regular logistic pattern, without any striking outliers, as shown in Subfigures 4.5e and 4.5a. The UK was the early European investor in secondary education, surpassing a 20% enrollment ratio by 1930, which is approximately triple the ratio reached by France at the same period. France did eventually catch up in 1970, thanks to a Δt that is over 20 years shorter in duration, and exceeding the UK in its secondary enrollment ratio by 1985.

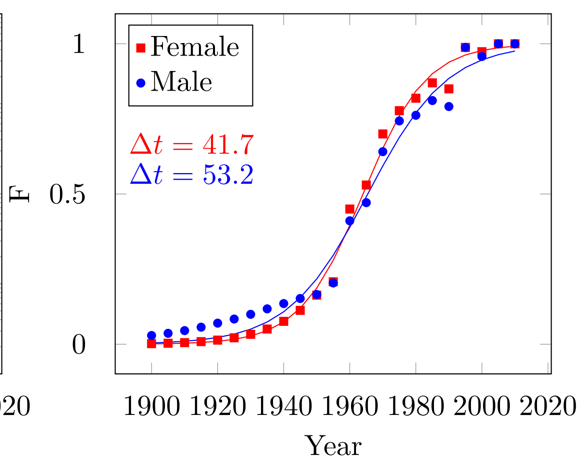
In the 20th century, and prior to World War II, the US had the highest enrollment ratio for secondary schooling in the world, benefiting from a period of high growth between 1915 and 1940. If we observe the actual data points for the US in Subfigure 4.5f, the only striking outlier is the period with a sudden and steep drop after World War II, but enrollments quickly recovered to explosive growth in the 1960s. The slope of the US's estimated logistic curve is much flatter than all other cases, this is a symptom of the very lengthy diffusion time, a Δt of 131.7 years. In the last couple of decades, the secondary enrollment ratio in the US has not reached the same saturation level as European countries (most have reached a 100% secondary enrollment ratio), this may be due to a high dropout percentage at the secondary level, and also due to the existence of homeschooling, which is legal in the US and not taken into account in the adjusted enrollment ratio.

Our last example in Figure 4.5 is Russia, who had a similar growth pattern to the rest of Europe, with a diffusion duration (Δt) of 42 years and a t_m reached in 1960. The major difference is the fall of their enrollment ratio from 1990 to 2010, after the collapse of the Soviet Union — a loss of nearly 10%. The logistic curve does not curve fit the saturation period closely, especially when fixing the K parameter to 1 (100%) like in Subfigure 4.5d. Thus prediction in such edge cases like Russia would be very speculative when taking into account the diffusion process over the whole duration; if we omit the last twenty years and only depict the growth period up to the initial stages of saturation, then the process in Russia is accurately described by the logistic model.

When observing the difference in the rate of diffusion between genders, the rate of secondary enrollments of the female population lagged behind the male population in the case of France until the 1960s. In Subfigure 4.6a, we can view the difference in slope between the two genders in the Fisher-Pry transform, the female slope is steeper which signifies that the diffusion duration for the female population was shorter. Indeed in numeric terms, the Δt for the female diffusion was 11 years shorter than the male diffusion at 41.7 years compared to 53.2 years. We expect this difference between genders to be more pronounced in primary enrollments, due to the primary enrollment diffusion happening earlier in time, in a period where compulsory education for the female population was not strictly enforced. This expectation is confirmed by the lower



(a) Fisher-Pry transform.



(b) Three-parameter logistic curve.

Figure 4.6: Secondary enrollment diffusion in France by gender. Female curve parameters: $b = 0.105$, $t_m = 1964.014$, $\Delta t = 41.7$ years, $K = 1$; male curve parameters: $b = 0.083$, $t_m = 1965.433$, $\Delta t = 53.2$ years, $K = 1$. Data source: [67].

estimated 19th-century female enrollment numbers compared to male enrollments in the data set of [67]. The aggregate (male and female population) Δt is approximately in the middle, at 47.3 years as shown in Subfigure 4.5a. The enrollment ratio was significantly higher for men up to the end of World War II, at which point the trend changed and women overtook men by 1960 (the point where the two Fisher-Pry linear functions meet). In this French example, when separating the population by gender, both curves are logistic in shape. The R^2 of the male logistic curve is 0.987, the R^2 of the female curve is 0.994. In conclusion, the initial stage of the diffusion process of the secondary enrollments of women was delayed compared to men but it benefited from a *catch-up effect*; both the male and female population have identical enrollment ratios at present.

4.1.6 The determinants of the diffusion

We can list a number of reasons[5] from our analysis that affected the spread of mass compulsory education laws and primary enrollments. This is not an exhaustive list of determinants, the understanding of the factors surrounding the rise of compulsory education in the West is complex, and intricate motives may elude us.

The diffusion of compulsory education did not take place before the 19th century due to the influence of the Church [65]. Overall, religious institutions had authority over the education systems in Western Europe up to the second part of the 19th century. With,

[5]Due to a lack of empirical evidence, we can not list political instability as a driver of compulsory education adoption, in the words of Ramirez and Boli: "There are no empirical analyses showing that political instability results in educational expansion." [65, p.155].

notably the Lutherans directly controlling the Prussian and then German educational system until the early 1900s [64]. The rise of compulsory education came only after the weakening of the clergy due to religious wars, and the rise of secular states after the influence of the Age of Enlightenment. But the religious presence in a country's educational system did not necessarily always negatively impact the educational legislature or enrollment rates according to the quantitative analysis in [64][6].

Countries with a *national church* (e.g., Sweden, Norway and Denmark) could adopt and enact state-driven mass education with less resistance because of the alliance between church and state, also benefiting from the *compatibility* (see Section 2.1.1) between a new state schooling system and the existing schooling infrastructure of the Church's schooling network. In addition, the Church already had experience with building an educational system: "Connections between state and church allowed the state to lean on prior religious organization of education when building a national system" [64, p.279]. In countries like France, the weakening influence of the clergy and aristocrats in state affairs paved the path for the adoption of compulsory education [64, 65]. As mentioned in Section 4.1.2 and 4.1.3, there were social groups who were actively opposed to compulsory education, notably those adverse to change and those who benefited from education remaining in private hands. Once the act of opposing compulsory education became politically untenable, priorities changed, and compulsory education became regarded as a method to educate the working class to avoid social unrest [65]. The *observability* of compulsory education was present and a possible factor in the adoption decision among European states; the de jure nature of compulsory education meant that it was a highly visible institutional change when introduced. Germany had a rapidly growing economy and it adopted nationwide compulsory education in the same year as its unification in 1871. This forced the UK to consider reforming its educational system, and whether to adopt compulsory education in order to improve British competitiveness [65]. Despite being the world economic leader of the time, and at freedom to operate differently from others [65], the UK reformed their educational system and adopted compulsory education nine years after Germany.

In addition to being a method of invigorating the national competitiveness of a country, educational reform was also used as a response to galvanize the nation after military conflicts [65]. We observed in Section 4.1.2 that many countries adopted compulsory education after a war: Denmark adopted compulsory education after its defeat in the Napoleonic Wars; Prussia and Austria-Hungary legislated educational reforms after the Seven Years' War; Germany adopted compulsory education when it unified after the Franco-Prussian War in 1871; France was only able to push through compulsory education reforms a decade after the defeat of Napoleon III in the Franco-Prussian War; Russia

[6] The goal of the study was to determine the significant independent variables that led to an earlier adoption of compulsory education laws; and to a fast expansion of primary enrollment rates. The principal factor for both dependent variables was the presence of a *national church*. The others independent variables in the regression analysis [64]: state revenue, urbanization, and Protestantism, had little impact on the dependent variables.

adopted mass compulsory schooling at the end of World War I in 1918 — a war in which they suffered mass casualties.

As described in the historical background of the diffusion (see Section 4.1.2), the construction of a state-driven compulsory educational system became an innovative and accepted way for state leaders to raise nationalist values and transform the masses into national citizens [65, 68]; in other terms, it became a nation-building effort [65, 43] to construct a national identity. Prior to its unification in 1871, Germany used to be split into many separate states, a national compulsory educational system was a method to build a homogeneous German nation [64, 43]. In the US, compulsory education also became a technique to integrate migrants after a period of mass migration, by the "Americanization" of the migrant children [43].

In an empirical study [64], Soysal & Strange concluded that enrollments were strongly influenced by the competitive dynamics between the different social groups (e.g., local, religious, noble, secular or bourgeoisie), who competed in providing schooling prior to state-run schooling. It was the competing forces between social groups and the state which led to the growth of mass schooling, and the lack of competition combined with a weak state structure which led to slow growth of enrollment rates in southern European countries.

Pertaining to the determinants of the re-invention of compulsory education, the compulsory period has been consistently increased over time across Europe because there is a focus on reducing school dropouts and turning a secondary level schooling degree into the minimum educational attainment [73]. This is because school dropouts are at higher risk of poverty and have a higher difficulty entering the job market resulting in unemployment and a dependence on welfare benefits [73, 74].

It is hard to estimate how large of an influence, the Industrial Revolution had on the diffusion of compulsory education. During the industrialization period in Britain, compulsory education did not exist; in 1841 only 4.9% of male workers and 2.2% of female workers were in an occupation requiring literacy [63]. At the time, skills were acquired through on-the-job training in the form of apprenticeships or other forms of labor experience. The on-the-job training period was not short compared to the years of formal education, trainees in cotton mills could spend up to ten years in training positions after being recruited at the age of 10 [63]. One would naturally assume that the technological advancements from the Industrial Revolution would lead to higher demand for an educated workforce, resulting in a push for a state-driven compulsory education system. But there is evidence of the opposite, a drop in literacy rates among the workers in certain new industries during the early Industrial Revolution before consequently rising [63]. This is explained as a result of the nature of the new industries: deskilled factory work compared to the previous handicraftsmen, "the shift from the artisan workshop to the factory was unskilled-labor using" [63, p.261]. In a modern example, the technological advancements from the shared economy have led to a new industry with unskilled jobs, such as on-demand delivery services and chauffeurs. Thus, new industries driven by technological advancement do not necessarily require higher educational attainment or increased skills. An increased demand for literacy or formal schooling could be supplied by the existing private schooling market, without the need for a public compulsory educational system.

Was the compulsory education adopted to directly combat child labor? Factory Acts were adopted with the intention of combating child labor in Britain, the compulsory education law only made it easier for inspectors to legally prove that children were working instead of attending school [76]. In 1835, 15.9% of the workforce in textile factories comprised children between the ages of eight and twelve, in 1885 (after the adoption of a compulsory education law) it was down to 8.9% [76]. According to Nardinelli [76], this drop was not due to any law changes (Factors Acts), but firstly due to the decreased need for secondary task workers — the type of job that children occupied — after technological improvements. Secondly, the drop in child labor was due to the rising real incomes of families during the 19th century, eliminating the family dependence on a child's income.

In our opinion, based on the evidence shown in this case study, conservatism (which can originate from a political, or religious source) in the US and Europe was the main barrier to the diffusion of compulsory education laws. It is within the definition of conservatism to oppose change: "Most current definitions of conservatism stress resistance to change as its most fundamental and prominent characteristic." [77, p.408]. The adoption of compulsory education laws is a radical institutional change, which had to surpass a conservatism barrier in order to spread. Those who tend to take a conservative position, are those who tend to be in a beneficial situation with the current rules of the game. At the time, the Church, and the various denominations were in a dominant position in Europe, with a lot to lose if radical institutional changes (a change in the rules of the game) were introduced. Another exemplary institutional innovation was the abolition of slavery, which was also hindered by those who benefited from slavery, known as anti-abolitionists.

4.2 The gold standard

In the mid-19th century, up to the beginning of World War I, the near majority of Western countries proceeded, one after another, to adopt the gold standard — a form of fixed exchange rate. In most industrialized cases, the adoption was a substitution process of a pre-existing silver or bimetallic[7] regime.

With the rapid growth of international trade, it became of great importance to find technological or institutional techniques to reduce transaction costs — a uniformization of currencies was an institutional innovation which had not been realized on the international scale yet[8] [79]. Due to its scarcity, indestructibility, and perceived long-term value, gold was in competition with silver, to become the international currency in a commodity-backed monetary regime, sometimes referred to as the *international gold standard*.

From the industrial revolution, emerged an industrially and financially powerful British Empire, who were to first to establish a gold standard in 1819[9] [81]. This eventually spurred a domino effect, and Britain's main trading partners eventually joined the gold standard, in order to improve trade flows and benefit from access to London's capital market [82].

For a standardized definition of the gold standard, which allows us to compare the adoption times between countries, we refer to Meissner's definition in [83], which considers a country to have adopted the gold standard if: "it established and adhered to a law which fixed a price between the domestic currency and a quantity of gold (and no other metals), mandated the free coinage of gold, and ensured the convertibility of notes into gold." [83, p.391-392]. *Convertibility* plays a crucial role in a commodity-based standard, as unrestricted convertibility signalled that a country was dedicated to adhering to the international gold standard, and it results in a stable exchange rate and lowered borrowing costs [84, 85].

In this case study, in an attempt to understand the diffusion process of a novel monetary technique, we analyze the diffusion of the gold standard (Section 4.2.3 and 4.2.3), within its historical context, during the prewar era: between 1854 and 1914 (Section 4.2.2). It is during this era (often referred to as the *classical gold standard era*) — a period of increased capital mobility where gold could freely flow among all countries on the gold standard [86] — that the sharp rise of the gold standard and simultaneous fall of the silver standard took place.

[7]A bimetallic regime is the concurrent adherence to both gold and silver standards, in contrast to a monometallic standard which is exclusively gold or silver. The price ratio between gold and silver has historically been around 15.5 ounces of silver to 1 ounce of gold prior to the classical gold standard era [78].

[8]it was attempted on a regional scale with the Latin Monetary Union, and the Scandinavian Monetary Union

[9]Other sources, such as [80] consider 1821 the year that Great Britain adopted the gold standard.

Table 4.3: Adoption and exit dates[1] of the gold standard (1854-1914).

Country	Gold standard	
	Adoption date	Exit date (if prior to 1914)
Austria-Hungary	-	-
Belgium	1878	-
Denmark	1873	-
Finland	1877	-
France	1878	-
Germany	1872	-
Greece	1910	-
Italy	1884	1894
Netherlands	1875	-
Norway	1873	-
Portugal	1854	1891
Spain	-	-
Sweden	1873	-
Switzerland	1878	-
Russia	1897	-
US	1879	-

[1] Data sources: [83, 82, 86, 87]

4.2.1 The data

Our focus is on the gold standard adoption dates of European countries, Russia and the US, starting in 1854 — the year that Portugal adopted the gold standard — until the outbreak of World War 1 in 1914. Our main sources for the adoption dates are [83] and [82]. The latter is also our source for the dates of convertibility suspension of several countries, which we have taken into account. In other words, if a country leaves the gold standard — we do not ignore it, such as Portugal suspending convertibility after 1891 — we remove that country from our list of adopting countries. See Table 4.3 for a summary of the adoption dates and also for exit dates. Great Britain is not taken into account in our data set, because it adopted the gold standard in a different period, over thirty years prior to 1854; it's exclusion does not have an effect on the temporal diffusion pattern but the logistic function would have to be corrected by a constant to compensate for the large initial value, had it been included.

In order to standardize our definition of the gold standard to allow cross-country comparison, we are using Meissner's definition [83]. For a nation to be considered as having adopted the gold standard, it must set a fixed exchange rate between their local currency and gold (exclusively, and not in addition to another commodity), they must offer gold coinage, and they must provide unrestricted convertibility. Thus, under this strict definition of the gold standard, we do not consider Austria-Hungary to have adopted a full gold-standard because the free convertibility of notes into gold was not ensured [83, 88, 89, 90]. A de jure adoption is not necessary, in many cases it was a de facto adoption of the gold standard.

With regards to how we calculate the diffusion, we are not simply using the cumulative number of adopting countries, because that would give equal weight to the industrialized and less industrialized countries and it would not be a representative pattern of the diffusion across Europe. Instead, we are using the GDP per capita of the adopting countries for each year between 1854 and 1914, in proportion to the total GDP per capita of all countries in the data set (for the year in question). Our data source for the GDP per capita time series is the Maddison Project [91]. We must use this GDP per capita proportion ratio in order to control for the growth pattern of the GDP per capita — this was a period where the GDP per capita of all nations concerned grew rapidly. We did not use the total GDP or population as metrics because the larger countries would drown out the adoption pattern of the smaller nations, who remain relatively important to understanding how this institutional innovation spread. In our opinion, the GDP per capita provides the most fitting balance.

4.2.2 The historical background (1854-1914) of the diffusion

In 1854, Great Britain was the world's economic powerhouse, with the highest GDP per capita [92]. They were the first country to adopt a gold-based monetary system over thirty years prior to the beginning date of our analysis [83]. In 1854, Portugal decided to join its main trading partner and creditor, Great Britain in adopting the gold standard. This move would reduce transaction costs on trade between the pair [82]; in addition, the adoption was facilitated by the low price of gold — due to the 1850s Gold Rush — and the pre-existing usage of British gold coins in the Portuguese monetary circulation [87]. Following a period of political and monetary instability, Portugal eventually abandoned the gold standard in 1891 [87]. Until the 1870s, no other European nation would join the gold standard.

To understand the adoption process of the gold standard, we must analyze the key role that other precious metal monetary regimes played. In reality, innovations rarely diffuse without competing forces. In this case, the gold standard was intertwined in a competition with the silver and bimetallic regimes. Trading with a country under the bimetallic regime meant you could pay your dues with either silver or gold, thus one would buy the slightly cheaper of the two metals depending on their purchasing power, which on an aggregate level would eventually drive the cheaper commodity's price back up to equilibrium, this (and governmental intervention) led to a stabilization of the gold-silver ratio under bimetallic regimes which were successfully enforced until the 1870s. The result was also that the price of silver was very stable — practically fixed — with its value fluctuating around 15-16 troy ounces of silver per troy ounce of gold as shown in Figure 4.7 [78, 79].

Table 4.4: Monetary groups in 1870

Gold standard	Silver standard	Bimetallic standard
Great Britain	Germany	US
Portugal	Austria-Hungary	France
	Sweden	Switzerland
	Denmark	Italy
	Norway	Spain
	Netherlands	Greece
	Finland	Belgium

[1] Data sources: [83, 79, 82, 86]

The crucial role in the early diffusion period of the international gold standard was not played by the earliest adopter, Great Britain, but their two European neighbors: Germany and France, who had just ended a war with one another (the German-Franco war) in 1871. A newly unified and economically growing Germany decided to join their economic rival, Great Britain by adopting the gold standard in 1872 [83, 79]. Before then, Germany was on a silver standard in a German monetary union with Austria-Hungary [93]. The costly transition[10] to the gold standard, meant that Germany had large stocks of silver that it had to dispose of on the international silver market which was led by France. Thus, Germany could not only take advantage of France's large war indemnity payment to help fund the transition but also take advantage of France's strong bimetallic regime — which was successfully maintained by France since 1803 despite market fluctuations of the prices of gold and silver — to sell its silver [79]. France's bimetallic market was large enough to absorb Germany's silver sell-off [94, 79], however, according to [79], France's decision to limit the silver coinage in 1873, was the catalyst which caused silver prices to immediately drop outside of the previously fixed pricing interval of 15 to 16 ounces of silver per ounce of gold as shown in Figure 4.7, and silver never recovered to its previous value against gold again. On the one hand, we agree that France's decision to limit the silver coinage was a key turning point in the downfall of silver as a currency. On the other hand, the American decision to suspend silver coinage in the same year — referred to by Milton Friedman as "The Crime of 1873" [81] — also played a large contribution in eliminating the silver and bimetallic standards.

So why did not France follow suit and immediately engage in the process of switching to the gold standard once Germany's intentions were clear? With the US having ended silver coinage, France would not be able to dispose of its silver at an adequate price as it was the last remaining major economic power on a silver or bimetallic regime, Paris played the international role of stabilizing the gold-silver price ratio prior to 1873 [81, 79]. Furthermore, their bimetallism regime had worked well, delivering a stable exchange

[10] Germany's transition to a new monetary standard during their unification process was complicated by the fact that every individual regional state (the largest being Prussia), which was now part of the unified Germany, used to have their own currency.

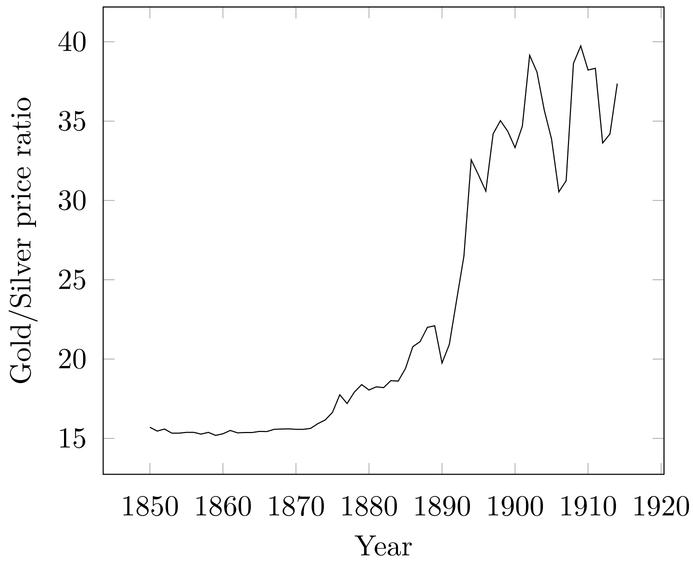


Figure 4.7: Historical gold-silver price ratio between 1850-1914 (ounces of silver per ounce of gold). Data source: [78].

and economic growth for the last 70 years, thus France held on and resisted change [79]. Indeed, France even intended on pushing the Germans to reverse their decision of leaving the silver standard according to [79], with their decision to suspend silver coinage being first, an anti-German move to hinder Germany's transition, and secondly a transitory policy to protect their bimetallism regime during a period of downwards pressure on silver. Eventually, the French and German governments had to accept the sizable loss of value in their silver holdings for the price of silver (in proportion to gold) dropped by 11% between 1873 and 1878, and by 25% loss by 1886 (see Figure 4.7).

France switched over to the gold standard in 1878 (under Meissner's definition of the gold standard [83]), along with Switzerland and Belgium; the trio were core countries of the Latin Monetary Union and with the switch, came the end of bimetallism's core in Western Europe. The Latin Monetary Union (LMU) was a group of European countries that attempted to lower financial transaction costs when trading among each other, by adhering to the same monetary standard, which was the bimetallism regime until 1873. The coinage of each member country was to be legal tender in all other member countries [80]. It's core members were: France, Belgium, Switzerland and Italy [81]. Belgium's circulation and monetary regime were strongly coupled to France's, with Belgium even relying on using French coins when it decided to interrupt its own coinage in the 1850s [80]. Thus it was inevitable that both countries would switch to the gold standard at the same time. Belgium's neighbor, the Netherlands had already adopted the gold standard two years earlier, in 1875 [83].

The smaller, Scandinavian countries (Sweden, Denmark, and Norway) all were able to substitute their silver-based monetary system with the gold standard in 1873, whilst

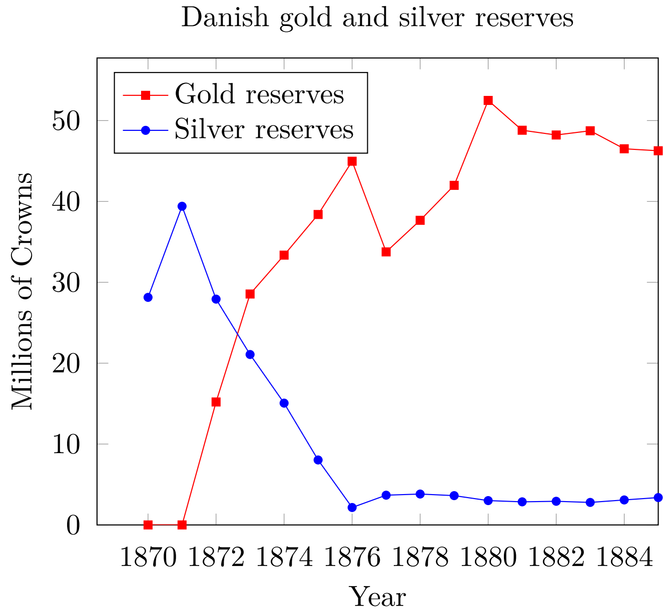


Figure 4.8: The gold and silver reserves of the Bank of Denmark between 1870-1885. Data source: [95, p.178]

forming the Scandinavian Monetary Union [93] shortly after Germany's decision to switch to gold [83]. They had less silver circulation and were, therefore, able to more easily switch over compared to Germany or France.

In Figure 4.8 we can observe the gold and silver reserves of the Bank of Denmark during the period of transition. In 1871, the Danish National Bank had no gold reserves (it was not necessary because Denmark was not on a bimetallic standard) and around 40 million Crowns in silver reserves. In 1872, the transition debuts with the bank acquiring 15.2 million Crowns worth of gold, and ending the year with 27.9 million Crowns worth of silver. In the next year, when the gold standard was adopted in Denmark, the gold reserves surpassed the silver reserves for the first time after reporting 28.5 million Crowns of gold and 21 million Crowns of silver. After the adoption of the gold standard, the Bank of Denmark kept progressively selling its silver reserves until it made up less than 10% of the total metallic reserves, its value floated around 3 million Crowns from 1876 up to 1885.

According to [93], changing to a new standard was principally due to the need to standardize the currencies between the three countries; the other main reason was to strengthen the regional alliance in terms of cooperation. There were several reasons the Scandinavian Monetary Union picked the gold standard instead of a bimetallic or silver monetary standard. Firstly, Germany had just switched from the silver to the gold standard and was in the process of getting rid of its silver circulation, which would pressure the price of silver downwards; postponing the decision to switch from silver to gold would be costly. Secondly, their main trading partners were Great Britain and Germany, who were both now on the gold standard — to follow suit would be economically beneficial in the long-run [93].

Austria-Hungary was a laggard in its pursuit of switching to the gold standard, it was still on the silver standard during the 1870s and 1880s whilst all other European neighbors had switched to a gold standard which meant that it sustained heavy losses due to the depreciation of silver — the silver sell-off meant that the central bank (OeNB) of Austria had to suspend silver coinage in 1879 [89]. Austria-Hungary adopted a de facto gold standard in 1892 (without gold convertibility) [88, 89].

Spain, who was on a bimetallic regime until 1883, was the only European country to never adopt the gold standard during the classical gold standard era, not even temporarily — this resulted in its national currency, the peseta to experience volatile exchange rates [82, 86]. Its government struggled with large public debt and yearly budget deficits in the 1870s and 1890s. Adhering to the gold standard was not feasible for Spain without large monetary reform [86]. The first reform would have been to engage in a sharp decrease in its fiat currency printing rate (the circulation of notes in Spain increased at a rapid pace, on average by ten percent in the 1890s [86]), the second reform would have been to increase their gold reserves. According to [86], the failure of the central bank of Spain to adopt monetary policies with the view of joining the rest of Europe on the gold standard resulted in a loss of foreign trade and GDP growth and high public debt.

Greece tried to adopt a gold standard in 1885, but it was unsuccessful and the regime only lasted several months with large gold outflows forcing the country to quickly switch to a fiat system in the same year [82]. After years of financial hardship from debt issues (including war indemnities owed to Turkey), Greece became the last European country to adhere to the gold standard during the classical era, with convertibility assumed in 1910, but it only lasted four years until the breakout of World War I [83, 82].

Italy was on a bimetallic regime along with other Latin Monetary Union members, they switched to a gold standard only six years after the other core members of the LMU had converted, in 1884. Their monetary regime only guaranteed partial convertibility [82] but it still adhered to the gold standard according to Meissner [83]. Governmental budget deficits and a lack of fiscal responsibility resulted in Italy abandoning the gold standard in 1894, convertibility was interrupted and never reintroduced during the classical era [82].

The United States of America had a bimetallism regime since the late 18th century with the Coinage Act of 1792, though the de facto standard drifted between silver and gold depending on the market price of gold and silver during the period in question [81]. The bimetallism monetary regime was temporarily interrupted during the American Civil War (1861-1865) [82, 81], and the pressure to finance the war saw the introduction of a paper currency named greenbacks, as a legal tender, with no backing to silver or gold [81].

As mentioned earlier, the United States eventually started to abandon silver in the 1870s, with the Coinage Act of 1873, which demonetized silver. The Treasury started the process of changing to a gold-based monetary regime by increasing their stocks of gold [81], and in 1879, the United States of America joined the international gold standard with an unrestricted convertibility of exclusively gold [83, 96, 82]. The complete abandonment of bimetallism by the US meant that the silver and bimetallic standards were truly dead on the international level after 1879.

Milton Friedman does not consider the American adoption of the gold standard in 1879 as an inevitable event despite all other industrialized countries adhering to a gold-based monetary regime. He recognizes a possible scenario where the United States would be on a bimetallic or silver standard, and Europe on a gold standard, with the United States acting as an international stabilizer for silver, and the gold-silver price ratio reflecting the exchange rate between the dollar and European currencies [81].

The American adoption of the gold standard had a large effect on the market dynamics of gold and silver. Firstly, the increased demand for gold from the U.S Treasury added another layer of buying pressure on top of the European powers who were at this point all on the gold standard [81]. Secondly, according to Milton Friedman, the discontinued role of the US as a silver absorbing nation had a dramatic downward effect on the price of silver. This gave rise to a cultural opposition referred to as the *free-silver* movement, including those with interests invested in silver and those unhappy at the deflationary nature of the gold standard (e.g., agrarian communities who benefit from an inflationary environment due to the high debt requirements of their activity), eventually leading to legislation such as the Bland–Allison Act — which required the U.S Treasury to buy millions of dollars worth of silver every month — and later, in 1890 the Sherman Silver Purchase Act [81].

The effect of the American adoption of gold on the price of goods was a deflationary economy: the rise of productivity during this period outgrew the rise of the supply of gold, and the increased international demand for gold made it scarcer, thus the price of goods and services in terms of gold were pressured downwards [81, 96]. Milton Friedman claims the deflationary environment was partially compensated by the rapid diffusion of the banking system that increased the money supply, which in his words "increased the amount of money that could be pyramided on each ounce of gold" [81, p.1170]. Increased gold production in the US after 1896 led to a worldwide inflation that helped ease the downward pressure on prices, and it was the result that the silver-free movement was advocating for — from then on, the silver movement faded [81].

The outbreak of the First World War in the summer of 1914 marked the end of the classical gold standard era. Nations needed to greatly increase their expenditures and finance the war, either through increased taxation or war bonds, and when that was insufficient, by printing paper money, thus the increased money supply complicating an eventual return to the gold standard [97, 89]. The majority of belligerent and even the non-belligerent nations (e.g., the Scandinavian countries) suspended convertibility in 1914, thus dropping out of the gold standard [98]. 1914 also marked the beginning of the collapse of the Scandinavian Monetary Union [98, 99] and the Latin Monetary Union

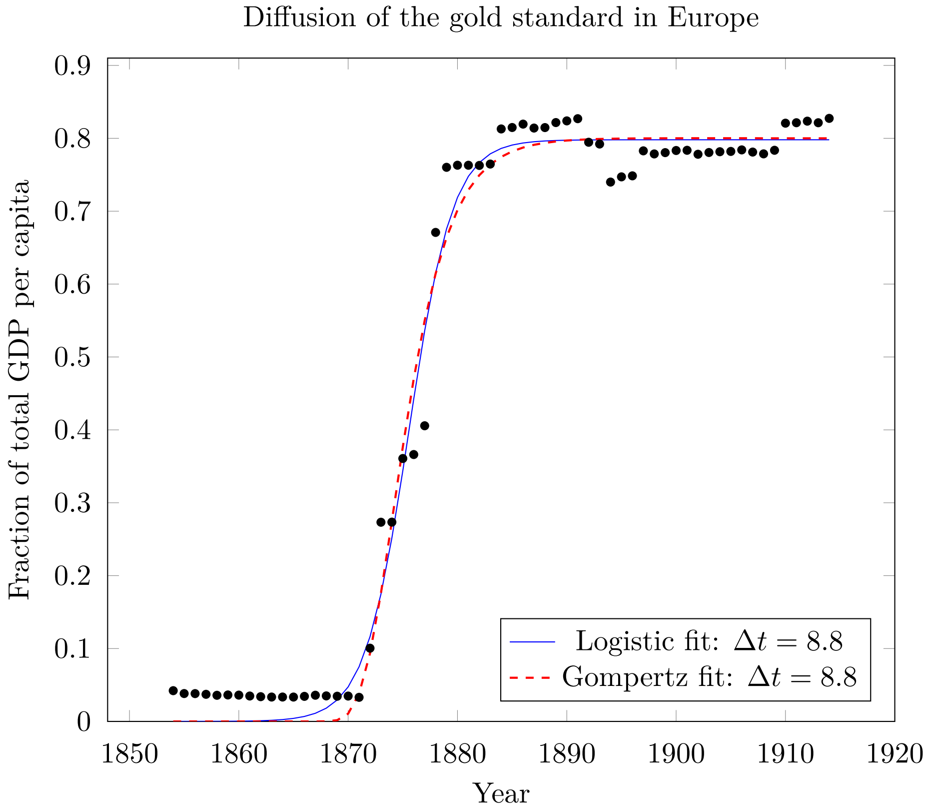


Figure 4.9: Diffusion of the gold standard in Europe (GDP per capita of adopting countries in proportion to total GDP per capita of all countries in data set). Time series from 1854 to 1914. Fitted 3 parameter logistic function in red ($K = 0.789$, $b = 0.497$, $t_m = 1875.56$ and $\Delta t = 8.836$) and a fitted Gompertz function in blue ($K = 0.8$, $b = 0.349147$, $t_m = 1874.19$ and $\Delta t = 8.834$). Data sources: [83, 82, 86, 87]

[100]. Countries imposed embargoes on the export of gold to block the outflow of gold during the war [97, 98] and as monetary regimes dramatically shifted, the international gold standard declined, from which it never recovered.

4.2.3 The temporal diffusion of the gold standard

The temporal diffusion pattern of the gold standard is logistic in shape, with the three-parameter logistic function best describing it (see Figure 4.9). The initial values above zero is a result of the early adoption by Portugal, followed by a period of growth at a superexponential rate in the 1870s with an inflection point in 1875 (t_m), and finally a period of saturation[11] around a K value ($K = 0.789$) of 79% of the total GDP per capita.

[11] If we were to consider Austria-Hungary as an adopting country of the gold standard, then saturation would reach close to 90% at its peak.

Our estimated logistic curve has a short diffusion duration, a Δt of 8.836 years. Indeed, in a 7-year interval between 1872 and 1879, the silver and bimetallism regimes were nearly completely substituted by the gold standard. All European countries — with the exception of Austria-Hungary, Italy, Greece, and Spain — adopted the gold standard prior to the US. The two drops in the actual data during the saturation period, in 1891 and 1894, are due to Portugal and Italy exiting the gold standard. Disregarding the drops, by assuming a unidirectional adoption process would lead to superior fitted parameters but it would be a gross simplification of the actual diffusion process.

The difference between the fittings of the logistic and Gompertz functions is marginal. The fitting of the logistic function surpasses the Gompertz in terms of R^2, with an R^2 of 99% for the former compared to 98.8% for the latter. The estimation of the K parameter is nearly identical for both curves, with a K of 0.798 and 0.8; the actual saturation level reached in 1913 is 0.821. The early adoption by Portugal meant that the diffusion curve is not asymmetrical, which could have made the Gompertz a better fit. The Δt of the two functions is near identical, at 8.836 and 8.834 years. A diffusion duration (Δt) under 20 years is categorized as *fast* according to Grübler et al. in [19].

The US and nine major Western European countries, including Germany and France, had a silver or bimetallism regime prior to their gold standard adoption in the 1870s. Thus, the diffusion process of the gold standard can be categorized as a *substitution/replacement* process [16], as opposed to a growth process. Grübler noted that substitution processes tend to have short diffusion times compared to growth processes, this case strengthens that position. Substitution processes are suited to be modeled by the Fisher-Pry transform (see Section 2.2.3).

In Figure 4.10, the Fisher-Pry transform of the three-parameter logistic diffusion is plotted. The steep slope of the linear function signals a fast diffusion duration, a process that took place almost entirely in the 1870s. As it is customary to do in Fisher-Pry transforms, the data points in the early and late stages of the process (which tend to be outliers in Fisher-Pry transforms) are ignored and not included in the plot. The data points are well described by the linear function given by the transform on a semi-logarithmic scale, which means the diffusion pattern is logistic in shape.

In the Fisher-Pry model, it is assumed that substitutions diffuse to complete saturation with $K = 1$, but in this case, the gold standard does not diffuse to 100% but to around 80%, thus the K of our model has been adjusted. Technological innovations have been shown to not diffuse completely (i.e., they do not reach 100% market share) [20], they start to get substituted by more recent and competing technologies during their saturation period, before the decline. The process for the gold standard diffusion is also not a simple unidirectional substitution process, some adopters exit the gold standard during the diffusion period. Finally, this case is not a binary substitution of the silver standard by the gold standard, but a diffusion process with multiple competing monetary standards: the gold standard in competition with fiat, silver, and bimetallic standards. We view the Fisher-Pry transform of the gold diffusion as a substitution process of the bimetallic and

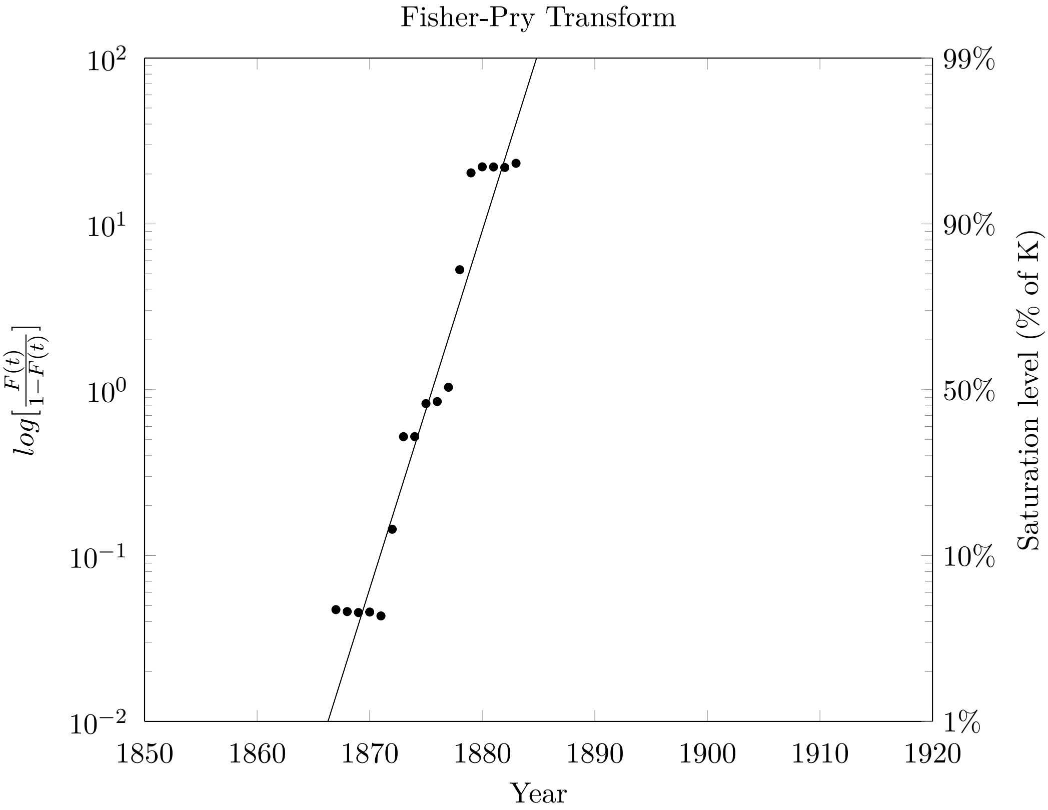


Figure 4.10: Fisher-Pry transform of the diffusion of the gold standard. Parameters from the three-parameter logistic function ($K = 0.789$, $b = 0.497$, $t_m = 1875.56$). Data sources: [83, 82, 86, 87]

silver standards for the gold standard; in other words, the $\frac{Y}{1-Y}$ adoption ratio is the ratio of gold standard adopters over non-adopters who are on a silver or bimetallism standard.

Let us evaluate the goodness-of-fit, the R^2 of 99% and estimated K which is close to the actual K value indicate a good fit. Concerning the residuals, no more than 5% of the standardized residuals should fall outside of $|2|$. In this case, we have three clear outliers out of 61 data points (plotted in Subfigure 4.11a: in 1873, 1877, and 1877 (ignoring the marginal outlier in 1879), hence the standardized residuals indicate a good fit for the data. The outliers are during the superexponential growth period, with large growth spikes within a small period, resulting in a larger variance and these outliers. The growth was not constant due to the unit of adoption of our analysis: countries. We can expect a more constant annual growth rate with smaller units of adoption. The early non-random pattern (which is expected in nonlinear models) in the standardized residuals indicates that the process started at an initial value higher than zero (with the early adoption by Portugal in 1854). In such cases, Grübler says in [16] to simply subtract the higher initial value, and reintroduce it into the model with a constant intercept.

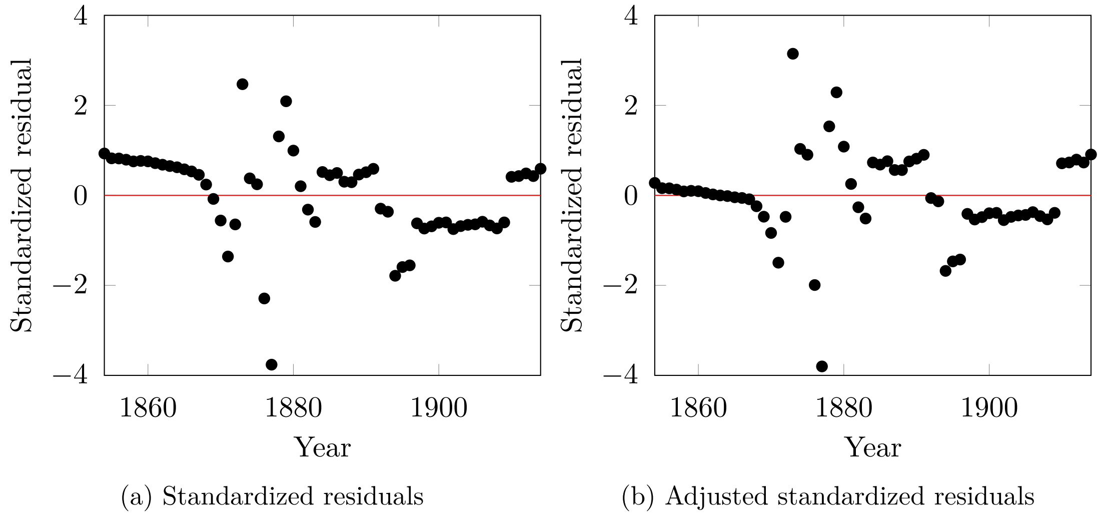


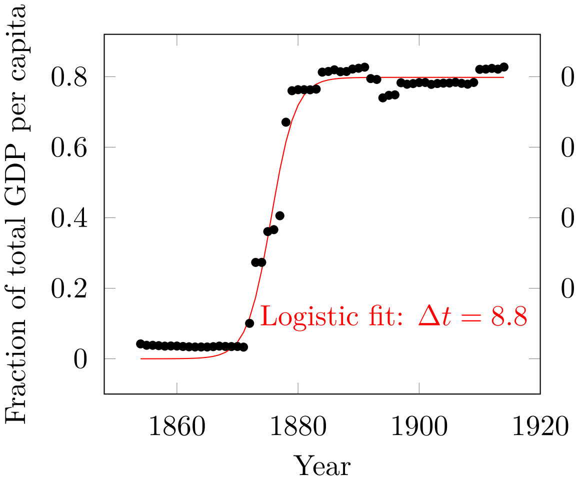
Figure 4.11: Standardized residuals of the gold standard diffusion with the three-parameter logistic function and the adjusted logistic function with an intercept.

When recalculating the parameters using an adjusted three-parameter logistic function with a constant intercept of 0.0329, the early stage fits better according to the residuals (non-standardized) and diffusion plot (see Figure 4.12). The standard error (σ) is marginally improved at 0.034 compared to 0.037 without the intercept (see the adjusted standardized residuals in Subfigure 4.11b). The difference is not significantly larger after adding the intercept because the better fit in the early stage is partially offset by a weaker fit in other stages. Concerning the diffusion duration, the intercept cancels out the adoption of Portugal, meaning a shorter diffusion duration, a 0.8-year reduction in the Δt.

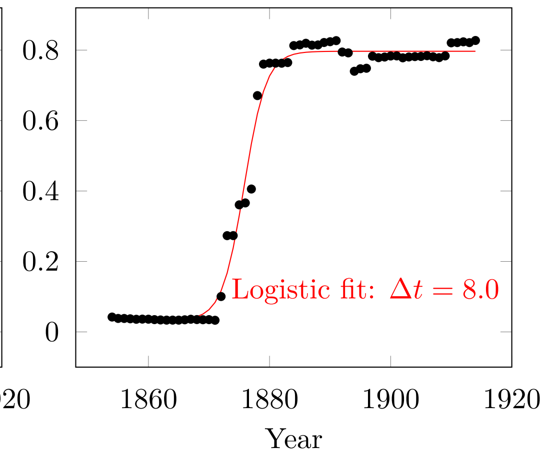
4.2.4 The spatial diffusion of the gold standard in Europe

The innovation center of the gold standard is Great Britain, starting in 1819 [81]. No other country joined until 1854 when its trading partner Portugal decided to adopt. Until 1872, these two coastal West European nations remained the single countries on a gold regime. As shown in Subfigure 4.13a, Europe was separated into three groups: on the West were Great Britain and Portugal on gold; in the center were the 'Latin' countries — Spain, France, Belgium, Switzerland, Italy and Greece — on a bimetallic standard; and in the East and North East were the countries on a silver standard such as Germany, Austria-Hungary, Scandinavia and the Netherlands.

After the turning point in 1872 (Germany's switch from silver to gold), the Nordic countries followed the next year by also abandoning silver in favor of gold as seen in Subfigure 4.13b. Thus, the diffusion spread to the most of Northern Europe. By 1878, the spread continued southwards, with Belgium, the Netherlands, France, and Switzerland all adopting. At this point, except Austria-Hungary, all countries north of Italy had joined



(a) Logistic function



(b) Logistic function with intercept.

Figure 4.12: Comparison between the logistic function with and without intercept for the diffusion of the gold standard. Parameters with intercept: $b = 0.610964$, $t_m = 1876.189$, $\Delta t = 7.19$ years, $K = 0.728853$; parameters without intercept and data source: see Figure 4.9.

the gold standard; the southern European countries (Italy, Spain, and Greece) had large public debt and yearly budget deficits making the adoption more difficult. By 1912 (see Subfigure 4.13d, all countries except Portugal, Spain, and Italy were on a gold standard[12], not out of choice but out of fiscal necessity. The final result during the classical gold standard era was a diffusion that spread centrally, eastwards and northwards from Great Britain, all the way East to Russia in 1897 [83], but it was never stably maintained in Southern Europe.

The spatial diffusion process can be categorized as partially *contagious* and partially *hierarchical* but it was an expansionary process as opposed to a relocation process (see Section 2.1.2). We categorize the process as hierarchical because there is a financial barrier to the adoption of the gold standard. Great Britain and Germany were among the earliest adopters; it is no coincidence that they were the two most economically developed European countries in the 1870s, with Germany overtaking France regarding GDP during this period [13]. They could afford the costly transition to a gold standard. Portugal was also an early adopter though they were not as industrialized, and they eventually had to exit the international gold standard in the 1890s prematurely.

The spread can be considered contagious, spreading with a 'neighborhood effect.' Many neighboring countries adopted simultaneously or within a close time interval. The reason for this, as explained in Section 4.2.2, is firstly due to the existence of monetary unions, and often neighboring countries were a part of the same monetary union. Secondly, as

[12] Austria-Hungary had a gold standard with restricted convertibility (see Section 4.2.2).

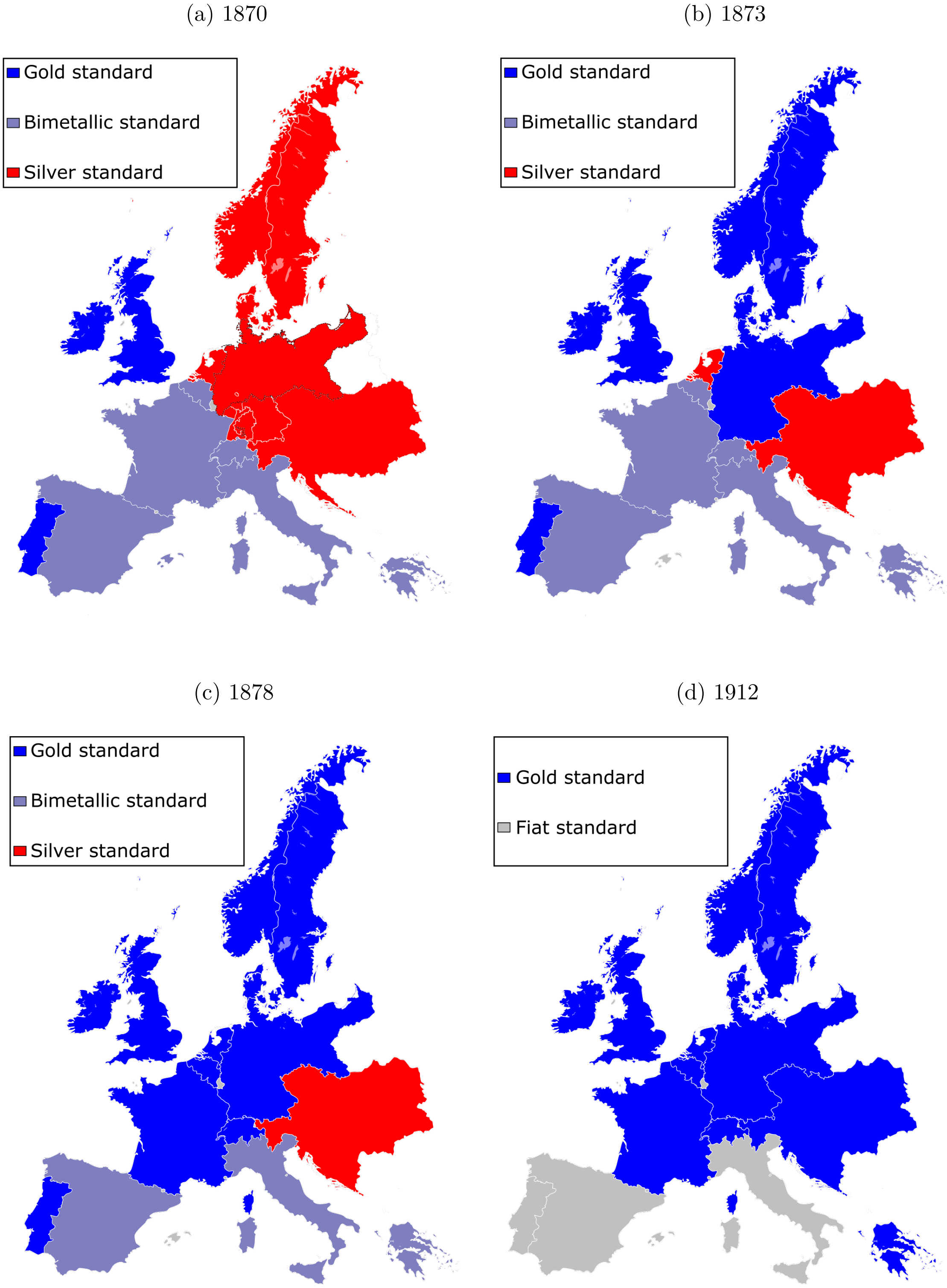


Figure 4.13: Spatial diffusion of the gold standard in Europe between 1870 and 1912. Note: Austria-Hungary was on a de facto gold standard without convertibility after 1896. Data sources: [83, 82, 88, 89, 86].

described in Section 4.2.5, there was a bandwagon effect with countries incentivized to adopt the monetary regime of their main trade partners; and as these trade partners were more commonly nearby neighbors as opposed to far away, the resulting spatial pattern of adoption resembles a contagious one.

4.2.5 The determinants of the diffusion

After our detailed review of the historical context of the diffusion and a temporal and spatial analysis of the diffusion, we attempt to infer with the literature the essential perceived advantages and disadvantages that affected the rate of diffusion and a nation's decision to adopt.

Positive determinants for the adoption

The spread of the gold standard across the Western World was likely driven by the incentive of access to the capital[13] markets of the industrialized economies of the time who were early adopters of the gold standard, notably the world's economic superpower of the time, Great Britain [82]. The adoption of the gold standard would act as a "seal of approval" of a government's ability to repay debts, and therefore aid a country in its attempts at securing foreign short-term and long-term loans [85, 98, 101, 82]. According to [85], this credibility allowed countries who adopted and maintained the gold standard, to save a 30 basis point spread (0.3% in interest) on their foreign loans prior during the classical gold standard era, Meissner also noted the importance of the reduced borrowing costs in affecting the decision to adopt [83]. A country's adherence to the gold standard represented a stamp of economic credibility because it implied a responsible fiscal and monetary policy, stable institutional environment [83] and a low inflation rate [101].

The fixed exchange rate between the currencies of the block of countries on the gold standard was to be perceived as an advantage, as it would lower transaction costs. Spain — the only European nation not to join the gold standard at any point during the classical era — suffered from volatile currency exchange rates [86]. The lowered transaction costs, thus facilitated trade flows among adopting countries, resulting in a network effect. The international gold standard was poised to benefit from a network effect (though each additional adopter would impact the price of gold and silver), with trade partners following each other in joining the gold standard in order to improve international trade by removing exchange rate uncertainty (e.g., smaller countries that are highly reliant on trade such as Portugal with Great Britain). Countries whose main trade partners were already on the gold standard, would in some cases, already have a significant amount of gold coinage circulation from their trading partner [87], thus facilitating the adoption. The advantage in adopting early or with the early majority was to avoid falling victim to the domino effect on the price of silver, the cost of switching could increase if delayed: downward pressure on the price of silver put pressure on silver and bimetallic nations to switch over. After the plummeting price of silver during the 1870s, adopting the gold

[13] International institutions that provide capital loans such as the World Bank and The International Monetary Fund (IMF) were not created until after The Second World War.

standard adoption could be perceived as a *preventive innovation* (see Section 2.1.1), to limit one's exposure to further depreciation of silver.

The argument that gold was adopted because it was cheaper to transport due to its higher density does not hold up, as demonstrated by Flandreau in [79]. Silver and gold were perfect substitutes concerning international transaction costs because the shipping costs during the classical gold standard era did not depend on the volume or weight of a shipment, but on its value [79]. Thus switching to gold would not have a negative effect on shipping costs.

Countries on the bimetallic standard already had gold convertibility, gold reserves and gold coinage in place (the 1850s Gold Rush resulted in the major European countries having large amounts of their currency in gold by the 1860s [79]) reducing switching costs; countries on the silver standard already had processes and technology in place for maintaining a commodity-based regime (such as minting and storing coins). In the diffusion framework of Rogers [7], this would be categorized as increased *compatibility* (see Section 2.1.1).

The *observability* of the consequences of being on a gold standard would have been skewed positively, due to the earliest adopter, Great Britain being the global economic superpower of the 19th century — and was only overtaken in terms of GDP per capita by the US towards the end of the classical gold standard era [67]. Hence the action of observing and trying to imitate the monetary regime of the economic leader of the time would have been a powerful draw towards the gold standard. Meissner [83] considers the British position on the gold standard as one of the most significant in influencing the global switch because it highly influenced Germany's decision to adopt, which then resulted in a domino effect (which is self-reinforced by the network effects).

Although it could be perceived as a disadvantage, there is an inherent *trialability* of the gold standard, with adopting countries able to — at any moment and without prior notice — permanently or temporarily exit the gold standard by suspending or restricting convertibility. Therefore the "seal of approval" of a country on the gold standard was not enough for creditors to turn a blind eye towards the financial situation of a country, as a country could exit the gold club at any moment if it did not have the means to maintain convertibility, which was often the case during a war. An example of trialability is when Greece experimented with the gold standard for less than a year between 1884 and 1885, but convertibility could not be maintained due to gold outflows. Greece eventually managed to join the gold standard twenty-five years later in 1910 and successfully maintained it for a four-year duration until World War 1 [82].

Negative determinants for the adoption

It is hard to say whether the deflationary nature of gold had an impact on the diffusion rate because gold was not always perceived as deflationary during the 19th century; it was seen as inflationary in the 1850s and 1860s after a supply shock of gold as a result of the Californian Gold Rush of the 1850s [79]. The low price of gold in the 1850s was a reason why Portugal was able to adopt the gold standard [87]. It is also incorrect to assume that a deflationary currency was viewed as negatively as today, as the Great Depression had not yet happened.

The deflationary nature of the gold standard made it attractive to creditors, but less attractive for debtors. Members of the agrarian industry — who have to borrow large amounts of money — were naturally opposed towards institutional innovations that were deflationary. The silver movement in the 1870s to 1890s, driven by the agrarians did have a political impact in the US — forcing the US government to buy significant amounts of silver. However, there is no evidence that they affected (by delaying) the country's decision or that of other developed countries the diffusion rate of the gold standard, because the silver agitation in industrialized nations happened only after a period of deflation [79], years after the majority of countries had already adopted gold. These opposing forces had no impact on the maintenance of the gold standard, and they were unsuccessful in making any developed nation switch back to a bimetallic or silver standard.

Due to the market dynamics of the precious metal prices when selling or buying large amounts, the substitution of a silver or bimetallic regime for a gold standard regime, was a costly and arduous process which discouraged early adopters [79]. If a country decided to switch from silver to gold, it would put considerable pressure on the price of silver depending on the quantity of silver a country had to sell off. It would also bring into question whether the remaining silver and bimetallic countries could absorb the sell-off, this would lead to a feedback loop that would drive more silver regime countries to abandon silver if the value of silver started to plummet. Before a de facto gold standard adoption, a country has to increase and maintain its stocks of gold, often financed through the silver sell-off, loans or public taxation [82]. As previously explained, Germany was able to fund its transition through the war indemnity that it imposed on France [79]. A country's national income was, therefore, an important factor in the adoption of the gold standard; industrialized nations such as Great Britain, Germany, and France had the finances to adopt faster than the lower-income countries such as Greece or Austria-Hungary [84].

Once the gold standard was adopted, there was also a cost in *maintaining* the commodity-based monetary system. The gold standard requires an active role by the country's central bank or treasury to maintain stability. Fixed exchange rates were maintained through market interventions by central banks selling or buying gold against their currencies, though the countries that adhered to the full gold standard had the convertibility of notes into gold coins that helped maintain the exchange rate through arbitrage traders [86]. The convertibility could stabilize the exchange rate as long as the country had a sufficient gold reserve.

Critics of the gold standard pointed out the high resource cost of maintaining a monetary system based on a commodity, such as American economist Milton Friedman, who estimated the cost of maintaining a (fully backed) gold standard or any other commodity-based currency to be around 1.5% of a country's GNP [102]. The value of precious metals like silver or gold is only valuable based on their perceived scarcity (ignoring the slight practical use of gold in dentistry), they cannot be eaten or used to build a house, and hence do not differ from paper money. Milton Friedman, therefore, was critical of the wasted resources necessary to base a monetary regime with paper currency back by a commodity, when the commodity could be skipped altogether with the same result and saving wasted resources. In other words, he questioned: why to go through the high effort of mining a precious metal from the earth, just to transport it across the world and keep it in a locked vault of the US Federal Reserve, if there exists a valid alternative that skips the commodity [96].

A disadvantage in adopting the gold standard or any other commodity-based standard is that it stops a country from freely being able to depreciate its currency's exchange rate, which can be important to balance trade deficits; the adoption of the gold standard also limits expansionary fiscal policies [101]. To finance the gold standard, Southern European countries such as Greece, Portugal, and Italy had to raise taxes and maintain fiscal discipline [82], which was a difficult task as these countries were burdened by the weak organizational structure of their states [64]. Also, the gold standard (and all commodity-backed regimes) restricts a country's money supply creation during a war. Wars are expensive, they require large amounts of initial spending, which makes it difficult to maintain a metallic regime while fighting in one, or right after. Moving onto a paper regime is an option for a state to finance a war, as it allows to increase the money supply without silver or gold backing rapidly. British and US both have examples of such conduct: interrupting specie convertibility and adopting a fiat currency to finance a war [81]. France is a counter example of such a response, by not adopting a paper regime during, or directly after the Napoleonic wars.

The experience of adopting the gold standard did not affect all countries equally. Depending on the stage of economic development of the country, the industrialized nations were less affected by short-term variations in product prices due to their mature diversified economies in 1870. The less developed countries, with less stable economies and higher sensitivity to unexpected price changes, had a harder time maintaining the gold standard [82]. We do not know if it was known at the time, as it could have had an effect on the adoption decision for the less industrialized countries.

4.3 Cross-case analysis

In this section, based on the two case studies analyzed in Section 4.1 and 4.2, we compare and contrast the findings to find common patterns in the diffusion process. The resulting theory is compared with the existing literature, to check if it validates the diffusion research by IIASA scholars as we attempt to determine whether institutional innovation diffuses in a similar pattern to technological innovations. Even though we are restricted to the two cases of institutional innovation examined in this work, we argue that the quality and relevance of the cases are more important than the number of cases — the adoption of compulsory education and the gold standard were *major* cases of institutional innovation.

Our case studies cover the two types of diffusion processes defined by Grübler (see Section 2.1). Our first study, on the diffusion of compulsory education, is a *growth* process, though nearly every diffusion process is partially a substitution process, in the words of Grübler "Hardly any innovation is introduced into a vacuum."[16, p.40]. The diffusion of compulsory education could be interpreted as a partial substitution of the pre-existing privatized schooling. However, we categorize it as a growth process because state-driven and *compulsory* schooling were novel; a whole network of schools, teachers and a state-driven curriculum had to be implemented with funding from the state. Our second case study is a typical example of a *substitution* process. After competing with the bimetallic, silver and fiat standards, the gold standard diffused to become the international monetary standard between the early 1870s and 1914, replacing the previous monetary standards that were employed.

Within the first case on compulsory education, we have subcases that can be compared with one another. First is the temporal diffusion of compulsory schooling laws among seventeen industrialized countries from Table 4.1 (excluding the US). Then, a subcase on the temporal diffusion of compulsory education laws in northern and southern US states. Finally, a subcase on the temporal diffusion of secondary enrollment ratios of the eighteen countries in Table 4.5. Concerning the spatial dimension, the European and American diffusions are studied separately and can be cross compared.

4.3.1 The temporal dimension

In both cases — the diffusion of compulsory education laws and the diffusion of the gold standard — the diffusion pattern is s-shaped and best described by the three-parameter logistic function. This primary finding is identical to the recurring conclusion drawn by Grübler et al. [9, 19] and in the words of Wilson "We found the logistic function consistently provided the best fit to the data." [61, p.386]. In the case of the gold standard, the logistic function with an intercept is marginally a better fit because our calculation of the diffusion does not start at zero. This can especially be the case when studying the diffusion of institutional innovations because we analyze on a country scale, and not an individual scale, causing a large initial value which can be better modeled by the addition of the intercept.

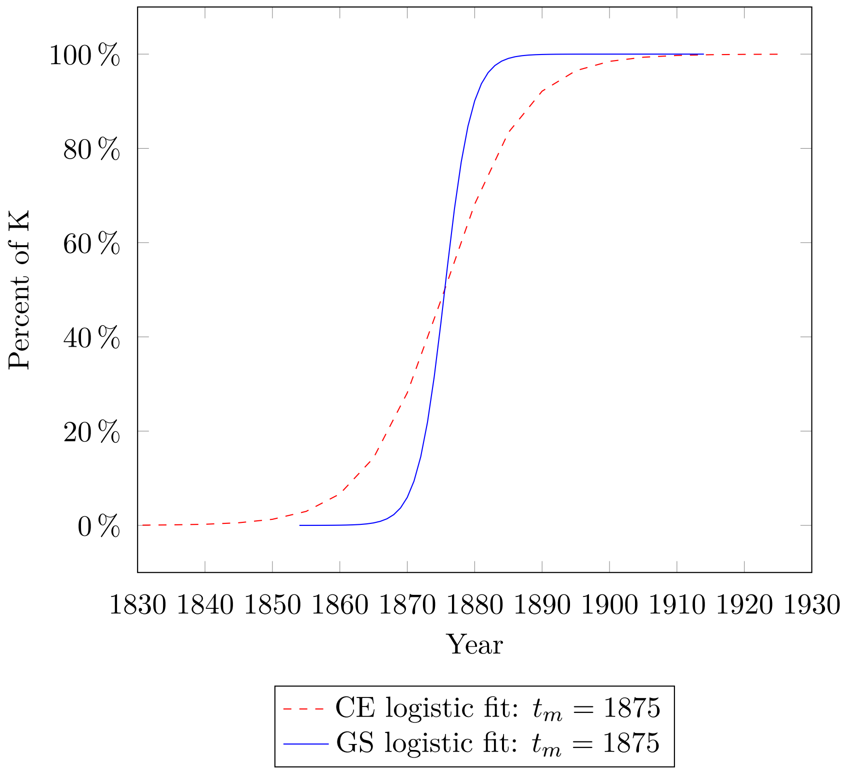


Figure 4.14: Comparison of the compulsory education in Europe and gold standard logistic curves (in percentage of their respective K). GS = Gold standard, CE = Compulsory education.

With regards to the diffusion of secondary enrollments within the compulsory education case study, only three countries out of eighteen do not surpass the 95% R^2 threshold for the logistic function, of which Portugal is only slightly under (94.6%). Germany (91%) had a disrupted history during the secondary level enrollment diffusion period, which explains a disrupted diffusion pattern. The USA (92.8%) has a long-spanning diffusion duration dating back to the early 1800s and experienced strong disruptions to its secondary enrollment ratio during and after the Second World War; without the outliers from the Second World War, the USA's diffusion pattern would be well fitted by the logistic curve. Overall, the diffusion pattern of secondary enrollment ratios is strongly s-shaped, with over nine out of eighteen countries having an R^2 over 99%, and six countries approximately over 98% (see Table 4.2).

In Figure 4.14, we observe an overlap of the compulsory education law adoption and gold standard curves, the striking similarity is that both curves have their inflection point in the same year, a t_m in 1875.52 for compulsory education and a t_m in 1875.56 for the gold standard. The 1870s were the peak years regarding institutional innovation adoptions for both cases.

Table 4.5: Duration of different diffusion processes[1]

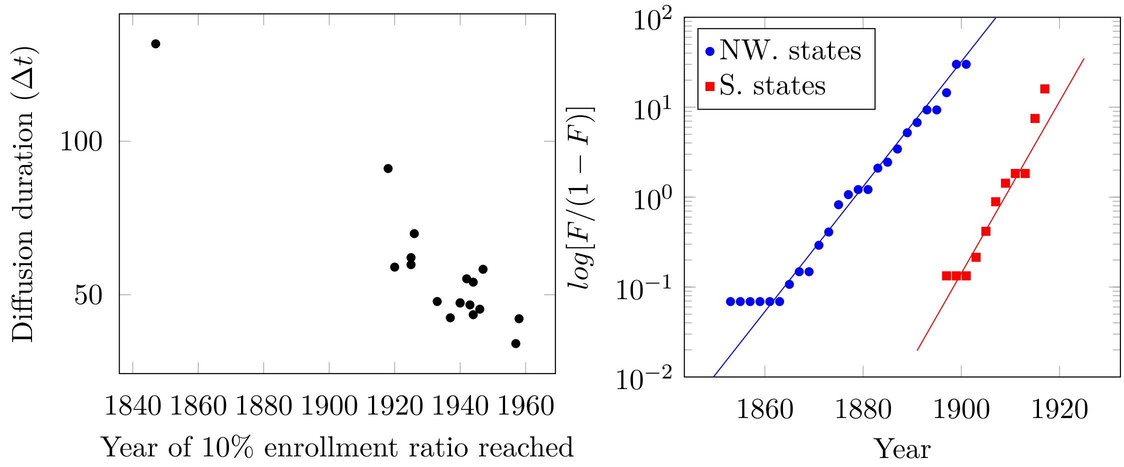
Duration	Example	Δt (yrs)	Diffusion (D) Substitution (S)	Ref.
Long > 40 yrs	Coal vs. traditional energy, World	110	S	[103]
	Railways, World	60	D	[16]
	Railways, France	47	D	[16]
	Secondary schooling enrollments, Europe[2]	54	D	4.1.5
	Secondary schooling enrollments, US	132	D	4.1.5
Medium 20-40 yrs	% households with radio, US	25	D	[104]
	Compulsory education laws, Europe	26	D	4.1.3
	Compulsory education laws, northern US	29	D	4.1.3
	Compulsory education laws, southern US	19	D	4.1.3
Fast < 20 yrs	Cars vs. horses, France	15	S	[16]
	Cars vs. horses, US	12	S	[32]
	Gold vs. silver & bimetallic standards, Europe and US	9	S	4.2.3

[1] Table adapted from [19, p.264], Δt values rounded to the nearest integer.
[2] Non-weighted average of the European countries in our data set (thus excluding Russia).

4.3.1.1 The diffusion duration of institutional innovation

Although the two cases reached their peak growth in the same year, the duration of their diffusion differs. The diffusion of compulsory education laws in Europe has a Δt of 25.77 years, the gold standard diffused in approximately one-third of that time duration, with a Δt of 8.83 years. The smaller Δt results in a steeper curve shown in Figure 4.14. For secondary level enrollment ratios, the average Δt is 57.7 years; if we disregard the USA the average Δt drops down to 53.3 years — this is the average diffusion rate of secondary enrollment ratios in Europe. The faster diffusion of the substitution process compared to the growth process validates Grübler's observation in [16] about substitution processes diffusing faster than growth processes because they tend to benefit from compatibility with the existing environment, system or infrastructure which was necessary for the previous technology to diffuse. In Table 4.5 we summarize the diffusion duration of our case studies, and in comparison with existing case studies. The diffusion durations of our cases are present across all duration categories, from a long diffusion duration with the secondary enrollment ratios to a fast diffusion duration with the gold standard. Hence, institutional innovation is not confined to a single diffusion duration category, and its diffusion duration is case specific.

For the spread of the gold standard, the necessary processes, financial systems, and institutions were already in place due to the bimetallic and silver standards that were



(a) Catch-up effect in secondary enrollments (b) Catch-up effect in compulsory education laws

Figure 4.15: Catch-up effect in the diffusion of secondary enrollment ratios and compulsory education laws in the US. Subfigure 4.15b adapted from [34, p.555]; data sources: [44, p.514-515], [67].

operational, allowing a possible swift substitution process. The main barrier to the rate of diffusion of the gold standard was a financial barrier as discussed in Section 4.2.5, and not cultural or institutional opposition as it was for compulsory education. Compulsory schooling and enrollment ratios could not diffuse as fast because it was a larger institutional change — the state schooling infrastructure still had to be developed and supplied especially for the poorer regions with compulsory attendance enforced, many teachers found and trained, and a state-driven curriculum created. We assume based on our case studies, ceteris paribus, the larger the change (less compatibility) caused by an institutional innovation, the slower the diffusion rate.

In Subfigure 4.15a, we can observe clear evidence of the catch-up effect in secondary level enrollment ratios. The catch-up effect is referred to as the *advantage of the backward* by Glaziev [35] because of the benefits of the learning effect for later adopters. The laggards to reach a 10% enrollment ratio had much a shorter diffusion duration compared to the larger Δt's of the early adopters. Despite the significant variance in the date of reaching a 10% enrollment ratio, the countries all reached saturation within a close period of time of each other. This synchronizing saturation is named the "The season of saturations" [16, p. 87] by Grübler. He describes with empirical evidence the catch-up effect in his work on the diffusion of infrastructure [16, 9], but an additional effect observed was that the later adopters also experienced a lower extent of their diffusion (lower K parameter), which is not observed in our cases. We also observed a catch-up effect in the spread of compulsory education laws in southern US states in Subfigure 4.15b. Their diffusion curve is 30 years behind the curve of their northern counterparts, but with a shorter diffusion duration (Δt) of 18.68 years, compared to 28.68 years for the northern states.

The contrast in the rate of diffusion is noticeable in the plot of the Fisher-Pry transform, the difference in the slope of the linear functions can be easily discerned. The diffusion process is thus a social learning process, with late adopters benefiting from the experience gained by the early adopters. This social learning process was observed in [16, 41, 35].

4.3.2 The spatial dimension

The spatial diffusion process in all cases were hierarchical and expansion processes, and partially contagious for the spread of compulsory education laws in the US and the gold standard. See Table 4.6 for a summary of the processes for the different cases. The hierarchical aspect is reoccurring in technological innovations with wealthier countries commonly being the early adopters. In institutional innovation, the hierarchical aspect is just as present, with innovations such as the gold standard requiring institutional and also financial stability to be adopted, promoting an adoption by the more industrialized countries first. The diffusion of compulsory education was a strongly hierarchical process because the innovation leapfrogged the regions where conservatism and conservative institutions such as the Catholic Church were firmly present.

The types of barriers in the two cases are distinct. Compulsory education laws encountered a cultural barrier — that of active resistance towards change by certain social groups. The gold standard had an economic barrier resulting in the previously described hierarchical process, requiring a responsible fiscal behavior and a developed financial system.

Table 4.6: Summary of spatial processes

Case	Expansion process	Relocation process	Contagious process	Hierarchical process
Compulsory education laws, EU	+++	-	-	+++
Compulsory education laws, USA	+++	-	++	++
Gold standard	+++	-	+	+++

The intensity of the spatial processes are rated qualitatively by the author on a scale from: absent (-), low (+), medium (++), to high (+++).

4.3.2.1 Scandinavian early adopters and southern laggards

We observe two groups of countries with close geographical proximity that have a particular pattern in their adoption times in the institutional innovations cases that we studied. The first group comprises the four southern European countries: Italy, Spain, Portugal, and Greece. This southern block suffered in the 19th century from a weak organizational structure in their states, fragile economies, and a high public debt [64], which highly affected the diffusion of the institutional innovations. Although three of them adopted a compulsory education law early on, their weak and fragmented states failed to implement a national school system within a considerate time-frame. The economic weakness of the southern group was even more illustrative in the diffusion of the gold standard. Spain never reached the necessary fiscal and organizational structure to adopt the gold standard [86], Italy and Portugal both had to leave the gold standard during the classical era due to their weak national finances, and Greece only joined the gold standard during a two-year period at the end of the international gold standard era. The second group comprises the Nordic countries: Denmark, Sweden, and Norway. In all case studies — regarding compulsory education laws adoption, enrollments ratios and the gold standard adoption — they were the early adopters. The reasons for their early adoptions were not the same in all cases, but their institutional stability after the Napoleonic Wars was a factor. To illustrate how much of a gap there was between the two groups, the earliest to reach the 90% primary level enrollment ratio threshold in our data set was Denmark, roughly around 1840; Sweden and Norway achieved this ratio in 1855 and 1890 respectively. It took Spain until approximately 1967 to reach the same 90% primary ratio, 127 years after Denmark; Greece until 1959, an 119-year gap; Portugal until 1955, an 115-year gap; and finally Italy, until 1927, an 87-year gap. In our case on the gold standard, the three Scandinavian countries successfully switched to gold from silver early on, in 1873, before silver started strongly depreciating.

4.3.3 Characteristics of the diffusion of institutional innovation

As it is with any change, institutional change is dependant on being ideologically accepted by the adopter. The opposition to change is strongly present in both case studies — the adoption of compulsory education and the gold standard were socially opposed by certain social groups during certain periods of their diffusion process. The opposition included conservative social groups for compulsory schooling and the Free Silver Movement for the gold standard. In both cases, the innovations were susceptible to social resistance despite the adoption decision mostly being an Authority Innovation-Decision. The opposing social groups were the ones who perceived the innovation to be a threat, and the success of their opposition depended on their degree of influence over the decision unit. This social resistance towards change is also present in technological innovations, Grübler modeled the attacks on farming equipment as a diffusion process in [9].

The *timing* of the diffusion of an institutional innovation is not by coincidence; both innovations required certain background conditions (i.e., other institutional or technological innovations to diffuse beforehand) to be met in order to diffuse. It is only after the 1800s that the necessary conditions were present: the spread of compulsory education would have been difficult before the French Revolution and the early diffusion of democracy in Europe, and it is only after the Industrial Revolution that the necessary conditions were met for the international gold standard to diffuse.

The perceived economic gain from adopting an institutional innovation may not be as obvious or quantifiable as with technological innovations. Thus the diffusion process depends to a greater extent on the act of imitation. In the diffusion of compulsory education laws, Great Britain only adopted compulsory education when it deemed necessary to follow suit of their rising economic rival Germany. In the gold standard, the imitative nature of the adoption process is even more explicit. A unified Germany decided to imitate the leading economic power of the time, Great Britain, in adopting the gold standard leading to an imitative action and bandwagon effect of the rest of Europe who was incentivized to adopt the monetary regime of their main trade partners despite high switching costs.

Re-invention — a concept in diffusion theory (see Section 1) described by Rogers [7] which concerns the modification of the original innovation by the adopter — is well present in institutional innovations. The form and components of an institutional innovation adopted in country A may not be the same as in country B, because there is flexibility involved, especially for the type of innovation that we study. Grübler refers to the concept as pervasive change: "The basic innovation that creates a new radical technology is follow by incremental changes that accrete around the basic innovation" [19, p.251-252]. Though the core idea of *compulsory* education diffused broadly, the actual implementation varied between countries and changed over time, such as the total compulsory period in years or the strictness of the enforcement. The gold standard also experienced incremental changes, notably in how national banks calculated how much gold reserves they were required to hold. Countries did not hold a pure gold standard — that of the quantity of money backed 100% by gold, but a fractional gold reserve often with a lower limit dictated by legislation or decision units which changed over time. Lowering the fractional gold reserve allowed the money supply to be increased as long as convertibility could be maintained, thus the national banks had a powerful tool to influence control over the money supply without having to increase their reserves in gold [96]. The gold reserve ratio differed between countries, and depended on the monetary policies of the central bank; certain policies required a higher reserve ratio. Sweden maintained a gold reserve ratio between 50% to 100% per the requirements of its intervention policies to preserve the gold standard [86]. The gold standard also contributed to a re-invention and growth of the banking system which diffused strongly during the classical gold standard era [14], which (we cited in Section 4.2.2) Milton Friedman described as "increased the amount of money that could be pyramided on each ounce of gold" [81, p.1170].

Another interesting observation is the reoccurring role that *war* played as a driver or barrier for institutional change. As discussed in Section 4.1.2 on multiple occasions such as in Germany and Great Britain, the adoption of compulsory education was a method of militaristic and industrial revitalization, to keep up or surpass competing nations [65]. In the case of the gold standard, war played a reversed effect with it acting as a barrier to adoption; countries exited or delayed their adoption of the commodity-based standards to finance the war. The major examples are Great Britain and the US, whom both had a temporary period of a fiat regime during the American Civil War and the Napoleonic Wars respectively before adopting the gold standard. France is a counterexample, having never left the bimetallic regime despite fighting half of Europe in the Napoleonic Wars. Due to the gold standard adoption being a costly adoption process as discussed in Section 4.2.5, it is evident that expensive episodes for countries such as war would have a large effect on the diffusion process.

CHAPTER 5

Conclusion

> *"Heterogeneity, not homogeneity, is the heart of things... Things are not born alike, they become alike."* [15, p.71]
>
> - Gabriel Tarde, *The Laws of Imitation*

5.1 Summary and discussion of the results

The objective of this work was to analyze the diffusion pattern of institutional innovation, and find out if it can be modeled by the logistic function. Our cases of institutional innovation were studied from a historical perspective, bringing to light the diffusion process of institutional change. The evidence in the quantitative analysis of our case study approach suggests that institutional innovations diffuse in the same pattern as technological innovations, in an s-shaped pattern best described by a three-parameter logistic function (with the option of an intercept). The diffusion rate is initially slow during the period of early adoption, and then the rate progressively accelerates during a period of mass adoption until reaching a point of inflection, at which point the rate decelerates while the laggards and those adverse to the innovation are the last to adopt, until saturation is reached.

The diffusion duration ($\Delta t's$) of our cases vary from long to short in duration depending on the nature of the diffusion (diffusion/substitution) and the degree of opposition towards the institutional innovation by influential social groups. In our first within-case analysis in Section 4.1 on the diffusion of compulsory education, the diffusion rate was of medium duration (26 years in Europe, 19 and 29 years in the southern and northern US) when measured by the enactment of compulsory education laws, and of long duration (54 years in Europe, and 132 years in the US) when measuring the secondary level schooling enrollment ratios. Even though compulsory education only started diffusing after the

Industrial Revolution, there is no evidence that it diffused out of the need or will to reduce child labor, or out of the necessity of an educated workforce demanded by the newly industrialized economy. However, there is evidence to suggest that the construction of a state-driven compulsory educational system became an innovative and useful way for state leaders to construct a national identity and transform the masses into national citizens [65, 68], notably in the US after a period of mass migration [43]. The diffusion of the gold standard, in contrast, was a substitution process and had a short diffusion duration (9 years), which is in confirmation of previous empirical results by Grübler [19] showing that substitution processes have a shorter diffusion duration because they are often compatible with the existing environment. The gold standard did not immediately spread throughout Europe after Britain's early adoption in 1821. Switching costs to the gold standard were high for countries that adhered to the silver or bimetallic standard, and the advantages of the gold standard were not convincing enough for countries to adopt prior the 1870s, as it was not necessarily a superior monetary regime to other commodity-based standards. The main advantage of joining the gold standard was access to the dominant capital market of the time which was Great Britain. Adhering to the gold standard meant better access to capital because it implied a responsible fiscal and monetary policy, and a stable institutional environment, thus a better ability to pay back debt. The turning point that propelled the diffusion of the international gold standard was in 1872, when the newly unified Germany decided to follow Great Britain in adopting the gold standard as their monetary regime, and fund their transition through the war indemnity that it imposed on France after the Franco-German War (1870-1871). The positive network effects from adhering to the same standard as the major economic powers, the bandwagon effect, and the downwards pressure on the value of silver from a nation selling its silver for gold stocks resulted in a positive feedback loop and a domino effect on the rest of Europe, incentivizing others to switch from the silver or bimetallic standard to the gold standard.

Institutional innovations are no different from technological innovations, in that they must surpass barriers to diffuse. The main type of barrier to institutional change in our cases was a social barrier (sometimes referred to as a cultural barrier), with a conservative stance taken by the social groups who perceived themselves to be negatively affected by the institutional change. A state-driven compulsory education system meant the Catholic Church and other social groups would lose their stronghold over the schooling system in the countries they held influence; it also meant that Southern US states would have to provide schooling for African American children, shortly after the abolition of slavery. The substitution of gold for silver, as the precious metal to which countries pegged their local currencies, eventually became negatively perceived by those with silver interests or the social groups who did not benefit from the deflationary nature of gold.

Concerning the spatial dimension of institutional innovation, the diffusion was principally hierarchical and expansionary as it is predominantly the case with technological innovations: wealthier countries could adopt and implement the innovation earlier than the less wealthy, especially in the case of the gold standard. Two groups of countries can be clustered together with similar adoption behaviors. The Scandinavian countries were early adopters in our case studies; they benefited from institutional stability after the Napoleonic Wars. The southern European countries were late adopters as a result of their weak organizational structures and their fragile economies in the 19th century [64].

We have shown evidence in our case studies that later adopters of institutional innovations benefited from a catch-up effect, learning from the experience of earlier adopters to efficiently use their resources to close the gap. The concept of re-invention — which is a modification of the original form of an institutional innovation to suit the adopter's needs — was observed: the actual implementation of the institutional innovation varied among the adopting countries, and it changed over time.

What are the implications of our results? First, our quantitative analysis of the diffusion duration (Δt) provides a measure of the rate at which an institutional innovation can spread across countries, and adds to the growing list of innovations that were measured in this manner. Although it is impossible to precisely predict the Δt of present and future institutional innovations, our Δt results can serve as a basis to help bound it. If we wrongly assume institutional innovations always to be positive changes, then the good news is that we have shown that institutional innovations can diffuse rapidly, such as it was the case of the international gold standard; the industrialized countries substituted their monetary regimes for a new one in less than nine years. There is, in fact, no limit to the speed at which the *rules of the game* can be changed. Technological innovations such as infrastructure or energy technologies have a limit to their diffusion rate because of their physical aspect; institutional innovations are essentially ideas, rules and organized processes and thus could theoretically spread as fast as information can spread.

Secondly, our results on the diffusion of compulsory education exposed the strong degree to which the diffusion of an institutional innovation can be blocked by opposing social groups. As changes to the rules of the game will benefit some and hurt others, we can expect most major institutional innovations to be opposed by certain groups to some extent. We conclude that, in addition to other factors, the diffusion rate of an institutional innovation is a function of the degree of influence of those opposing it; the stronger the influence the opposition holds, the slower the diffusion rate will be. The gold standard was able to diffuse rapidly because the degree of opposition by those in a position of influence was weak, and the movement against it only built up decades after it finished diffusing when its deflationary effects were first being felt.

Thirdly, our conclusion on the three-parameter logistic function (with the option of an intercept) being the best suited to describe the diffusion process of this type of innovation, can serve as a reference for modeling historical cases in future work. Also, to the extent of our knowledge, this is the first study where the central focus is on *institutional* change, and the diffusion of several cases of institutional innovations was analyzed and cross-compared, on temporal and spatial dimensions. Studies on the diffusion of innovations that are institutional remain scarce, and are often not differentiated from *social* or *cultural* innovations.

Finally, our results imply that institutional changes happen similarly to technological changes. This adds to the evidence that the diffusion pattern in which new (successful) ideas are adopted is logistic in shape, and the diffusion process is a social learning process [16]; this observation was already described by Tarde in the late 1800s.

5.2 Limitations and future work

A limit to our case study approach is the generalization of the resulting theory: It is somewhat confined to the specific context of the case. We can not assume that institutional innovations diffuse at the same rate in the 19th century as in the 21st century. Thus, a future study on whether the diffusion rate of institutional innovation has sped up with time due to globalization and the internet would be of interest. The number of case studies we analyzed and cross-compared remains limited, we only studied two major cases of institutional innovation, but our case study on compulsory education contained several subcases which were used for cross comparison. Our analysis also remained limited to Europe and the US. In our case on the diffusion of compulsory education. We disregarded tertiary education because it did not fall under the scope of *compulsory* education. A quantitative analysis of the diffusion of universities and their enrollment rates would be a valuable case study in any future work.

Although we attempted to avoid selection bias in the selection of cases, we admit there is an inherent survivorship bias in diffusion studies, in that often, only cases where the innovation successfully diffused are chosen, and the majority of cases — which are failed cases of innovation — are ignored. The vast majority of innovations do not grow enough to have a diffusion pattern worth analyzing. Eventually, a broader set of case studies to analyze and cross compare — a mixture of successful, failed, positive and negative innovations — would be beneficial to confirm a more definitive pattern of institutional innovation.

In future work relating to the 19th century, it would be of interest to analyze what other institutional innovations diffused after the Industrial Revolution with an inflection point (t_m) in the 1870s. The intensity of institutional innovations over time would also be of interest, has the frequency of these innovations diminished since the end of the industrial revolution? Were there periods in history with much higher numbers of institutional innovations?

As the world of the 21st century will probably experience another revolution, that of robotics and automation while coping with climate change and the demands of an increasing population. In the words of Schumpeter on the future of our capitalist economic system: "Capitalist enterprise, by its very achievements, tends to automatize progress, we conclude that it tends to make itself superfluous — to break to pieces under the pressure of its own success." [105, p.134]. It is paramount to explore the necessary institutional innovations required to respond to the challenges that will be encountered. The hope is that beneficial institutional innovations will diffuse globally at the same rate as the international gold standard: fast.

List of Figures

List of Tables

Bibliography

[1] A. Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations*, 25202. Printed at the University Press for T. Nelson and P. Brown, 1827.

[2] D. C. North, „Institutions“, *Journal of Economic Perspectives*, vol. 5, no. 1, pp. 97–112, Mar. 1991.

[3] D. C. North, „Economic performance through time“, *The American economic review*, vol. 84, no. 3, pp. 359–368, 1994.

[4] M. Filzmoser, „Be innovative at innovating: An interdisciplinary comparison of innovation processes“, *Economic and Social Development: Book of Proceedings*, p. 612, 2014.

[5] R. Coombs, P. Saviotti, and V. Walsh, *Economics and technological change*. Rowman & Littlefield, 1987.

[6] R. Peres, E. Muller, and V. Mahajan, „Innovation diffusion and new product growth models: A critical review and research directions“, *International Journal of Research in Marketing*, vol. 27, no. 2, pp. 91–106, 2010.

[7] E. M. Rogers, *Diffusion of Innovations, 5th Edition*. Free Press, 2003.

[8] A. Grübler, „Diffusion: Long-term Patterns and Discontinuities“, in *Diffusion of Technologies and Social Behavior*, N. Nakićenović and A. Grübler, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1991, pp. 451–482, ISBN: 978-3-662-02700-4.

[9] A. Grubler, „Time for a change: On the patterns of diffusion of innovation“, *Daedalus*, vol. 125, no. 3, pp. 19–42, 1996.

[10] D. C. North, *Institutions, institutional change and economic performance*. Cambridge university press, 1990.

[11] A. Grubler, „Technology and Global Change: Land Use, Past and Present“, IIASA, Laxenburg, Austria, IIASA Working Paper, Jan. 1992.

[12] D. Acemoglu, S. Johnson, and J. A. Robinson, „Institutions as a fundamental cause of long-run growth“, *Handbook of economic growth*, vol. 1, pp. 385–472, 2005.

[13] A. Maddison, „Statistics on world population, GDP and per capita GDP, 1-2008 AD“, *Historical Statistics*, pp. 1–36, 2010.

[14] L. Jonung, „Swedish experience under the classical gold standard, 1873-1914“, in *A Retrospective on the Classical Gold Standard, 1821-1931*, University of Chicago Press, 1984, pp. 361–404.

[15] G. De Tarde, *The laws of imitation*. H. Holt, 1903.

[16] A. Grubler, *The Rise and Fall of Infrastructures: Dynamics of Evolution and Technological Change in Transport*. Heidelberg: Physica-Verlag, 1990.

[17] A. M. Turing, „Computing machinery and intelligence“, *Mind*, vol. 59, no. 236, pp. 433–460, 1950.

[18] A. Grübler, *Technology and global change*. Cambridge University Press, 2003.

[19] A. Grübler, N. Nakićenović, and D. G. Victor, „Dynamics of energy technologies and global change“, *Energy Policy*, vol. 27, no. 5, pp. 247–280, 1999.

[20] C. Marchetti and N. Nakicenovic, „The Dynamics of Energy Systems and the Logistic Substitution Model“, IIASA, Laxenburg, Austria, IIASA Research Report, Dec. 1979.

[21] L. Barreto and R. Kemp, „Inclusion of technology diffusion in energy-systems models: some gaps and needs“, *Journal of cleaner production*, vol. 16, no. 1, S95–S101, 2008.

[22] E. Mansfield, „Technical change and the rate of imitation“, *Econometrica: Journal of the Econometric Society*, pp. 741–766, 1961.

[23] C. Wilson, „Meta-analysis of unit and industry level scaling dynamics in energy technologies and climate change mitigation scenarios“, International Institute for Applied Systems Analysis (IIASA), Interim Report IR-09-29, Aug. 2009.

[24] C. M. Christensen, „Exploring the limits of the technology S-curve. Part I: component technologies“, *Production and Operations Management*, vol. 1, no. 4, pp. 334–357, 1992.

[25] J. H. Ausubel, „Rat Race Dynamics and Crazy Companies: The Diffusion of Technologies and Social Behavior“, in *Diffusion of Technologies and Social Behavior*, N. Nakićenović and A. Grübler, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1991, pp. 1–17, ISBN: 978-3-662-02700-4.

[26] A. Debecker and T. Modis, „Determination of the uncertainties in S-curve logistic fits“, *Technological Forecasting and Social Change*, vol. 46, no. 2, pp. 153–173, 1994.

[27] J. C. Fisher and R. H. Pry, „A simple substitution model of technological change“, *Technological forecasting and social change*, vol. 3, pp. 75–88, 1971.

[28] P. S. Meyer, J. W. Yung, and J. H. Ausubel, „A primer on logistic growth and substitution: the mathematics of the Loglet Lab software“, *Technological Forecasting and Social Change*, vol. 61, no. 3, pp. 247–271, 1999.

[29] N. Bento, „Historical Diffusion of Mobile Phones and Its Impact on Energy Demand: Findings and Outlook“, IIASA, Laxenburg, Austria, IIASA Interim Report, May 2012.

[30] C. Frydman and R. E. Saks, „Executive Compensation: A New View from a Long-Term Perspective, 1936-2005“, National Bureau of Economic Research, Tech. Rep., 2008.

[31] T. Büttner and A. Grubler, „The birth of a "green generation"? Generational dynamics of resource consumption patterns“, *Technological Forecasting and Social Change*, vol. 50, no. 2, pp. 113–134, 1995.

[32] N. Nakicenovic, „The automobile road to technological change: diffusion of the automobile as a process of technological substitution“, *Technological Forecasting and Social Change*, vol. 29, no. 4, pp. 309–340, 1986.

[33] L. B. Mohr, „Determinants of Innovation in Organizations“, *The American Political Science Review*, vol. 63, no. 1, pp. 111–126, 1969.

[34] H. E. Pemberton, „The Curve of Culture Diffusion Rate“, *American Sociological Review*, vol. 1, no. 4, pp. 547–556, 1936.

[35] S. Y. Glaziev, „Some General Regularities of Techno-Economic Evolution“, in *Diffusion of Technologies and Social Behavior*, N. Nakićenović and A. Grübler, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1991, pp. 295–315, ISBN: 978-3-662-02700-4.

[36] T. Hägerstrand, *The propagation of innovation waves*, ser. Lund studies in geography : Series B, Human geography ; 4. Lund: Lund Univ. Press [u.a.], 1952.

[37] A. R. Pred and T. Hägerstrand, *Innovation diffusion as a spatial process*, 1967.

[38] P. Gould, „Spatial diffusion: Commission on College Geography“, *Association of American Geographers, Washington, DC*, p. 80, 1969.

[39] H. P. Young, „The dynamics of social innovation“, *Proceedings of the National Academy of Sciences*, vol. 108, no. Supplement 4, pp. 21 285–21 291, 2011.

[40] ——, „Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning“, *The American economic review*, vol. 99, no. 5, pp. 1899–1924, 2009.

[41] L. M. Brown and H. P. Young, „The diffusion of a social innovation: executive stock options from 1936–2005“, *Work. Pap., Dep. Econ., Univ. Oxford*, 2014.

[42] R. L. Hamblin and J. L. L. Miller, „Reinforcement and the Origin, Rate and Extent of Cultural Diffusion“, *Social Forces*, vol. 54, no. 4, pp. 743–759, 1976.

[43] O. Bandiera, M. Mohnen, I. Rasul, and M. Viarengo, „Nation-Building Through Compulsory Schooling During the Age of Mass Migration“, Suntory, Toyota International Centres for Economics, and Related Disciplines, LSE, STICERD - Economic Organisation and Public Policy Discussion Papers Series 057, Apr. 2015.

[44] B. of Education, *Biennial Survey of Education, 1916-18*, Bulletin, ser. 1919, No.88. 1921, vol. 1, p. 748.

[45] G. Modelski and G. Perry, „Democratization from a Long-term Perspective“, in *Diffusion of Technologies and Social Behavior*, N. Nakićenović and A. Grübler, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1991, pp. 19–35.

[46] P.-F. Verhulst, „Notice sur la loi que la population poursuit dans son accroissement Correspondance mathématique et physique 10: 113-121“, Retrieved 09/08, Tech. Rep., 2009.

[47] P. F. Verhulst, „Recherches mathématiques sur la loi d'accroissement de la population.“, *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles*, vol. 18, pp. 14–54, 1845.

[48] T. R. Maltus, *An essay on the principle of population.* Cosimo, Inc., 2006, vol. 1.

[49] C. P. Birch, „A new generalized logistic sigmoid growth equation compared with the Richards growth equation“, *Annals of Botany*, vol. 83, no. 6, pp. 713–723, 1999.

[50] J. Berkson, „Application of the logistic function to bio-assay“, *Journal of the American Statistical Association*, vol. 39, no. 227, pp. 357–365, 1944.

[51] D. P. Helmbold, J. Kivinen, and M. K. Warmuth, „Worst-case loss bounds for single neurons“, *Advances in neural information processing systems*, pp. 309–315, 1996.

[52] P. A. Geroski, „Models of technology diffusion“, *Research policy*, vol. 29, no. 4, pp. 603–625, 2000.

[53] J. Han and C. Moraga, „The influence of the sigmoid function parameters on the speed of backpropagation learning“, in *International Workshop on Artificial Neural Networks*, Springer, 1995, pp. 195–201.

[54] N. Bento, „New Evidence in Technology Scaling Dynamics and the Role of the Formative Phase“, IIASA, Laxenburg, Austria, IIASA Interim Report, Jun. 2013.

[55] G. E. Box, „Science and statistics“, *Journal of the American Statistical Association*, vol. 71, no. 356, pp. 791–799, 1976.

[56] H. Kwasnicka, R. Galar, and W. Kwasnicki, „Technological substitution forecasting with a model based on biological analogy“, *Technological Forecasting and Social Change*, vol. 23, no. 1, pp. 41–58, 1983.

[57] G. E. Box, N. R. Draper, *et al.*, *Empirical model-building and response surfaces.* Wiley New York, 1987, vol. 424.

[58] K. M. Eisenhardt, „Building theories from case study research“, *Academy of management review*, vol. 14, no. 4, pp. 532–550, 1989.

[59] M. Filzmoser, *Diffusion of organizational and institutional innovations*, Vienna University of Technology, Institute of Management Science, Research Proposal, Jun. 2016.

[60] B. Ryan and N. C. Gross, „The diffusion of hybrid seed corn in two Iowa communities.“, *Rural sociology*, vol. 8, no. 1, p. 15, 1943.

[61] C. Wilson, A. Grubler, N. Bauer, V. Krey, and K. Riahi, „Future capacity growth of energy technologies: are scenarios consistent with historical evidence?“, *Climatic Change*, vol. 118, no. 2, pp. 381–395, 2013.

[62] A.-N. Spiess and N. Neumeyer, „An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach“, *BMC pharmacology*, vol. 10, no. 1, p. 6, 2010.

[63] D. Mitch, „The role of human capital in the first Industrial Revolution“, *The British Industrial Revolution: An Economic Perspective*, pp. 267–307, 1993.

[64] Y. N. Soysal and D. Strang, „Construction of the first mass education systems in nineteenth-century Europe“, *Sociology of education*, pp. 277–288, 1989.

[65] J. Boli, F. O. Ramirez, and J. W. Meyer, „Explaining the origins and expansion of mass education“, *Comparative education review*, vol. 29, no. 2, pp. 145–170, 1985.

[66] W. J. Collins and R. A. Margo, „Historical perspectives on racial differences in schooling in the United States“, *Handbook of the Economics of Education*, vol. 1, pp. 107–154, 2006.

[67] J.-W. Lee and H. Lee, „Human capital in the long run“, *Journal of Development Economics*, vol. 122, pp. 147–169, 2016.

[68] J. A. R. Marriott and C. G. Robertson, *The evolution of Prussia: the making of an empire*. The Clarendon press, 1965.

[69] Aristotle, *Politics*. Aeterna Press, 2015.

[70] O. Feldbæk, „Denmark and the Treaty of Kiel 1814“, *Scandinavian Journal of History*, vol. 15, no. 3-4, pp. 259–268, 1990.

[71] M. de Noailles, *The Life and Memoirs of Count Molé, 1781–1855*, 1923.

[72] Eurydice. (Dec. 2015). Compulsory education in Europe – 2015/16, [Online]. Available: `https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/Publications:Compulsory_education_in_Europe_%5C%E2%5C%80%5C%93_2015/16`.

[73] S. J. Cabus and K. De Witte, „Does school time matter?—On the impact of compulsory education age on school dropout“, *Economics of Education Review*, vol. 30, no. 6, pp. 1384–1398, 2011.

[74] Bundeskanzleramt Österreich. (Sep. 2016). Ausbildungspflicht für Jugendliche bis 18, [Online]. Available: `https://www.help.gv.at/Portal.Node/hlpd/public/content/194/Seite.1940281.html`.

[75] R. Forstall, „Population of states and counties of the United States: 1790 to 1990, United States Department of Commerce, Bureau of the Census“, *Population Division*, 1996.

[76] C. Nardinelli, „Child labor and the factory acts“, *The Journal of Economic History*, vol. 40, no. 4, pp. 739–755, 1980.

[77] W. Kendal and G. W. Carey, „Towards a Definition of" Conservatism"“, *The Journal of Politics*, vol. 26, no. 2, pp. 406–422, 1964.

[78] L. H. Officer and S. H. Williamson, *The Price of Gold, 1257-2014.* 2017. [Online]. Available: `http://www.measuringworth.com/gold/`.

[79] M. Flandreau, „The French crime of 1873: An essay on the emergence of the international gold standard, 1870–1880“, *The Journal of Economic History*, vol. 56, no. 04, pp. 862–897, 1996.

[80] F. X. Diebold, S. Husted, and M. Rush, „Real exchange rates under the gold standard“, *Journal of Political Economy*, vol. 99, no. 6, pp. 1252–1271, 1991.

[81] M. Friedman, „The crime of 1873“, *Journal of Political Economy*, vol. 98, no. 6, pp. 1159–1194, 1990.

[82] M. D. Bordo and A. J. Schwartz, „The specie standard as a contingent rule: some evidence for core and peripheral countries, 1880-1990“, National Bureau of Economic Research, Tech. Rep., 1994.

[83] C. M. Meissner, „A new world order: explaining the international diffusion of the gold standard, 1870–1913“, *Journal of International Economics*, vol. 66, no. 2, pp. 385–406, 2005.

[84] B. Eichengreen, M. Flandreau, *et al.*, „The geography of the gold standard“, University of California at Berkeley, Tech. Rep., 1994.

[85] M. Obstfeld and A. M. Taylor, „Sovereign risk, credibility and the gold standard: 1870–1913 versus 1925–31“, *The Economic Journal*, vol. 113, no. 487, pp. 241–275, 2003.

[86] P. Martin-Acena, E. Martinez-Ruiz, and P. Nogues-Marco, „Floating against the Tide: Spanish Monetary Policy 1870–1931“, in *The Gold Standard Peripheries*, Springer, 2012, pp. 145–173.

[87] A. P. Duarte and J. S. Andrade, „How the Gold Standard functioned in Portugal: an analysis of some macroeconomic aspects“, *Applied Economics*, vol. 44, no. 5, pp. 617–629, 2012.

[88] T. Schreiber, „Monetary History of the Austro-Hungarian Empire (1867–1914)“, in *The Experience of Exchange Rate Regimes in Southeastern Europe in a Historical and Comparative Perspective*, Nationalbank, Oesterreichische, 2007, pp. 412–418.

[89] H. Handler, „Two centuries of currency policy in Austria“, *Monetary Policy & the Economy*, no. 3, pp. 61–76, 2016.

[90] K. Dowd, R. H. Timberlake, *et al.*, *Money and the Nation State: The Financial Revolution, Governement and the World Monetary System.* Transaction Publishers, 1998.

[91] A. Maddison, „The Maddison-Project“, 2013. [Online]. Available: `http://www.ggdc.net/maddison/maddison-project/home.htm`.

[92] J. Bolt and J. L. Zanden, „The Maddison Project: collaborative research on historical national accounts“, *The Economic History Review*, vol. 67, no. 3, pp. 627–651, 2014.

[93] M. Bergman, S. Gerlach, and L. Jonung, „The rise and fall of the Scandinavian Currency Union 1873–1920“, *European Economic Review*, vol. 37, no. 2-3, pp. 507–517, 1993.

[94] S. E. Oppers, „Was the worldwide shift to gold inevitable? An analysis of the end of bimetallism“, *Journal of Monetary Economics*, vol. 37, no. 1, pp. 143–162, 1996.

[95] E. Atkinson, *Report Made by Edward Atkins... to the President of the United States, Upon the Present Status of Bimetallism in Europe. October, 1887*, 34. US Government Printing Office, 1887.

[96] M. Friedman, *Capitalism and freedom*. University of Chicago press, 2009.

[97] L. Crabbe, „The international gold standard and US monetary policy from World War I to the New Deal“, *Federal Reserve Bulletin*, no. Jun, pp. 423–440, 1989.

[98] L. Jonung, „The Scandinavian Monetary Union 1873 1924, in, Cottrell, P., G. Notaras and G. Tortella,(ed.), From the Athenian Tetradrachm to the Euro“, *Studies in European Monetary Integration, Ashgate, Aldershot*, 2007.

[99] I. Henriksen and N. Kærgård, „The Scandinavian currency union 1875–1914“, in *International monetary systems in historical perspective*, Springer, 1995, pp. 91–112.

[100] K.-H. Bae and W. Bailey, „The Latin Monetary Union: some evidence on Europe's failed common currency“, *Review of development finance*, vol. 1, no. 2, pp. 131–149, 2011.

[101] M. D. Bordo and H. Rockoff, „The gold standard as a “good housekeeping seal of approval”“, *The Journal of Economic History*, vol. 56, no. 02, pp. 389–428, 1996.

[102] M. Friedman, *Essays in positive economics*. University of Chicago Press, 1953.

[103] N. Nakicenovic, „Growth to limits: Long waves and the dynamics of technology; IIASA draft report“, PhD thesis, 1984.

[104] U. S. B. of the Census, *Historical statistics of the United States, colonial times to 1970*, 93. US Department of Commerce, Bureau of the Census, 1975.

[105] J. Schumpeter, *Capitalism, Socialism and Democracy*, 5th Revised ed. Routledge, Apr. 1994.