Geo-information Science and Remote Sensing

Thesis Report MSc Geo-Information Science

MAPPING BIBLIOGRAPHIC DATA

Book production and Censorship in the Early Modern Low Countries

Bart Waterreus

July 1, 2025

WAGENINGEN
UNIVERSITY & RESEARCH

Mapping Bibliographic Data

Book production and Censorship in the Early Modern Low Countries

Bart Waterreus

Registration number: 1448315

Supervisors:

Arno Timmer

Sandra de Pleijt

Thesis Report submitted in partial fulfilment of the degree of Master of Science

at Wageningen University and Research Centre,

The Netherlands.

July 1, 2025

Wageningen, The Netherlands

Thesis code number: GRS80436

Wageningen University and Research Centre

Laboratory of Geo-Information Science and Remote Sensing

The printing press is regarded as one of the most influential inventions of human history. Related work claims that the geographical production, distribution and circulation of books are significant to understanding the dissemination of knowledge and ideas. However, no other related published study considers the location of Authors.

This study introduces a new dataset containing geocoded Authors which we compare with library catalogs to explore spatial and temporal patterns in the geography of the book in the Low Countries, using a centroid approach proposed by Lucas Koren. The findings show different patterns observed in the late fifteenth and early sixteenth century from which the Library Catalogs align closely with the narrative from historical research, when the Author data fails to so.

Furthermore, we investigate effect of censorship. Historical research considering the geography of Protestantism within the Low Countries reveals clear spatial and temporal trends. However, when mapping patterns of works targeted by Catholic censorship controlled by total printing volume, the results fail to show any interpretable pattern that is obtained from the literature.

Last, this study investigates whether the printing location of works can be explained by the location of their Authors and whether it has been subjected to censorship within a discrete choice modelling context. While publishing at distant locations was a known strategy to evade censorship, our findings provide no statistical evidence that censored work is printer further away than non-censored work. Also, only considering censorship and authors locations results in significantly worse choice prediction performance than assuming random chance. This suggests that other unknown variables motivate the choice of printing location.

Contents

1. Introduction

The Renaissance spread to early Modern Europe in the late fifteenth and sixteenth century. The increasing degree of literacy and reliance on written records demanded a less costly method of production of literary works. This eventually led to Johannes Gutenberg demonstrating the practicality and efficiency of a movable printing press between 1448 and 1450 [1]. He showed that his invention could produce high quality copies significantly cheaper and faster than traditional manuscript writing [2], causing the price of books dropping with 65 percent between 1450 and 1500 in Europe [1]. It entirely changed the way knowledge would diffuse allowing new ideas to spread quickly with more significant impacts [3][4][5], democratizing the access to written material [1] and "touching on and often penetrating almost every sphere of activity" [6, p. 6].

In this research we emphasize on the study of 'place' as a geographical dimension of early modern book culture [7]. This is often referred to as "The Geography of the book" [8][9][7]. Its study claims that their geographical production, distribution, and circulation are significant to understanding the dissemination of knowledge and ideas [10]. A new phase integrating spatio-temporal dimensions with GIS tools has enabled new research areas to the geography of the book [7]. Distinguishing itself as a multidisciplinary research field, GIS has brought opportunities at both micro and macro levels using existing print databases about print culture as well as providing new means to enrich these same databases [8][9][7]. Examples are using book production as a proxy for advanced literacy skills as indicators of human capital [11] and using the adoption of the printing press as an explanation for city growth, population growth [12], technological change and localized knowledge spillovers [1].

Historical knowledge production is usually researched with mixing aspects from growth theory with elements from economic geography [13]. Evidence on nationwide scales shows relationships between the growth of book editions per capita and growth capabilities explaining economic growth divergences between countries [11] and the positive effect of population density on human capital formation [13]. Other studies at lower scale levels show that cities which are earlier connected to railroad networks experience specialization and an 11% increase of knowledge production [14]. New work considering authors locations shows that authors death counts per city proxy for historic economic activity, with robust correlations with existing growth estimates [15].

1.1. Historical Context

Authorities quickly became aware that the printing press was accelerating the spread of movements challenging the established regimes [16], and censorship legislation was initiated in various ways to control the presses. More specifically, we emphasize Repressive censorship, which consists of works declared forbidden after being written, produced and distributed. Any individuals involved could be punished with fines, banishment, or death sentences. [17] and [18] contain many examples of Dutch authorities attempting to censor printed material speaking of critique towards the reigning regimes. On an international level, the Vatican established lists with forbidden authors and works to condemn all written material perceived as "Dangerous" to the Catholic church [19]. These lists are regarded as the Catholic church's primary weapon against the spread of the new ideologies during the reformation [18]. Historical research has proved that freedom of religion was vital to many individuals in Early modern Europe. More specifically, it is described as the most crucial push and pull factor for relocating individuals and businesses in the book industry [20][21].

While censorship was undoubtedly present, the regulation system was often far from perfect. One of the ways to bypass legislation was to smuggle illegally printed material, which was printed elsewhere because regulations only targeted the domestic production of books [22]. It is well documented that the Dutch Republic was known for having a well-developed printing culture [11], which was relatively open compared to other countries [16][17]. Due to the fragmented political structure, it was nearly impossible to maintain adequate censorship legislation, let alone enforce it in practice [16]. Earlier studies have indicated that these relatively tolerant conditions increased the attractiveness of migrants and upper-tail human capital [23] to generate economic revenue [16][17]. The Low Countries were a home to many international newspapers and provided a host for many works of dubious religious, philosophical, and moral content in the Republic.

Some writers even claimed that the Republic was the "Mecca of Writers", which was certainly not the case. Although many offenses would go unpunished due to the absence of a central system, effective censorship was applied at the local level. Decisions made at a higher level of government were often easily ignored at the lower, municipal level [16]. However, the freedom of the press was relatively guaranteed for works in foreign languages printed for export. The Dutch publishers and book traders were known to be major suppliers of literary smuggling goods, generating high economic revenues [17]. Therefore, certain authorities had little incentive to establish a strong censorship system to keep monetary conditions favorable [17][16].

Because effective censorship was mainly applied on the local level, significant differences in its enforcement are reported in spatial and temporal dimensions [17, p. 31-44]. Dominant Stadhouders had considerable leverage in influencing censorship trials, which they would steer to their interests. This made the enforcement of censorship regulations largely dependent on time, place and person [16].

2. Academic Context & Research needs

Despite the recognition of the geography of the book and and the growth of the so-called "Historical GIS" [24], applications are limited. Baten & van Zanden [11] use the output of printed books based on the number of titles or editions from short title catalogs, but these are aggregated to a nationwide scale. Dittmar [1] provides evidence that cities that early adopted the printing press had maintained growth advantages over extended periods running from 1500 1700 and even 1500 1800, but do not involve any printing volume.

New work considering authors locations by Chaney [15] and Lucas Koren [25] provides evidence that spatio-temporal patterns of economic development can be validated using author death counts per year, and serve as a proxy for economic growth. Another study by Chiopris [14] considers both Author and Publication data, showing that cities that are connected to the railroad network in Germany from 1835 1914 experienced increased specialization with an 11% increase in knowledge production. Her research is focused on the clustering of specific knowledge and the mobility of authors instead of general economic development.

Censorship, among other variables such as speed, costs, duties, and fraud, has been widely acknowledged as an essential additional variable to consider when studying the distribution of books [7]. Recently published studies show the interest in the effect of the Catholic reaction to the spread of protestant ideas in the 16th century [23][26][27][28]. These three studies all study the effect of works, authors and publishers listed on the Index Librorum Prohibitorum, which is considered a primary tool in the church's arsenal aimed at censoring the production and circulation of heretical publications. They conclude that the indexes were successful in reducing the printing volume of heretic works [23], reducing the share of scholars pursuing revolutionary ideas, altering the path of knowledge development [26][23][28], encouraging printers to print from non-indexed authors [23], and reduce market shares of printing firms targeted by its enforcement [27].

Only Becker et al. [23] studied the geography of the indexes. They study the effect of limiting the printing volume of listed authors by sampling cities within a 500 km range. They find larger coefficients for near-located cities than far-located cities, but find no significant differences [23]. Also, these studies only consider data about the locations where works are published and printed; they do not include data about the places where authors are active in the analysis.

The Low Countries are known as the printing house of Europe between 1461 and 1800 [11], pressing the most literate works for the international market [16][17] with books among the most lucrative export items [29]. Besides earlier findings by Koren, this has only been covered by traditional historical research.

3. Research Objectives

As can be read in section 1 and 2, geocoded Author data and Publication data are used for historical economic growth analysis and as a proxy for human capital. The main objective of this research is to assess similarities and differences in their use. We provide special attention to the influence of censorship.

This is broken down into the following research questions:

1. What are the Spatial-Temporal patterns of the geography of the book from 1450 - 1800 in the Low Countries according to previous research?
2. What differences can be observed when mapping Spatial-Temporal Patterns with Mean and Median Centroids when using the Author's death locations and Publishing locations?
3. To what extent can we model Authors' Printing Location choice behavior with the Authors' Activity Location and Censorship Data?

This research builds on earlier accomplishments achieved by Lucas Koren [25], who investigated if economic development patterns could be mapped with author locations data from Eric Chaney [30] in the Low Countries and Great Britain. In this study, his findings are extended by also deploying his proposed centroid methods on publication data from the Short Title Catalogs Netherlands and Vlaanderen (STCN [31] & STCV [32]) to see how the results of each of the data sources correspond to findings from earlier historical studies.

Furthermore, this research assesses the relationship between the activity locations of individual authors from Chaney's data and their spatial choice of printing location. This is integrated within a Discrete Choice Modeling framework, where the authors are considered decision makers when choosing a specific printer as a first step in distributing their work. Due to the authors' and printers' resolution and observable characteristics, we approach this in an aggregated choice context, where the Authors chose a particular city as a printing location.

Within this framework, we integrate censorship by identifying which authors have been exposed to censorship using the works of Debunjanda et al. [33][34] and Weekhout [35]. As can be read in the introduction, censored authors have worked around censorship enforcement by printing their work outside the reach of enforcing institutions. As reported, the absence of a strong central government made smaller regions and cities relatively autonomous, with local people in power frequently pursuing their interests. From this, we hypothesize that these unique power dynamics incentivize authors writing dubious content to have their work printed further away. As setting up a communication and transportation network for the possibility of printing work at distant locations was very costly, authors had little incentive to do so when it was not necessary [17]. In the modeling framework, we test this behavioral hypothesis that censored authors chose cities that are further away than authors who are not subjected to censorship.

4. Data Sources

4.1. Author data

The author data which will be used is a dataset recently created by Eric Chaney [30]. The construction is based on the VIAF (Virtual International Authority File) database [36], which links authority files of various participating libraries together into clusters to provide a uniform name for the authors. These clusters provide detailed information about the authors in question, including birth, death and active years. They were geo-referenced using different 6 national library catalogues, wikipedia and the Google maps API. The complete overview of the construction of this dataset can be read in [30]. For this research only authors georeferenced to the Netherlands and Belgium were extracted, resulting in 18026 individuals with at least one complete field of place of activity or death location.

4.2. Publication and Printing press data

The dataset containing the different publications from 1500 - 1800 is the Short-Title Catalogue Netherlands (STCN) [31]. It contains detailed information about 220.000 literary works which have been pressed in the Netherlands or in the Dutch language from 1460 - 1801. This information contains in which year the work was published; the printer who has pressed the work and in which city this printer was located.

4.3. Censored books

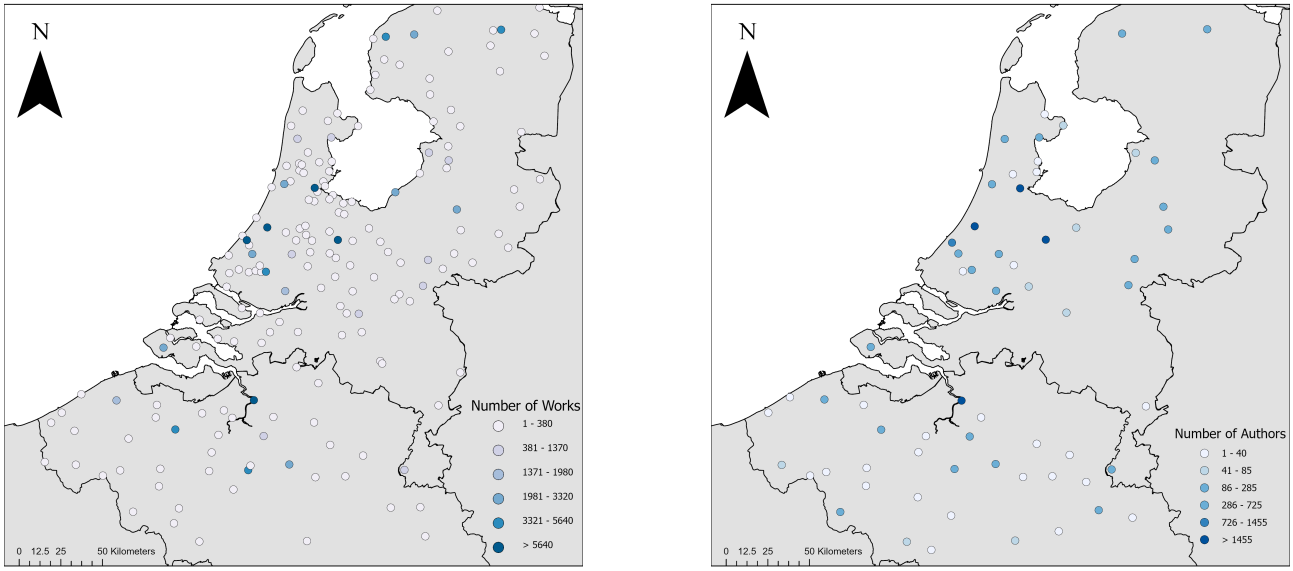
We use two different sources to identify which works and authors have been targeted by censorship. The Index Librorum Prohibitorum series of De Bujanda et al. has identified all the books prohibited by the Catholic Church. Volumes X [33] and XI [34] together contain all books censored after 1500 until 1966. These volumes list about 5200 censored works written by a total of 3000 different authors. Accompanying these, I use Weekhout's most recent study investigating book censorship in the Dutch Republic, which identified 263 censored works [35].

4.4. Matching data

The STCN & STCV are not linked to the VIAF database. These have to be matched together to determine the place of activity for the authors in the catalogs. For this, I rely on earlier work done by Julius Koschnick. He has identified matches between both catalogs and the VIAF database, taking author names and alternative spellings, death years, and median publication dates into account. When encountering duplicate matches only based on author names, he created a tie-breaker method with two approaches. Difference between years of death between the catalogs and VIAF, and a synthetic approach with the difference between year of death and median publication year.

4.5. Study Area

In this study we restrict the data to only include works that are printed anywhere in the Low Countries and Authors that have died in the Low Countries from 1450 - 1800. This is explained in greater detail in section 5. The spatial distribution of both data sets can be seen in figure 1.



(a) Total number of printed works per location from the STCV & STCN, 179 unique locations

(b) Total number of authors deaths per location from the Chaney data, 65 unique locations

Figure 1. Study Area.

Source: Landsgrenzen, Esri Nederland, 2025 (Flevoland removed)

In figure 1a we can see that most of the works over the entire period concentrate in Amsterdam (40461), Antwerp (15601), Leiden (19182), The Hague (14346) and Utrecht (12735). The same five places count the most number of Authors deaths, but in a different order: Amsterdam (3152), Leiden (1975), Utrecht (1868), Antwerp (1456) and The Hague (1268). Furthermore, figure 1b shows that the spatial distribution of authors is more evenly distributed than the number of printed works. Important to note is that these figures display the total sum over the whole period from 1450 - 1800. The temporal distribution is discussed in section 5.4. The STCN and STCV have many smaller settlements included than the Chaney data, which is logical since the Chaney dataset was constructed for analysis at a continental scale. The only cities with a reasonable printing volume that are not not present at all in Chaneys data are Franeker (the blue dot close to shore in East Friesland) and Harderwijk (blue dot at the center south shore of the Zuiderzee).

5. Pre Processing

Before the making of any map and doing any type of computation for the results of this thesis, significant data augmentation needs to be done because of the formatting of the data sources. A graphical explanation of the pre-processing workflow can be seen in figure 2:

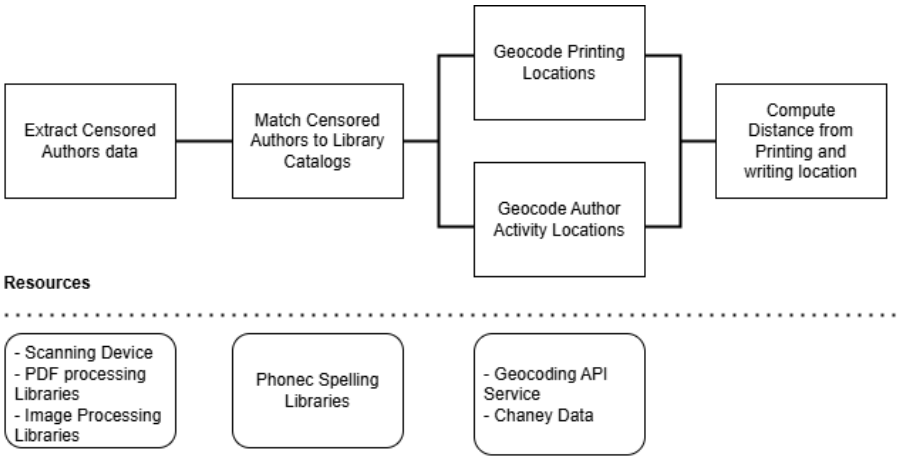


Figure 2. Preprocessing Workflow

5.1. Extract Censored Authors Data

Debujanda Vol. XI [34] & Weekhout [35] contain alphabetically ordered lists which are available online and Debujanda Vol. X [33] is provided as a physical copy by Utrecht University Library from which I scanned the lists of prohibited works that were listed under being printed in the Low Countries. The digital PDF's and scanned pages are processed by `Python` PDF and Image text extraction libraries. Since these libraries did not always produce clean correctly formatted text, I manually corrected the imperfections. After cleaning the plain text, I used regular expressions to extract the relevant data. The works of Debujanda provide a comprehensive overview of relevant details of the authors. All the authors listed are accompanied by alternative name spellings, birth and death years (if known) and the dates when a work was mentioned on a specific index. Unfortunately the combined works of Weekhout & Knuttel (as provided by Weekhout [35]) are not as comprehensive. The only relevant information that could be extracted about the Authors are their last names and first letters, which makes matching much more tricky. They did also provide information about the years when specific works were prohibited and whether this was of a Political, Political-Religious, Religious, Political-Philosophical or Moral motive.

5.2. Matching Censored Authors Data to Library Catalogs

Common problems when working with historical data are alternative spellings for individual authors, duplicate names and variations of birth and death years. To cope with the problem of alternative name spellings, the library catalogs have included known alternative spellings of authors. However, it was often the case that spellings of authors as listed in the banned books sources were not included in the library catalogs. To deal with these issues, it is common to use some sort of phonetic name matching algorithm [37]. After a discussion with Julius Koschnick, a colleague from supervisor Sandra de Pleijt, we agreed that using phonetic spelling algorithm NYSIIS would be a suitable solution as they have good experiences using this algorithm on names from the Dutch Library catalogs.

NYSIIS is an alphabetic algorithm, which is easy to implement and has proven itself to be successful for an extensive series of record linking studies [37]. I used this algorithm to match the author names and variances found in the works of Debunjanda to the authors listed in the STCN & STCV and manually check and filter out false positives, since the total number of matches was not to large. I did the same for the authors listed in Weekhout, but evaluated matches much more critically since only the first letter of the first names of the authors were known.

5.3. Geocoding

5.3.1. Printing Locations

As with the authors names, spelling variations are also very common for locations mentioned in the library catalogs, especially the places of publication mentioned in the STCV. In this catalog, all place variations are mentioned which includes spelling and its variations from old, country specific language. Most of these places could easily be replaced with their modern versions, as the locations identifiers had a tree type structure including a country code, with their root naming the modern place name. This was not the case for all location names. Sometimes, I could easily recognized the correct place name, especially for the remaining place names from the Netherlands and Belgium, but often I could not.

To find their corresponding modern place names, I used three online resources, . For some instances I found duplicates and it was not clear which place was the correct one. Also some places could not be found. These instances are discarded from the data. Some places were labeled as fictive place or clear false addresses where the correct location could not be resolved. These were also discarded. For the STCN the same was done, places where the right location was not clear after searching it on the internet were discarded.

The remaining place locations were geocoded using a free version of the Here api [38]. Sometimes the response include multiple cities with the same name. These were manually checked to pick the right one. If it was unclear which town had to be picked out after some research, the location was discarded.

The sources used for determining the right place when duplicates showed up after using the API or place names were unrecognizable are google search and Wikipedia.

5.3.2. Author Activity locations

As explained in section 4, for pointing the authors city of activity, I use the dataset created by Eric Chaney [30]. Using the VIAF matches provided Julius Koschnick, I am able to extract a portion of the STCN and STCV where the VIAF identifiers match with the same identifiers as present in the Chaney data.

5.4. Overview of the Datasets

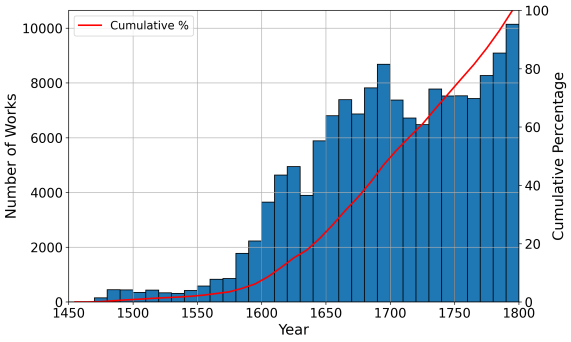
To summarize, I excluded all records from the Title Catalogs which had a fictive place or missing place of publication, or the correct place could not be resolved. From the works that could be geocoded, 9,688 works that were not printed in the Low Countries were removed. After this we removed the publications which printing year was unknown or unclear. The unclear ones had ranges from 10 years to entire centuries, which were all excluded from the data for convenance, since this group only consisted of 9348 publications. The remaining 148143 works were all used for the calculation of the

centroids. From these, 2071 were censored. From all of the works 35595 were written by Authors mention in the Chaney data.

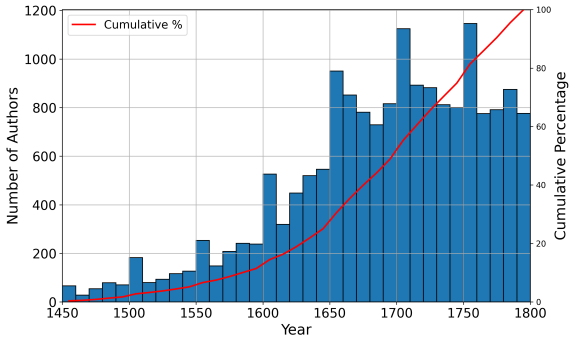
The Chaney data used in this research is slightly different from Korens [25]. Since he was working with Author clusters in the Low Countries and in Great britain, he dropped all authors geocoded to small settlements with less than 10.000 inhabitants. This is not done for this research to get as many matches with the Authors included in the STCN and STCV. All authors that were geocoded to a location in the Low Countries were kept, which consisted of 18026 different Authors.

For the data used for estimating and predicting with the Discrete Choice model, we exclude all works which authors can not be matched with the authors mentioned in the Chaney data. This left a total of 34925 observations which were printed in 93 different locations in the Low Countries. In total, 1139 of these works were written by an Author that was subjected to Censorship before or in the same year as the work was printed.

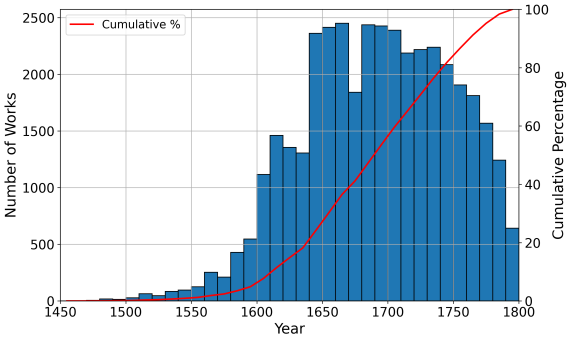
The temporal distribution of each of the subsets can be seen in figure 3:



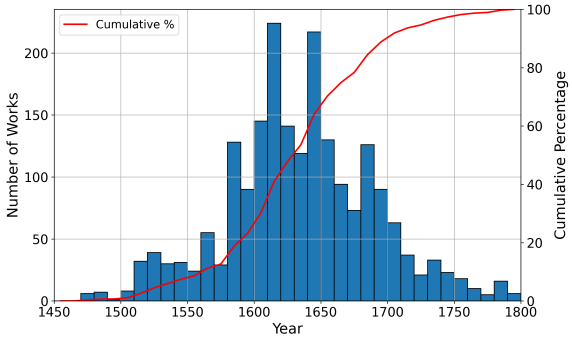
(a) All Works printed in the Low Countries, 148143 total



(b) Authors from Chaney Data in the Low Countries, 18026 total



(c) Works written by Authors from the Chaney data, 35595 total



(d) Censored works Printed in the Low Countries, 2212 total

Figure 3. Histograms with Cumulative percentage graphs for the Data Subsets

As can be seen, the available data is very skewed. From all the works that have been pressed in the Netherlands within the temporal range covered by this study, 90% is is pressed after 1600, after the fall of Antwerp. The Chaney data distribution seems to be spread out a bit more evenly with about 15% of the Authors deaths before 1600 compared to about 10% of the STCN and STCV. Furthermore we can see that the number of works produced by the authors in the Chaney data stagnate significantly in the second half of the 18th century. We can see that almost all works that could be matched with the censorship sources are printed in from 1580 - 1700 with only about 10% in the 18th century.

6. Methodology

6.1. Mean and Median Centroids

Mean and Centroids were calculated for comparing the trends observed from the printing locations, authors death locations and printing locations of censored works in particular. This method was proposed by Koren [25]. Both centroids are calculated for a set of observation points that are grouped together based on a specific time window. This creates two sets of observation points, one formed by the Chaney data and the other formed by the works listed in the STCN and the STCV, with each point having coordinates (x_i, y_i, t_i). First, observations sharing the same coordinates are grouped by unique set of (x_i, y_i) pairs and t_i, resulting in total counts per location per year. These counts are then normalized by the sum of all observations for a given year of t, which are assigned as weigh w_i. The mean and median centroid methods are applied twice on the censored works, with the second time additionally weighting the w_i with the total volume of works printed per city for a particular year to eliminate higher printing volumes obscuring the effects of only the censored works.

Given a set of points P with coordinates (x_i, y_i, t_i) associated with weights w_i for $i = 1, 2, ..., n$, non overlapping subsets are created with a temporal interval of 10 years, which we we call $P_{timestep}$. For each $P_{timestep}$ we calculate the mean centroid $\bar{C}$ with $(\bar{x}, \bar{y})$ and the median centroid $\tilde{C}$ with $(\tilde{x}, \tilde{y})$

The mean centroid $\bar{C}$ is derived by computing the mean x and y coordinates $(\bar{x}, \bar{y})$ weighted by the weights:

$$\bar{C} = (\bar{x}, \bar{y}) = \left(\frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}, \frac{\sum_{i=1}^{n} y_i w_i}{\sum_{i=1}^{n} w_i} \right) \tag{1}$$

For the median centroid $\tilde{C}$, the median x and y coordinates $(\tilde{x}, \tilde{y})$ are computed separately. A particular subset is split into two arrays referred to as V_x and V_y with values v_i and weights w_i for $i = 1, 2, ...n$. These are sorted in ascending order of v_i keeping their weights w_i. After this the cumulative sum of w_i is computed to find the median $(\tilde{x}, \tilde{y})$ by searching their arrays for half the total sum of weights $(0.5W_n)$. It is not always the case that the exact half of the sum of weights directly corresponds with one of the values in the searched array. Therefore, for each of the arrays we return the two indices that just below and above of 50% of the cumulative weight called i_{lower} and i_{upper}. If these are the same, we directly take the corresponding index. If they are different there are four options. Chose i_{lower}, i_{upper}, the higher corresponding weight w_i of i_{lower} or i_{upper}, or calculate the weighted mean. As recommended by Koren [25], in this situation we calculate the weighted mean as it aligns with the implementation of a regular median. This is calculated using equation:

$$\tilde{v} = \frac{v_{lower}(W[i_{upper}] - 0.5W_n) + v_{upper}(0.5W_n - W[i_{lower} - 1]}{W[i_{upper}] - W[i_{lower} - 1]]} \tag{2}$$

6.2. Choice Modeling

6.2.1. General Discrete Choice Model Theory

In general, discrete choice modeling is known as a quantitative evaluation of individual decision making [39][40], derived in a Random Utility (RUT) framework [41][42][43][40][44][45]. In a choice situation, decision maker i independently considers m_i mutually exclusive alternatives that form the choice set C_i, which may vary among decision makers. Each of the alternatives is assumed to be independent of all other possible alternatives. Decision maker i assigns a perceived utility, or attractiveness, U_{ij} to each alternative j in choice set C_i and is assumed to choose alternative j that maximizes this utility. This perceived utility depends on the observable characteristics of the decision makers and alternatives: $U_{ij} = U(V_{ij})$, where V_{ij} is the vector of attributes relative to alternative j and decision maker i [42][41][46]. For the sake of this study, we describe the utility as follows. Author i obtains a certain level of utility of printer j at time t as U_{ijt} with $j = 1, 2, 3...N_j$ from C_{it}. The alternatives forming choice set C_{it} are the printers that are active at time t [40].

Because of incomplete information, some aspects of utility are not directly observable. Therefore the utility of printer j to author i at time t consists of two components [45]:

$$U_{ijt} = V_{ijt} + \epsilon_{ijt} \tag{3}$$

V_{ijt} is the deterministic (explainable) component of utility and ϵ_{ijt} is the stochastic (inexplicable) component of utility [45]. This is treated as a random vector because we do not know $\epsilon_{nj} \forall j$. With assumptions about the joint density of this vector, we can make statements about the decision maker's choice in probabilistic terms [40].

$$P_{jit} = \text{Prob}(\text{Author}_{\,i} \text{ choses printer}_{\,j} \text{ at time}_{\,t} \mid \text{choice set } C) \tag{4}$$

6.2.2. Spatially Aggregated Choices

In contrast to general discrete choice modeling, spatial choices provide a much more complicated background than the a-spatial context in which the discrete choice modeling framework was developed [44]. Effects of Spatial Dependence and Spatial Heterogeneity could cause spatial correlations among individual decision-makers or alternatives [47][41][48], which violate assumptions of the general discrete choice model that decision makers act independently, and that all alternatives are independent. This can lead to inconsistent parameter estimates and unrealistic forecasts when models are not adapted accordingly [44][47].

A Nested Logit Model is the most widely used method to relax the independence of alternatives assumption when dealing with spatial choice. This method divides the choice set into mutually exclusive nests, allowing correlation across alternatives that share a nest [49][50][51][52][49][53][54]. When dealing with a large number of different alternatives, it might be helpful to approach the choice as the selection of an aggregated zone, instead of the choice of a specific elemental alternative [55][56][57]. In many ways, this represents a nested logit model, but the difference is that we only pay attention to the choice at the nest level [58]. Since the Chaney Data and the STCN and STCV catalogs are geocoded to the city level, aggregating the data and using the unique cities as analysis zones is a valid approach.

For the aggregated choice model, we consider the following utility function at the nest level as recommended by [58]:

$$U_{nest} = \mu_{nest} \log \left(\sum_{i \in nest} \exp(U_{ijt} \mu_{nest}) \right) + \epsilon_{ijt} = V_{ijt} + \mu_{nest} \log(N_{nest}) + \epsilon_{ijt} \tag{5}$$

With V_{ij} representing the deterministic part of the nest's utility, which assumes that all printers in each nest are homogeneous with the same systematic utility. This is a valid assumption since we cannot observe any attributes to model heterogeneity within nests. The number of printers forming the nest is directly observable. From the STCN and STCV, we mark the printers' activity span by their first and last appearance. Using these, we can directly observe how many printers in each city were active in a given year. This is included in the function as $\log(N_{nest})$. For this study, the defined zones have a systematic political boundary [57]. As mentioned in section 1, due to the fragmented political structure, legislation and enforcement targeting the book industry greatly varied among different cities [17, p. 31- 44][16]. To consider this kind of varying policies, μ_{nest} is included in the equation as recommended by [58].

6.2.3. Model estimation

Within the Discrete Choice modeling framework, the unknown parameters of the model, $\beta_1, \beta_2, \ldots, \beta_n$, are estimated using Maximum Likelihood Estimation (MLE). As explained in section 6.2.1, we can not directly observe the utility that motivates the observed choices. Using the observed choices and explanatory variables $x_1 + x_2 \ldots, x_n$ from the data, MLE provides a way to estimate the parameters of the utility function. The likelihood function represents the probability of observing the actual choices given the model structure and the observed explanatory variables. By maximizing this function, the parameter values of $\beta_1, \beta_2, \ldots, \beta_n$ are calculated that maximize the probability for the actual observed choices for the given data. For a more comprehensive explanation we point to Part 1 section 3.7 of [43].

The only explanatory variable considered in the vector V_{ijt} is the author's activity city. This is incorporated by computing the Euclidean distance between the active city and each alternative city as a cost-distance variable $dist_{ijt}$. As recommended by Bierlaire [59] instead of using the distance in Meters, we convert it to Its natural logarithm, since the difference between alternative cities becomes less relevant the further they are away from the Authors location. As the number of alternatives for the models estimation and euclidean distance can be 0, we add 1 to all the values to avoid negative infinite values.

Including the topic of the publications with K-1 dummy coding [60] and including censorship as a binary variable was also tested, but this resulted in identification errors, suggesting that there is not enough variability in the data to estimate the parameters [61]. Therefore, the publication's topic is left out, and censorship is incorporated as follows. The option to include a categorized variable for motive of censorship as variability could be observed from the data from Weekhout, section 5.1, but this not tested because of the same reason. Only 45 matches were identified between the Weekhout data and the library catalogs, meaning that 99.7% of the observations had a religieuse motive of censorship.

To model the effect of censorship on the distance, the model's parameters are first estimated using all the observations. Second, the observations are separated into two groups, one with observations where the author has not been subjected to censorship (yet) and one where they have. Following the

estimation of the parameters on both these groups, as is also done by Hammadou et al. for various subgroups [51].

The estimation of the parameters is done using the `Python` package `Biogeme` [62]. This package is convenient because it has built in tools for variables with unique values for specific individual and alternative pairs, as is the case for the number of active printers and the distance between Authors activity cities and printing alternatives. The package is also resource efficient as it mainly functions as a data wrapper in `python` compiling all calculations in a `C++` environment.

6.2.4. Validation

The estimation of the models parameters is done in a $K = 5$ fold validation setup with a 80/20 estimation and test split. To assess the models performance, we first provide the frequently used goodness of fit statistics Likelihood Ratio Test (LRT) and Rho-square (ρ^2). These are both dependent on the Log-Likelihood, which is the logsum of the predicted probabilities of the actual observed choices from the data, we write this down as $\log(L)$.

The general idea of the LRT is to investigate whether a simpler model is sufficient to explain the data. This is done by introducing linear restrictions on the parameters, which form the null hypothesis, that the restricted (simpler) model is correct. If the test rejects this hypothesis, it suggests that the unrestricted (more complex) model model provides a significantly better fit and the degrees of freedom [63]. In our case, the restricted model involves forcing all parameters to 0 to assume equal probability for all alternatives, and trying to reject this with the alternative model that do not have this [64][65].

Since the equal probability is equal to random chance it ususally performs really poorly. Alternatively, the null model can be replaced by a model only including the alternative specific constants, which is often referred to as the initial model [65]. However, due to the high number of alternatives in this research estimating all the alternative specific constants is infeasible [66], therefore the only option is using the null model as benchmark. Last, the higher the number of alternatives, the worse the null model performs. Therefore, this statistic needs to be interpreted with caution.

The test is calculated as follows:

$$\Lambda = -2(\log(L_0) - \log(L_1)) \sim \chi^2_{(K_U - K_R)} \tag{6}$$

- $\log(L_0)$ is the log likelihood of the restricted model
- $\log(L_1)$ is the log likelihood of the unrestricted model
- K_R is the number of parameters in the restricted model
- K_U is the number of parameters in the unrestricted model

The Rho-square is calculated as follows:

$$\rho^2 = 1 - \frac{\log(L_1)}{\log(L_0)} \tag{7}$$

It is simply to explain the proportion of variance explained by the unrestricted model compared to the null model scaled between 0 and 1. This looks similar to the famous r^2 but is generally a lot

lower [67]. This can also be adjusted to the number of estimated parameters as the $\bar{\rho}^2$, where K is the number of estimate parameters in the full model [68]:

$$\bar{\rho}^2 = 1 - \frac{\log L_1 - K}{\log L_0} \tag{8}$$

With the estimated parameters from the train split, the perceived utility equations are solved and choice probabilities of the alternatives are simulated for the test split with the following softmax function [69]. Given utilities $\{U_j\}_{i=j}^{n}$, the choice probability P for alternative j from choice set C is:

$$P_j = \begin{cases} \frac{e^{U_j}}{\sum_{j \in C} e^{U_j}} & \text{if alternative } j \text{ is available} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

As explained in section 6.2.2, the number of active printers in alternative city j depends on given publishing year t. If the number of active printers at given year t is 0, alternative j is unavailable. If so, alternative j is masked out to not interfere with the probability space and set to 0 after.

Alongside the predictions we generate alternative probabilities which are equal for all available alternatives allows us to calculate the LRT and ρ^2 to compare the performance of the estimated models models. Accompanying these we calculate two more metrics to assess performance, the mean distance from predicted to observed location and the percentage of correct predictions.

Last, we compare the performance of the model estimated on the entire set with the models estimated on the subsets. Treating the model estimated on the entire subset as $\log(L_0)$ and the total log likelihood of the subsets as $\log(L_1) = \log(L_{censored}) + \log(L_{not\ censored})$ we can do a final LRT and ρ^2.

7. Dynamics from Literature

7.1. Deventer

Under the influence of the 'Modern Devotion', the book industry in Deventer flourished in the 15th century [70]. The city had all the qualities for success: five yearly international fairs, a strong book tradition dating back to 1377 [71], and the presence of its Latin School, which had a broad reach of attracting scholars and students [70][71]. The printing press also came quickly, with the first starting its operation in 1477. This quick adoption and the presence of the school led to Deventer growing out into the important city for book production in 1500 and was economically the strongest city of the Low Countries [70]. Deventer's decline was most likely related to a series of conflicts between the Burgundians and Habsburgs in the first half of the 16th century [72].

7.2. Antwerp

The work of other historical studies claims that in the first quarter of the 16th century, Antwerp and its surrounding areas grew to the most important international trade hub by land and sea [73][21]. In 1543-1545 when the central government established a 1% tax on all export revenue, the metropolitan area of Antwerp was responsible for 75% of the total sum coming from the Low Countries [21], with Amsterdam taking second place, contributing only 5% [20]. Until 1560, the circumstances in Antwerp were perfect. A steady supply of production materials, an abundant supply of labor [20][21][18] and close proximity to the University of Leuven [20][18] provided outstanding opportunities for the printing houses. From 1500-1540, more than half of all active printers in the Low Countries had established themselves within Antwerp's city walls [18]. Starting from 1560, Antwerp's decline occurred as rapidly as its growth. The city experienced mass emigration of human and economic capital due to the intervention of the Spanish forces, and especially the coming of the Duke of Alva. The main concern the Spanish brought was restrictions on the freedom of religion, with rough punishment. Catholicism ended having the upper hand, which forced protestants to flee, which is considered the main motive of the emigrations of printers [18][21][74][75]. Before 1585, most relocating printers went to England and Germany, and after 1585 most of them settled in the Northern Low Countries with most relocating to The Hague, Leiden, Rotterdam and Amsterdam [18]. The emigrations of 1585-1589 are also considered to be the most significant with Antwerp's population dropping with about 50% from 1585 - 1589 [76][21]. Research covering the whole of the South report an accelerated exodus of 100,000 emigrants primarily relocating to the Northern Low Countries [77].

7.3. Holland

The relocating migrants coming from the south are considered the main driver of the economic prosperity the cities of Holland experienced during the 17th and 18th centuries, as they brought along their knowledge and capital [74][75]. Historical demographic research has proven that around 1600 about 70% of Amsterdam's population were migrants. For the total population of Leiden and Haarlem, this was about 18% and 60% for the combined population of Rotterdam, Delft, and The Hague [21]. This meant that the book production industry of Holland flourished, which hosted about 80% of all the printers during the 17th century [35, p. 42 & 43]. Having the opportunity to prepare any move within four years from 1585, most relocating printers could sell their belongings and properly organize the trip so they could get right back to business [21][75]. Furthermore, there is no discussion among Historians of geography of book production in the 18th century, they all widely acknowledge Amsterdam as the leading city producing by far the most literary work when the newly formed Dutch

Republic had grown into a nation of global power overtaking Venice and Basel as leader of the worlds book production [78].

7.4. Protestantism in the Low Countries

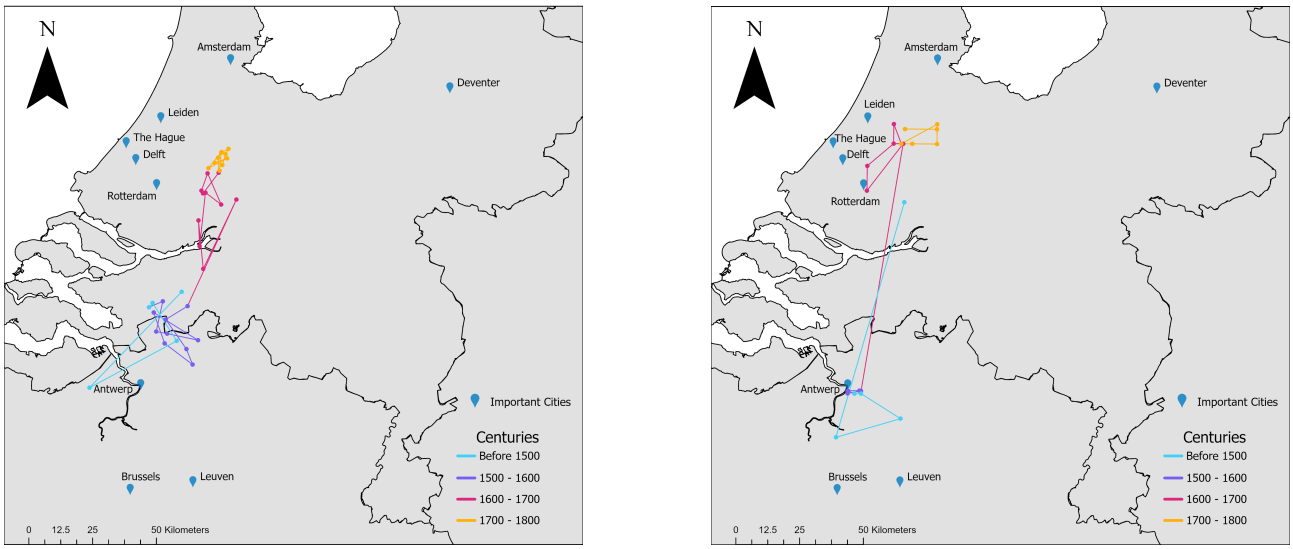
As also explained by Antwerp's emigrations above, religion was one of the most important aspects of life in early modern times in which the Low Countries population participated as fully as anywhere else in Europe [79]. Historical evidence suggest that reason why the Low Countries were such a fertile region for the reformation is because of its fragmented political nature. This caused a relative absence of a powerful suppressive force against the new ideologies allowing them to flourish [17, p. 31-44][16].

Antwerp's position is described as the the birthplace and center of the new religious movements of the reformation in the early 16th century [80][17, p. 24], with at least 12 Latin and 10 Dutch versions of Luthers work pressed in the city between 1520 and 1522. Other historians describe Antwerp's role as the gathering place of religious malcontents, a place of refuge for religieuse migrants [81, p. 173], and as a distributor of heretical literature [82, p. 270]. Due to the increasing commercially and industrially oriented population Antwerp was by tradition the safest city for the Low Countries heretics [81, p. 173]. Starting in the early 16th century as fertile grounds for Lutheranism [17, p. 24], Antwerp acted around 1560 as the Low Countries metropolis of Calvinism [81, p. 173]. This changed after the Fall of Antwerp in 1585 when the Spanish conquered the city together with most of the South of the Low Countries giving Catholicism the upper hand [21]. Historical evidence describes the following period in the Southern Low Countries as one with strict censorship on the print industry targeting all works supporting the ideologies of the reformers [76][17]. The fleeing Calvinsits would mostly relocate to the region of Holland, in which the Catholics would still by far be the biggest group. By 1650, half of the entire population of the Northern Low Countries is reported to have become Protestant [83].

8. Results

8.1. Mean & Median centroids

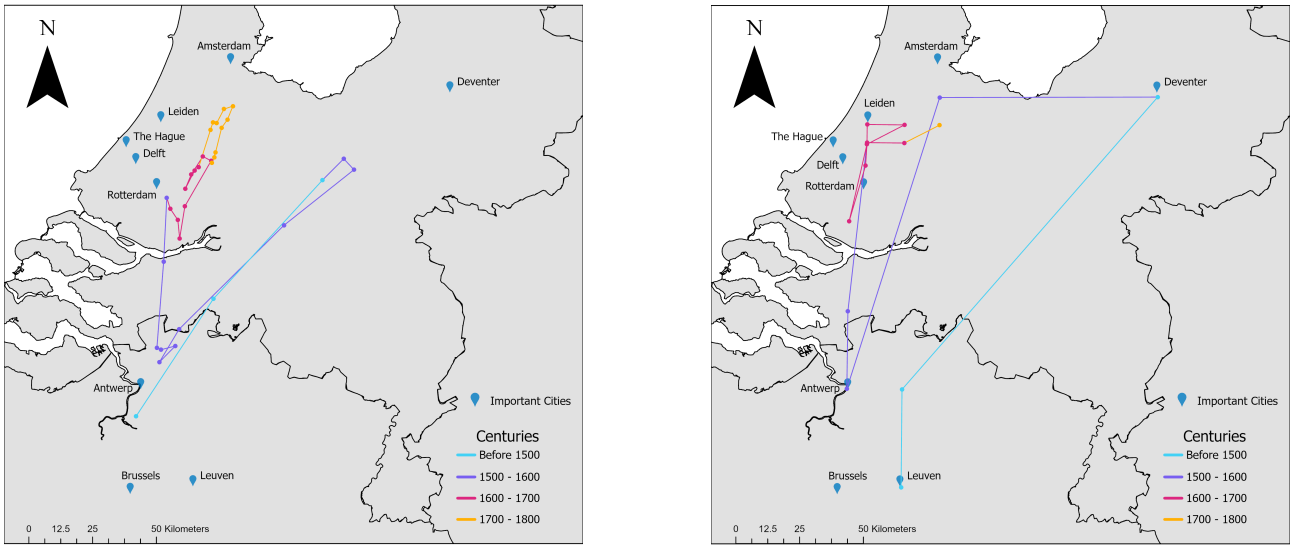
The reproduced findings of Koren can be seen in figure 4. As he also observed, we can see a clear trend from South to North, with a first shift of mean centroid first moving from Antwerp towards Holland in the early 17th century and nudging closer towards Amsterdam until the end of the 18th century [25].



(a) Mean Centroids (b) Median Centroids

Figure 4. Recreation of Koren's findings using the Chaney data
Source Country Borders: Landsgrenzen, Esri Nederland, 2025 (Flevoland removed)

When using the locations of printed works of the library catalogs to calculate the mean and median centroids, the movements of the centroids over time look different, as can be seen in figure 5.



(a) Mean Centroids (b) Median Centroids

Figure 5. Results from all works printed in the Low Countries
Source Country Borders: Landsgrenzen, Esri Nederland, 2025 (Flevoland removed)

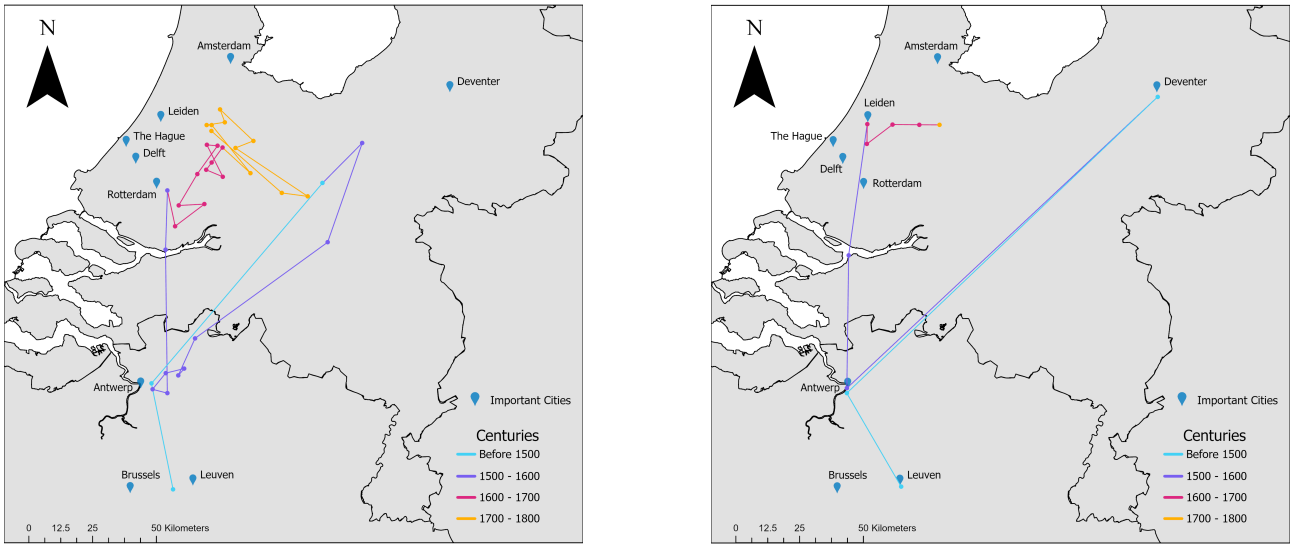
The overall trend from the end of the late 15th century to late 18th century is similar to figure 4. The centroids move from the region of Antwerp towards Holland and end up in Amsterdam. Although the final stretch in the 18th century, with the centroids moving closer towards Amsterdam is more clearly visible, especially when comparing the Mean Centroid figures.

The difference between the two data sources is the difference in movement in the early period until halfway into the 16th century. The median centroids originating in Antwerp dash in North-East direction towards Deventer in the late 15th century and return towards Antwerp around halfway into the 16th century.

Another clear difference can be seen in the late 16th century. The purple dashed line represents an abrupt change in mean centroid from 1580-1590 to the centroid of 1590 - 1600. The centroids in figure 4 move towards Holland much later in the early 17th century.

Works by Authors in the Chaney data

Directly comparing figure 5 and 4 is not logical. As explained in section 5.3.2, only a portion of the authors mentioned in the STCN & STCN could be matched to the Authors in the Chaney data. To make a fair comparison, the centroids are re-recreated with the works in which the authors appeared in the Chaney data, which can be seen in figure 6.



(a) Mean Centroids

(b) Median Centroids

Figure 6. Results from all works printed in the Low Countries with authors that appear in the Chaney data Source Country Borders: Landsgrenzen, Esri Nederland, 2025 (Flevoland removed)

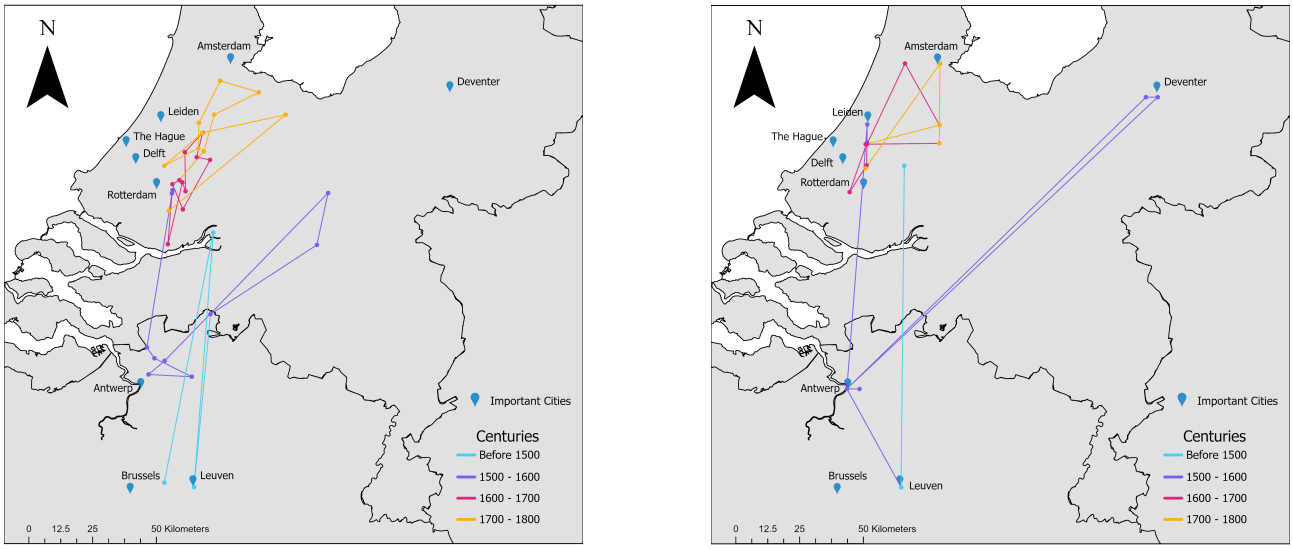
Interestingly, the same attractiveness of Deventer can be seen, but it seems to be a little less present when compared to figure 5. The mean centroids for the first half of the 16th century move back to Antwerp sooner in figure 6 than in figure 5. Furthermore, the transition from Antwerp to Holland in the late sixteenth century and the pattern in movement towards Amsterdam in the 17th century are very similar in both figures. For the case of the 18th century, where 4 and 5 show similar movement of the centroids nudging closer towards Amsterdam and appearing to form a dense cluster. Figure 6 is very different as the centroids are very far apart from each other compared to the other figures.

If we compare the Median centroid maps of figure 5 and figure 6, the overall movement looks very similar, except that the median centroid moves from Deventer to Amsterdam before moving back to Antwerp in figure 5. This is likely due to how the median centroid is calculated. As explained in

section 6.1, the median coordinates are calculated separately by sorting the longitude and latitude arrays in ascending order and calculating the cumulative sum of weights.

Censored Works

To visualize the geography of censored material, we also use the centroid methods, which can be seen in figure 7.



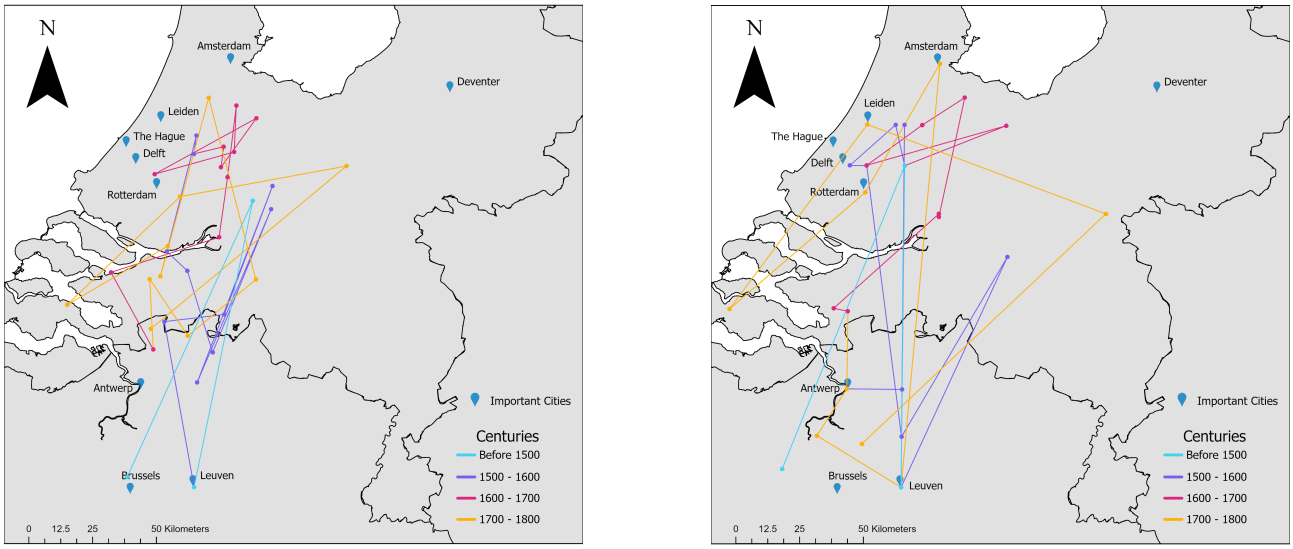
(a) Mean Centroids

(b) Median Centroids

Figure 7. Results from all censored workes printed in the Low Countries
Source Country Borders: Landsgrenzen, Esri Nederland, 2025 (Flevoland removed)

The centroids before 1500 are very scattered. Starting near Brussels and Leuven, they move up towards Amsterdam and then move back. In the early 16th century, we see a similar movement as in figure 6 towards Deventer, moving towards Antwerp in the mid 16th century, and movement to Holland in the late 16th century. The only apparent difference is that the mean centroids of figure 6 move further towards Deventer and earlier in the 16th century. The mean centroids in the 17th century show a different pattern when compared. They don't seem to move gradually towards Amsterdam, but they tend to scatter around and move back towards Rotterdam in the later half of the 17th century. The movements in the 18th century fail to display any clear patterns as they scatter all over Holland over the century.

When we also weigh the centroids by the total number of printed per city per year, observing a pattern becomes much more challenging, as shown in figure 8. The centroids move back and forth until 1560, when they move towards Holland. They keep themselves relatively centered next to Delft and The Hague directly underneath Amsterdam until 1670 after which the seem to move in all kinds of directions without any observable patterns.



(a) Mean Centroids

(b) Median Centroids

Figure 8. Results from all censored workes printed in the Low Countries weighted by total printing volume per city per year
Source Country Borders: Landsgrenzen, Esri Nederland, 2025 (Flevoland removed)

8.2. Discrete Choice Modeling

8.2.1. Estimation Full observation Set

First we elaborate on the estimated parameters using the complete set of observations, the full estimation reports can be found in the appendix table 3. The estimated parameters across all the folds are consistent, precise and statistically significant with a P-value of 0 for both parameters. The estimated coefficient for the natural logarithm of distance in meters averages around 3.834, with an average standard error of 0.02806. The estimated parameter for the natural logarithm of the number of active printers is even more precise and consisted with the coefficient for all folds of 1.2 with an average standard error of 0.03164. Like the parameter estimates, the goodness of fit statistics are also very consistent across the five folds with the ρ^2 and $\bar{\rho}^2$ averaging out to 0.521.

This score is considered very high since a score of 0.3 is often referred to as a 'decent model fit of a discrete choice model' [64]. However, it is important to note that this informal goodness of fit statistic is calculated by benchmarking the model against the equal probability model, also known as the null model, which often performs very poorly.

8.2.2. Estimations with Censored and Non Censored subsets

To test the effect of an author being censored, we split the total of observations into two separate groups and estimate the parameters separately. For the non censored authors group the parameter estimates are almost identical to the estimation for the full set with the parameter of number of active printers being 1.99 for all five folds and the distance parameter averaging out at -3.864. Both estimations are also very consistent and precise with near identical estimated values and standard deviations averaging to 0.0285 for the distance parameter and around 0.00438 for the active printers parameter. Both are also statistically significant with P-Values of 0. The ρ^2 and $\bar{\rho}^2$ are also almost identical with both averaging out at 0.523.

The estimations for the censored authors subset is very different from the other two instances. As can be seen in table 1:

Fold	Parameter	Value	Std error	t-test	P-value	Null LL	Final LL	ρ^2
1	β active printers β distance	1.36 -5.24	0.0416 0.196	38.2 -37.1	0 0	-4129.198	-1693.691	0.59
2	β active printers β distance	1.33 -5.09	0.0338 0.138	39.4 -36.8	0 0	-4129.198	-1768.367	0.572
3	β active printers β distance	1.35 -5.16	0.0351 0.14	38.4 -36.8	0 0	-4129.198	-1742.702	0.578
4	β active printers β distance	1.33 -5.1	0.0343 0.14	38.7 -36.4	0 0	-4129.198	-1777.619	0.57
5	β active printers β distance	1.34 -5.17	0.035 0.14	38.3 -37	0 0	-4133.731	-1733.502	0.581

Table 1. Parameter Estimation Summary of censored authors observations group

The parameter estimates are still relatively consistent and significant, but much less precise as indicated by the standard errors of the parameters. The most interesting phenomenon are the values of the distance parameter. The estimated parameters show a much larger penalizing effect of the cost-distance of alternative cities than for non censored authors. However, the sample size of the observations by censored authors is way smaller than the full estimation and non censored set with only 911 observations per fold compared to 27028 for the non censored subset. Nevertheless, the number of observations is still enough to estimate highly significant effects. The ρ^2 and $\bar{\rho}^2$ are averaging out to 0.578 indicating a slightly better fit than the previous estimated models.

8.2.3. Predicted Probabilities

As explained in section 6.2.4, for each of the folds, 20% of the observations is reserved as a test subset for the prediction of probabilities. Because the estimations and the fit is fairly consistent throughout the folds for each of the three models, we present prediction results for the entire set of observations for each of the models in table 2.

Metric	Full Sample	Non-Censored	Censored
Correct Predictions (%)	44.36	44.63	50.13
Average prob correct predictions	0.9960	0.9969	0.9917
Average prob incorrect predictions	0.0019	0.0012	0.0018
Nearst city with Active printers Predictions (%)	100	100	100
City of Residence (%)	98.26	98.57	98.24
Mean Euclidean Distance (M)	62,846	62,057	55,982
Null Log-Likelihood	-140,564	-136,290	-4,454
Final Log-Likelihood	-300,878	-291,713	-13,307
Likelihood Ratio Test (LRT)	-320,628	-310,846	-17,707
ρ^2	-1.1405	-1.1404	-1.9879
Number of Observations (N)	34,925	33,786	1,139

Table 2. Prediction performance metrics for the full dataset and its censored/non-censored subsets.

Due to the large negative effect of log(distance) in the utility function, all the models predicted the highest probability to the closest city with at least one active printer in every choice situation with an average probability that is very close to 1 for each prediction. This was almost every single time the authors city of residence. The correct percentage of predictions and average Euclidean Distance between Observed and Predicted City suggest slight benefits from splitting the observations into the different subsets.

Unlike the in sample estimation statistics discussed, these metrics show very different results for the out of sample predictions. The Null Log-Likelihood, Final Log-Likelihood and Likelihood Ratio Statistic for each of the three models suggest that forcing equal probabilities results into a much better fit with the LRT being highly negative for all three models. This results in a negative ρ^2 which is about -1.14 for the full and non-censored models and almost -2 for the censored model, which suggest that the model estimated on the censored subset has by far the worst fit out of all three models.

To directly compare the prediction results of the model from the full sample with the results of the predictions on the subsets, we do a last LRT. With the Final Log-Likelihood from the full sample as benchmark (unrestricted model) with the sum of the Final Log-Likelihoods from the predictions of the subsets (alternative model) we can test if this implies a better fit. This has a LRT of -8286.80 with a ρ^2d of -0.0138, which indicates a slightly worse fit.

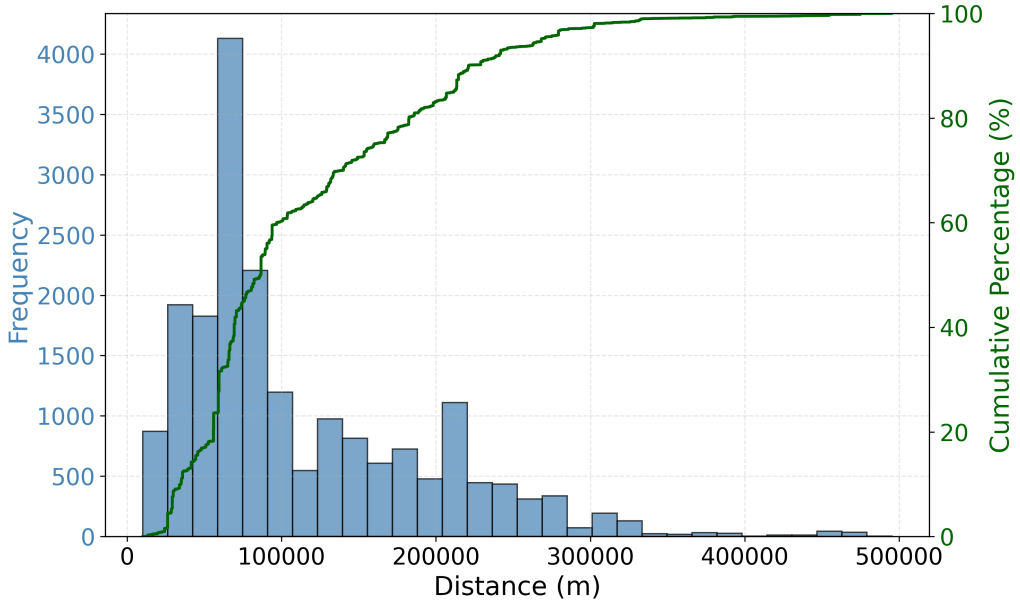


Figure 9. Distribution of observations not printed in the Authors city of residence, total of 19534 observations.

Looking at figure 9 we can see that most of the works are printed within a 100 kilometer range, when not printed in the city of residence. From these, only 106 were correctly predicted by the model.

9. Discussion

The clear message that speaks from early modern times in the Low Countries is that it consisted of complicated economic, religious, and political dynamics [21]. Cities have experienced rapid growth as well as decline. These frequent shocks forced individuals to adapt to new circumstances frequently. Demographic research concludes that the population of low countries has adopted a hypermobile nature due to these shocks. Population statistics confirm this, indicating that many of Holland's most significant cities consisted of Migrants around 1600 [21]. Evidence from this and related studies shows that authors and book printers had indeed adopted this hypermobile nature. Mainly driven by push and pull factors related to freedom and constraints of faith [21][18], they moved themselves and their business several times relocating to different Countries [18][20].

9.1. Centroids

Using bibliographic data, spatial-temporal patterns can be observed that closely align with qualitative descriptive studies considering the Geography of the Book. Looking at figures 4, 5, and 6, we can see a shift from south to north, which is related to the rise and fall of Antwerp. In all figures, we can observe it emerging from 1500 - 1625 into Europe's most important city and its decline in the late 17th century. As Koren also explained [25], research suggests that Antwerp's economy was already in decline before 1585. The results from this study show that using Author and publication data both fail to show these earlier waves of emigration. This is obvious as the reviewed literature (section 7.2) claims that most of these migrants moved to England and Germany, making them unobservable when restricting the data to the Low Countries. The fall of Antwerp is most clearly observable in figure 5 and 6. The last two median centroids from the 16th century that moved to Holland consist of the data from 1580 - 1599. These strongly support the described significance of the migration waves after the fall of Antwerp (section 7.2). Interestingly, in figure 4, the mean centroids move to the North much later, starting in the seventeenth century.

The most obvious difference from the figures is the position of Deventer. As explained in section 7.1, Deventer had an important position in the Low Countries' book culture around 1500, clearly recognizable in figure 5 and 6. The period of downfall and the rise of Antwerp overlap, which explains the move of the mean centroids. Nothing of this can be seen in figure 6. The only limitation that could be found is that he only georeferenced authors that were associated with a settlement that had at least one growth estimate in the union of Bairoch et al. [84], Bosker et al. [85], and Chandler [86]. Still, when the centroids were calculated exclusively using publications written by authors that appear in the Chaney data in figure 6, the importance of Deventer around 1500 is clearly visible.

The inability of the Works written by Authors in the Chaneys dataset to display trends that confirm the rise and importance of Amsterdam can be explained by the temporal dimension of Chaneys' dataset stopping at 1800. All possible authors that could have had their work printed in the Low Countries in the late 18th century who would have died later in the earlier periods of the 19th century are not included. Therefore, as more and more authors die approaching the 19th century, the number of observations to compute the centroids decreases (histogram 6) which could explain the decrease in density of the centroids.

Despite the minimal subset of censored works, the centroid method shows very similar patterns as described in the literature, with Antwerp displayed as an important place for conspiring protestants in the middle of the 16th century [81, p. 173]. The fleeing protestants who settled mostly in Holland after the Fall of Antwerp are also clearly visible. However, as we normalize the number of censored works

by the total number of printed works per city per year, we lose almost all observed patterns (figure 8). One could still argue to see some kind of concentration of censored works in Antwerp in the middle of the 16th century, with a movement towards Holland that could be linked to the Fall of Antwerp and the relatively high concentration of protestants in the cities of Holland. This period also contains the vast majority of observations, summing up to about 50% according to the cumulative graph in histogram 3d. As the number of observations drop, the centroids tend to move in all directions when weighted by total number of printed editions. It is also observed that they frequently go back to the South, which is perceived to be under strict catholic control. This shows that they do not have a very systematic meaning by themselves when visualizing patterns.

Something of an explanation for the low number of observations is the fact that Protestantism has grown substantially over the course of 1600 - 1800, with the literature explaining that the Protestants had a strong position of power, especially in Holland, even suppressing Catholics [81]. This could imply that the Catholic church had limited power and control of the printing culture of the Republic, failing to monitor and censor works actively. Other research suggests that the Catholic inquisitors had a lot of informants who provided accurate information, suggesting the opposite [17]. Another frequently reported way of dealing with repressive censorship was for the Authors and printers to publish and print works anonymously or use fictive or wrong printing locations [17]. This likely does have some role for the limited number of observations and records included in all the different subsets considered in this research. The STCN and STCV claim they include over 250000 unique works of literary content, of which only 148143 could be geocoded to places in the Low Countries.

Other data limitations are caused by the unclear nature of historical data. During the preprocessing, we ran into similar issues Koschnick faced when matching the STCN and STCV to the VIAF database. Many duplicate matches were found when matching the Authors from the Censorship Sources to the Catalogs, from which about 50% could not be resolved and had to be discarded entirely. Koschnick was also forced to do this for many cases, limiting the matches that could be found between the Chaney data and the Catalogs, which further reduced the available observations. Last, the printing industry of the Low Countries was largely oriented to producing literary works for international export, which makes it reasonable to assume that a large share of printed works were not written domestically. This is mentioned explicitly for the case of Deventer, as the city specialized in producing Latin textbooks meant for international Education [20]. Also, it is very plausible that the geocoded works of the STCN and STCV contained works for which the Author is unknown or for which any information about the death, activity, or birth year is missing. Chaney [30] reported that if this was the case, the author would be discarded from the database, limiting the number of matches.

9.2. Choice modeling

The estimation results described in section 8.2.1 and 8.2.2 generally indicate a good fit to the data, while the prediction results described in section 8.2.3 show the exact opposite for the test data. At first, this indicates that the generalization of the models is terrible. When tested on reserved data, the percentage of predictions seems okay. Still, the general goodness of fit statistics indicate awful performance compared to random chance with negative LRTs and ρ^2. This means the models generally suffer from overfitting and don't handle previously unseen data very well. Looking at the distribution of works in figure 9, we can see that the distribution of works not printed in the Authors' city of residence is left-skewed, with about 16% of works being printed within 100km. This is very obvious given the high coefficient of the logarithm of the distance in meters as reported by the parameter estimations. With the coefficient for active printers not having a substantial enough effect on the

utility function, almost none of these observations could correctly be predicted.

First, as explained above, the printers are reported to have moved their operations elsewhere, with some relocating multiple times. It is highly plausible that the authors also adopted this hypermobile nature, which imposes a lot of fundamental issues. These movements are not included in the Chaney data as the activity location is reported as a single place. Therefore, it is highly likely that the writing location and distance to alternative cities are incorrect for all works considered. We cannot assess this error as it is not observable with the available data.

Another possible issue that came up is the availability of the alternative cities for the estimation of the models parameters. The way the number of active printers was coded into the model data is by calculating the log plus 1 to avoid infinitely negative values. This resulted into cities with 0 active printers being present in the estimation data. The Null Log-Likelihood calculated during the models parameter estimation does take these into account. When the probabilities were predicted on the test data, these alternatives were masked out and not included when calculating the Null Log-Likelihood with equal probability for all available alternatives. This difference could be an explanation for the difference in the goodness of fit statistics for the estimation and predictions.

From the goodness of fit statistics when comparing the model estimated on the full sample with the model's estimate from the different subsets, we can observe that the statistical fit is slightly worse with a LRT of-8286.80 ρ^2 of -0.0138. This implies that the slight increase of the correct predictions and the decrease of the Mean Euclidean Distance is most likely caused as an overfitting effect on the data by reducing its heterogeneity. The reason that the model performs so badly is most likely of how the LRT is calculated. When calculating the logsum of the probabilities of the observed choices, the large number of tiny probabilities is quite large for 50% or more of all predictions for each of the three estimated models.

10. Concluding Remarks

As explained in section 1, this research contributes to the study of the "Geography of the Book" by comparing the authors' locations with the publishing locations of their work. The inclusion of Authors' locations in historical spatial-economic research is very new. First contributions come in the forms of a preliminary discussion paper by Chaney [15], a PHD draft by Chiopris [14], and as a Master's Thesis by Koren [25]. As explained in the works of Chaney and Chiopris, hand crafting and geocoding a high resolution Author location dataset is a precise and time consuming process, that relies on web scraping and manual search & evaluation of online sources such as Wikipedia to assign relevant cities to individual authors [30][14]. These conditions are certainly less favorable than the geocoding of works, as printers in the Low Countries were often obligated to include their address in produced works to be held accountable [17][16]. These data implications and the absence of a single published scientific study including author locations highlight the current novelty of this topic.

The literature reviewed in this study concludes that there is generally a consensus among historians about the print industry's general spatial and temporal trends in the Low Countries. They declared Deventer to be the first dominant city around 1500. After that, Antwerp takes the crown, growing into one of the biggest trading hubs in the world around 1560, with Holland taking over after the Spanish conquered most of the South and caused a lot of human and business capital to fuel the flourishing industry of Holland. This is considered one of the most critical factors for the new Dutch Republic to grow into the world leader in producing literary works in the 17th and 18th centuries.

This study highlights apparent differences in the spatial and temporal patterns of the Geography of the Book in the Low Countries when comparing author and publication data. Building on Korens work [25], we apply his proposed centroids-based method on author death locations and places of publication. A striking example is Deventer, a book production hub in the late 15th century that peaked in 1500 [70][71]. Its prominence is reflected in centroids based on publication data (Figures 6 and 5), but not in those derived from author locations (Figure 4). Furthermore, the publication data registers the significant historical disruption known as "The Fall of Antwerp" in 1585 (section 7.2) more clearly than the Author data. On the other hand, the author's data seems to align more closely with research about the rise of Holland and specifically Amsterdam in the 17th & 18th centuries than their respective publications. This makes it challenging to crown either as more suitable. We also attempt to observe trends when computing the centroids, just using the printing locations to see if we can make any statements about the geography of censorship. As can be seen in Figure 8, the patterns almost entirely disappear when the centroids are controlled for the total printing volume per city per year, preventing any sort of meaningful interpretation about the geography of censorship in the Low Countries.

Second, we model whether censorship influenced authors' choice of publication city using a Discrete Choice framework, testing if censored authors published farther from their residence to evade regulation in the politically fragmented Low Countries. Statistics of the models' estimation suggest the opposite, reporting a higher, strongly significant penalizing effect of distance on publication choice for censored authors than non-censored authors. The probabilistic predictions show that for a large chunk of observed publications, about 45 - 50% the place of publications is always the closest city with at least one active printer. However, as the model always predicts the closest option with extremely high confidence, the model offers no further explainability of the choice behavior, ultimately making its performance worse than random chance. Regarding censorship, we show that splitting the observations into censored and non-censored subgroups to reduce heterogeneity shows signs of overfitting, making the model's predictions worse when applied to reserved unseen data. This could be caused by a significantly smaller sample size of 1,139 censored observations compared to 33,786 Not Censored. From these results, we can conclude that the distance and censorship on their own are not suitable to model the spatial location choice for printing.

11. Recommendations

First, as explained in the conclusions, the novelty of including Authors' publications in historical spatial-economic research exhibits the need for more scientific attention. This study highlights a clear difference when mapping the spatial and temporal patterns using Authors' locations and Publication locations in the Low Countries. As also noted by Koren [25], the dimensionality of the data causes issues for the interpretation of the visualizations. Color coding the lines between the centroids allows better understanding of the observed trends along the timeline. Still, when the centroids are more dense within century groups, the overview of the timeline becomes very difficult. This is a classic issue when visualizing dynamics on a static map [87]. Displaying animations instead would ease the interpretation.

Second, Chaney's dataset was designed to use cities' yearly author death counts as a high-frequency measure of historic economic activity at a continental scale for both the entirety of Europe and the Middle East & North Africa regions [15]. While he does not limit the data to large author-intensive cities, directly comparing it to the region-specific STCN & STCV imposes a clear difference in the goals and use of the datasets. Therefore, it is very likely that modeling the centroids using an Author

dataset created explicitly for the Low Countries would display different results. A future study that includes as many Authors included in the STCN & STCV would also allow testing the Chaney dataset's robustness when using it for smaller scale applications.

Third, the use of the mean and median centroid methods obscures entirely the underlying distribution of the data, as is also noted by Koren [25]. He proposed a spatial-temporal kernel density estimation method to consider the distribution. Applying this on the STCN & STCV data would offer additional comparability between the Author and publication data.

Regarding the discrete choice modeling results, the nested choice model has received a lot of criticism for its capabilities to incorporate spatial autocorrelation in choice modeling. The boundaries of spatial clusters are more likely to be perceived as "fuzzy" than discrete zonal borders, which can cause problems [44].

Several methods have been proposed to model spatial effects more accurately. They typically involve the use of some Spatial Weight matrix to account for spatial correlations among individual decision makers and alternatives [44][41][54]. However, these methods are known to require even more computational resources than the aggregated choice model proposed in this research. Exploring these methods with appropriate resources would be recommended.

Second, in this research, we imply that the Authors act as rational decision makers with complete information. They are assumed to know all active printers within the Low Countries. Communication and transportation over large distances were costly in Early Modern Europe [17], which questions whether it is correct to assume that every Author would be aware of all the theoretically available alternative cities. Also, we consider the Low Countries' borders according to the current country borders of the Netherlands and Belgium. This excludes any alternative cities that could be relatively close by in Germany or France, which are not taken into account. This suggest that experimentation with the other ways to define the authors choice sets could be further explored.

Third, as can be read in the introduction, it was relatively easy to print controversial works in the Low Countries compared to other European countries. Findings from the literature, for example, confirm that exiles from England settled on the mainland, producing copies which would then be smuggled back into Great Britain to evade the British licensing act [22]. As the greatest relative freedom of religion was experienced in the Low Countries, an Author from the Low Countries who the Catholic Church would censor would probably have an easier time pressing his work domestically than anywhere else. Applying the model to data on a larger scale including options for censored foreign authors to print their works in the Low Countries would be suitable to assess the role of the printers providing illegal works for the international market as is clearly exhibited in the reviewed literature.

Last, the results show that the Active location of the authors and whether they have been subjected to censorship do not provide enough deterministic capability to explain the choice behavior to outperform a random chance baseline model. This means that other unobserved factors are responsible for the choice of printing, which would need to be explored in future research. As can be read in section 7, the reviewed literature reported that the decline of the economy of certain cities and regions was often caused by armed conflicts, causing human and business capital to migrate elsewhere. The characteristic of alternative cities related to being involved in these conflicts could be a good starting point.

12. Generative AI Statement

During the process of this thesis I have used generative AI as an assistant for helping with coding problems, LaTeXformatting, translations, brainstorming and a little bit of writing feedback.

References

[1] Jeremiah Dittmar. "Book prices in early modern Europe: an economic perspective". In: *Buying and Selling*. Brill, 2019, pp. 72–87.

[2] Chris Browne. "Gutenberg, chemistry and the advancement of knowledge". In: *Australian Journal of Chemistry* (2023).

[3] Steven Kreis. "The printing press". In: *The History Guide. URL: http://www. historyguide. org/intellect/press. html (: 02.04. 2018)* (2004).

[4] Lotte Hellinga. "The Gutenberg Revolutions". In: *A Companion to the History of the Book* (2019), pp. 377–392.

[5] Eltjo Buringh and Jan Luiten Van Zanden. "Charting the Rise of the West: Manuscripts and Printed Books in Europe, a long-term Perspective from the Sixth through Eighteenth Centuries". In: *The Journal of Economic History* 69.2 (2009), pp. 409–445.

[6] Brenda M Hosington. "Introduction: Translation and print culture in early modern Europe". In: *Renaissance Studies* 29.1 (2015), pp. 5–18.

[7] Daniel Bellingradt and Jeroen Salman. "Books and Book History in Motion: Materiality, Sociality and Spatially". In: *Books in Motion in Early Modern Europe: Beyond Production, Circulation and Consumption*. Springer Nature, 2017, pp. 1–11.

[8] Bertrum H MacDonald and Fiona A Black. "Using GIS for spatial and temporal analyses in print culture studies: some opportunities and challenges". In: *Social Science History* 24.3 (2000), pp. 505–536.

[9] Innes M Keighren. "Bringing geography to the book: charting the reception of Influences of geographic environment". In: *Transactions of the Institute of British Geographers* 31.4 (2006), pp. 525–540.

[10] Innes M Keighren. "Geographies of the book: Review and prospect". In: *Geography Compass* 7.11 (2013), pp. 745–758.

[11] Joerg Baten and Jan Luiten Van Zanden. "Book production and the onset of modern economic growth". In: *Journal of Economic Growth* 13 (2008), pp. 217–235.

[12] Jan Luiten Van Zanden. "Chapter Six. The Philosophers And The Revolution Of The Printing Press". In: *The Long Road to the Industrial Revolution*. Brill, 2009, pp. 177–203.

[13] Raouf Boucekkine, David De la Croix, and Dominique Peeters. "Early literacy achievements, population density, and the transition to modern growth". In: *Journal of the European Economic Association* 5.1 (2007), pp. 183–226.

[14] Caterina Chiopris. *Spatial Networks and The Diffusion of Ideas*. Tech. rep. Accessed: 2024-10-15. Harvard University, 2024.

[15] Eric Chaney. *MODERN LIBRARY HOLDINGS AND HISTORIC CITY GROWTH*. CEPR Discussion Paper DP14686. Preliminary results. Centre for Economic Policy Research, 2020. URL: `https://papers.ssrn.com/sol3/Delivery.cfm?abstractid=3594320`.

[16] Joris van Eijnatten. "Van godsdienstvrijheid naar mensenrecht. Meningsvorming over censuur en persvrijheid in de Republiek, 1579-1795". In: *Bijdragen en mededelingen betreffende de geschiedenis der Nederlanden* 118 (2003), pp. 1–21.

[17] Marita Mathijsen. *Boeken onder druk: Censuur en pers-onvrijheid in Nederland sinds de boekdrukkunst*. Amsterdam University Press, 2011.

[18] Johannes Gerardus Carolus Antonius Briels. *Zuidnederlandse boekdrukkers en boekverkopers in de Republiek der Verenigde Nederlanden omstreeks 1570-1630: Een bijdrage tot de kennis van de geschiedenis van het boek. Met in bijlage bio-en bibliografische aantekeningen betr. Zuid-en No-*

ordnederlandse boekdrukkers, uitgevers, boekverkopers, lettergieters, etc., en andere documenten. Vol. 6. Brill, 2024.

[19] Daniel Melo and Júlia Leitão de Barros. "Politics, culture, and religion in modern times: the Catholic Church and the restructuring of censorship." In: *Estudos em Comunicação* 37 (2023).

[20] Sandra Langereis. *De woordenaar: Christoffel Plantijn, s werelds grootste drukker en uitgever (1520-1589)*. Uitgeverij Balans, 2014.

[21] Jan De Meester. "Gastvrij Antwerpen? Arbeidsmigratie naar het zestiende-eeuwse Antwerpen". PhD thesis. University of Antwerp, 2011.

[22] Isabelle Fernandes. *Publish and Perish: The Practice of Censorship in the British Isles in the Early Modern Period*. Vernon Press, 2020.

[23] Sascha O Becker, Francisco J Pino, and Jordi Vidal-Robert. "Freedom of the press? Catholic censorship during the counter-reformation". In: (2021).

[24] Ian N Gregory and Alistair Geddes. *Toward spatial humanities: Historical GIS and spatial history*. Indiana University Press, 2014.

[25] Lucas Koren. *Mapping the onset of Modern Economic Growth: A regional approach using authors and library holdings*. 2024.

[26] Fabio Blasutto and David De la Croix. "Catholic censorship and the demise of knowledge production in early modern Italy". In: *The Economic Journal* 133.656 (2023), pp. 2899–2924.

[27] Stefano Comino, Alberto Galasso, and Clara Graziano. "Censorship, industry structure, and creativity: evidence from the Catholic inquisition in Renaissance Venice". In: *The Journal of Law, Economics, and Organization* (2024), ewae015.

[28] Matías Cabello. "The counter-reformation, science, and long-term growth: A black legend?" In: *Science, and Long-Term Growth: A Black Legend* (2023).

[29] Andrew Pettegree and Arthur Der Weduwen. *De boekhandel van de wereld: drukkers, boekverkopers en lezers in de Gouden Eeuw*. Atlas Contact, 2019.

[30] Eric Chaney. *Online Data Appendix to Modern Library Holdings and Historic City Growth*. 2024. URL: `https://sites.google.com/site/ericjustinchaney/working-papers`.

[31] Nederland Nationale Bibiotheek. *Short-Title Catalogue Netherlands (STCN)*. 2025. URL: `https://www.kb.nl/over-ons/diensten/stcn`.

[32] Vlaamse Erfgoed Bibliotheken. *Short-Title Catalogue Vlaandren (STCV)*. 2025. URL: `https://stcv.be/`.

[33] Jesús Martínez de Bujanda. *Thesaurus de la littérature interdite au XVIe siècle. Auteurs, ouvrages, éditions avec addenda et corrigenda. Centre détudes de la Renaissance*. Vol. 10. Éditions de lUniversité de Sherbrooke, Librairie Droz, 1996.

[34] Jesús Martínez de Bujanda. *Index librorum prohibitorum: 1600-1966*. Vol. 11. Médiaspaul, Librairie Droz, 2002.

[35] Ingrid Marlies Weekhout. *Boekencensuur in de Noordelijke Nederlanden. Een verkennend onderzoek naar de vrijheid van drukpers gedurende de 17e eeuw*. 's-Gravenhage: SDU Uitgevers, 1998.

[36] VIAF. *VIAF, Virtual International Authority File*. 2025. URL: `https://viaf.org/`.

[37] Chakkrit Snae et al. "A comparison and analysis of name matching algorithms". In: *International Journal of Applied Science. Engineering and Technology* 4.1 (2007), pp. 252–257.

[38] *HERE Geocoding Search*. `https://www.here.com/platform/geocoding`. Accessed: 2025-02-15.

[39] Milad Haghani, Michiel CJ Bliemer, and David A Hensher. "The landscape of econometric discrete choice modelling research". In: *Journal of choice modelling* 40 (2021), p. 100303.

[40] William Greene. "Discrete choice modeling". In: *Palgrave Handbook of Econometrics: Volume 2: Applied Econometrics*. Springer, 2009, pp. 473–556.

[41] Oleg A Smirnov. "Modeling spatial discrete choice". In: *Regional science and urban economics* 40.5 (2010), pp. 292–298.

[42] Michel Bierlaire. "Discrete choice models". In: *Operations research and decision aid methodologies in traffic and transportation management*. Springer, 1998, pp. 203–227.

[43] Kenneth E Train. *Discrete choice methods with simulation*. Cambridge university press, 2009.

[44] Kingsley E Haynes and A Stewart Fotheringham. "The impact of space on the application of discrete choice models". In: *Review of Regional Studies* 20.2 (1990), pp. 39–49.

[45] Thomas Reutterer and Christoph Teller. "Store format choice and shopping trip types". In: *International Journal of Retail & Distribution Management* 37.8 (2009), pp. 695–710.

[46] Ennio Cascetta. "Random Utility Theory". In: *Transportation Systems Analysis: Models and Applications*. Boston, MA: Springer US, 2009, pp. 89–167. ISBN: 978-0-387-75857-2. DOI: 10.1007/978-0-387-75857-2_3. URL: https://doi.org/10.1007/978-0-387-75857-2_3.

[47] Ipek N Sener, Ram M Pendyala, and Chandra R Bhat. "Accommodating spatial correlation across choice alternatives in discrete choice models: an application to modeling residential location choice behavior". In: *Journal of transport geography* 19.2 (2011), pp. 294–303.

[48] Kazuaki Miyamoto et al. "Discrete choice model with structuralized spatial effects for location analysis". In: *Transportation Research Record* 1898.1 (2004), pp. 183–190.

[49] Liya Yang, Guo Zheng, and Xiaoning Zhu. "Cross-nested logit model for the joint choice of residential location, travel mode, and departure time". In: *Habitat International* 38 (2013), pp. 157–166.

[50] Pavlos S Kanaroglou and Mark R Ferguson. "Discrete spatial choice models for aggregate destinations". In: *Journal of Regional Science* 36.2 (1996), pp. 271–290.

[51] Hakim Hammadou et al. "How to incorporate the spatial dimension in destination choice models: The case of Antwerp". In: *Transportation Planning and Technology* 31.2 (2008), pp. 153–181.

[52] M Elmorssy and TH Onur. "Modelling departure time, destination and travel mode choices by using generalized nested logit model: discretionary trips". In: *dimensions* 9 (2020), p. 10.

[53] Thomas A Knapp, Nancy E White, and David E Clark. "A nested logit approach to household mobility". In: *Journal of Regional Science* 41.1 (2001), pp. 1–22.

[54] Jose-Benito Perez-Lopez, Margarita Novales, and Alfonso Orro. "Spatially correlated nested logit model for spatial location choice". In: *Transportation Research Part B: Methodological* 161 (2022), pp. 1–12.

[55] Yanyan Chen, Zheng Zhang, and Tianwen Liang. "Assessing urban travel patterns: An analysis of traffic analysis zone-based mobility patterns". In: *Sustainability* 11.19 (2019), p. 5452.

[56] Lee-Fang Chow et al. "Development and evaluation of aggregate destination choice models for trip distribution in Florida". In: *Transportation research record* 1931.1 (2005), pp. 18–27.

[57] Eric J Miller. "Traffic analysis zone definition: Issues & guidance". In: *Travelling Modelling Group; University of Toronto: Toronto, ON, Canada* (2021), pp. 1–22.

[58] Travel Forecasting Resource. *Choice Models with Aggregate Choices*. https://tfresource.org/topics/Choice_Models_with_Aggregate_Choices.html [Accessed: (03-04-2025)]. 2025.

[59] Michel Bierlaire. *Model specification*. 2023. URL: https://biogeme.epfl.ch/sphinx/auto_examples/sampling/specification.html#sphx-glr-auto-examples-sampling-specification-py.

[60] Andrew Daly, Thijs Dekker, and Stephane Hess. "Dummy coding vs effects coding for categorical variables: Clarifications and extensions". In: *Journal of choice modelling* 21 (2016), pp. 36–41.

[61] Michel Bierlaire. *PythonBiogeme: a short introduction*. 2016. URL: https://transp-or.epfl.ch/pythonbiogeme/documentation/pythonfirstmodel/pythonfirstmodel.html.

[62] Miachel Bierlaire. *Biogeme*. https://github.com/michelbierlaire/biogeme. Version 3.12.14. 2024.

[63] Michel Bierlaire. *Testing 6.4 Likelihood ratio test*. 2017. URL: https://transp-or.epfl.ch/courses/dca2017/slides/v641.pdf.

[64] Patricia L Mokhtarian. "Discrete choice models ρ2: A reintroduction to an old friend". In: *Journal of choice modelling* 21 (2016), pp. 60–65.

[65] Michel Bierlaire. *6.4.2 Applications of the likelihood radio test*. 2018. URL: https://www.youtube.com/watch?v=ILn_QCe4TU0&t=404s (visited on 06/10/2025).

[66] Emma Frejinger, Michel Bierlaire, and Moshe Ben-Akiva. "Sampling of alternatives for route choice modeling". In: *Transportation Research Part B: Methodological* 43.10 (2009), pp. 984–994.

[67] Giancarlos Parady, David Ory, and Joan Walker. "The overreliance on statistical goodness-of-fit and under-reliance on model validation in discrete choice models: A review of validation practices in the transportation academic literature". In: *Journal of Choice Modelling* 38 (2021), p. 100257.

[68] Michel Bierlaire. *A short introduction to Biogeme*. 2023. URL: https://transp-or.epfl.ch/documents/technicalReports/Bier23.pdf.

[69] Deepankar Singh. *From Logits to Probabilities: Understanding Softmax in Neural Networks*. 2024. URL: https://medium.com/ai-enthusiast/from-logits-to-probabilities-understanding-softmax-in-neural-networks-3ebea2e95cfe (visited on 06/10/2025).

[70] Suzan Folkerts and Garrelt Verhoeven. *Deventer Boekenstad: twaalf eeuwen boekcultuur aan de IJssel*. Amsterdam University Press, 2020.

[71] Hendrik DL Vervliet. "Deventer, Richard Pafraet, 1509". In: *Post-Incunabula en Hun Uitgevers in de Lage Landen/Post-Incunabula and Their Publishers in the Low Countries: Een bloemlezing gebaseerd op Wouter Nijhoffs LArt typographique uitgegeven ter gelegenheid van het 125-jarig bestaan van Martinus Nijhoff op 1 januari 1978/A selection based on Wouter Nijhoffs LArt typographique published in commemoration of the 125th anniversary of Martinus Nijhoff on January 1, 1978*. Springer, 1978, pp. 118–119.

[72] Uknown. *Geschiedenis van Deventer*. 2025. URL: https://nl.wikipedia.org/wiki/Geschiedenis_van_Deventer.

[73] Ann Marynissen. "De lange weg naar een Nederlandse standaardtaal. Een beknopte geschiedenis van de standaardisering van het Nederlands". In: Acco, 2017. Chap. 3.

[74] A de Bruycker and D van Netten. "'Zodat mijn verbanning tegelijk jouw straf is.'Bloei, verval en migratie van wetenschap in de Republiek en de Spaanse Nederlanden". In: *BMGN-Low Countries Historical Review* 123.1 (2008), pp. 3–30.

[75] Gustaaf Asaert. *1585: De val van Antwerpen en de uittocht van Vlamingen en Brabanders*. Vol. 2. Lannoo Uitgeverij, 2004.

[76] Karel Porteman and Mieke B Smits-Veldt. "Een nieuw vaderland voor de muzen: geschiedenis van de Nederlandse literatuur 1560-1700". In: *(No Title)* (2009).

[77] Johan Decavele. "De eerste protestanten in de Lage Landen". In: *Geloof en heldenmoed* (2004).

[78] Jetze Touber. "Brieven uit Rome: de geleerdenreis van Hendrik Copes, 1695-'96". In: *ROMA AETERNA* 3.1-2 (2015), pp. 118–129.

[79] Christine Kooi. *Reformatie in de Lage Landen, 1500-1620*. Prometheus, 2023.

[80] René Boumans. "De getalsterkte van katholieken en protestanten te Antwerpen in 1585". In: *Revue belge de philologie et d'histoire* 30.3 (1952), pp. 741–798.

[81] Ludovicus Jacobus Rogier. *Geschiedenis van het Katholicisme in Noord-Nederland in de 16e en de 17e eeuw*. Vol. 2. Urbi et orbi Amsterdam, 1947.

[82] Rob van Roosbroeck. *Geschiedenis van Vlaanderen: deel IV, Vlaanderen in de 16de eeuw*. Vol. 4. Antwerpen: Standaard, 1939.

[83] Johannes Antonius De Kok. "Probleemstelling en premissen". In: *Nederland op de breuklijn Rome-Reformatie: numerieke aspecten van protestantisering en katholieke herleving in de noordelijke Nederlanden 1580-1880*. Radboud University, 1964, pp. 3–68.

[84] Paul Bairoch. *Cities and economic development: from the dawn of history to the present*. University of Chicago Press, 1988.

[85] Maarten Bosker, Eltjo Buringh, and Jan Luiten Van Zanden. "From baghdad to london: Unraveling urban development in europe, the middle east, and north africa, 800–1800". In: *Review of Economics and Statistics* 95.4 (2013), pp. 1418–1437.

[86] Tertius Chandler. *Four thousand years of urban growth: An historical census*. Edwin Mellen Press Ltd, 1987.

[87] Dana Klimeová and Eva Ocelíková. "Spatial-temporal modeling and visualization". In: *International Journal of Mathematical Models and Methods in Applied Sciences* 6.1 (2012), pp. 149–156.

13. Appendix

13.1. Research Data

Alongside this report, all the source code, data used for this thesis and created results is supplied, except for Chaney's author data as I am not allowed to share it. This is agreed upon with second supervisor: Sandra de Pleijt. Alongside the "Data" and "Results" folder, all subfolders contain the scripts used for this Thesis.

13.2. Estimation Report

Fold 1	Value	Std Error	T-Test	P-Value
beta_distance	-3.84	0.0281	-137	0
b_active_printers	1.2	0.00435	277	0
Null log likelihood	-126640.8			
Final log likelihood	-60568.43			
Likelihood Ratio Test	132144.8			
Rho-square	0.522			
Adjusted Rho-square	0.522			

Sample Size: 27940

Fold 2	Value	Std Error	T-Test	P-Value
beta_distance	-3.82	0.028	-136	0
b_active_printers	1.2	0.00434	277	0
Null log likelihood	-126640.8			
Final log likelihood	-60724.69			
Likelihood Ratio Test	131832.3			
Rho-square	0.52			
Adjusted Rho-square	0.52			

Sample Size: 27940

Fold 3	Value	Std Error	T-Test	P-Value
b_distance	-3.82	0.028	-136	0
b_active_printers	1.2	0.00433	277	0
Null log likelihood	-126640.8			
Final log likelihood	-60798.04			
Likelihood Ratio Test	131685.6			
Rho-square	0.52			
Adjusted Rho-square	0.52			

Sample Size: 27940

Fold 4	Value	Std Error	T-Test	P-Value
b_distance	-3.84	0.0281	-137	0
b_active_printers	1.2	0.00434	277	0
Null log likelihood	-126640.8			
Final log likelihood	-60586.71			
Likelihood Ratio Test	132108.2			
Rho-square	0.522			
Adjusted Rho-square	0.522			

Sample Size: 27940

Fold 5	Value	Str Error	T-Test	P-Value
b_distance	-3.85	0.0281	-137	0
b_active_printers	1.2	0.00435	276	0
Null log likelihood	-126640.8			
Final log likelihood	-60620.06			
Likelihood Ratio Test	132041.5			
Rho-square	0.521			
Adjusted Rho-square	0.521			

Sample Size: 27940

Table 3. In-Sample parameter estimation reports of individual folds using the full observations set