Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Diego Israel Rojas Baez	Date

Essays in Network Econometrics and Data Science: Theory and Empirics

By

Diego Rojas-Baez

Doctor of Philosophy

Economics

Kim P. Huynh, Ph.D.
Advisor

David T. Jacho-Chavez. Ph.D.
Advisor

Christoph Breunig, Ph.D.
Committee Member

Elena Pesavento, Ph.D.
Committee Member

Accepted:

Kimberly Jacob Arriola, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Essays in Network Econometrics and Data Science: Theory and Empirics

By

Diego Israel Rojas Baez

M.A., Emory University, 2020

Advisor: Kim P. Huynh, Ph.D.
Advisor: David T. Jacho-Chavez. Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Economics
2022

Abstract

Essays in Network Econometrics and Data Science: Theory and Empirics

By Diego Israel Rojas Baez

The increasing availability of social and economic interactions has sparked the academic effort to use models taking into account the effect that peers have on the decision of economic agents. One challenge to tackle on the way is how to approach the massive amount of data that firms and governments collect on economic interactions. Another challenge is how to use this large amount of information to model and better understand economic interactions with peer effects, Understanding that interactions between can have different natures. The first essays of this dissertation provide insights on how to use data science to approach using big data in administrative records. The third essay offers a way to model multilayer networks, networks with different types of connections, using the characteristics of the economic agents and a dyadic regression.

The first chapter describes a unique large administrative set of records from the Bank of Canada's Currency Inventory Management Strategy. The data from a single note inspection procedure generates a sample of 900 million banknotes. After defining the duration of the banknote circulation, several data science techniques are employed to describe the duration patterns across different banknotes: K–prototype clustering, Network graphs, and statistics, among others. A hazard model estimates the survival curve of banknote circulation cycles based on their clusters and characteristics.

The second chapter explores an intervention of The Bank of Canada in the year 2017. It introduced a change in the imagery of forty million CAD 10 banknotes. The chapter sets to explore whether slight differences in the appearance of a banknote can elicit a change in the behavior of economic agents enough to measure it. Specifically, we construct a measure of the duration of the circulation cycle of a note using proprietary data of the Bank of Canada.

The third chapter extends the primary tool to model network formation: the dyadic regression. This chapter proposes that dyadic regression provides a sound framework to model multilayer network formation under standard regularity conditions. This paper introduces an extended version of a gravity model that can accommodate a larger network class that allows for multiple layers of connections.

Essays in Network Econometrics and Data Science: Theory and Empirics

By

Diego Israel Rojas Baez

M.A., Emory University, 2020

Advisor: Kim P. Huynh, Ph.D.
Advisor: David T. Jacho-Chavez. Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Economics
2022

Acknowledgments

I want to thank David Jacho-Chavez and Kim P. Huynh, who have excelled in their jobs as chair and co-chair of my committee. The continued heartfelt feedback, supervision, and advice pushed me to be better, to show the world. It would be an understatement to say they used all the tools and resources available to help me keep going in the direction they thought best for me. I also want to show my appreciation for Jezamin Lim for being a source of wisdom and knowledge about navigating the labor market.

I want to thank Elena Pesavento for her wise advice during challenging times and never stop believing in me and my potential. Elena's words were always a source of solace and direction when the weight of the doctoral program seemed too heavy to carry. Chris Breunig, for his excellent job as committee member. I want to thank Daniel Levy, whose positivity and enthusiasm helped me believe in myself and never give up.

I want to thank the support provided by the Bank of Canada of the data and computer tools. This support was fundamental for developing the first and second chapters of this dissertation. I thank my co-advisor, Kim P. Huynh, for advocating and investing the requisite support for these projects. I want to acknowledge the efforts of the Bank of Canada staff, including Valerie Clermont, Ben Fung, Ted Garanzotis, Janet Hua Jiang, Mireille Lacroix, Andrew Marshall, Ramesh Paskarathas, Phil Riopelle, and Nathalie Swift, Julia Zhu and the Single Note Inspection data from the Currency Inventory Management Strategy and the use of the Bank of Canada's Digital Analytical Zone Microsoft Azure Cloud. The views expressed in this dissertation are those of the author. No responsibility for them should be attributed to the Bank of Canada.

I want to thank the organizers of the 2019 Advances in Econometrics workshop on the "Econometrics of Networks" co-hosted by the Research Centre of the Lucian Blaga University of Sibiu and the National Bank of Romania for the financial support to attend the event and present my work. Thanks to the Co-editors and organizers of the conference: Áureo de Paula (University College London), Elie Tamer (Harvard University), Marcel-Cristian Voia (Laboratoire d'Économie d'Orléans) for providing early feedback on my research. Special thanks to Marcel-Cristian Voia for his hospitality during my visit to Romania

In addition, I want to thank the Bank of Canada for providing the financial support to attend the 2019 Microeconometrics: Survey Methodology and Data Science conference organized by the Bank of Canada (BoC) in Ottawa from 26-27 September 2019. Thanks to Heng Chen and Kim P. Huynh for organizing an excellent conference that brought together a diverse set of researchers, including Harry Crane, who heavily influenced my understanding of network statistics and complex structures like multilayered networks.

I want to thank the administrative staff members of the Department of Economics who took great care of us. Stephanie White for her patience and readiness to lend a hand in solving administrative issues. Renee for her unending kindness and patience, helping us comply with the critical and minor requirements equally.

I want to thank the unconditional love and trust I received from my parents. Through their unrelenting efforts, struggle, and sacrifice, I have climbed a steep path against all the

odds. My mother patiently lending an ear to my unending rambles was a quintessential part of keeping my mind on track. I also want to thank Victoria, my wife, whose continued emotional support and help have sustained me through this endeavor.

I want to acknowledge that none of this would be possible if not for the wonderful members of my class cohort. I feel fortunate to have shared a classroom with such a group of brilliant and wholehearted human beings. They have been a continuous source of inspiration. Thanks to Juan Estrada for being an intellectual partner in crime, exploring the world of networks, pushing me to grow by simply being him, and providing a counterpoint to my overambitious imagination. Kaylyn Renee Sanbower for being my family's link to America's traditions and welcoming us into her traditions. Thanks to Cheng Ding for being a wonderful human being whose heart cannot be eclipsed by his enormous intellect. Thanks to Yisroel Cahn for being a wonderful friend.

Contents

1 Survival Analysis of Bank Note Circulation: Fitness, Network Structure, and Machine Learning **1**

1.1 Introduction . . . 2
1.2 Single Note Inspection Data . . . 4
1.2.1 Institutional Background . . . 4
1.2.2 Structure of the IMS dataset . . . 5
1.2.3 Sample description . . . 10
1.3 The Network and Spatial Patterns of Bank Notes . . . 15
1.3.1 The Cycle Duration and Bank Note Fitness . . . 15
1.3.2 Money Circulation Network . . . 19
1.4 Banknote Clusters . . . 25
1.5 Hazard Model for Bank Notes . . . 32
1.6 Conclusion . . . 40

2 On the Effect of Changing the Appearance of Money: Evidence from a Country-Wide Event **43**

2.1 Introduction . . . 44
2.2 Data . . . 46
2.2.1 The intervention . . . 46
2.2.2 Data . . . 49
2.3 Identification . . . 51
2.3.1 Decision frames and Money . . . 51
2.3.2 Identification strategy . . . 54
2.3.3 Estimating equation . . . 57
2.4 Results . . . 59
2.4.1 Duration patterns . . . 59
2.4.2 Estimates . . . 63
2.5 Conclusions . . . 66

3 Multilayer Dyadic Gravity Models **73**

3.1 Introduction . . . 74
3.2 Mulilayer Dyadic Data Model . . . 78
3.2.1 Multi-layer Gravity Model . . . 80
3.3 Estimation . . . 85
3.4 Numerical Experiments . . . 90
3.5 Empirical Application . . . 96
3.5.1 Data . . . 97
3.5.2 Results . . . 100
3.6 Conclusions . . . 102

List of Figures

1.1 Bank Note Life Representation 9
1.2 Sample Distribution over Time 12
1.3 Characteristics of the Matched Sample 14
1.4 Bank Note Duration vs. Fitness Status 15
1.5 Duration Distribution 18
1.6 Duration by Region and Fitness Status 19
1.7 Geographical exchange 21
1.8 Directed Graphs for Bank Note Circulation Between Regions by Denomination (In-degree) 22
1.9 Directed Graphs for Bank Note Circulation Between Regions by Denomination (Out-degree) 24
1.10 Distribution of Bank Notes over Clusters 28
1.11 Proportion of Bank Notes by Cluster Across Regions 32
1.12 Kaplan-Meier Estimates by Denomination 33

2.1 Comparison of CAD 10 notes after and before the intervention 48
2.2 Survival curves of treatment and control group 58
2.3 Imputed duration by year and denomination 60
2.4 Kaplan Meier - Survival probability estimates by denomination 61
2.5 Survival probability estimates by denomination and treatment period 63
2.6 Averate Treatment effect on the treated by parametric model 64
2.7 Cox-Snell residuals by parametric model 65
2.8 Comparison of CAD 10 notes after and before the intervention 67
2.9 Comparison of CAD 10 notes after and before the intervention 68
2.10 Comparison of CAD 10 notes after and before the intervention 72

3.1 Multilayer Network Examples 81
3.2 Q-Q plot simulation results for 3 layers 95
3.3 Q-Q plot, simulation results: Poisson Specification 114
3.4 Q-Q plot, simulation results: Negative Binomial $\phi = 1$ 114
3.5 Q-Q plot, simulation results: Negative Binomial $\phi = 5$ 114
3.6 Q-Q plot, simulation results: Negative Binomial $\phi = 10$ 114
3.7 Q-Q plot, simulation results: Negative Binomial $\phi = 10$ 115
3.8 Q-Q plot, simulation results: Lognormal $\sigma_{ij}^{lk^2} = 1$ 115
3.9 Q-Q plot, simulation results: Lognormal $\sigma_{ij}^{lk^2} = \lambda_{ij}^{lk^{-1}}$ 115
3.10 Q-Q plot, simulation results: Lognormal $\sigma_{ij}^{lk^2} = 1 + \lambda_{ij}^{lk^{-1}}$ 115
3.11 Q-Q plot, simulation results: Lognormal $\sigma_{ij}^{lk^2} = \lambda_{ij}^{lk^{-2}}$ 115

List of Tables

1.1 Fitness Dimensions Collected by Scanners 7
1.2 Salient Features of Cluster Assignment 30
1.3 Accelerated Failure Time: Bank Note Duration 40

2.1 Descriptive statistics of the sample 50
2.2 AFT regression: 4 specifications 71

3.1 Simulation Results for a 3 layer DGP 94
3.2 Summary Statistics 99
3.3 Multilayer gravity estimates for Trade, Tourism, and FDI 101
3.4 Simulation Results for a 3 layer DGP 111
3.5 Simulation Results for a 4 layer DGP 112
3.6 Multilayer gravity estimates for Trade, Tourism, and FDI (specification 2) . 116

Chapter 1

Survival Analysis of Bank Note Circulation: Fitness, Network Structure, and Machine Learning[†]

Diego Rojas, Juan Estrada, Kim P. Huynh, David T. Jacho-Chávez

Abstract

The efficient distribution of bank notes is a first-order responsibility of central banks. We study the distribution patterns of bank notes with an administrative dataset from the Bank of Canada's Currency Inventory Management Strategy. The single note inspection procedure generates a sample of 900 million bank notes in which we can trace the length of the stay of a bank note in the market. We define the duration of the bank note circulation cycle as beginning on the date the bank note is first shipped by the Bank of Canada to a financial institution and ending when it is returned to the Bank of Canada. In addition, we provide information regarding where the bank note is shipped and later received, as well as the physical fitness of the bank note upon return to the Bank of Canada's distribution centres. K–prototype clustering classifies bank notes into types. A hazard model estimates the duration of bank note circulation cycles based on their clusters and characteristics. An adaptive elastic net provides an algorithm for dimension reduction. It is found that while the distribution of the duration is affected by fitness measures, their effects are negligible when compared with the influence exerted by the clusters related to bank note denominations.

[†]We thank Áureo de Paula, Jean-Frédéric Demers, David Drukker, Ben Fung, Ted Garanzotis, Harry J. Paarsch, Ramesh Paskarathas, Marcel Voia, and participants of the Advances in Econometrics conference on the 'Econometrics of Networks' organized by the National Bank of Romania and the Faculty of Economic Sciences - Lucian Blaga University, Sibiu on May 16–17, 2019 for their various comments and suggestions on an earlier draft. We also acknowledge the efforts of Valerie Clermont, Ted Garanzotis, Mireille Lacroix, Andrew Marshall, Phil Riopelle, and Nathalie Swift and the use of the Bank of Canada's Digital Analytical Zone Microsoft Azure Cloud. We thank Meredith Fraser-Ohman for providing excellent editorial assistance. We thank Ruth Eilon from the Bank of Israel for providing valuable feedback. The views expressed in this paper are those of the authors. No responsibility for them should be attributed to the Bank of Canada.

1.1 Introduction

Underlying and sometimes unseen network structures shape the observed outcomes across different markets. Human behaviours like crime, substance abuse, and educational achievement, among others, are significantly affected by the social network spanned by each individual (see, e.g., Jackson (2008) and references therein). More complex structures developed by individuals, such as financial systems, are also governed by network interactions. Exploring the network characteristics that affect an individual's behaviour provides important insights on the role that networks play in determining overall outcomes.

One particular market where the network structure is the cause as well as the consequence of human social interaction is the market of money circulation in a country. In Hobbes' (1996) Leviathan, there is a comparison of the economic structure of a nation as a human body, and bank notes are like the blood cells in the human circulatory system. The flow of bank notes among agents is a necessary condition for economic transactions to occur; therefore, the flow of bank notes literally fuels the economic engine of a nation. Following the circulatory system analogy, it is important to understand how this fuel that is being distributed among agents arises. This task is undertaken here in a novel way, going beyond the bloodstream and studying the cells themselves through the lens of big data analytics.

Our research uses the Bank of Canada's Currency Inventory Management Strategy (IMS) study of bank notes. IMS uses high-speed scanners to collect detailed images from each bank note in the Bank's possession. This results in a unique dataset containing records that track the events throughout the circulation cycle of every bank note issued by the Bank

of Canada, as well as information about when the bank note was created, where it was shipped to, and when it was returned to the Bank. For this paper, we have a sample of over 900 million bank note unique scan records. The scan records span from August 2017 to July 2018. This dataset allows us to build networks at two levels: the region and the financial institution.[1] Specifically, for the latter, we examine the financial institution that receives the bank note from the Bank and the financial institution that returns it to the Bank.

We first present an exploratory analysis of this unique dataset addressing the bank note circulation patterns that can be identified. K-prototypes clustering is then implemented to classify bank notes together into clusters of bank note types. These clusters are then used as features when fitting a hazard model for the duration of circulation of bank notes in Canada. For this paper, the object of analysis is the duration of the first circulation cycle of the bank notes. The circulation cycle, or stay in the market, begins with the shipment of the note to the financial institution and ends when the note is returned to the Bank of Canada. An elastic net algorithm is then used to select the best accelerated failure time (AFT) model for circulation duration. We find that while the distribution of the duration is affected by the scanned fitness measures, their effects are negligible compared with the influence exerted by the clusters associated with bank note denominations.

The paper is organized as follows: Section 1.2 describes how the data have been collected, their structure, and contained information. Section 1.3 provides insights about the recovered network structure implied by the data, while Section 1.4 explains how machine learning community detection (clustering) algorithms are used here to uncover bank note types. Section 1.5 provides the results from fitting a hazard model, while Section 1.6 con-

[1] Due to confidentiality reasons, we do not name them.

cludes.

1.2 Single Note Inspection Data

1.2.1 Institutional Background

The Bank of Canada supplies financial institutions (FIs) with the bank notes they require to satisfy public demand through the country's Bank Note Distribution System (BNDS); see Bilkes (1997) for details.[2] At the centre of BNDS are the Bank of Canada's two cash processing centres, also known as Agency Operations Centres (AOCs), located in Montréal and Toronto.

The two AOCs are responsible for the distribution and secure storage of Canadian currency, mechanized note processing of financial institution deposits, and destruction of unfit note deposits. Under the distribution system agreement, these AOCs supply notes to FI Regional Distribution Centres (RDCs) located in 10 key cities across Canada, known as Regional Distribution Points (RDPs).

The inventory of bank notes is held at the RDCs and is distributed by the member FIs throughout their networks. In addition, the RDCs also supply bank notes to other FIs that are not part of the BNDS. The inventories are managed through an automated system used by the Bank of Canada and FIs, known as the Note Exchange System (NES), to facilitate the distribution, withdrawal and deposit of bank notes.

The NES maintains the record of inventories across all RDCs in Canada. The NES is

[2]We thank Jean-Frédéric Demers (Director of the BNDS) for sharing his knowledge and the internal document of the BNDS which forms the basis of Section 2.1 Institutional Background.

used to identify and reallocate surplus inventories between RDCs, receive and fulfill withdrawal orders from FIs, and send instructions to the cash-in-transit (CIT) companies. In addition, the NES provides a tool to forecast demand, and transmits the end-of-day balance to the Bank of Canada's general ledger.

The Bank of Canada maintains control over the flow and quality of bank notes by removing unfit notes from circulation and supplying new or fit notes as required. FIs ship their unfit notes to their RDCs, which are transferred by CIT companies and by air to the AOCs for processing. Paskarathas et al. (2017) provides a graphical representation of the BNDS system.

1.2.2 Structure of the IMS dataset

Recently, the Bank of Canada instituted single note inspection, which allows for a digital analysis of all bank notes in the IMS. The IMS is a unique dataset containing records that track the events throughout the duration of every note issued by the Bank. The collection of digital images for every note in the IMS has been part of the Bank of Canada's efforts to transition from a regularly implemented method of tracking bank notes based on sampling with human supervision to a more technologically advanced framework. Until late 2015, the Bank of Canada regularly sampled around 4 million notes from its inventory to try to assess the quality, or fitness, of batches of bank notes. The new system allows the Bank of Canada to track the movements of 22 million notes per month in a regular period, and up to 45 million notes per month in a peak period.

The new system uses high speed scanners to collect detailed images from each bank

note. This allows the Bank of Canada to observe 22 fitness dimensions for each scanned note. From these recorded characteristics, the Bank of Canada can calculate statistics that summarize the wear and tear that each bank note has endured. A description of each of these 22 dimensions can be found in Table 1.1. The work of Paskarathas and Balodis (2019) uses the same set of fitness features and principal component analysis to try to establish associations among those note characteristics. The fitness status of note i is determined as follows: for each dimension d_{ik}, where $k = 1, \ldots, 22$, we observe the indicator function $\phi_{ik} = \mathbf{1}\{d_{ik} \geq l_k\}$. If $\phi_{ik} = 1$ we say that the note i is considered unfit on dimension k. For example, if any note has a "front stain" level greater than 8, we consider the note unfit on the dimension "front stain." Then, a note is considered unfit if $\max\{\phi_{i,1}, \ldots, \phi_{i,22}\} = 1$.

Table 1.1: Fitness Dimensions Collected by Scanners

Category	**Parameter**	**Description**	**Threshold**
Holes	Hole	Sum of holes (sq mm)	3
Tears	Sum Open Tear	Sum of all open tears (sq mm)	4
	Max Open Tear	Longest open tear (mm)	2
	Sum Closed Tear	Sum of all closed tears (sq mm)	8
	Max Closed Tear	Longest open tear (mm)	8
Corner Faults	Missing Corner	Sum of all missing corners (sq mm)	15
	Folded Corner	Sum of all folded corners (sq mm)	70
Edge Faults	Missing Edge	Sum of missed edge (sq mm)	5
	Folded Edge	Sum of folded edge (sq mm)	5
Tape	Tape	Sum of tape area (sq mm)	100
Foil	Missing Foil	Missing foil area (sq mm)	40
	Sum Foil Scratch	Sum of foil scratch lengths (mm)	150
	Max Foil Scratch	Longest foil scratch length (mm)	20
Ink Wear	Ink Wear - Front	Ink wear front (levels 0-15)	8
	Ink Wear - Back	Ink wear back (levels 0-15)	8
Graffiti	Graffiti - Front	Graffiti level - total front (levels 0-15)	8
	Graffiti - Back	Graffiti level - total back (levels 0-15)	8
	Graffiti - Window	Graffiti level - clear window/foil (levels 0-15)	8
Stain	Stain - Front	Staining level front (levels 0-15)	8
	Stain - Back	Staining level back (levels 0-15)	8
Crease	Crease	Creasing/Crumpling level (levels 0-7)	5
Soil	Soil - Back	Soiling level back (levels 0-15)	8
	Soil - Front	Soiling level front (levels 0-15)	8

Note: Information is collected via high-speed scanners located in various Bank of Canada deposit centres once a bank note in circulation is deposited back in the Bank of Canada by a financial institution.

Additionally, the data contain time stamps for: i) the shipment of the bank note from a Bank of Canada deposit centre; and ii) the receipt and scan of the bank note once it leaves circulation and is deposited back with the Bank of Canada by an FI. The data also contain geographical information regarding the region of the deposit centre that shipped/scanned the bank note, and which FI required/deposited the note. These variables allow us to build a variable to describe what we call a "cycle," which is central to the analysis in this research. A cycle can be defined as the set of six elements:

1. Origin: the distribution centre from which each bank note was shipped.

2. Destination: the distribution centre where the bank note was deposited back.

3. Requesting FI.

4. Depositing FI.

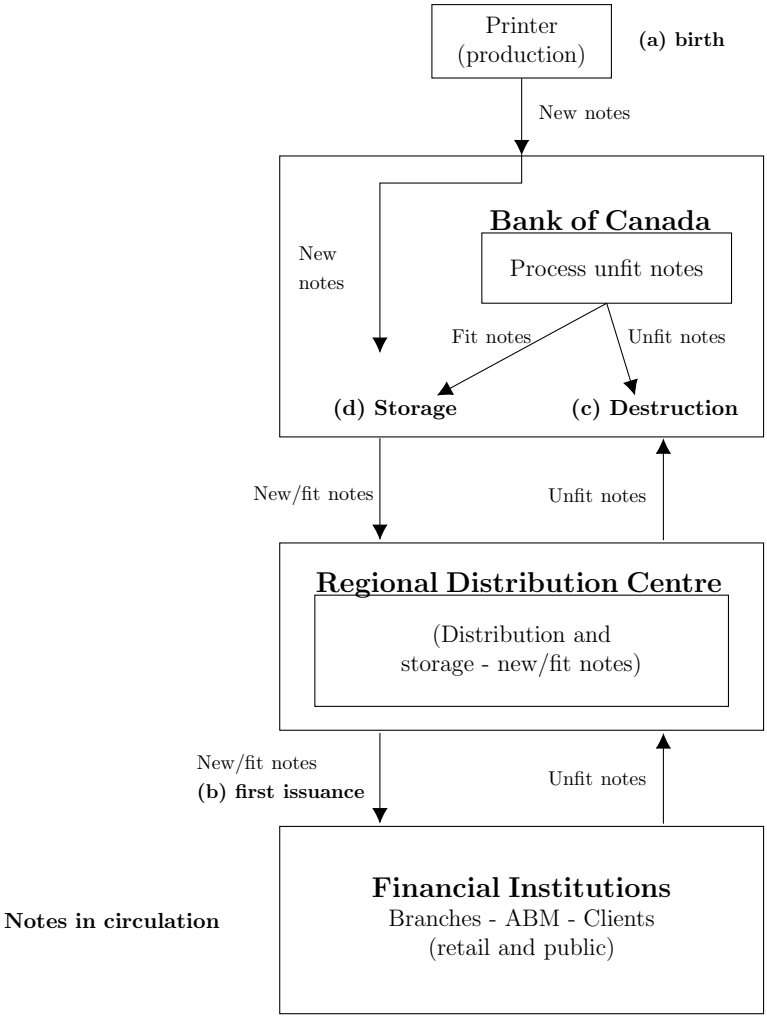
5. Ship time stamp.

6. Scan time stamp.

Each bank note can have zero or more cycles associated with it. For example, a bank note that was shipped but has not been deposited back lacks three elements of the cycle, hence it has zero cycles associated with it. A bank note can also exhibit more than one cycle if it was shipped to the market and deposited back more than once. A graphical description of the life of a note is depicted in Figure 1.1.

In the cash industry, the *lifetime* of a note is defined as beginning when a note is first produced, at the moment of printing, and ending the moment it is destroyed. An important part of a bank note's life cycle is the time it spends in circulation (circulation cycle in Figure 1, bottom rectangle). Therefore, our analysis focuses only on the first time a bank note is put into circulation and when it is received back from circulation, i.e., its first circulation cycle.

Figure 1.1: Bank Note Life Representation



Note: A typical "cycle" any bank note would pass through while in circulation. This figure is reproduced from Paskarathas et al. (2017). The cycle can be tracked through each bank note's unique serial number. The life of a note is represented by the events (a)-(d). This paper focuses only on the first cycle of circulation; in other words, only the first time the note is shipped to financial institutions and returned to the Bank of Canada. Source: Paskarathas et al. (2017).

The data in its pure form is a set of relational databases linked by a unique identifier: the bank note serial number. The construction of the cycle and its duration involves pairing records through the serial number on each of the tables on the relational dataset. After the matching process is performed, we retrieve the elements of a cycle associated with each bank note. Therefore, we can describe such cycles and build networks at three levels:

1. Region level: This is the less complex network where the edges are formed by the paths among 10 regions. These paths are constructed using data of the deposit centre region where the note was shipped, and the one where it was deposited back.

2. FI level: This has a medium level of complexity, and involves the financial institution that received the bank note from the Bank of Canada, and the financial institution that deposited back the bank note. For this case, a financial institution is assumed to be an independent unit in each of the regions that correspond to the deposit centre. For example, if Bank A has a presence in regions X and Y, then for the purpose of this network, this represents two nodes, Bank (A,Y) and Bank (A,X). This yields a network with 40 nodes.

3. Bank note as nodes: This level involves changing the unit observation to the bank note. In other words, each node in the network is given by a unique note. The edges are built around the paths they travelled, time frames, and other characteristics. This yields a network with around 6.7 million nodes.

Due to confidentiality requirements, only the region-level network is described in the next section. This paper analyzes solely the region level network. The analysis of more granular network structures is left for future research.

1.2.3 Sample description

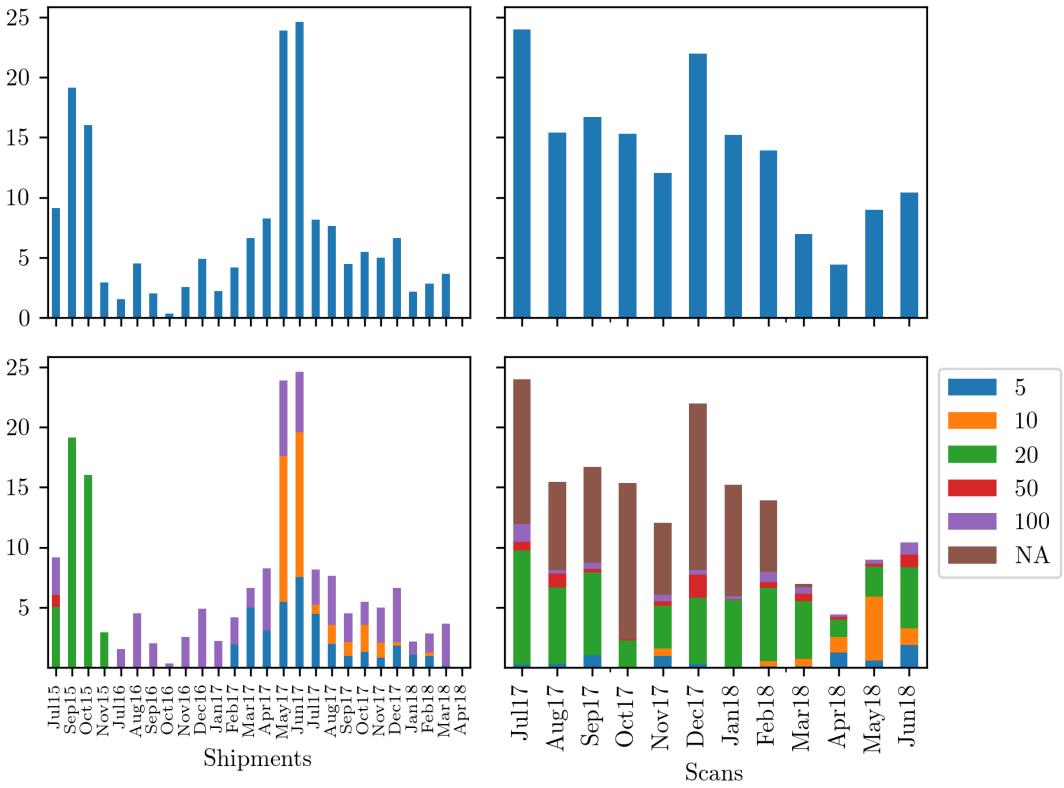
The sample obtained for this analysis includes several data tables that provide a wide variety of information on individual notes. The shipment and scan data tables contain the most relevant information. The first contains around 600 million records of bank note shipments

that range in time between July 2015 and April 2018; the second consists of nearly 300 million records of scanned notes corresponding to the period from August 2017 to July 2018. The first step in the analysis of this data is to eliminate possible duplicates in the data. We use the serial number of the note along with other characteristics to make sure we obtain unique records of notes up to a specific date.

The evolution of shipments and scans of the sample over time is depicted in Figure 1.2. The upper panels show only the evolution of shipments and deposits over time. The lower panels of Figure 1.2 show how scattered across time the shipments are by denomination in this sample. This is not surprising, since shipments are made in big lots. The deposits, which are the result of market factors, have a more consistent evolution; however, it is important to notice that the portion of notes whose denomination could not be identified is quite sizeable. In both samples, the 20-dollar notes have the greatest share. However, we cannot observe any shipment of 20-dollar notes in our sample after 2015. This is because the tracking of shipments of 20-dollar notes was not active after 2017.[3] We face a similar issue with the 50-dollar notes. Because the shipments that are tracked are in a different period than that of the other denominations in the sample, we decided not to include the 20-dollar and the 50-dollar notes in our sample.

[3]The tracking of shipments is still in a pilot phase; it is being rolled out for several denominations in different periods of time.

Figure 1.2: Sample Distribution over Time



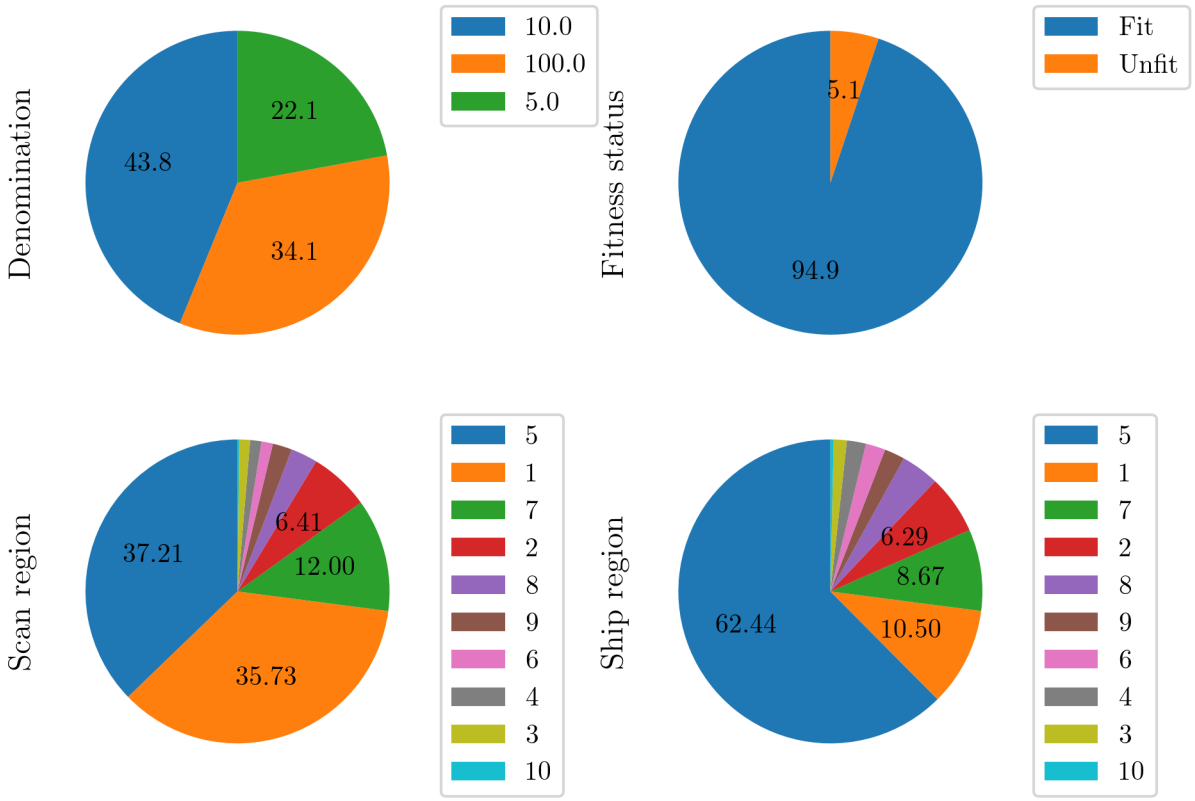
Note: The distribution of deposits and shipments behave differently over time. Both 20- and 50-dollar note shipments are more likely to occur before 2017 because the tracked shipments are in a different period.

In addition to the previous information, the corresponding characteristics of the scan data were also obtained. There are roughly 900 million records for the information including all denominations. However, after the data cleaning process, the sample reduces to a total of 179 million records of shipments and a total of 165 million notes deposited. The final sample allows us to match all the events in which a bank note is involved, through its unique serial number and other relevant characteristics. In turn, this provides the opportunity to build the geographical paths where each bank note travelled. After matching, the data consist of around 6.7 million observations. We then filter for those notes that have more than one cycle. This is the main difference from the analysis performed by Paskarathas et al. (2017),

where they allow for both right censoring and the possibility of multiple cycles. In contrast to this analysis, they focus on the total time it takes for a note to be considered unfit for circulation. Instead, we keep in the sample only those notes that have appeared once, at most. The particular group of notes that is left out is a small portion of data; around 98% of the notes are observed in the sample once, at most.

The 6.7 million matched notes provide a large amount of data that can be analyzed in several different ways. However, duration is the central feature of this research. Figure 1.3 shows the composition of the matched sample. First, the upper left panel shows the denomination prevalence. All three denominations are represented by a relevant number of notes in the final sample, and the most prevalent denomination is the CAN$10 note. This means that our analysis will have validity for any of those denominations. Next, on the upper right panel, we find the proportion of notes that have been deemed unfit by the scan process. The sample shows around 5% of the notes have been deemed unfit. While the ratio seems small, in practical terms this represents a sample of around 300,000 bank notes from which we can derive inference. The lower left panel shows the percentage of notes that were deposited back in each of the 10 regions. The largest shares correspond to regions 1, 5, and 7, which in turn corresponds to the regions that are able to draw more notes than the others. Since the Bank of Canada cannot control where these notes are deposited back, one reasonable explanation for this pattern is that it is a consequence of market transactions. Finally, the lower right panel shows the share of notes that were shipped to each region. The largest share of notes was sent to region 5. The notes are sent to the market in each region according to the FIs' requirements. Interestingly, observe that for the case of region 1, the share of notes sent to that region is larger than the portion of notes that it deposits back.

Figure 1.3: Characteristics of the Matched Sample



Note: Every denomination in the sample has sizeable share.

This result is a sign of movement of notes between regions that is not put into action by the Bank of Canada.

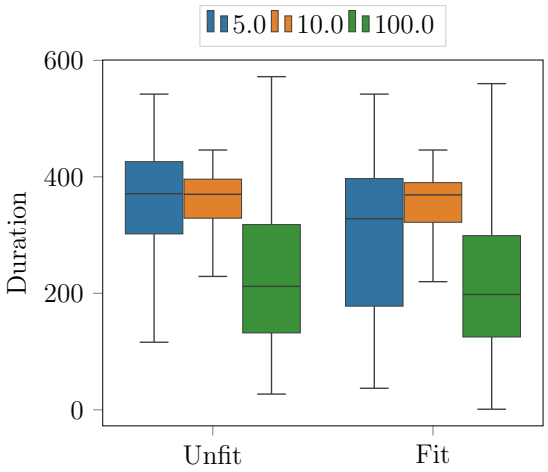
The elements shown in Figure 1.3 play key roles in the analysis regarding the geographical movements of the notes. The paths that the notes have travelled can be characterized by these four elements. This analysis constitutes a unique exercise to better understand the underpinnings of bank note circulation in the Canadian economy. Most importantly for this document, these elements will be crucial in understanding the determinants of the duration of circulation of a note in the market.

1.3 The Network and Spatial Patterns of Bank Notes

1.3.1 The Cycle Duration and Bank Note Fitness

The most notable feature when characterizing a bank note cycle is the time it takes to be completed, which we call the duration. Specifically, we focus on the time that passes between when a note is first shipped from the Bank of Canada and when it is first returned. We can analyze the distribution of the duration with respect to different groups of bank notes. For example, if we group bank notes by denomination, we can observe how their time in the market differs between groups. Additionally, we can group the bank notes by their fitness status. Figure 1.4 gives a sense of the distribution of duration when we consider these two dimensions.

Figure 1.4: Bank Note Duration vs. Fitness Status



Note: Graph is based on the sample of 6.7 million bank notes matched from every source, with at least one cycle. Duration is measured in days, while "fit" and "unfit" follow the Bank of Canada's official definitions based on the recorded 22 dimensions of fitness. The boxplot shows the lower and upper quartile values of the data with a line at the median, and the whiskers extend from the box to show the range of the data.

One can draw several implications from Figure 1.4. First, the median duration of a

banknote is longer for small denominations (5- and 10-dollar notes) compared with the large denomination (the 100-dollar note). This result is true regardless of the fitness status. Second, unfit bank notes, at the median, regardless of their denomination, show similar or larger duration values— e.g., the unfit 5-dollar notes have around the same median as the fit 5-dollar notes; the unfit 10-dollar notes have a larger median than the unfit 10-dollar notes; and the fit and unfit 100-dollar notes also have around the same median. Finally, the location and shape of the distribution of duration seem to follow denomination-specific patterns, i.e., changes in the location and shape distribution of duration are less noticeable across fitness status.

The first finding is related to the nature of the usage of bank notes. Lower denominations (CAN$5 and CAN$10) are involved in a higher number of everyday transactions. Therefore, it is not surprising that the CAN$5 bank notes have a shorter stay in the market than the CAN$10 bank notes. Since there is a high number of transactions, we can conjecture that the CAN$5 bank note should be more likely to be deemed unfit than a CAN$10 note at any point in time. A similar statement could be made when comparing CAN$5 with CAN$100, or CAN$10 with CAN$100. However, the CAN$100 bank notes' duration distribution behaves differently than expected.

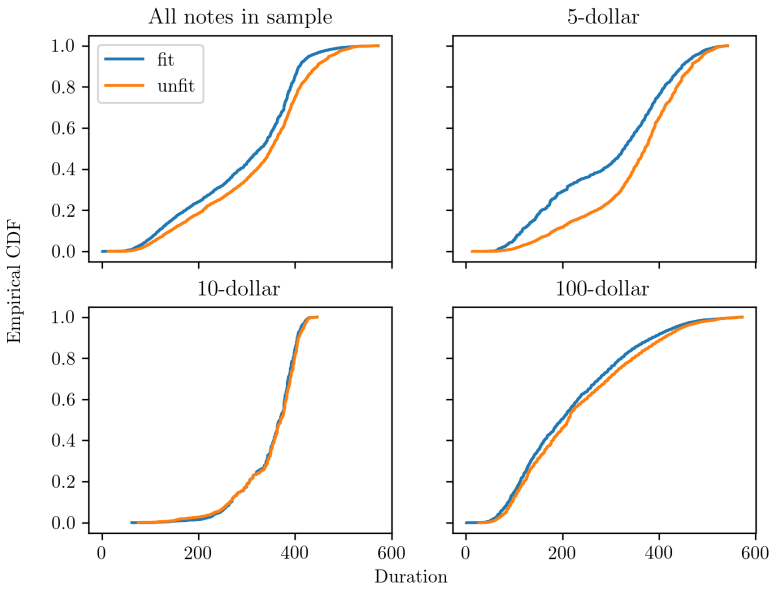
The CAN$100 bank notes show a duration distribution that is vastly more dispersed than that of other denominations. Also, the duration distribution of these notes is considerably shifted towards zero with respect to the other two high value denominations. The median duration of the CAN$100 bank notes is even lower than that of the CAN$5 bank notes. One possible explanation is the different logistics and management of CAN$100 bank notes. For example, CAN$100 bank notes are not available at automated teller machines

(ATMs); consumers would need to obtain these notes directly from the physical bank branch. This different management could end up accelerating the life cycle of the CAN$100 bank notes beyond the factors that explain the behaviour of the other denominations. Jiang and Shao (2019) pose a possible explanation for differential velocity based on retailer behaviour. Retailers, due to operational costs, are more likely to require lower denomination bank notes to make change for their customers. These retailers would keep lower denominations in situ, and use the large notes to make large deposits with financial institutions.

The second finding is more aligned with what one would expect. The bank notes in the sample deemed fit have, at the median, a shorter stay on the market than those deemed unfit. This result is true for the CAN$5, CAN$10, and CAN$100 bank notes. The pattern holds beyond the median. To demonstrate this more clearly, Figure 1.5 shows the cumulative distribution of duration for each denomination and by fitness status. Notice that fit note duration dominates the unfit for CAN$5 and CAN$100. In other words, for every percentile in the distribution of duration, the duration of fit notes is shorter than unfit notes. For CAN$10 the difference is difficult to observe on the graph. The results show that every percentile difference is significant, with different signs depending on the percentile, but close to zero. Hence, we cannot reject the hypothesis that distributions of the two samples are equal. This difference could be significant due to the large sample that induces very small standard errors in the estimations. It is worth mentioning that the CAN$100 duration shape is different than other denominations. It has a more concave distribution, which suggests a different behaviour in the market.

The last finding is related to the role that denomination plays in determining the length of stay in circulation. A suitable hypothesis is that the main driver of changes in duration is

Figure 1.5: Duration Distribution

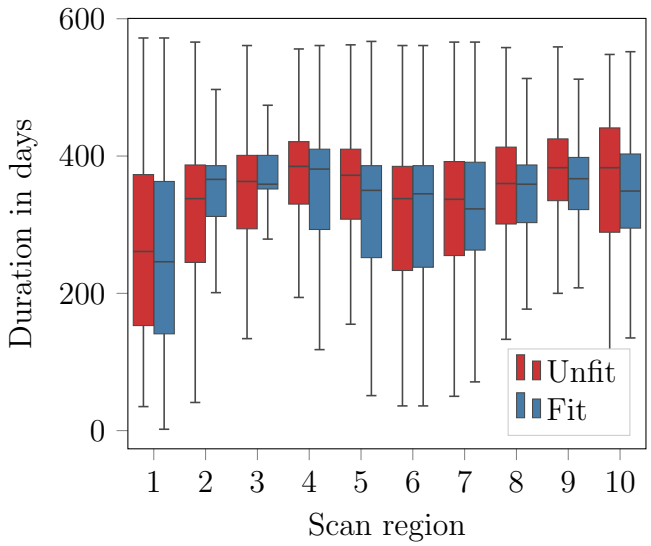


Note: Graph is based on the sample of 6.7 million bank notes matched from every source, with at least one cycle. Duration is measured in days, while "fit" and "unfit" follow the Bank of Canada's official definitions based on the recorded 22 dimensions of fitness.

the denomination of the bank note. While fitness does affect the duration distribution, these changes are relatively small when compared with the changes induced by denomination. For instance, Figure 1.4 clearly shows how every distribution across the fitness status changes only marginally, yet the change is dramatic across denominations. One possible explanation for this behaviour is related to the reasoning presented before for the first finding: the use of notes and the profile of the users change dramatically with denomination.

Finally, we take a geographic approach to this analysis and break down the duration by region and fitness status. The results are presented in Figure 1.6, which shows heterogeneous patterns of duration across regions. Region 1 shows the most distinct pattern: for this region the whole distribution is shifted towards zero and it has a higher dispersion in duration than any other region. This behaviour can be caused by a situation where the notes scanned in this region have a shorter stay in the market compared with every other region. Without

Figure 1.6: Duration by Region and Fitness Status



Note: Graph is based on the sample of 6.7 million bank notes matched from every source, with at least one cycle. Duration is measured in days while "fit" and "unfit" follow the Bank of Canada's official definitions based on the recorded 22 dimensions of fitness. The numbers from 1 to 10 in the horizontal axis are associated with the deposit centre region where the note was scanned.

disclosing the name of this particular region, we can assert that there is no compelling economic reason to believe that this region could have such a different behaviour than other nearby regions. The importance of understanding the paths travelled by the notes, hence the network they trace, becomes self-evident.

1.3.2 Money Circulation Network

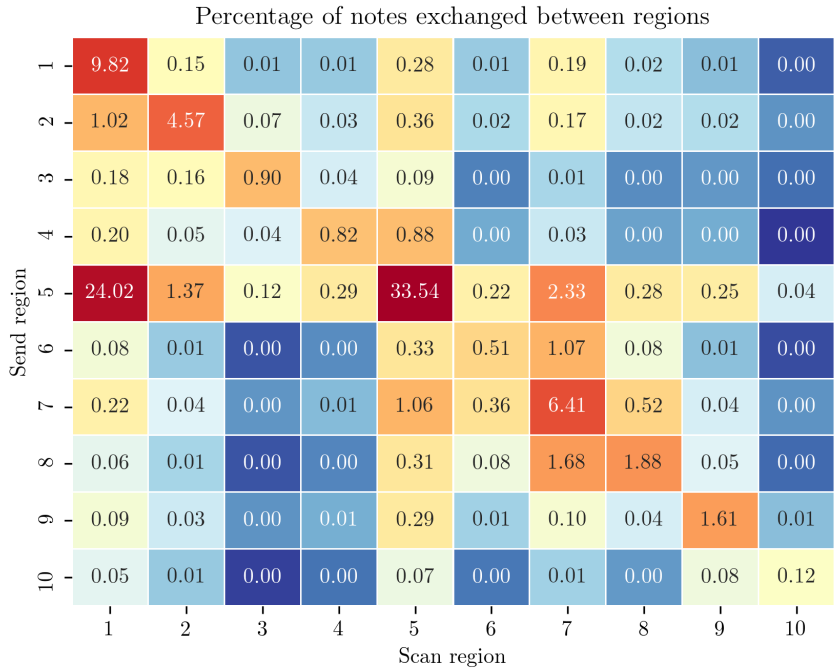
The previous sections have focused on the description of the velocity of bank notes' circulation, quality, and denomination heterogeneity. An additional analysis can be performed regarding the circulation patterns through the network formed by regions. As described in Section 1.3, the network spanned by different actors involved in the distribution of bank notes allows us to analyze the flow of notes on the edges of the network. This circulation network is a directed graph composed of non-reciprocal relationships. Furthermore, the quantity of

notes that are being sent and received across different regions may be totally asymmetric. It is important to mention that the adjacency matrix that serves as a base for the construction of these networks has a diagonal of non-zeros.

The first way to visualize this is through an actual adjacency/transition matrix. Figure 1.7 shows the share of notes that were sent between regions. The vertical axis lists each region i to which notes were sent. The horizontal axis lists each region j from which notes were redeposited and scanned. Then, each cell (i, j) represents the share of notes that were sent to region i and were redeposited at region j. It is natural to observe larger shares along the main diagonal of the matrix. In other words, notes are more likely to travel between regions that are closer to each other. Additionally, notice that the main diagonal concentrates around 60% of all note flows. The implication then is that the majority of the flows follow a geographic sensible pattern, i.e., the notes stay in the region where they were shipped originally. However, notice that a sizeable portion of notes, around 24%, are actually sent from region 5 bounded to region 1. This movement, which escapes the proximity logic, represents a physical movement where the notes travelled through the economy between two places that are geographically far apart.

A second way to approach the same problem is to represent the flow of notes as a network. We can map the flow matrix we observed in Figure 1.7 into a graph by letting the set of nodes be the set regions, and a set of hyperedges be the connections between regions. For example, a note that travelled from region i to j is going to be an element of the hyperedge (i, j). Then we can characterize each hyperedge by aggregating outcomes on each of the notes that belong in it. For the sake of tractability, we define weights for the edges as follows: the percentage of notes that one node is transferring to another as a proportion

Figure 1.7: Geographical exchange



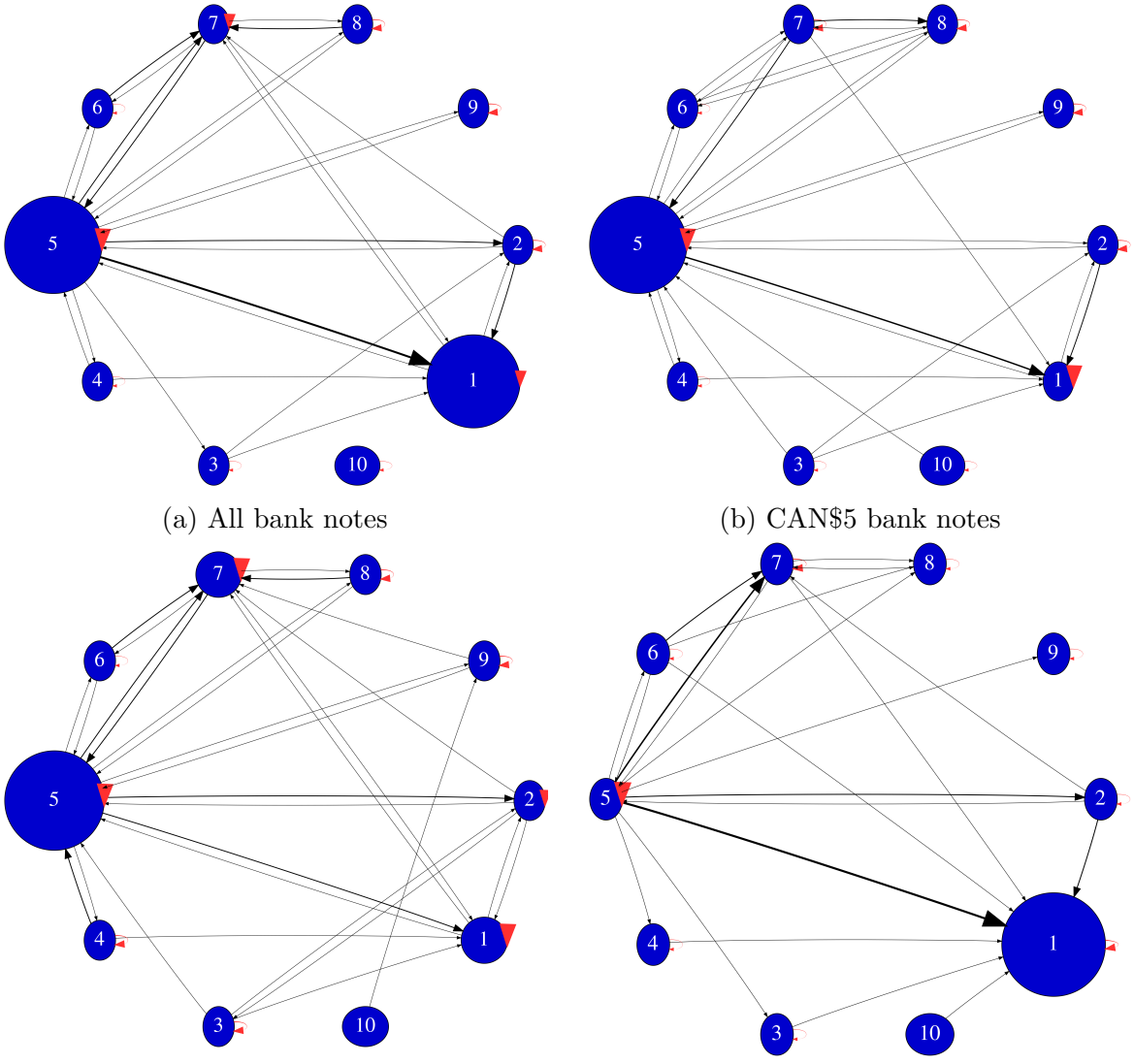
Note: Graph is based on the sample of 6.7 million bank notes matched from every source, with at least one cycle. Each cell represents the portion of notes sent between regions. The larger the share, the closer the colour resembles deep red. The lower the share, the closer the colour resembles deep blue.

of the total quantity of notes transferred in the network. Figures 1.8 and 1.9 depict the circulation patterns across the 10 distribution centres that are observed in the sample. It is important to remember that these flows do not imply the direct shipment of notes between RDCs. Instead, they imply movements of notes caused by market transactions—notes that were originally sent to node i are being collected back in node k. The number associated with each node gives the code representing each region.[4] The size of the nodes is a representation of the in-/out-degree statistic of the node. For both graphs, the darker edges represent a higher volume of transactions between nodes. In Figure 1.8, the bigger nodes are the regions receiving relatively more bank notes. Conversely, the bigger nodes in Figure 1.9 are the

[4]The computations for network structures and statistics were performed using the `networkx`, and the graph by `PyGraphviz` modules for Python 3.7.

regions from where relatively more bank notes are moving away. For instance, the fact that region 1 in Figure 1.8 is bigger for the CAN$100 bank notes means that large quantities of CAN$100 bank notes are being scanned in that region after being shipped to regions like 6 or 7. The principal destination of the notes can be inferred by the darkness of the edge.

Figure 1.8: Directed Graphs for Bank Note Circulation Between Regions by Denomination (In-degree)



(a) All bank notes

(b) CAN$5 bank notes

(c) CAN$10 bank notes

(d) CAN$100 bank notes

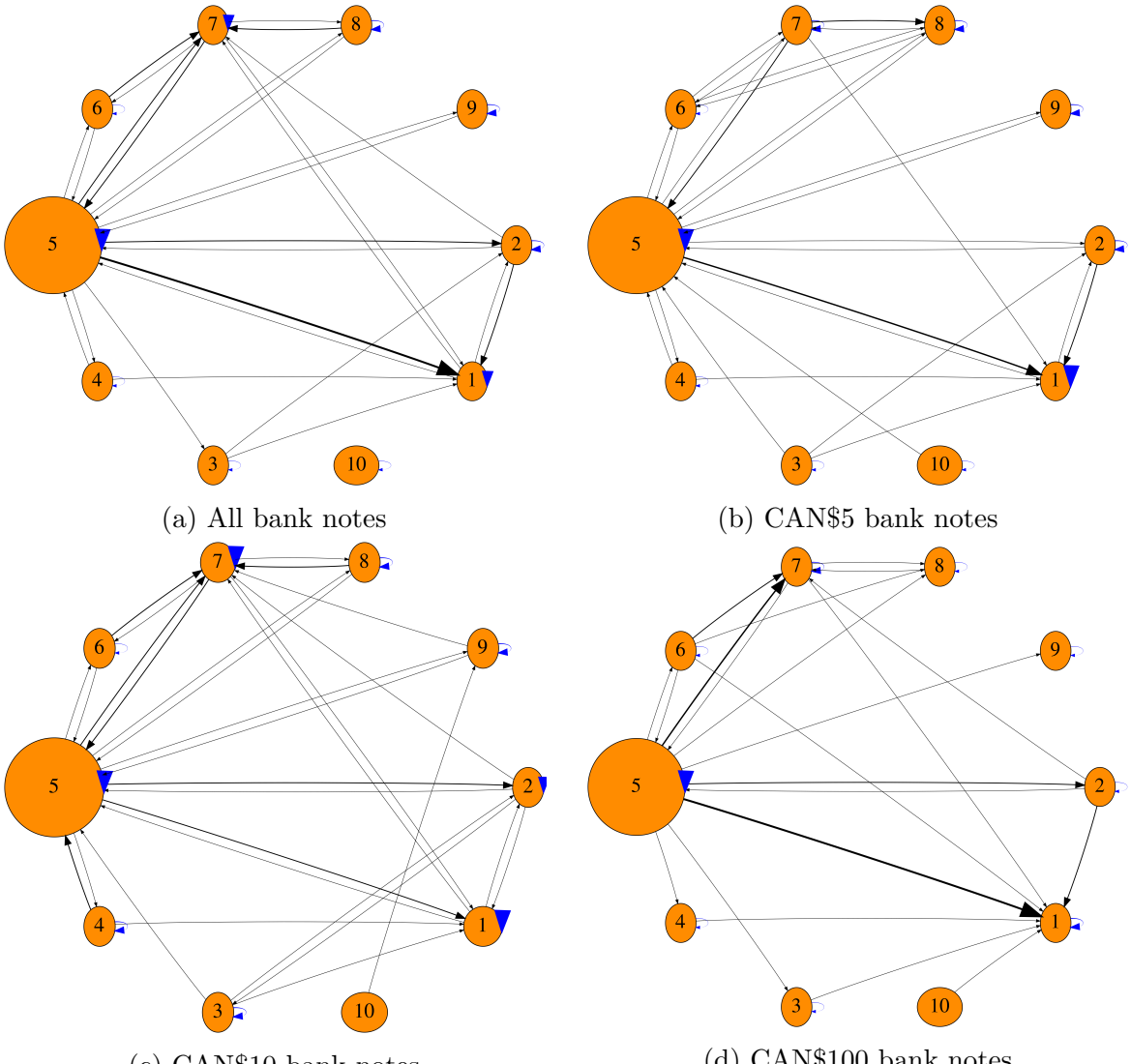
Note: The size of the nodes in each plot is given by a multiple of the "in-degree." The edges represent the share of all notes. They are coded in levels from thinnest to thickest following: less than 0.1% (transparent), 0.1% to 1%, 1% to 5%,5% to 10%, and greater than 10%.

From the two graphs, interesting patterns appear. First, when analyzing the behaviour across all notes, regions 1 and 5 are the recipients of most bank notes. The pattern does not change when we break down the analysis by denomination. Notice that in the graphs for CAN$5 and CAN$10, we observe the presence of more edges. Furthermore, these edges depict several numbers of bipartite relations between region 5 and the rest of nodes. This relationship is a clear indication that this region has a high centrality in the network representing the flow of these bank notes. However, this relationship is not the case when we analyze the CAN$100. The CAN$100 bank notes have a clearly different behaviour pattern in comparison to other denominations. Regions 1 and 7 are consistently receiving CAN$100 denomination bank notes from all other regions. Most importantly, notice that such flows do not lead to important flows in the reverse direction. Combining these findings with the information coming from the shares of notes staying and moving across regions in the matrix of shares in Figure 1.7, we can argue that a large share of notes, mostly CAN$100, travel in one direction, towards region 1, and stay there until deposited. This means that at least one-quarter of the notes travel a large geographical distance, with no particularly compelling economic reason, towards a single point that seems to only attract notes and not send them back. Intuitively, this would lead to the implication that there is a large number of economic transactions where region 1 absorbs CAN$100 notes from all regions, but very few where the opposite happens.

When we examine the out-degree flows, notice that the most important region across all bank notes is region 5. The pattern remains the same across denominations. This suggests that region 5 is a particularly important economic centre, because it circulates most of the bank notes in the economy. Interestingly, the role of region 1 diminishes in terms of out-

degree. Intuitively, this means that this node plays a small role when it comes to sending bank notes to other regions, specifically in the case of CAN$100. To explain these flows, the next section uses data-driven techniques in order to use the characteristics of each bank note on the hyperedges.

Figure 1.9: Directed Graphs for Bank Note Circulation Between Regions by Denomination (Out-degree)



(a) All bank notes

(b) CAN$5 bank notes

(c) CAN$10 bank notes

(d) CAN$100 bank notes

Note: The size of the nodes in each plot is given by a multiple of the "out-degree." The edges represent the share of all notes. They are coded in levels from thinnest to thickest following: less than 0.1% (transparent), 0.1% to 1%, 1% to 5%,5% to 10%, and greater than 10%.

1.4 Banknote Clusters

This section explores the characteristics of bank notes in more depth. To do this, we impose some structure on a highly dimensional dataset, and in doing so we uncover patterns that may relate to the network flows we documented in Section 1.3. The network we observed consists of nodes, regions, edges, and flows of bank notes. Hence characterizing the bank notes is equivalent to characterizing the flows of bank notes among regions. By classifying bank notes by their characteristics, we are intrinsically creating heuristic relations among them. These relations are expressed as the condition of belonging to a particular set of bank notes called a *cluster*. The main purpose of clustering bank notes is to explore the configuration of each cluster regarding its bank note characteristics. This empirical exercise allows us to gain insights into the way bank notes are performing in the economy, and provides crucial control information for the estimation of conditional hazard models in the next section.

We can sum up clustering in its most basic conception as minimizing the distance among nodes inside the cluster while maximizing the distance between clusters. The most popular clustering method is "K-means," (see, i.e., Jain (2010) and reference therein), which was discovered independently in several fields (Steinhaus (1956); Lloyd (1982); Ball and Hall (1965); MacQueen (1967)). In its simplest form, this algorithm classifies observations using the distance of each observation to some K-means on the feature space. The algorithm first picks K observations from the data, and labels them as *centroids*. Then, one calculates the distance of every point in the dataset to each centroid. For each centroid there is a cluster;

each observation is assigned to the closest centroid. From this classification we can obtain a mean squared error (MSE). Next, a new centroid is computed as the mean of the just-constructed clusters. The clustering repeats itself until the MSE converges. For the specific dataset we are using, we cannot directly use such a method. This is because a large number of variables in our dataset are categorical variables. The K-means method has only been proven to converge locally in the context of continuous variables. To take this possibility into account, a modification to the K-means algorithm is needed. There are several possibilities to tackle this task; see Szepannek (2018) and references therein. However, we use the strain of methods that follow the *K-prototypes* algorithm in Huang (1998), particularly an `R` implementation by Szepannek (2016). For completeness and following the description in Szepannek (2018), we now summarize the method. The dataset consists of n observations and a set of p random variables, x_i, $i = 1, \ldots, n$. To start the algorithm we assume there exist k prototype observations μ_j, $j = 1, 2, \ldots, k$. We can construct $c_{ij} \in C_{n\times k}$ a binary partition matrix where $\sum_{j=1}^{k} c_{ij} = 1$. Then we can choose the partition matrix that minimizes the objective function:

$$\min_{C_{n\times k}} \sum_{i=1}^{n} \sum_{j=1}^{k} c_{ij} \cdot d\left(x_i, \mu_j\right), \tag{1.1}$$

where the distance function $d(\cdot)$ is defined as:

$$d\left(x_i, \mu_j\right) = \sum_{m=1}^{q} (x_i^m - \mu_j^m)^2 + \lambda \sum_{m=q+1}^{p} \mathbb{I}(x_i^m \neq \mu_j^m). \tag{1.2}$$

The first term concerns the first q continuous variables with the square of the Euclidean distance. The second term concerns the $p - q$ categorical variables. The latter, the count

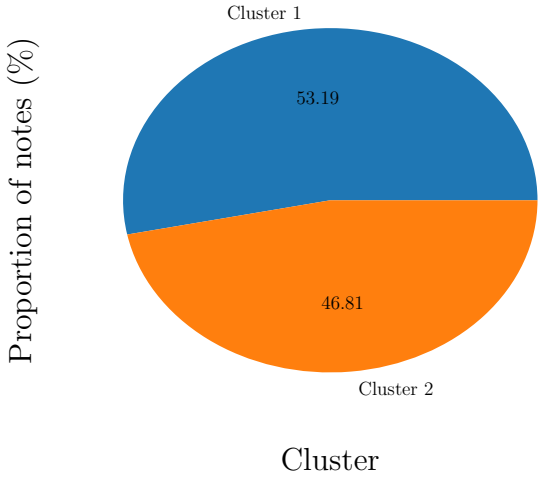
of mismatches, is penalized by a parameter λ which assigns less or more importance to the categorical data. In our implementation, the parameter is chosen "optimally" based on the data according to the guidelines in Huang (1998) and Szepannek (2018).

The clustering would be indicative that notes have grouped, almost naturally, into well-defined groups according to their characteristics. Therefore, the choice of number of clusters becomes very important. In other words, we need a criterion to specify the number of clusters used in the algorithm. Following common practices in the literature, we performed the procedure using several numbers of clusters ranging from 1 to 15. We also computed several metrics of precision of the method, e.g., C-index, Gamma index, Dunn index, among others; see Charrad et al. (2014). Along with these criteria we also used the number of clusters that explained the most significant changes in the sum of squared distances. This method is known as the *elbow test*.

After computing the clustering scenarios, the tests coincide in two clusters as good choices for K. To perform K-means clustering, we have used our sample of matched notes. The matrix $\mathcal{X}$ considers 60 features in total. The matrix has been first standardized in order to improve the performance of the K-prototypes algorithm. The features include: fitness measures (see Table 1.1), origin of the bank note, and destination of the bank note, among others. The results for the clustering exercise are shown in Figure 1.10. The clusters are somewhat evenly sized, with 53.1% and 46.8% of the notes on clusters 1 and 2, respectively. This classification allows us to benefit from the rich structure of the data to form groups, and after classification it also allows us to analyze a selection of features in each cluster. In this sense, we have employed this technique as a sort of dimensional reduction technique.[5]

[5]The procedure was implemented using the module `clustMixType` in `R`.

Figure 1.10: Distribution of Bank Notes over Clusters



Note: This graph is the result of applying the K-means algorithm to the sample of 15 million matched bank notes that have completed one cycle at most. The clustering is done over the space of 60 features of the bank notes, including fitness measures, geographical, and spatial information.

The high dimensionality of the data poses a challenge to try to characterize each cluster in a meaningful way. Our approach is to select the features that are most relevant to the purpose of this analysis. The selected features to analyze in each cluster are:

1. **Duration**: period of time, in days, between the moment the note is shipped from the Bank of Canada and the moment it is deposited back into the Bank of Canada.

2. **Scanned region**: share of bank notes that were scanned in each destination. As previously discussed, the destination of the bank notes is one of the traits that has shown more interesting results in Sections 1.2 and 1.3.

3. **Denomination**: specifically, we examine the proportion of bank notes by denomination within each cluster, as the denomination structure is a main predictor of bank note duration.

4. **Year of scanning**: proportion of bank notes scanned in 2017 or 2018. These are the two years for which deposit information is available.

5. **Year of shipment**: proportion of bank notes shipped in 2017 or 2018. These are the two years for which deposit information is available.

6. **Fitness status**: proportion of bank notes deemed fit by the IMS data study after being deposited.

The features we have selected respond to two criteria. They must be conceptually compelling and have shown relevancy in the description of the data. The description of each cluster, based on the previously mentioned features, is shown in Table 1.2. One observes that the mean duration is very different for each cluster; coincidentally, the mean duration decreases with the label of the cluster. Dispersion of the duration measurement is lower within the first cluster. An additional interesting trait picked up by the cluster classification is the diversity of the scan location of the bank notes. There is a region that is more prevalent in each cluster. In the case of cluster 1, almost half of the notes were scanned in region 5. In cluster 2, almost 60% of the notes were scanned in region 1. In terms of denomination, the structure is different. Cluster 1 has a large proportion, around 80%, of CAN$10 notes. Meanwhile, cluster 2 has around 70% of CAN$100 notes. The remainders, in both clusters, are composed of CAN$5 notes. Regarding the scan year, cluster 1 is composed almost exclusively of bank notes scanned in 2018. Cluster 2 has a larger portion, around 30%, of notes that were scanned in 2017. The shipping of notes in the sample was almost exclusively done in 2017. Regarding the fitness status of the bank notes, we can see that cluster 1 has a larger prevalence of fit notes, around 97%, compared with 92% in cluster 2.

Table 1.2: Salient Features of Cluster Assignment

		Cluster	
		1	2
Duration	Mean	337.42	248.68
	Std. Deviation	78.03	125.31
Scan region	1	0.1550	0.5872
	2	0.0884	0.0365
	3	0.0171	0.0052
	4	0.0125	0.0114
	5	0.4931	0.2345
	6	0.0136	0.0106
	7	0.1596	0.0751
	8	0.0334	0.0228
	9	0.0260	0.0144
	10	0.0014	0.0022
Denomination	5.0	0.1482	0.3040
	10.0	0.8119	0.0132
	100.0	0.0398	0.6828
Scan year	2017	0.0648	0.2934
	2018	0.9352	0.7066
Shipment year	2017	0.9957	0.9424
	2018	0.0043	0.0576
Fitness status	Unfit	0.0295	0.0761
	Fit	0.9705	0.9239

Note: This table shows selected features of the clusters, obtained by performing K-means clustering on the 15 million matched bank notes in the sample. The patterns shown in this table allow us to construct profiles for the notes that belong to each cluster. The most salient traits are that the duration dramatically changes along with the denomination; also, the denomination composition almost exclusively segregates bank notes of one denomination to a specific cluster.

Table 1.2 shows the following profiles:

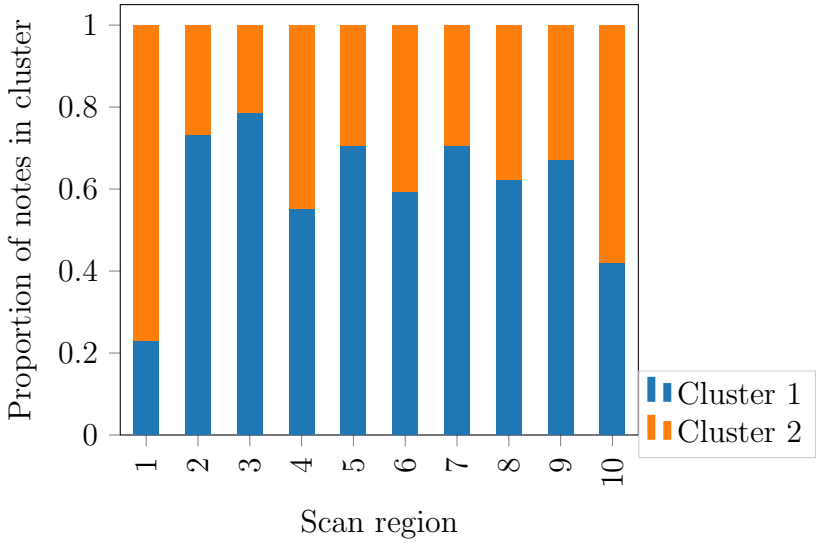
- **Cluster 1 -**[$10 - **slower**]: longest stay in the market, relatively lower dispersion; primarily scanned in region 5 in 2018. Composed of a majority of CAN$10 bank notes. Shipped almost entirely in 2018. With a 97% share of fit banknotes.

- **Cluster 2-**[$100 - $ **faster**]: shorter stay in the market, relatively larger dispersion; primarily scanned in region 1. Composed of CAN$100 bank notes. Shipped almost entirely in 2018. With a 92% share of fit bank notes.

Notice that the network we are trying to characterize has, until now, a geographical structure. Each deposit centre or region is considered a node. Hence we can characterize each node according to the notes it receives. Also, consider that in the link between our network and the bank notes, up to this point, the bank notes can be interpreted as elements in the hyperedges between regions. We can think of the deposit centres and the flow of notes as macro features of the network, while the bank notes themselves are micro features that characterize the macro measures. In this sense, and regarding the patterns observed in Section 1.3, it is important to analyze whether the classification of the bank notes into clusters is uniform across regions, i.e., across nodes. This could inform us about the cluster label and characteristics of bank notes that are more prevalent in regions that have relatively higher/lower flow of bank notes in/out. We break down the bank notes according to where they were scanned and obtain the proportion of notes per cluster within each region. Results are shown in Figure 1.11.

In terms of bank note flow, one of the most interesting patterns mentioned before is the strong connectivity between regions 1 and 5. The in-degree graph in Section 1.3 showed us a pattern of bank note flow that is not clearly explained by economic activity between these regions. In general, while heterogeneity does exist, it is possible to distinguish two groups of regions: regions 2 through 9 are predominantly composed of notes in cluster 1, while the notes in regions 1 and 10 mostly belong to cluster 2. This finding, along with the

Figure 1.11: Proportion of Bank Notes by Cluster Across Regions



Note: This graph breaks down the total number of matched bank notes by cluster and region where they were scanned. The heterogeneity in the proportions for each region shows how the flow of bank notes among regions differs greatly in terms of the types of transactions among regions and within each region.

network description, allow us to confirm that region 1 is something of an economic puzzle itself. Taking into account the profile we built for cluster 2, this analysis shows that even from a data-driven analysis, the particular behaviour of region 1—which mainly receives CAN$100—is not similar to any other region's behaviour. These insights come from a quick overview of the heterogeneity of the distributions of the clusters.

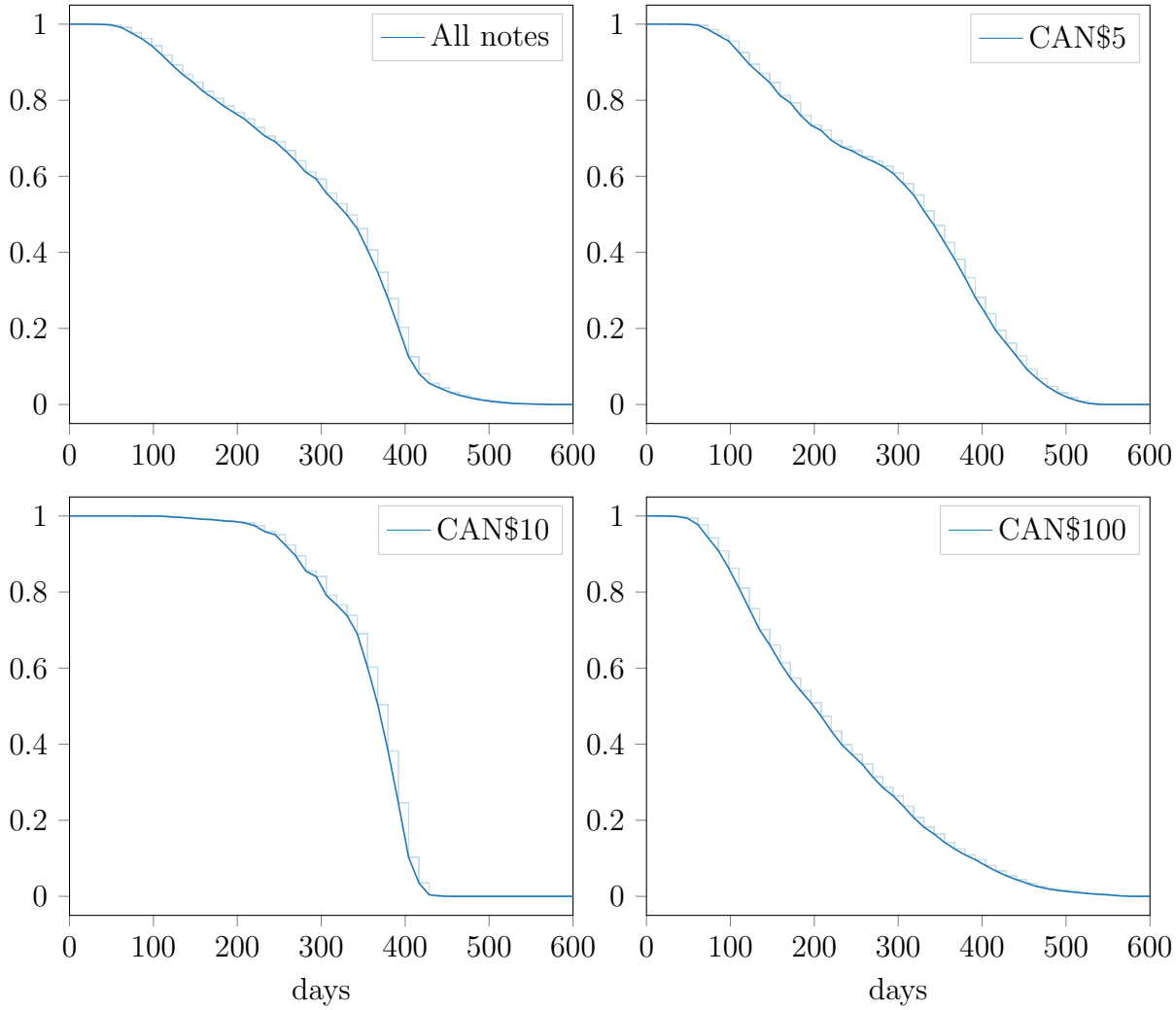
1.5 Hazard Model for Bank Notes

In this section, we analyze what factors influence a bank note's duration in circulation. The analysis in Section 1.3 showed how changes in the duration of stay in the market are correlated to the denomination of the bank note. [6] We start by plotting the survival curves of a bank note cycle.[7] In this context, the duration of a cycle of a bank note starts to be

[6] Paskarathas et al. (2017) conducted a study of the life cycles of polymer versus cotton substrate.
[7] The computations were performed using the `lifelines` module in Python 3.7.

measured at the time of shipment and ends at the time of scan. We are only concerned with the bank notes in the sample, hence we do not take into account any right or left censoring.[8] Figure 1.12 displays the Kaplan-Meier (KM) estimator of the survival curves for each denomination.

Figure 1.12: Kaplan-Meier Estimates by Denomination



Note: This graph shows the Kaplan-Meier estimator of the survival curve for the duration of the bank notes' stay in the market. The estimate is based on the individual duration of each bank note. Each panel represents a subset of the sample according to the denomination of the bank note. The pattern that emerges is that the survival curve differs greatly, depending on the denomination. The notes with a faster rate of failure in their curves are the CAN$10 bank notes. The CAN$5 and the CAN$100 bank notes show a slope that is less steep.

[8]Since we are analyzing only one-cycle-complete spell durations, we can ignore left and right censoring as per our sample description in Section 1.2.

The results are consistent with the findings in Section 1.2. The duration in circulation of the CAN$5 notes starts to decay after a period of around 100 days. Their failure rate steadily increases from that point onward. A similar case can be built around the KM estimate for CAN$100 notes. However, notice that in this case the curvature reverses, meaning the notes are deposited back faster in comparison to the CAN$5 bank notes. Interestingly, the most noticeable behaviour is that of the CAN$10 notes. These notes start to be deposited back after 200 days in circulation, twice as long as the CAN$5. This result is unexpected, since these two notes should have some degree of substitutability in the market. Another interesting feature is the very pronounced slope of the survival curve in comparison to the other denominations. We observe that the CAN$10 survival rate goes from one to zero in a tight period between 200 and 400 days after the original shipping date.

We can analyze these observed features in the framework of the AFT models, see, e.g., Kalbfleisch and Prentice (2002). This class of models allows for the duration to be "accelerated" by observable features of the units. In this case, we can observe how the duration changes with respect to the features of the note. Since we are assuming neither left nor right censoring, the specification can be estimated through the implementation of the ordinary least squares (OLS) estimator. This in contrast to other work such as Paskarathas et al. (2017), who observe right censoring and have to experiment with different model specifications and distributional assumptions. Our specification for this model is then:

$$\ln(t_i) = x_i'\beta + \varepsilon_i = f_i'\psi + k_i'\gamma + z_i'\phi + \varepsilon_i, \tag{1.3}$$

where t_i represents the duration that bank note i stays in circulation, f_i is a vector containing

i's recorded features, k_i represents a set of dummies from the cluster classification, z_i is a set of interactions of origin, destination, timestamps, and the features included in f_i, and ε_i represents the standard regression error. In total, the number of regressors in this model is 360.

There is an issue of near multicollinearity given the sparsity of the data and the high number of categorical variables included as features. This problem tends to inflate the coefficients of the model and shrinks its standard errors. To avoid these problems, we implement statistical learning regularization methods. These methods impose a penalization on the model's objective function so that the most relevant coeffcients can be selected. Popular methods within this class include ridge regression and lasso. The elastic net proposed by Zou and Zhang (2009) combined with double machine learning by Chernozhukov et al. (2018) is used here instead. This provides an automated model selection mechanism as well as automatic inference in the resulting model.

The elastic net method, proposed by Zou and Hastie (2005),[9] can perform better than lasso, or ridge, because it allows for grouped selection. For example, the lasso tends to select one variable from a group and ignore the others that are highly correlated with it. In contrast, elastic net regularization encourages grouping effects while stabilizing the L_1-regularization path. This approach tends to numerically break under conditions of near-multicollinearity. This is the case of our data. We then use a two-step procedure (see, e.g., Zou (2006)) that provides a more numerically stable approach to the selection of parameters using our data, i.e., Zou and Zhang's (2009) version of the adaptive elastic net. For the case of a linear regression, following Zou and Zhang (2009), assuming the error is normally distributed with

[9]The model was implemented in the package `glmnet` in R.

variance σ^2, it solves the problem:

$$\widehat{\boldsymbol{\beta}} = (1+\lambda_2/n)\left\{\arg\min_{\beta}\|\mathbf{ln}(\mathbf{t_i}) - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda_1^*\sum_{j=1}^{p}\widehat{w}_j\left|\beta_j + \lambda_2\|\boldsymbol{\beta}\|_2^2\right|\right\}, \tag{1.4}$$

where $|\cdots|$ and $\|\cdot\|$ represent the L_1– and L_2–norms, respectively, and $\widehat{w}_j$ comes from a first-step implementation of the elastic net estimator. Specifically, the values $\widehat{w}_j$ are given by $\widehat{w}_j = (|\hat{\beta}_j^{\star}|)^{-\gamma}$. The coefficients $\widehat{\beta}_j^{\star}$ are the coefficients of the elastic net estimation process in the first step.

The tuning parameters $\lambda = (\lambda_1^*, \lambda_2)'$ are the only ones left to be chosen. Specifically, λ_1^* is a regularization parameter that measures the "strength" of the L-1 penalization term in the objective function. When $\lambda_1^* = 0$, estimating the model is equivalent to estimating a ridge regression. The other parameter, λ_2, determines the "strength" of the L-2 penalization term in the objective function. If we set $\alpha = \lambda_2/(\lambda_1^* + \lambda_2)$, one can see how close we are from a ridge or lasso regression, i.e., $\alpha = 1$ means that the resulting estimator is equivalent to using a lasso estimator; when $\alpha = 0$ the fitted model is equivalent to fitting a ridge regression model. The results use the λ that minimizes the MSE subject to $\alpha = 0.95$ in a 10-fold cross-validation on a grid of 100 λs.[10]

To obtain inference, we used a novel approach developed by Chernozhukov et al. (2018). The estimation commonly referred to as double machine learning uses Neyman–orthogonal moments/ scores in order to achieve consistent estimators of the linear model coefficients. This estimator grants asymptotic normality of the coefficients, hence it allows for the esti-

[10]The value of α was chosen based on the speed of convergence of the resulting MSE.

mation of confidence intervals for some parameters of interest. This method is more suitable in comparison to the traditional way to use regularization methods, which report confidence intervals using a post-selection linear regression. However, the cost of such a convenient feature is that we can only make these inferences on a subset of the coefficients. For these exercises, and in view of previous findings, the variables we are interested in are: fitness status, denomination, and cluster label. This allows us to discern between the influence of quality, denomination, and the characteristics associated with the profiles we mentioned in the previous section. The double machine learning algorithm implemented closely follows the algorithm in Chernozhukov et al. (2018) and was coded up using the `R` package `glmnet`. The implementation of the adaptive elastic net estimator uses 10 folds each. Additionally, an OLS is estimated along with a "naive elastic net." The last one corresponds to fitting an elastic net model to select relevant variables, and then using OLS on the selected set to estimate standard errors.

The result of the specification of the adaptive elastic net procedure is a list of four coefficients along with their standard errors. The naive elastic net fits a log-normal linear model that takes into account only those variables selected by the method. The OLS includes all variables which are not multicollinear. In the original model, the changes were modelled using a comparison case in order to avoid perfect multicollinearity that stems from the use of categorical variables and aliasing. The base case considered for this exercise was a CAN$5 bank note scanned in the month of January in the year 2017, shipped in the month of January of the year 2015, travelled from region 1 to region 1, belonged to cluster 1, was considered fit, and had no damage in any of the fitness measures.

The results of this exercise can be found in Table 1.3. Robust standard errors are used to

construct the z statistics and tests. All coefficients in the table are significant at a 99% level of confidence. The coefficients are transformed to represent the rate of acceleration using an exponential function. First, notice that the ADA-Enet implementation does show differences with respect to the other two estimation methods. However, these differences are very small. This means that after we implemented a procedure to avoid near-multicollinearity, both OLS and Naive Enet are close enough to the more robust results. In this particular case, the size of the data could play a fundamental role in making the problems of near-multicolinearity shrink. Since the three methods show a relatively similar size and sign, we interpret the coefficients for one of the implementations.

The coefficients shown in Table 1.3 follow the findings from previous sections. Take, for example, the coefficient on the CAN$100; it significantly decreases the duration with respect to the base case scenario, CAN$5 notes. The opposite happens with the CAN$10 coefficient; in this case the duration is positively affected, growing almost 2% with respect to the CAN$5. These results differ from previous ones since they are able to partial out the effect of around 280 other variables through the implemented selection and projection methods. The results confirm that the CAN$100 denomination shows a pattern of higher velocity. The explanation for this particular phenomenon is unclear. On one hand, it could be that preferences to use larger denominations for larger transactions increases the velocity of the CAN$100. On the other hand, it could be that retailers are not willing to deal with large denominations. Something that could shed light on this matter would be to repeat the same analysis including CAN$20 and CAN$50 notes. However, the present sample does not allow for that.

Another interesting insight that comes from Table 1.3 is the sign of the fitness status

category. Results suggest that compared with the base category—a note with no fitness problems—changing the status to unfit does not dramatically alter the duration of the note. It does, however, have a significant effect of 0.4% over duration. The dwindling size of the effect of this variable can mean that the actual time when a note is deemed unfit is almost irrelevant to its duration or that it is reflected in the cluster.

The opposite happens when considering the cluster dummies. Notice how the labelling of the clusters has a very clear effect on the bank note duration with respect to the base case scenario. The size of the coefficients is higher in comparison to those of the fitness measures. The base case was built on CAN$5 notes that were split evenly across clusters. Remember that cluster 2 was primarily composed of CAN$100 bank notes, and the stay in the market of CAN$100 bank notes was particularly shorter than other denominations, especially the higher denominations. Also, cluster 2 was the most prevalent cluster label in region 1, where the anomalous behaviour was present. Hence, the results shown in Table 1.3 are consistent with our previous findings.

The analysis performed in this section corroborates the previous findings. The most important determinant of note duration is the denomination of the bank note. Other characteristics, and their interactions, can play a role in determining other moments of the duration distribution; however, the location of the mean is affected mostly by the denomination. While these results are not causal, they imply that any hidden factors, if they exist, that affect the duration are most likely denomination-specific.

Table 1.3: Accelerated Failure Time: Bank Note Duration

	A-Enet		OLS		Enet	
	Coef.	A. F.	Coef.	A. F.	Coef.	A. F.
Intercept			5.8400	343.7793	5.8330	341.3813
			(0.0017)		(0.0015)	
CAN$10	0.0141	1.0142	0.0191	1.0193	0.0189	1.0191
	(0.0005)		(0.0005)		(0.0005)	
CAN$100	-0.0436	0.9573	-0.0428	0.9581	-0.04308	0.9579
	(0.0002)		(0.0002)		(0.0002)	
Unfit	-0.0043	0.9957	-0.0045	0.9955	-0.00455	0.9955
	(0.0003)		(0.0003)		(0.0003)	
Cluster 2	-0.0241	0.9762	-0.0247	0.9756	-0.02467	0.9757
	(0.0002)		(0.0002)		(0.0002)	

Note: This table shows the results of estimating an AFT model with adaptive elastic net (A-Enet), OLS, and naive elastic net (Enet), on the log-duration with respect to around 280 features and interactions, to select relevant regressors. To obtain the standard errors (SEs) enclosed in brackets, Chernozhukov et al.'s (2018) double debiased machine learning algorithm is performed. Coef. stands for the estimated coefficients, while Acc. Factor stands for their exponential. All coefficients shown in the table are significant at the 99% level. The base case considered for this exercise was a CAN$5 bank note scanned in the month of January in the year 2017, shipped in the month of January of the year 2015, travelled from region 1 to region 1, belonged to cluster 1, was considered fit, and had no damage in any of the fitness measures.

1.6 Conclusion

Recent survey evidence has highlighted the decreased usage of cash for point-of-sale transactions in Canada (see Henry et al. (2018))), while Engert et al. (2019) document that cash demand in Canada, measured as cash in circulation relative to GDP, has been stable for decades and has even increased in recent years. To understand the potential difference in transaction versus non-transaction demand for cash, we exploit the IMS network data to uncover bank note flows in the economy. One of the most important components of these flows is the physical exchange and movement of bank notes. Understanding the characteristics of such a network, and decomposing each component of its units, allows us to better understand the way economic transactions occur in an economy. Exploiting the uniqueness

of the IMS data from the Bank of Canada, we first explore such a type of network. While several features were described, the one that took the central role in this research was the bank notes' length of the stay in the market, i.e., the duration in circulation.

Our analysis of the IMS data finds that one determinant of the duration of bank notes in the market is the group of fitness measures of the bank note, which are statistically (not necessarily economically) significant. We also find that the denomination structure of the bank note is economically and statistically significant, after we analyzed the network structure traced by the bank notes' travels. However, the nodes that behave as principal senders and receivers persist across all denominations. We compute the weighted in-degree and out-degree statistics for each node in this directed network to analyze the importance of each geographical region in Canada. In addition, we use the share of flows to analyze the importance of each edge of the network. We impose structure through K-prototype clustering on the data to classify the bank notes according to their features. Not surprisingly, even with a battery of confounding variables, the bank note denomination emerged as one of the key clustering determinants.

To obtain more compelling evidence, we analyze the distribution of duration through survival analysis. The unconditional analysis, with a Kaplan-Meier estimator of the survival curve, shows that there are clear, different, and denomination-specific survival curves. Next, we use an accelerated failure time model with an elastic penalty to assess the main drivers of the duration. The results show that the fitness measures are not the main determinants of bank note duration but the cluster labels associated with bank note denominations are.

Our results are novel in that we find the duration in circulation of a bank note is related to denomination-specific clusters. This result implies that certain bank notes are involved

in specific types of transactions that make certain denominations more or less likely to be deposited back to the Bank of Canada. Understanding denomination structure will help to understand payment choice; see Chen et al. (2019). Due to the non-causal nature of our results, we can only speculate for now; however, denomination-specific patterns seem to govern this variable. There are a variety of extensions that can be performed. For example, research could be undertaken that focuses on the spatial and temporal features of the data to understand if there are underlying unobserved factors that may confound the results. These are left for future research.

Chapter 2

On the Effect of Changing the Appearance of Money: Evidence from a Country-Wide Event[†]

Diego Israel Rojas Baez

Abstract

The Bank of Canada, Canada's monetary authority, in the year 2017, implemented an exogenous intervention on its currency. It introduced a change in the imagery of forty million CAD 10 banknotes. This intervention did not change the note's purchasing power. We set to explore whether slight changes in the appearance of a note can elicit a change in the behavior of economic agents enough to measure it. Specifically, we construct a measure of the duration of the circulation cycle of a note using proprietary data of the Bank of Canada. Using an adaptation of difference in differences for survival data and an accelerated failure time specification, we estimate the change in the duration of a ten-dollar banknote when the change was introduced.

[†]I want to thank Juan Estrada, Janet Hua Jiang, Kim P. Huynh, David Jacho-Chavez, Ramesh Paskarathas for providing insights and feedback about this document. I also want to acknowledge the efforts of Valerie Clermont, Ted Garanzotis, Mireille Lacroix, Andrew Marshall, Phil Riopelle, and Nathalie Swift and the use of the Bank of Canada's Digital Analytical Zone Microsoft Azure Cloud. The views expressed in this paper are those of the author. No responsibility for them should be attributed to the Bank of Canada.

2.1 Introduction

The Bank of Canada, Canada's monetary authority, in the year 2017, implemented an exogenous intervention on its currency. It introduced a change in the imagery of the CAD 10 note. This intervention did not change the note's purchasing power. According to microeconomic theory this intervention should not change the decisions of households and firms in Canada. Following the framework introduced by Tversky and Kahneman (1981); Kahneman and Tversky (1979), I find evidence that a change in the appearance of money can affect the decisions of agents. The facts we present were observed in a greater context than a controlled experiment. The intervention takes place in a developed country, none the less, in one of the seven largest economies in the world.

Several authors have explored the relation between money and decision frames (Shafir et al., 1997; Mishra et al., 2006; Di Muro and Noseworthy, 2013; Raghubir and Srivastava, 2002, 2009, 2008; Alter and Oppenheimer, 2008; Fehr and Tyran, 2001; Noussair et al., 2012; Cannon and Cipriani, 2006). They present evidence supporting the theory, which they observe through small samples of individuals in experimental settings. This document explores the relation between money and rational choice using a data set that spans all economic agents' decisions in the Canadian economy. In doing so, it expands the sample size considerably. Hence, adding external validity to the arguments of previous authors.

Related works explore different ways that framing effects take place. The work of Mishra et al. (2006), in the context of experiments, show that when presented with the same amount of money expressed by different denominations, people tend to change their

willingness to pay. Moreover, they show that people are more likely to keep a single bill of a high denomination than a large number of smaller denominations. The works of Boeschoten and Fase (1992); Tschoegl (1997); Raghubir and Srivastava (2009); Raghubir et al. (2017), also present evidence in favor of this "Denomination Effect".

One of the first studies to approach the physical appearance of money is the work of Bruce et al. (1983). Using experimental evidence, the authors show that physical traits can affect the subjective valuation that people assign to physical currency without knowing their actual value. Also, in this context Alter and Oppenheimer (2008), use experimental evidence to show that people change their spending behavior when presented with more familiar currency forms. More recently, Di Muro and Noseworthy (2013) show that notes showing more wear and tear are more likely to be spent. Moreover, the study shows that this physical appearance might be able to override the denomination effect.

I use data from the Information Management Strategy (Bilkes, 1997). I construct the duration of the circulation cycle of the notes present in the Canadian economy; specifically, those printed between 2011 and 2019. I departs from other similar studies on the duration of the circulation of money in Canada Paskarathas and Balodis (2019); Paskarathas et al. (2017); Rojas et al. (2020) because it is not concerned about the duration phenomenon itself. Instead, it is interested in finding the effect of the intervention previously mentioned. In doing so, it finds evidence supporting the deviation from rational choice. The identification strategy follows closely that of a difference in difference model, profiting the existence of information on the duration of other denominations to isolate the effect of the intervention. It achieves identification while addressing the issues natural to censored data. In that sense, it borrows heavily in terms of notation from Sant'Anna (2020); however, it does not achieve

a nonparametric identification like the mentioned work. It diverges from Sant'Anna (2020) because it imposes a parametric structure and more stringent assumptions about the relations between the treatment, the censoring, and the outcome. The estimation is done under the Accelerated Failure Time model using several parametric specifications.

The organization of this document is the following: Section 2.2 provides a brief description of the data and the intervention; Section 2.3 provides the identification strategy; Section 2.4 gives a brief analysis of the survival data considered for this study, and the result for the estimation. Finally, Section 2.5 offers the conclusions.

2.2 Data

2.2.1 The intervention

As Canada's monetary authority, one of the functions of the Bank of Canada is the design, production, and distribution, of currency. The bank has introduced several changes to the design of currency in different themed series throughout its history. In 2017, the Bank of Canada introduced a new graphic design for the CAD 10 note. The note was issued to commemorate the 150th anniversary of the confederation of Canada. Among other features, the commemorative note was the first note portraying a Canadian woman and an Indigenous Canadian person. To implement such change the Bank of Canada hired experts in public opinion research and consultation. It also allowed direct access for Canadians to comment their opinions in a website dedicated for such matter. It also conducted two rounds of focus group and testing.

Two surveys were conducted before the notes were into print. The first online survey, among 1707, Ipsos Reid Canadian Household Panelists.Its purpose was to find themes that Canadias found representative of their heritage. The second survey, among 2007 Ipsos Canadian Household Panelists, conducted between 29 December 2014 and 6 January 2015, which purpose was to ranks the themes that emerged from the first round.

Additionally, the Bank of Canada made an awareness study consisting of 3 surveys that measured the public awareness of the banknote, as well as interest in obtaining and retaining it. Very important insights were gained from this study for example: half (52%) of Canadians were aware of the note around a month before it was release. Only 34% of Canadians had seen the new design by the time of its release. About 52% percent of the respondents had not came across the new note by September of 2017. And only a 20% of the respondents actually were in possession of one of these notes. From that 20% 42% reported having spent every new note, the rest mostly kept one of the notes and planned to keep it. All of these results are strong indications that the new bank notes would have a non neutral effect on the decisions of consumers and firms.

Forty million notes were printed and put into circulation between June the first and July the first. These notes were not a replacement for the previous notes. The notes new notes had the same purchasing power of the notes already circulating on the economy, and they were perfect substitutes. It is essential to notice that no other notes design was issued or distributed after 2017 until the last quarter of 2018.

Figure 2.1 shows the change introduced in the CAD 10 banknote. The intervention did not affect the note's: size, material, color palette, security band position, the position of the denomination label, and the braille inscription position. There were four changes introduced:

Figure 2.1: Comparison of CAD 10 notes after and before the intervention

(a) Regular CAD 10 note (b) Commemorative CAD 10 note

Note: The note on panel (a) is the note that was introduced by the bank with the frontier series in 2013. Panel (b) shows the commemorative edition introduced in June 2017. The latter is introduced on the ocassion of the 150th anniversary of the confederation of the Commonwealth of Canada.

1. imagery depicted on the front of the note,

2. imagery depicted on the back of the note,

3. maple leaf transparency removed,

4. change of image on the security band.

This document argues that these changes only affect the point of reference for economic agents. Under the framework of rationality, in the absence of a change in the note's purchasing power, these changes should not affect individuals' decisions. Using the words of Kahneman and Tversky (1979), the intervention of the monetary authority effectively introduces a different decision frame in the economy. Therefore, for rational agents, the two notes depicted in figure 2.1 should be effectively exchangeable. In turn, this implies that any outcome of the notes, across decision frames, should be virtually equal. Any finding that shows such effect in any outcome is evidence that individuals' decision, in this case, is not entirely rational.

2.2.2 Data

The data set used for this project is part of the Information Management Strategy data set. This data set allows to monitor the distribution and inventories of notes in Canada (Bilkes, 1997). The data sample for this project contains records for the printing and recollection of the years 2011 through 2019. For each note, we can observe the date of printing, the date of collection, the fitness status, the regional distribution center, and the financial institution. Additionally, we can observe the metadata of the note: denomination, series, among others. The data set includes all five denominations of notes circulating in Canada: CAD 5, CAD 10, CAD 20, CAD 50, and CAD 100.

We match printing and collection records through the note's serial number to track the notes' duration. We call a circulation cycle the process of printing, shipping, using for transactions, and collecting each note, in that order. For this document, we define duration as the time, measured in days, elapsed between the printing date and the collection date. This document focuses on the first cycle of the circulation of the note. We do so to capture the agents' first response to the appearance of the new notes. A note that has been captured and recirculated is subject to new external factors in the subsequent circulation cycles. After the records match through the serial number, the total number of notes remaining in the sample is 1.7 billion. We build a sample using stratification according to the year of printing, the collection region, the denomination, and fitness status. By denomination, a "censored" stratum is assigned to notes that do not have a collection date up to the data's administrative close. The final sample for work considers a 1% sample on each of the strata; this yields a total sample size of 17.2 million notes.

Table 2.1: Descriptive statistics of the sample

		CAD 5	CAD 10	CAD 20	CAD 50	CAD 100
Number of notes per year	2011	-	-	-	-	200,367.0
	2012	-	-	227,317.0	204,048.0	-
	2013	220,878.0	113,929.0	1,813,060.0	-	-
	2014	986,134.0	481,326.0	1,507,292.0	-	961,181.0
	2015	51.0	500,818.0	1,917,154.0	845,791.0	-
	2016	583,510.0	-	976,365.0	-	406,711.0
	2017	-	421,658.0	243,593.0	-	1,227,300.0
	2018	-	413,862.0	409,754.0	697,637.0	-
	2019	193,320.0	109,513.0	1,616,727.0	-	-
Percent of Uncensored		26.2	22.3	36.9	24.4	14.2
Percent of Unfit		2.5	2.7	3.3	1.3	1.8

Note: The numbers shown in this table are based on a total 17 million note stratified sample. This represents 1% of the total administrative records on the IMS system between 2011 and 2019.

The distribution of the notes across denominations and year of printing is illustrated in table 2.1. The notes that appear more consistently on the sample are CAD 10 and CAD 20. Both denominations are the ones that are more likely to be dispensed by ATMs. The share of twenty dollar notes is overwhelmingly higher than that of other denominations. The denomination that appears less in the sample is CAD 50. The prevalence of CAD 20 throughout the years in the sample says that this denomination is suitable for making comparisons. Notice that the denomination of interest, CAD 10, appears does not appear in the sample in the years 2016, 2012, and 2011. Hence our analysis will have to focus only on those years that have records for CAD 10.

There is a portion of notes in this dataset for which we do not observe a collection event before the study's administrative closure. This censored data represents a non-negligible amount of records, around 71% of the total number of notes. However, this number is not evenly distributed across the sample. We can observe a regular feature in survival studies: the closer the notes were printed to the closure date of the study,2019-09, it is more likely

to have censored observations. For example, we show in the supplemental material that the year 2019 shows no collection of notes at all. For this reason, we exclude the notes printed in 2019 from the analysis in the coming sections.

Across denomination, the censoring varies too. Table 2.1 shows the denomination with the largest portion of uncensored notes is CAD 20, around 33%; for CAD 10 notes, the portion is around 25%. Both denominations have enough uncensored observations to perform estimation of survival models; however, we will need some modeling assumptions to account for such censoring.

Last, we see that in our sample, around 2.7% of the CAD 10 notes are unfit for circulation after their first cycle in the market; a 3.3% portion of the CAD 20 are in the same circumstances. Same for all other notes. In the supplemental material, we show that the distribution of the fitness status is similar across denominations, once we condition on the year of printing.

2.3 Identification

2.3.1 Decision frames and Money

The question of whether the appearance of banknotes can affect their outcomes has a theoretical justification. The work of Kahneman and Tversky (1979) is famously know for the introduction of prospect theory. While Tversky and Kahneman (1981) introduces the concept of a "decision frame". Briefly, a frame is the subjective conception of a choice that an individual has when presented with it. The key difference between this and the usual

rational choice is that different frames, perceptions of the same choice problem, can affect the individual's decision. Therefore, under prospect theory, the same individual, facing the same choice, could choose different options under different frames.

These findings are of particular interest when dealing with goods that are considered numeraires, like banknotes. The intrinsic value of money has been the subject of discussion from the dawn of the field (Humphrey, 1974). There are several ways on how theoretical microeconomics approach the role of money in economic agents' choices (Wallace et al., 1998; Walsh, 2017). However, these frameworks make their cases based on the assumption that the individual can make a precise and coherent judgment on money's value. Meaning, these are still inside the framework of the rational choice. In this document, I take a more agnostic approach. I use the data from the intervention in order to gain empirical evidence in support of the existence of decision frames.

In order to understand how I do this, lets introduce an example. Begin by assuming that any economic agent's choice regarding money can be understood through a utility function. We limit our example to the static, riskless case for illustration; however, these statements should also hold under dynamic and risky choices. Let the utility function of individual i be defined as $u_i^{(f)} : \mathbf{X} \times M : \mathbb{R}_+$. The set $\mathbf{X} \subseteq \mathbb{R}_+^L$ is the set of possible bundles of goods that the person can choose, and $M \subseteq \mathbb{R}_+$ is the amount of money that a person can choose to have. The super index $f \in \mathcal{F}$ denotes the frame, or subjective perception, under which individual i values her bundle. Under rational choice, we should expect that:

$$\forall f, f' \in \mathcal{F} : u_i^{(f)}(\mathbf{x}, m) = u_i^{(f')}(\mathbf{x}, m). \tag{2.1}$$

Equation (2.1) implies that no matter what the point of view of individual i, the utility perceived for any pair (**x**,m) should be the same. For example, take a two-frame case where an individual is presented with the bundle $(\mathbf{x}, 10)$. Under frame number one, the individual is given the amount **x** of goods and two five-dollar bills. Under frame two, the individual is given the amount **x** of goods and one ten-dollar bill. If we assume there is no transactional gain from the change in denomination, we should expect the utility of both bundles to be the same, and the individual is indifferent concerning the change of frame. The prospect theory version of this problem, however, states that:

$$\exists\, \tilde{\mathbf{x}} \in \mathbf{X},\ \tilde{m} \in M,\ f, f' \in \mathcal{F} \quad s.t. \quad u_i^{(f)}(\tilde{\mathbf{x}}, \tilde{m}) \neq u_i^{(f')}(\tilde{\mathbf{x}}, \tilde{m}). \tag{2.2}$$

Notice that the statement only requires the existence of a counterexample to be true. Back to our example, this would mean that for a particular denomination and bundle, equation (2.2) to hold. If that was the case, then the implication regarding the valuation of money is that subjective perceptions can change the way people behave when faced with decisions related to money, even when the total quantity (purchasing power) is held constant.

Let $\mathcal{G}(\cdot)$ be any function such that any one of its arguments is the utility function $u_i^{(f)}$. Fixing all the other arguments of $\mathcal{G}$, for equation (2.1) to hold, we need the following:

$$\forall\, \mathbf{x} \in \mathbf{X},\ m \in M,\ f, f' \in \mathcal{F} \quad : \mathcal{G}\left(u_i^{(f)}(\mathbf{x}, m)\right) = \mathcal{G}\left(u_i^{(f')}(\mathbf{x}, m)\right). \tag{2.3}$$

The most obvious case is when the function $\mathcal{G}$ is an identity. Then the statement in equation (2.3) reduces to the rationality condition. I argue that I can understand the outcomes of the notes as a particular case of $\mathcal{G}$. Hence we should be able to test whether

equation (2.3) holds with these outcomes using standard causal inference methods.

2.3.2 Identification strategy

In very few words the hypothesis to test is whether introduction of physical appearance of banknotes has an effect it's outcomes. I argue that such change is expected to happen in a world where the theory of decision frameworks have been proven to work. I try to isolate the effect of such changes by profiting from the exogenous nature of intervention, and using nationwide data. Doing so provides large scale evidence in favor of framing effects in a more general setting, rather than a small experiment in a very peculiar setting. In this section, I aim to establish a framework to evaluate whether a big scale, exogenous intervention, on the physical appearance of money circulating in an economy can lead to a non-negligible change in the behavior of the economic agents.

In order to assess the effect of cosmetic changes in the CAD 10 note, the duration of the first circulation cycle is chosen as the outcome to measure. Define duration as the time, measured in days, elapsed between the printing date and the collection date. As mentioned before, we focus on the duration of the first circulation cycle only. The circulation duration is a relevant quantity for the monetary authority for operational purposes. For an efficient distribution of notes across the economy, the Bank of Canada needs to be aware of the notes' duration on circulation to replenish inventories of notes efficiently.

The intervention, described in section 2.2, allows us to observe the CAD 10 notes printed before and after the intervention. Since there is no date for which the two notes enter into the survival study simultaneously, this raise the concern that they might be affected by different

external factors specific to each group. I propose using a model that works very similarly to a difference in difference set up to address this. First, find another denomination or group of denominations that did not undergo a cosmetic change simultaneously as the CAD 10. Then, we expect that external factors before and after the moment of the intervention, June 2017, similarly affected all denominations. Hence, we can difference out those external factors and identify the effect of the intervention.

Let the duration of a note printed at time t be the random variable Y_t. For the sake of exposition let $t = 0, 1$, where $t = 0$ is the time before the intervention happened and $t = 1$ the opposite. Because we only observe data of collection until the present day, Y_t is a latent variable. We observe the variable Q_t, which is the duration time Y_t if the note is collected before the study's closure or a censoring time C_t. Then, let $D = 1$ if the note is CAD 10, or $D = 0$ if it belongs to the untreated group. Finally, let $\delta = 1$ if the note is censored and $\delta = 0$ otherwise. If we adopt the potential outcome notation, then we can express Q_t as:

$$Q_t = \overbrace{\left[Y_t^1(1-\delta) + C_t^1(\delta)\right]}^{Q_t^1} D + \underbrace{\left[Y_t^0(1-\delta) + C_t^0(\delta)\right]}_{Q_t^0}[1-D] \tag{2.4}$$

In relation to our example, Y_t, is a special case of the $\mathcal{G}$. The change in appearance is the change in framing. Hence we can compare Y_t across frames to know if a relation like equation (2.3) holds. Then, the objective is to find the difference in the duration between the altered and non-altered CAD 10 notes. It follows that, in this context, we call this quantity the *treatment effect on the treated*. The document will focus on the *average treatment effect on the treated* from now referred to as "ATET". This can be expressed as $E\left[Y_1^1 - Y_1^0 | D = 1, X = x\right]$. The conditioning on covariates are omitted from this point for the sake of economizing

notation. An ATET of zero means that equation (2.3) holds; an ATET different from zero implies that equation (2.3) does not hold.

In order to establish identification, the standard assumptions of a difference in difference model are necessary. Additionally assumption we make an assumption that allows us to address censoring. Following Lechner et al. (2011), the assumptions are:

A1 **Stable unit treatment value**. Which means that one, and only one, of the potential outcomes of cycle duration is observed for all banknotes in the population.

A2 **Exogeneity.** Observables are independent of treatment. In the case of this intervention is trivial since intervention is exogenous.

A3 **No backwards effect.** no effect of treatment in pre-treatment periods. This means that the change in the physical appearance of the notes did not affect the outcomes of notes in the pre-treatment period.

A4 **Common Trend.** The differences in the expected potential outcomes of the banknotes, before the intervention, are unrelated to belonging to group that received the physical appearance change. Again this assumptions is satisfied by the exogenous nature of the intervention.

A5 **Common Support.** Observations with characteristics X exist for treated and untreated before and after the intervention.

A6 **Uniformative Censoring.** Censoring process is distributed at random across the sample and treatment groups.

Provided all of these conditions hold, then it can be shown that the average treatment effect can be expressed as:

$$\begin{aligned} ATET =& E\left[Y_1^1 - Y_1^0 | D = 1\right] \\ =& E\left[Q_1 | D = 1, X = x\right] - E\left[Q_1 | D = 0, X = x\right] + \\ &+ E\left[Q_0 | D = 0, X = x\right] - E\left[Q_0 | D = 1, X = x\right]. \end{aligned} \tag{2.5}$$

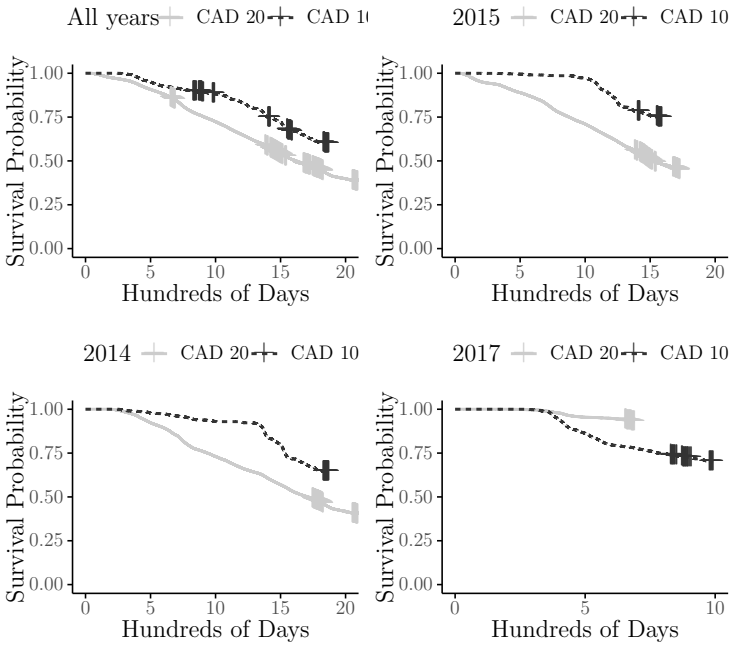
The equation (2.5) is similar to the definition of the ATET in a difference in difference model, except that we use the censored observed outcome Q_t. I will assume a parametric distribution of the estimation error, along with the assumption that the effect of treatment is homogeneous in time.

The most important assumption for the identification strategy to hold is that of parallel trends. The absence of a greater number of periods to analyze does not allow us to test rigorously for this assumption. However, in the period selected, the survival curve conditional on the year of printing can be estimated. If the common trend assumption holds, it must be that these curves behave similarly in the periods before treatment, 2014, and 2015. The estimates for such curves are shown in figure 2.2. The results support the parallel trend assumption. The sudden switch in the survival curves in 2017 makes a stronger case for the existence of a possible effect of monetary authority intervention.

2.3.3 Estimating equation

To estimate the average treatment effect in equation (2.5), we use a log-linear specification of the duration time. This choice is not arbitrary; such specification is a well-known model

Figure 2.2: Survival curves of treatment and control group



Note: The survival probability curve is estimated by the product limit estimator. The crosses on each curve represent time points where censored data is present. The numbers shown in this figure are based on a total 17 million note stratified sample. This represents 1% of the total administrative records on the IMS system between 2011 and 2019. This particular sample excludes the year 2019 because that year shows a 100% censoring.

in survival analysis, the Accelerated Failure Time model. This model provides a framework that allows us to deal with the censored observations by assuming a parametric distribution of the error term. The model that is estimated has the following specification:

$$T = \ln Q = \mu + \beta' \mathbf{X} + \sigma W. \tag{2.6}$$

Q is the imputed time variable, $\mathbf{X}$ is the matrix of covariates, and W_i term is assumed to follow a parametric distribution and to be mean independent of the covariates. Univariate analysis of the distribution of Q showed that the most relevant parametric distributions to be considered are log-normal, log-logistic, exponential, and Weibull. To estimate the average treatment effect, equation (2.6) must have the following form:

$$T_i = \mu + \beta_1' d_i + \beta_2' \tau_i + \rho d_i \times \tau_i + \beta_3' \mathbf{entry}_i + \sigma W_i, \tag{2.7}$$

.

The d_i is a dummy indicating that the note i is a CAD 10 note, τ_i is a dummy variable indicating that note i was printed after or before the intervention; and, **entry** is a vector of entry date dummies. Following Lechner et al. (2011), under this specification of equation (2.7) the coefficient ρ is an estimator of the average treatment effect. The estimation of such a model must be carried out by maximum likelihood to consider the censoring issue.

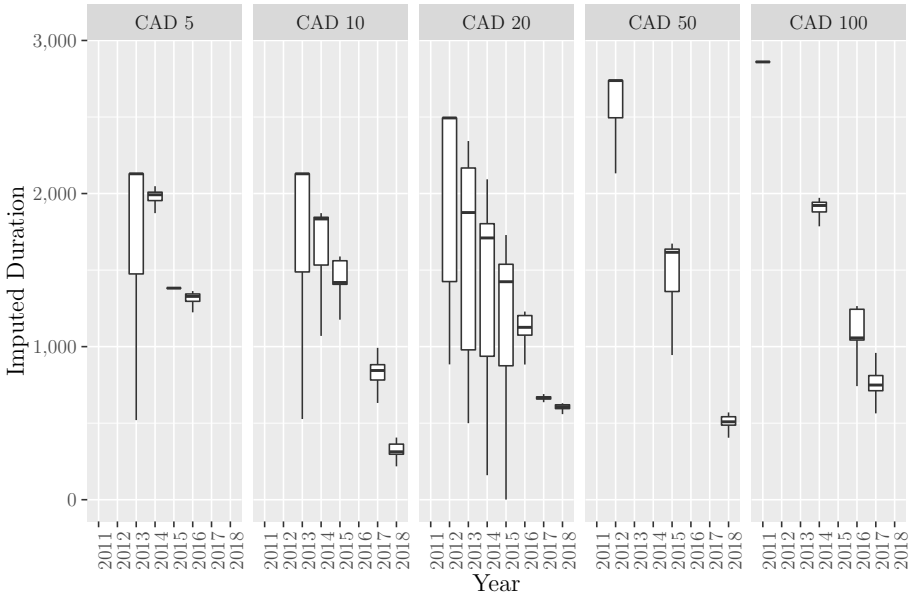
2.4 Results

2.4.1 Duration patterns

In order to construct the duration measure, an imputation has to take place. Remember, the duration is defined as the difference in days between the collection date and the printing date. For the data where no collection date is observed, the study's closure date is imputed as the collection date. The data tracks the notes that were part of this process with a censoring indicator. Figure 2.3 give us a first look at the constructed variable. The duration tends to diminish as the time of printing approaches the final date of the study. This feature is a usual trait in studies with different entry dates of individuals. The more exciting feature we find when comparing between denominations. For example, CAD 100 shows a tighter distribution of the circulation duration than all other denominations. CAD 5 shows a unique pattern and does not reproduce the trend observed in other denominations. Notice that the

duration distributions, conditional on each year, seem to follow a somewhat similar pattern. This evidence suggests that CAD 20 can be a good control group in our design.

Figure 2.3: Imputed duration by year and denomination

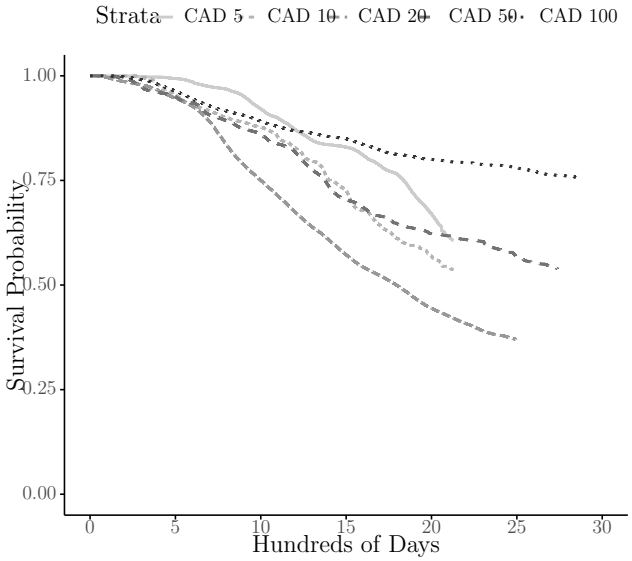


Note: The duration distribution observed uses the administrative closure of the study as the collection data for the censored data. Each panel shows a denomination's duration conditional on every year. The numbers shown in this figure are based on a total 17 million note stratified sample. This represents 1% of the total administrative records on the IMS system between 2011 and 2019. Additionally the notes of 2019 have been filtered since the censoring on that year reaches 100%.

It is good to switch perspectives towards the study of survival probabilities to account for censoring and understanding the duration distribution better. While describing these quantities is not the central object of this document (see, for example Kalbfleisch and Prentice (2011); Klein and Moeschberger (2006) for a detailed analysis), the survival probability is the one minus the probability of failing at any point of time in a study. For example, for this document's case, the survival probability is one minus the probability of being collected back at the bank of Canada at a particular time. To estimate consistently the survival curve, which describes the survival probability at all times, we need an estimator that can

deal with censoring. We use the K-M estimator introduced by Kaplan and Meier (1958) to estimate the survival curves for each denomination on the sample. Figure 2.4 shows that the two higher denominations, CAD 50 and CAD 100, have very different survival curves to that of the CAD 10. These two are becoming flatter as time increases, suggesting that these notes are used for value storage and hoarding. This finding would be consistent with the observation of Boeschoten and Fase (1992) in the Netherlands. The curves of the lower denominations seem to decay faster. This finding is also consistent with the transaction role that these notes are believed to have. Hence, we can argue that any control denomination that we pick for CAD 10 must belong in the lower denominations group.

Figure 2.4: Kaplan Meier - Survival probability estimates by denomination



Note: The survival probability curve is estimated by the product limit estimator. The crosses on each curve represent time points where censored data is present.The numbers shown in this figure are based on a total 17 million note stratified sample. This represents 1% of the total administrative records on the IMS system between 2011 and 2019. This particular sample excludes the year 2019 because that year shows a 100% censoring.

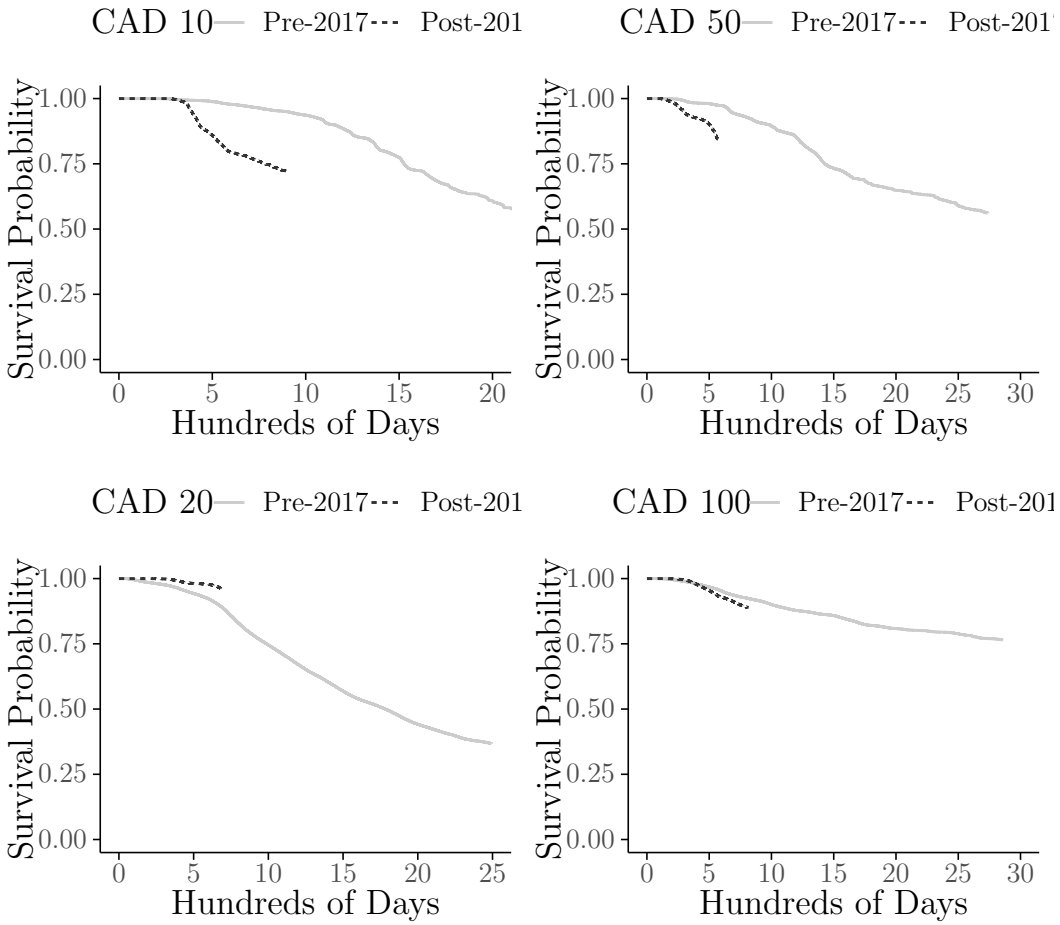
The identification strategy outlined in section 2.3 needs full support shared between treatment groups. In other words, if we are to condition on the entry date to the study, the

date of printing, then we need the denominations we use for comparison also to be observed in the same time ranges. Hence, as shown in table 2.1, the CAD 5 notes are not available in the year 2017, the year of the intervention. Therefore the strongest candidate so far, as the control denomination, is the 20 CAD note.

Finally, to support the hypothesis that an effect might exist on the duration of the CAD 10 notes duration, it is possible to estimate each denomination's survival curves before and after the intervention happened. If there is no effect of the intervention, it is reasonable to expect that the curves after the intervention behave almost identically to the pre-intervention period's curves. Figure 2.5 shows the results of this exercise. The survival curve of CAD 10 notes is noticeably steeper when comparing after and before the intervention. The same cannot be said of the survival curves for the other denominations. For example, the CAD 100 seems to behave almost identically between periods. The other two denominations show changes in their survival curves but with considerably smaller changes in their survival probabilities.

The results shown here are merely a visual cue that points to the effect of the intervention in the CAD 10 circulation cycle duration. A more robust analysis will be carried out in the next section to implement an estimation of the ATET. Notice that the evidence shown in this section points to the CAD 20 to be the control denomination that will allow to difference out the external factors affecting both denominations.

Figure 2.5: Survival probability estimates by denomination and treatment period



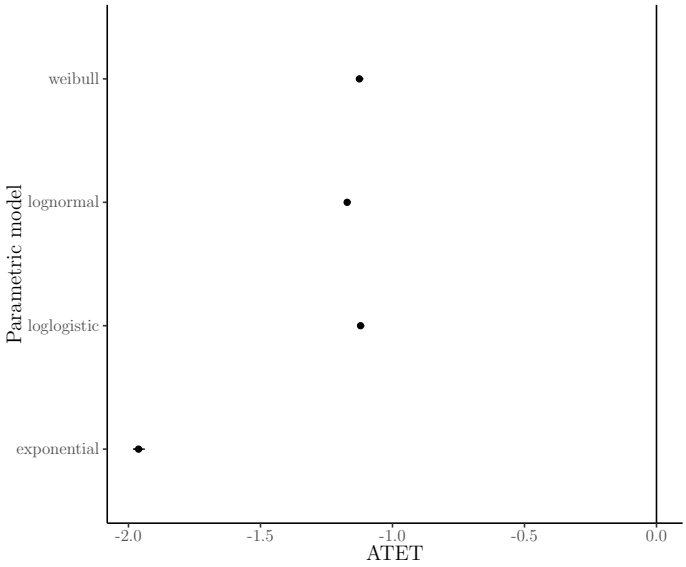
Note: The survival probability curve is estimated by the product limit estimator. The crosses on each curve represent time points where censored data is present. The categories consider the date of June 2017 as the threshold to split each denomination sample. The numbers shown in this figure are based on a total 17 million note stratified sample. This represents 1% of the total administrative records on the IMS system between 2011 and 2019. This particular sample excludes the year 2019 because that year shows a 100% censoring.

2.4.2 Estimates

The results of the implementation using four parametric distributions are shown in figure 2.6.

The results show, at a very high level of confidence, that the CAD 10 notes that were the target of the intervention have, on average, a lower duration than the regular notes. This

Figure 2.6: Averate Treatment effect on the treated by parametric model

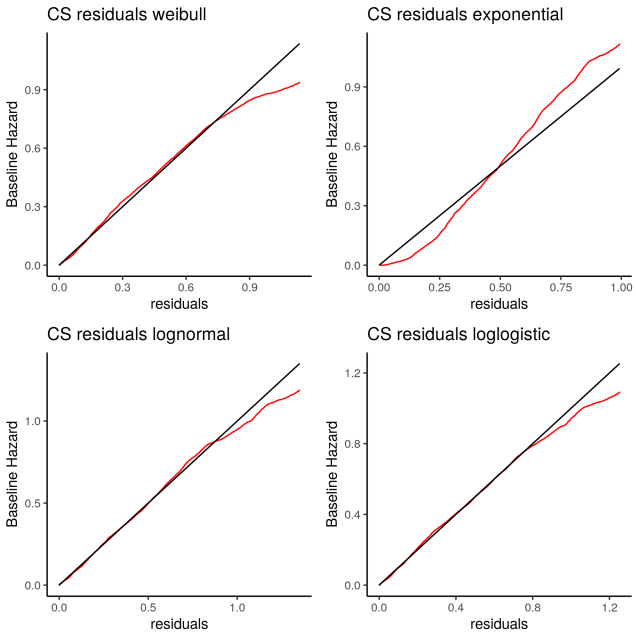


Note: The results are obtained from 4 different parametric specifications. Only the ATET coefficient is shown. Confidence intervals are included in the plot. The numbers shown in this figure are based on a total 17 million note stratified sample. This represents 1% of the total administrative records on the IMS system between 2011 and 2019. This particular sample excludes the year 2019 because that year shows a 100% censoring.

finding holds regardless of the parametric assumption imposed for the estimation. Concretely, taking back the results from the log-linear scale, the average treatment effect is a reduction of 67.5% of the duration of the circulation cycle at best and an 85% reduction at worst.

As a means to assess the performance of the specifications, regarding the parametric distributions, we calculate the Cox-Snell residuals. Following Klein and Moeschberger (2006), if the fitted model does a good job predicting the data, it is expected that the Cox-Snell residuals have an exponential distribution. Then, a plot of residuals against the residuals' cumulative hazard should be a straight line with a slope equal to one. The results of this exercise are shown in figure 2.7. The exponential model is the one that has the worst performance. This result is compatible with the atypical behavior of the estimator concerning the other three models. It is also compatible with the performance assessment according to

Figure 2.7: Cox-Snell residuals by parametric model



Note: The Cox-Snell residuals are calculated as shown in Klein and Moeschberger (2006), chapter 12. A perfect fit on the 45 degree line is evidence of a great performance of the fitted model. The numbers shown in this figure are based on a total 17 million note stratified sample. This represents 1% of the total administrative records on the IMS system between 2011 and 2019. This particular sample excludes the year 2019 because that year shows a 100% censoring.

the Akaike and Bayesian information criterion, where the exponential model ranks last. The other three models do a reasonably good job in earlier times but tend to perform worse later. Visually, it is challenging to choose which model performs best. The information criteria rank as the best the log-logistic model, followed by the log-normal. The Weibull model is third, closer to the exponential model. The last two models are the ones that admit a proportional hazard representation. More unsatisfactory performance of these is compatible with exploration analysis that rejected the proportional hazard hypothesis in the supplemental material.

2.5 Conclusions

This document uses data from an exogenous intervention in Canada to show that the duration of the circulation cycle of the CAD 10 subject of the intervention is reduced. These results also support the implication that individuals act according to the subjective perception of the choice presented to them. The existence of non-zero average treatment effects of the intervention on the treated is evidence supporting this claim. For these results to hold, the assumptions made are reasonable because of the nature of the observation units. However, we assume a strict structure of the relation between the censoring, treatment, and outcome. This assumption is reasonable as long as the treatment has homogeneous effects. An extension that allows for heterogeneous effects of the intervention is a path for future research,

The results found are in line with the theoretical expositions of Tversky and Kahneman (1981); Kahneman and Tversky (1979). They are also consistent with the empirical evidence observed in experimental settings like Di Muro and Noseworthy (2013); Raghubir and Srivastava (2008, 2002); Alter and Oppenheimer (2008).

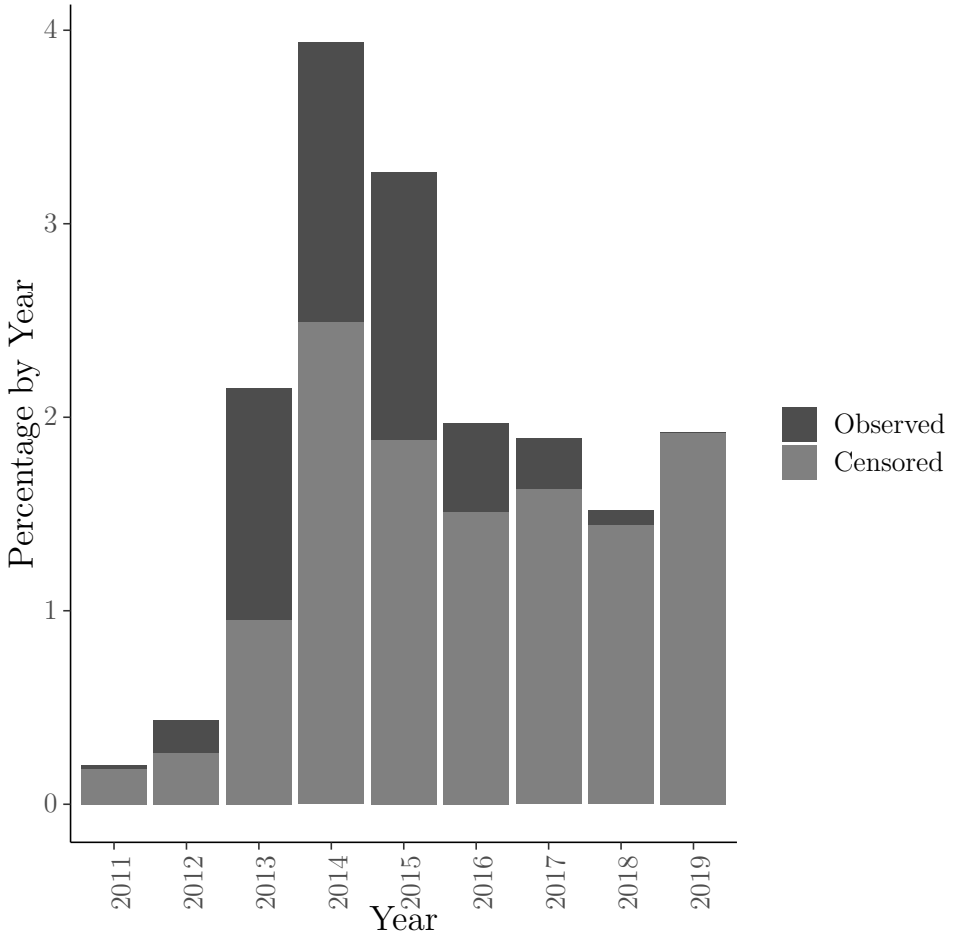
Finally, these results suggest that currency interventions that do not affect their purchasing power can have unintended consequences. Notably, this document finds a reduction of 67% of the duration of the CAD 10 notes' circulation cycle. This finding is just evidence of one side of those consequences; other consequences may be related but cannot be observed.

Appendix

Descriptive Misc

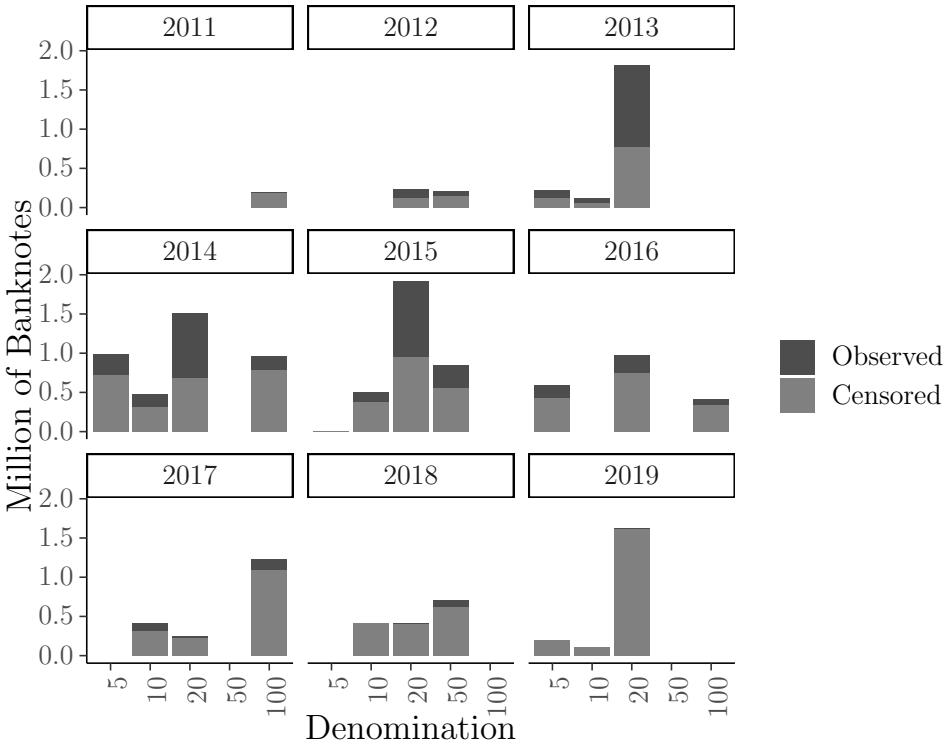
The following figures illustrate features of the sampole.

Figure 2.8: Comparison of CAD 10 notes after and before the intervention



Note: The note on panel (a) is the note that was introduced by the bank with the frontier series in 2013. Panel (b) shows the commemorative edition introduced in June 2017. The latter is introduced on the ocassion of the 150th anniversary of the confederation of the Commonwealth of Canada.

Figure 2.9: Comparison of CAD 10 notes after and before the intervention



Note: The note on panel (a) is the note that was introduced by the bank with the frontier series in 2013. Panel (b) shows the commemorative edition introduced in June 2017. The latter is introduced on the ocassion of the 150th anniversary of the confederation of the Commonwealth of Canada.

Duration Patterns

The following models provide show the full sets of results.

	Dependent variable:			
	inputdur			
	Weibull	*exponential*	*survreg: lognormal*	*survreg: loglogistic*
	(1)	(2)	(3)	(4)
treat	0.683***	1.130***	0.693***	0.665***
	(0.004)	(0.007)	(0.004)	(0.004)
ftreat	1.243***	2.614***	0.948***	1.091***
	(0.011)	(0.020)	(0.008)	(0.010)
itreat	−2.010***	−3.561***	−1.729***	−1.852***
	(0.012)	(0.021)	(0.009)	(0.011)
factor(bn_denom_id)20	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)

factor(year)2014:factor(month)1	0.195***	0.091***	0.118***	0.127***
	(0.003)	(0.004)	(0.003)	(0.003)
factor(year)2015:factor(month)1	0.185***	0.170***	0.143***	0.161***
	(0.003)	(0.004)	(0.003)	(0.003)
factor(year)2017:factor(month)1	−0.467***	−0.557***	−0.452***	−0.451***
	(0.012)	(0.021)	(0.012)	(0.012)
factor(year)2014:factor(month)2	0.241***	0.164***	0.188***	0.197***
	(0.003)	(0.005)	(0.003)	(0.003)
factor(year)2015:factor(month)2	0.251***	0.279***	0.242***	0.253***
	(0.003)	(0.005)	(0.003)	(0.003)
factor(year)2017:factor(month)2	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
factor(year)2014:factor(month)4	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
factor(year)2015:factor(month)4	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
factor(year)2017:factor(month)4	0.069***	0.075***	0.081***	0.073***
	(0.004)	(0.007)	(0.004)	(0.004)
factor(year)2014:factor(month)5	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
factor(year)2015:factor(month)5	−0.277***	−0.565***	−0.206***	−0.222***
	(0.014)	(0.024)	(0.016)	(0.014)
factor(year)2017:factor(month)5	0.039***	0.035***	0.039***	0.040***
	(0.004)	(0.007)	(0.004)	(0.004)
factor(year)2014:factor(month)6	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
factor(year)2015:factor(month)6	−0.149***	−0.345***	−0.081***	−0.103***
	(0.005)	(0.009)	(0.005)	(0.005)
factor(year)2017:factor(month)6	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
factor(year)2014:factor(month)7	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
factor(year)2015:factor(month)7	0.199***	0.252***	0.228***	0.217***
	(0.004)	(0.007)	(0.005)	(0.004)
factor(year)2017:factor(month)7	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
factor(year)2014:factor(month)8	−0.335***	−0.759***	−0.395***	−0.319***
	(0.006)	(0.010)	(0.006)	(0.006)
factor(year)2015:factor(month)8	−0.194***	−0.361***	−0.601***	−0.416***
	(0.003)	(0.005)	(0.004)	(0.003)
factor(year)2017:factor(month)8	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
factor(year)2014:factor(month)9	−0.085***	−0.350***	−0.032***	−0.034***
	(0.005)	(0.008)	(0.005)	(0.005)
factor(year)2015:factor(month)9	−0.079***	−0.173***	−0.387***	−0.185***
	(0.003)	(0.005)	(0.003)	(0.003)

factor(year)2017:factor(month)9	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
factor(year)2014:factor(month)10	0.188***	0.137***	0.060***	0.137***
	(0.003)	(0.005)	(0.003)	(0.003)
factor(year)2015:factor(month)10	0.004	−0.025***	−0.131***	−0.019***
	(0.003)	(0.005)	(0.003)	(0.003)
factor(year)2017:factor(month)10	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
factor(year)2014:factor(month)11	0.229***	0.209***	0.173***	0.220***
	(0.003)	(0.005)	(0.003)	(0.003)
factor(year)2015:factor(month)11	0.107***	0.155***	0.134***	0.122***
	(0.003)	(0.006)	(0.004)	(0.004)
factor(year)2017:factor(month)11	−0.960***	−1.668***	−0.702***	−0.843***
	(0.012)	(0.021)	(0.008)	(0.011)
factor(year)2014:factor(month)12	0.074***	−0.011**	−0.042***	0.0003
	(0.003)	(0.005)	(0.003)	(0.003)
factor(year)2015:factor(month)12	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
factor(year)2017:factor(month)12	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	7.469***	7.736***	7.359***	7.292***
	(0.002)	(0.004)	(0.003)	(0.002)
Observations	5,071,841	5,071,841	5,071,841	5,071,841
Log Likelihood	−19,491,707.000	−19,819,156.000	−19,464,048.000	−19,463,902.000
χ^2 (df = 37)	317,617.800***	273,384.200***	479,327.200***	378,720.400***

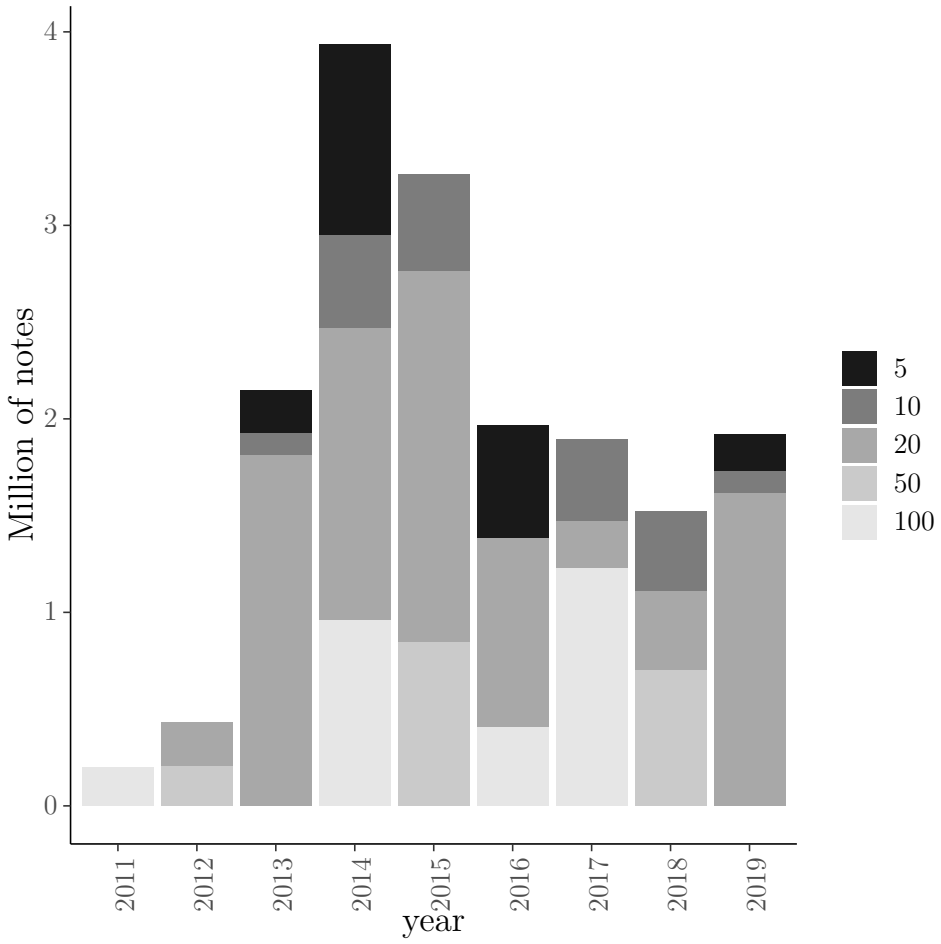
Note: *p<0.1; **p<0.05; ***p<0.01

Table 2.2: AFT regression: 4 specifications

	Dependent variable:			
	inputdur			
	Weibull	*exponential*	*survreg: lognormal*	*survreg: loglogistic*
	(1)	(2)	(3)	(4)
treat	0.452***	0.738***	0.634***	0.525***
	(0.001)	(0.002)	(0.001)	(0.001)
ftreat	0.842***	1.658***	0.912***	0.850***
	(0.012)	(0.021)	(0.012)	(0.012)
itreat	−1.111***	−2.014***	−1.209***	−1.128***
	(0.005)	(0.009)	(0.004)	(0.005)
factor(year)2015	−0.084***	−0.031***	−0.113***	−0.073***
	(0.001)	(0.001)	(0.001)	(0.001)
factor(year)2017	−0.406***	−0.264***	−0.451***	−0.431***
	(0.012)	(0.020)	(0.012)	(0.011)
factor(bn_denom_id)20	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	7.650***	7.836***	7.453***	7.422***
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	5,071,841	5,071,841	5,071,841	5,071,841
Log Likelihood	−19,539,167.000	−19,854,499.000	−19,562,681.000	−19,525,853.000
χ^2 (df = 6)	222,697.300***	202,698.600***	282,062.700***	254,817.100***

Note: *p<0.1; **p<0.05; ***p<0.01

Figure 2.10: Comparison of CAD 10 notes after and before the intervention



Note: The note on panel (a) is the note that was introduced by the bank with the frontier series in 2013. Panel (b) shows the commemorative edition introduced in June 2017. The latter is introduced on the ocassion of the 150th anniversary of the confederation of the Commonwealth of Canada.

Chapter 3

Multilayer Dyadic Gravity Models[†]

Diego Israel Rojas Baez

Abstract

The reduced cost of communications and transport increased the number and types of connections between economic agents. Generally, Economics models networks as two-dimensional arrays, overlooking the multiple types of links between agents. The primary tool to model network formation in this framework is the dyadic regression, for example, the gravity trade model. I propose that dyadic regression provides a sound framework to model multilayer network formation under standard regularity conditions. This paper introduces an extended version of a gravity model that can accommodate a larger network class that allows for multiple layers of connections. Furthermore, I propose a GMM estimator that is consistent and asymptotically normal. I use country data for trade, tourism, and investment to estimate the homophily parameter of a country-level multilayer network.

[†]I want to thank Juan Estrada, Ana Maria Herrera, Kim P. Huynh, David Jacho-Chavez, Elena Pesavento, Christopher Breunig for providing insights and feedback about this essay. The views expressed in this paper are those of the author. No responsibility for them should be attributed to the Bank of Canada.

3.1 Introduction

International trade, friendship, followers, or messages are examples of connections between economic agents. Recent developments in Econometrics (Graham, 2020a) and Microeconomics (Jackson and Wolinsky, 1996) show that the influence that peers can enact on the decisions of economic agents is highly relevant. The reduced cost of communication, brought by the accelerated growth of internet platforms, has increased the number and types of connections between economic agents. The result is that networks have evolved into complex systems where agents connect multiple times across different aspects. For example, two people might connect through family and a different type of relation like work. Considering only one of these connections, or assuming the connections are interchangeable, discards information about the network. Until recently, network science and economics have overlooked that networks are collections of heterogeneous connections and have restricted their analysis to the realm of networks that summarized the relations between agents into two-dimensional graphs. In this context, the dyadic regression is the primary tool to model and predict the formation of bilateral connections between agents, for example, the gravity model of international trade. Can a dyadic regression model capture the behavior of a more complex structure that considers the different natures of connections between agents? This paper is the first attempt to answer such a question. I propose that a model similar to the gravity model can predict the dyadic formation of connections of different natures. Furthermore, I provide a consistent GMM estimator for the homophily parameter.

Kivelä et al. (2014) and Boccaletti et al. (2014) are extensive literature reviews on multilayer networks across disciplines. As noted in both works, the concept of a multilayer network as a collection of connections of different types already attracted social scientists several decades ago. However, until recently, network science has put aside the very relevant fact that connections among nodes in a system can have different nature. According to Kivelä et al. (2014) the more traditional approach to study networks, referred to as monolayer from now on, uses graphs as abstractions of the systems they intend to describe. Under this scenario, the system is a set of nodes and connections; the latter can be weighted, directed, or undirected. According to Boccaletti et al. (2014) this approach ignores the possibility that connections between agents can have different nature. Consider, for example, the case of a network of family and friends. Under the traditional approach, all connections in this network are indistinguishable regardless of whether they originate from a friendship or family relation. Instead, a multilayer approach to this scenario proposes that each of these interactions is happening in a different dimension. Boccaletti et al. (2014) argues that including these considerations can improve the analysis of real-world networks; moreover, not including them can be detrimental.

In economics, the study of networks itself is a relatively novel topic. According to Graham (2020b), economists have historically "avoided" the study of networks. However, there is a recent surge in theoretical and empirical work around this subject. The same author argues that the factors contributing to an increased interest in networks include the increased availability of data and the recent developments in random graph theory. Regarding the use of multilayer network structures, economics is no different from other fields. The workhorse tools abstract networks to the traditional monolayer graph approach in both theoretical and

empirical work. This way to operate has proven incredibly successful, bringing in milestone contributions like Jackson and Wolinsky (1996), or Manski (1993).

Nevertheless, all these contributions operate under the assumption that the network involved in the data generating process is a set of connections of the same nature. This assumption is primarily to keep the problem tractable and that the field implicitly assumes that only one type of connection is relevant to a particular problem. For example, for education outcomes, the econometrician thinks that the relevant connections are friendships among classmates, implicitly ignoring other types of connections that could occur between them. In a vast sense, this is equivalent to an omitted variable bias. Suppose that there are different types of networks in the population that influence the decisions of economic agents. A well-specified model has to consider the heterogeneity in the types of connections. Alternatively, one can think that the true network spanned among individuals is a complex object with different types of connections. By reducing all the connections to a unique type, one effectively loses potentially relevant information about such a network.

In this context, dyadic regression is one of the most used tools to model network formation in economics. The work of Graham (2020a,b) presents a deep dive into the literature around dyadic regression, theory, estimation, and inference. In addition, Graham (2020b) provides a taxonomy of fields where economists are using the dyadic regression model to tackle economic questions. To name a few, these include international trade, corporate governance, production networks, research collaboration, risk-sharing across households, insurer-provider and referral networks for healthcare, and employment search. The latest contributions to this model focus on estimation and inference improvementsAronow et al. (2017); Menzel (2019); Tabord-Meehan (2019); Graham (2020b). The first use of dyadic models

in economics is the gravity model (Tinbergen, 1962) for international trade. The model is dyadic because it approaches trade as a bilateral decision that depends only on the countries that take part in the trade. The intuition to this approach is very ingenious: countries that are closer and have higher levels of GDP are more likely to trade among them. This model and its following iterations focus on the trade network and do it with much success. However, countries do not connect through trade only, but through a myriad of aspects like technology, migration, investment, and others. Hence, while the network we observe is sufficient to model trade, the true face of the complex network that connects countries remains hidden. In order to understand how countries exert influence on each other, a broader framework that can take on complex structures is necessary. This document proposes that such a framework is that of multilayer networks.

In a broad sense, the contribution of this paper is to provide a way to estimate dyadic regression models in the context of multilayer networks. Hence, providing a way to model the formation of more complex structures that allow for heterogeneous types of connections. In order to illustrate the use of the estimator, I use an extended version of the gravity model that considers connections between countries in several layers. As in the monolayer case, the intuition is that similar countries are more likely to be connected, and exporter and importer fixed effects enter the equation. Nevertheless, I consider a complex network with multiple layers in this case. For the empirical application, I include two additional layers: investment and tourism flow, along with trade flows. The multilayer gravity model allows for connections between nodes in any layer and across them. In this setting, an additional source of shock appears as a layer-specific fixed effect. The way I approach this problem is to extend the network formation model that is present in the work of Jochmans (2017) to a larger

class of relations between agents. Specifically, I show that under similar conditions to those imposed in Jochmans (2017), I can construct a consistent GMM estimator for the similarity parameter in a multilayer network formation model. Moreover, I make the conjecture that such estimator is asymptotically normally distributed. I provide simulation evidence to show the effectiveness properties of the model under several scenarios. Additionally, I provide an empirical application to model a worldwide multilayer trade, tourism, and investment network.

3.2 Mulilayer Dyadic Data Model

In order to describe the new estimator, one needs to briefly describe how to deal with dyadic data models in a setting where only one-layered networks are present. A typical application of such models in economics is the gravity model (Tinbergen, 1962) for international trade. The model is dyadic because it approaches trade as a bilateral decision that depends only on the countries that take part in the trade. The intuition to this approach is very ingenious: countries that are closer and have higher levels of GDP are more likely to trade among them. Anderson and van Wincoop (2003) adds multilateral resistance terms in the form of country-specific shocks in order to improve the model estimates to the observed trade patterns. Silva and Tenreyro (2006) argues for a Poisson pseudo maximum likelihood estimation of the multiplicative model of gravity instead of the standard log-transformed version to avoid biases in the estimation due to Jensen's inequality. Helpman et al. (2008) introduces importer and exporter specific shocks. The consequence of such addition is the problem of incidental parameters in nonlinear panels. Charbonneau (2017) uses conditional maximum likelihood

to extend the solutions to the incidental parameter problem for the single fixed effect case to a two-way fixed effect framework. In a different approach, Jochmans (2017) introduces a GMM estimator that relies on the construction of conditional moments that difference out the two-way fixed effects. The estimator I propose extends the findings of Jochmans (2017) to accommodate an extensive class multilayer network.

Typically, in dyadic regression, the connection between two nodes is understood as a multiplicative model. Following the exposition of Jochmans (2017) one can represent such models as:

$$Y_{ij} = \varphi\left(R_{ij}; \theta_0\right) U_{ij}$$

Where Y_{ij} is the dyadic relation between two nodes i and j, and φ is a measurable function known up to a parameter θ_0. The disturbance U_{ij} is takes the form of a product: $U_{ij} = A_i B_j V_{ij}$. Where A_i and B_j are node-specific permanent unobserved effects, and V_{ij} is the idiosyncratic dyad disturbance. The latter is independent across i and j. R_{ij} is a vector of observed characteristics specific to each connection. In the particular case of the gravity model for trade, each dyad Y_{ij} is the trade flow from country i to j, and A_i and B_j are importer and exporter unobserved permanent shocks. The vector of parameters θ_0, also known as the homophily parameter, is the object that the econometrician tries to estimate. Intuitively, this parameter conveys the importance of similarity between nodes for the connection level.

Conventionally, one can estimate the parameters of this class of models using a Poisson pseudo-maximum likelihood (Gourieroux et al., 1984). More recently, in the light of the

evidence of potential bias in the estimation (Fernández-Val and Weidner, 2016), approaches like Charbonneau (2017) and Jochmans (2017) are the alternatives to achieve sound inference by using differences, in moment conditions, or likelihood functions, to cancel out the effect of individual fixed effects. For the model of multilayer dyadic regression, the developments of Jochmans (2017) are the building block upon which the estimator builds.

The GMM approach that Jochmans (2017) produces relies on the construction of moment conditions that difference out the effect of the unobserved, node-specific, permanent fixed effects. Specifically, the author constructs the following moment condition:

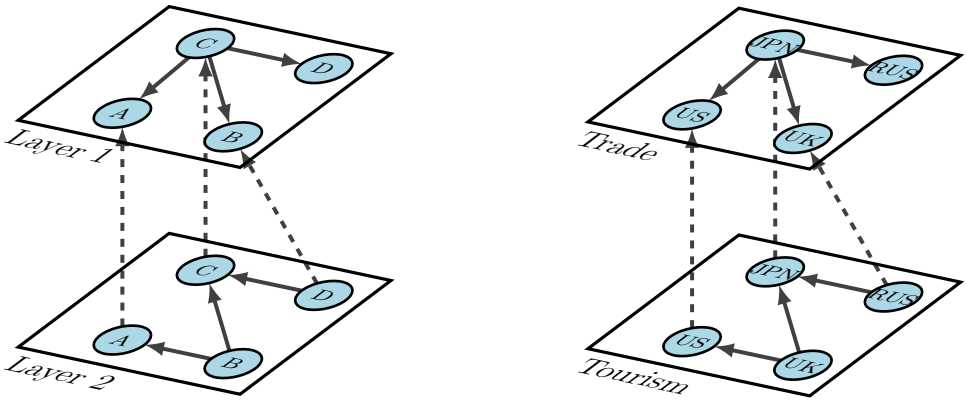
$$E\left[\phi\left(R_{ij}, R_{ij'}, R_{i'j}, R_{i'j'}; \theta_0\right)\left(U_{ij}U_{i'j'} - U_{ij'}U_{i'j}\right)\right] = 0$$

Which holds for all unique choices for (i, i') and (j, j'). The total number of unique choices is defined as, $\varrho = \begin{pmatrix} n \\ 2 \end{pmatrix}\begin{pmatrix} n \\ 2 \end{pmatrix} = \frac{n^2(n-1)^2}{4}$. Using the sample analog of such condition, and choosing an proper instrument $\phi\left(R_{ij}, R_{ij'}, R_{i'j}, R_{i'j'}; \theta_0\right)$, the author is able to achieve consistent estimation and derive its asymptotic properties.

3.2.1 Multi-layer Gravity Model

Next, I take the developments of Jochmans (2017) and use them in a different setting where nodes in the network can connect in different aspects or dimensions. In the example of the gravity model, this can be achieved, for example, by adding other types of flows that occur between countries, like tourism or investment. First, it is necessary to introduce the notion of multilayer networks and define the family of networks that the model can cover.

Following, Boccaletti et al. (2014), a multi-layer network is a pair $\mathcal{M} = (g, \mathcal{C})$ where



(a) Multilayer Network (b) Multilayer Country Network

Figure 3.1: Multilayer Network Examples

$g = \{G_\alpha; \alpha \in \{1, \ldots, M\}\}$ is a class of graphs, $G_\alpha = (X_\alpha, E_\alpha)$. X_α is the set of nodes in G_α and E_α is the set of connections between nodes in the same graph. On one hand, the set of connections that occur between nodes in the same graph are known as "intra-layer" connections. On the other, this framework allows for "inter-layer" connections, i.e, edges between the nodes in different layers. The latter is formally described by the set $\mathcal{C} = \{E_{\alpha\beta} \subseteq X_\alpha \times X_\beta; \alpha, \beta \in \{1, \ldots, M\}, \alpha \neq \beta\}$. An example of a multilayer network with 4 nodes and two layers is depicted in the left panel of Figure 3.1. Notice that intra-layer connections are depicted with solid lines, while inter-layer connections are represented by dashed lines.

For the gravity model example, depicted on the right panel of Figure 3.1, the first layer is trade, and a second layer of connections could be tourism. The interlayer connections represent interactions between the countries that lie across the two layers. One example could be taxes retained for purchases made by tourists. More generally, the multilayer network for the gravity model has the elements of the general description. The set nodes, i.e., the set of countries. The set of intra-layer connections like trade, tourism, investments.

The set of inter-layer connections is bit trickier, but we might think of these as tax treaties, or free migration agreements. In practice, any set of interlayer connections between two layers can be thought as an additional layer by itself. However, for conceptual reasons, specific to the use case, sometimes is better to keep this representation. Notice that I can label each one of these dimensions where the countries are connected as types of connections themselves. Each type of connection has its corresponding layer, or graph. Effectively, I am observing several graphs where the same nodes are present, but where each graph symbolizes a different type of connection.

For the purpose of this model I limit my attention to those multilayer networks where the set of nodes is the same for all graphs and I allow for different connections between them, both intra and inter layer. In order to account for the modifications in the setup, I need to change the model to allow for existence of unobserved permanent shocks that occur at the level of the layer. I also need to allow for several types of connections. The model I propose to model such type-indexed interactions between countries is shown in equation (3.1)

$$Y_{(i;l)(j;k)} = \varphi\left(R_{(i)(j)}; \theta_0\right) U_{(i;l)(j;k)} \tag{3.1}$$

In this case $U_{(i;l)(j;k)} = A_{(i;l)}B_{(j;k)}V_{(i;l)(j;k)}$. Again $A_{(i;l)(j;k)}$ and $B_{(i;l)(j;k)}$ are the node permanent unobserved effects. $V_{(i;l)(j;k)}$ is the dyad idiosyncratic disturbance, which is independent across $(i;l)$, $(j;k)$, and j. In a sense this change of notation is analogous to keep the notation from the mono-layer case but including unobserved effects for each one of the layers involved in the connection. When I deal with an intra-layer connection $k = l$ and $k \neq l$ for inter-layer connections.

An important feature that the model preserves is the idea that the homophily parameter is the same across layers and connections. This means that I consider that the proximity in characteristics affects all types of connections with the same intensity. To gain some perspective lets go back to the multilayered gravity model. This means that country specific observed characteristics have the same effect on the intensity of the connection. For example, the distance between countries will have the same effect on the flow of goods, and people, or investment. The source of heterogeneity in the connection type comes exclusively from the layer fixed effect.

In a similar fashion as Jochmans (2017) I impose a conditional mean restriction

$$E\left[V_{(i;l)(j;k)} \mid R_{11}, \ldots, R_{nn}\right] = 1$$

which in turn implies that $E\left[U_{(i;l)(j;k)} \mid R_{11}, \ldots, R_{nn}\right] = A_{(i;l)}B_{(j;k)}$. Then if it follows that $E\left[V_{(i;l)(j;k)}V_{(i;l)'(j;k)'} \mid R_{11}, \ldots, R_{nn}\right] = 1$ With this I can build the moment conditions analogously to what I observed on the mono layer case. Simply by permuting the indexes involved in the moments I can artificially get moment equations that difference out both layer, and node specific fixed effects. For this to happen a necessary condition is that, analogously to the mono-layer case, the errors of intra-layer and inter-layer connections to be independent, or conditionally independent.

Notice that the term $A_{(i;l)}$ is simply the interaction of a node level fixed effect $A_{(i)}$ and the layer specific fixed effect $C_{(l)}$. In the same fashion, $B_{(i;l)}$ is simply the interaction of a node level fixed effect $B_{(j)}$ and the layer specific fixed effect $D_{(k)}$. Hence, another way to represent the dyadic connection is the disturbance term $U_{(i;l)(j;k)}$ is to explicitly write:

$$U_{(i;l)(j;k)} = A_i B_j C_l D_k V_{(i;l)(j;k)}$$

Which changes the expression for the conditional mean assumption:

$$E\left[U_{(i;l)(j;k)} \mid R_{11}, \ldots, R_{nn}\right] = A_i B_j C_l D_k$$

. The intuition behind is that the individual connection disturbance is unrelated to any of these fixed effects. For nodes in different layers this would imply that the moment conditions can be constructed as follows:

$$E\left[U_{(i;l)(j;k)} U_{(i';l')(j';k')} \mid R_{11}, \ldots, R_{nn}\right] = A_i A_{i'} B_j B_{j'} C_l C_{l'} D_k D_{k'}$$
$$E\left[U_{(i';l')(j;k)} U_{(i';l')(j;k')} \mid R_{11}, \ldots, R_{nn}\right] = A_i A_{i'} B_j B_{j'} C_l C_{l'} D_k D_{k'}$$

Notice that the four connections involved must come from at least two different layers. This setup collapses to the representation of Jochmans (2017) when the number of layers in the network is just one. Once these conditions are found, then you can take the between difference these two to obtain the conditional moment condition:

$$E\left[U_{(i;l)(j;k)} U_{(i';l')} - U_{(i';l')(j;k)} U_{(i';l')(j;k')} \mid R_{11}, \ldots, R_{nn}\right] = 0$$

which in turn leads to the unconditional moment condition:

$$E\left[\phi\left(R_{ij}, R_{ij'}, R_{i'j}, R_{i'j'}; \theta_0\right)\left(U_{(i;l)(j;k)} U_{(i';l')} - U_{(i';l')(j;k)} U_{(i';l')(j;k')}\right)\right] = 0 \tag{3.2}$$

3.3 Estimation

In order to estimate this in a GMM setting I first need to find the sample moment for the equation (3.2). In a close fashion to Jochmans (2017) I let $U_{(i;l)(j;k)}(\theta) = Y_{(i;l)(j;k)}/\varphi(R_{ij};\theta)$. Define the set of all combinations of l, l', k, k' that build moment conditions that satisfy (3.2) as Λ. Then, the empirical moment at a given value of theta θ is obtained with the following expression:

$$\begin{aligned} s(\theta) =& \varrho^{-1} \sum_{(l,l',k,k')\in\Lambda} \sum_{i=1}^{n} \sum_{i<i'} \sum_{j=1}^{n} \sum_{j<j'} \phi(R_{ij}, R_{ij'}, R_{i'j}, R_{i'j'};\theta) \\ & \times (U_{l;ij}(\theta)U_{l;i'j'}(\theta) - U_{l;ij'}(\theta)U_{l;i'j}(\theta)) \end{aligned} \tag{3.3}$$

The GMM estimator:

$$\theta_n = \arg\min_{\theta\in\theta} s(\theta)'\Omega_n s(\theta)$$

The set Θ is the parameter space, and Ω_n is a weight matrix that must be positive definite.

Next I need to define the set of assumptions that allow for the estimation of the parameter. First I need to define some auxiliary terms that will be used in the phrasing of such conditions:

Let $\lambda_1 \equiv (l,k) \in \Lambda_1$, and $\lambda_2 \equiv (l',k') \in \Lambda_2$, where $\lambda_1 \cup \lambda_2 \in \Lambda$. Then, define w_{ij}^{λ} and τ_{ij} as:

$$\tau(R_{ij};\psi) = \varphi'(R_{ij};\theta)/\varphi(x_{ij};\theta)$$

$$w_{ij}^{\lambda_1} = \frac{4}{(n-1)^2} \sum_{(l',k')\in\Lambda_2} \sum_{i'\neq i} \sum_{j'\neq j} \phi\left(R_{ij}, R_{ij'}, R_{i'j}, R_{i'j'}; \theta_0\right) A_i A_{i'} B_j B_{j'} C_l C_{l'} D_k D_{k'}$$

$$\Sigma = -\lim_{n\to\infty} \frac{1}{n^2} \sum_{(l,k)\in\Lambda_1} \sum_{i=1}^{n} \sum_{j=1}^{n} E\left[w_{ij}^{\lambda_1} \tau_{ij}\left(R_{ij}; \theta_0\right)'\right]$$

First I impose the usual conditions to make the problem tractable.

Assumption 3.3.1 *Regularity conditions:*

1. *The set Θ is compact and the true value of the homophily parameter, θ_0, is an interior point.*

2. *ϕ and φ are continuously differentiable in the parameter θ with derivatives ϕ' and φ'*

3. *$\Omega_n \to^p \Omega$, where Ω is a positive definite matrix.*

Next, I define conditions for the identification of the homophily parameter θ_0:

Assumption 3.3.2 *Rank conditions:*

With $\bar{s}(\theta) = \lim_{n\to\infty} s(\theta)$, $\|\bar{s}\left(\theta_k\right)\| \to 0$ implies $\|\theta_k - \theta_0\| \to 0$ for any sequence of vectors $\{\theta_k\}$ from $\mathcal{S}$. The matrix Σ has maximal column rank.

The following assumption makes the case for an independent sampling of the agents involved in the network. In other words, I assume that I sample n independent agents and observe the connections among them. This assumption allows for the existence of dependence between connections, which is particularly relevant in the case of multiple connections of different types.

Assumption 3.3.3 *Sampling conditions: A number n of observations are sampled independently at random from a population.*

Next, I turn my attention towards the conditions to achieve consistency through conventional asymptotic theory.

Assumption 3.3.4 *Boundary conditions:*

1. *There exist finite constants β_u and β_ϕ, independent of θ, such that $E\left[\left\|U_{ij}^{lk}(\theta)\right\|^8\right] < \beta_u$ and $E\left[\left\|\phi\left(R_{ij}, R_{ij'}, R_{i'j}, R_{i'j'}; \theta\right)\right\|^8\right] < \beta_\phi$ for all θ in Θ*

2. *Addtionally, C_l, D_l, A_i, and B_i are finite for all i and all l*

3. *There exists a finite constant β_σ such that $E\left[V_{ij}^4 \mid R_{11}, \ldots, R_{nn}\right] < \beta_\sigma$.*

4. *the conditional variance σ_{ij}^2 is positive and has finite fourth-order moment.*

Finally, to establish the variance estimator, it is necessary that the Jacobian of the moment conditions in the population is well behaved. Since the Jacobian depends only on the paramenter and the covariates, this assumption is almost exactly the same as in Jochmans (2017) .

Assumption 3.3.5 *Jacobian Boundedness condition:*

There exist finite constants β_τ and $\beta_{\phi'}$, independent of θ, such that $E\left[\left\|\tau\left(R_{ij}; \theta\right)\right\|^8\right] < C_\tau$ and $E\left[\left\|\phi'\left(R_{ij}, R_{ij'}, R_{i'j}, R_{i'j'}; \theta\right)\right\|^8\right] < C_{\phi'}$ for all θ in Θ.

Provided all this assumptions hold, I can demonstrate that:

Theorem 1 ***Consistent Estimation****: If assumptions one through four hold, then I have that:*

$$\theta_n - \theta_0 \xrightarrow{p} 0 \qquad as \quad n \to \infty$$

Theorem 1 tells us that even under the presence of independence among nodes of different types, and across layers, I can achieve consistent estimation of the parameter of similarity. The details of the proof follow closely the work of Jochmans (2017), and can be found in the supplementary material for this document. The intuition is rather simple, the variance of the estimator averages overall all possible combinations of dyads across layers, hence it averages over $O(n^8)$ connections. However, any connections that do not share a dyad or a layer index are considered to be independent. In the average this amounts to a zero contribution to the variance of the estimator. The number of dyads that do not share an index is of the order $O(n^7)$. Hence the variance will grow with the order $O(n^-1)$.

Next I need to derive the asymptotic distribution of the estimator. Following Jochmans (2017), I exploit the properties of U-Statistics to approach this task. The estimator portrayed in (3.3) is similar in structure to that of Jochmans (2017), but in the order of the U-statistic. Therefore, the treatment of Jochmans (2017) translates straightforward to my estimator. First, notice that by properties of the U-statistic we know that conditional on $R_{1,1}, \ldots, R_{n,n}$ the estimator at it's true value, $s(\theta_0)$, is asymptotically equivalent to its Hajek projection:

$$ns(\theta_0) = 1/n \sum_{(l,k)\in\Lambda_1} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}^{\lambda_1} (V_{ij}^{\lambda_1} - 1) + o_p(1). \tag{3.4}$$

Where the Hajek projection is the first element of the RHS of equation (3.4). It follows that the elements of the sum in equation (3.4) are independent conditional on $R_{1,1}, \ldots, R_{n,n}$; moreover, they are by construction mean zero. Hence, the variance of the estimator evaluated at θ_0 is bounded by a sequence $O(n^{-2})$. Which in turn provides: $n||\theta_n - \theta_0|| = O_p(1)$. It must be then, by a conditional central limit theorem (Rao, 2009) , it must be that:

$$ns(\theta_0) \rightarrow^d N(0, \Gamma),$$

$$\Gamma = \lim_{n\to\infty} \frac{1}{n^2} \sum_{\lambda_1 \in \Lambda_1}^{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E\left[w_{ij}^{\lambda_1} {w_{ij}^{\lambda_1}}^T {\sigma_{ij}^{\lambda_1}}^2 \right]$$

The next step in Jochmans (2017) is to use a first order expansion of the GMM function around the true parameter. For this to work, the Jacobian of the population moment conditions must be well behaved. Provided assumption five holds, then it must be that the Jacobian $S(\theta_n) \rightarrow^p \sigma$. The first order approximation yields:

Conjecture 3.3.1 *(asymptotic normality). If assumptions 1 to 5 hold and* Γ *is positive definite, then*

$$n\left(\theta_n - \theta_0\right) \xrightarrow{d} N(0, \Upsilon)$$

as $n \to \infty$*, where the covariance matrix is* $\Upsilon = \left(\Sigma'\Omega\Sigma\right)^{-1}\left(\Sigma'\Omega\Gamma\Omega\Sigma\right)\left(\Sigma'\Omega\Sigma\right)^{-1}$.

The proof of this statement follows from the proof for the second theorem of Jochmans (2017). The logical process laid out only has to accomodate for the fact that now we allow for more layers and the estimator I propose is a U-statistic hihger order that the proposed by Jochmans (2017).

To estimate the variance we use consistent estimators of the elements of the asymptotic variance.

$$\Upsilon_n = \left(S_n' \Omega_n S_n\right)^{-1} \left(S_n' \Omega_n \Gamma_n \Omega_n S_n\right) \left(S_n' \Omega_n S_n\right)^{-1}$$

where $S_n = S(\theta_n)$ is the Jacobian of the empirical moment conditions and Γ_n is an estimator of Γ:

$$\Gamma_n = \frac{1}{n^2} \sum_{\lambda_1 \in \Lambda_1}^{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \hat{\gamma}_{ij}^{\lambda_1} \hat{\gamma}_{ij}^{\lambda_1 T}$$

$$\begin{aligned} \hat{\gamma}_{ij}^{\lambda_1} =& \frac{4}{(n-1)^2} \sum_{\lambda_2 \in \Lambda_2}^{n} \sum_{i' \neq i} \sum_{j' \neq j} \phi\left(R_{ij}, R_{ij'}, R_{i'j}, R_{i'j'}; \theta_n\right) \\ & \times \left(\hat{u}_{ij}^{\lambda_1} \hat{u}_{i'j'}^{\lambda_2} - \hat{u}_{ij'}^{\lambda_1} \hat{u}_{i'j}^{\lambda_2}\right) \end{aligned}$$

Provided statement 1 holds, consistency of the variance estimate follows by assumptions 4 and 5, i.e., the bounds assumed for the moment conditions. I do not provide proof for statement 1 as it

3.4 Numerical Experiments

The results from the previous section provide a sound framework to estimate dyadic regression in the context of multilayer networks. This section provides a battery of results that illustrate the efficacy of the estimator under ideal conditions. The numerical experiments follow closely those in Jochmans (2017). While the estimator allows for a flexible functional form, this section focuses on the widely used exponential regression model. A notable

example of this family of models is the gravity model.

It is essential to mention that there is no clear benchmark to measure against the estimator's performance because this is the first approach to model a multilayer network. The alternatives include estimating the original GMM model inspired by Jochman or a modification of the Poisson Pseudo-Maximum-Likelihood that tries to incorporate the idea of data types. However, since these models are ill-equipped to face a structure that escapes their original purpose, a performance comparison would be futile. Hence, the indicators for performance are the small sample behavior results. For evidence of consistency, one can consider bias and variance decrease as the number of nodes grows. Q-Q plots that point towards normality support the conjecture about the asymptotic behavior.

The simulations consider three specifications for the data generating process, two for count data, and one for a continuous outcome. For the first part, count data is highly likely to show in national data. On the other, continuous data is highly prevalent in individual studies. The first scenario is a Poisson model where the conditional mean and variance are the arrival rate: $\lambda_{ij}^{lk} = \exp\{R_{ij}\theta_0\}A_iB_jC_lD_k$. In the second scenario, the data comes from a Negative Binomial distribution. It is possible to express such a distribution as a mixture of Poisson models, where the arrival rate is distributed gamma with positive shape and scale parameters ϕ, and $p_{ij}^{lk} = (1 + \lambda_{ij}^{lk}/\phi)^{-1}$, then by allowing ϕ to take several values one can control for different levels of over-dispersion since the variance is $var(y_{ij}^{lk}|R_ij) = \lambda_{ij}^{lk} + \phi\lambda_{ij}^{lk^2}$. The simulation uses a model for the continuous outcomes where the error originates from a log-normal distribution. This grants positive outcomes: $y_{ij}^{lk} = \lambda_{ij}^{lk}\varepsilon_{ij}^{lk}$ where εij^{lk} is distributed log normal with mean $-1/2ln(1+\sigma_{ij}^{lk^2})$ and variance $ln(1+\sigma_{ij}^{lk^2})$. Where $\sigma_{ij}^{lk^2} \in \{1, \lambda_{ij}^{lk^{-1}}, 1+\lambda_{ij}^{lk^{-1}}, \lambda_{ij}^{lk^{-2}}\}$ This flexible form allows for four cases homoskedastic errors,

Poisson type errors, negative binomial type errors, and homoskedastic outcomes respectively.

In order to generate the conditional mean, the simulation draws the fixed effects, for individuals, from a bivariate normal distribution, allowing for correlation across individuals. Similarly, the process obtains the layer-specific fixed effects. Then, it generates the regressors R_ij allowing them to depend only on the individual fixed effects. Allowing the observable characteristics to be correlated with the layer fixed effects would mean that a connection's observable characteristics are specific to each layer. For example, the distance between countries would be layer-specific in the gravity model. Hence, this way to specify fixed effects for a layer is not very realistic. The final output is an outcome variable, two regressors, one continuous, and one binary.

The results use one of the specifications of ϕ proposed in Jochmans (2017). The GMM estimator uses the instrument $(R_{ij} - R_{ij'}) - (R_{i'j} - R_{i'j'})$. Jochmans (2017) modifies the optimal instruments suggested in the work of Chamberlain (1987). In the setup of this simulation, the instrument remains unchanged since the observable characteristics R_{ij} do not vary across layers. The experiments run a total of a thousand draws for each of the four sample sizes considered: 25, 50, 100, and 200 nodes. Notice, however, that while the number of nodes is 25, the total number of connections grows much faster in a multilayered network.

An essential distinction in this setup is the number of layers in simulation. In order to observe the behavior of the estimator under a different number of layers, the results consider three and four layers for each sample size. Notice that adding an extra layer to the model simultaneously adds several interlayer types of connections. For example, when there are three layers, this opens the possibility of interlayer connections between these three layers. Effectively, six types of connections are possible, three interlayer and three intralayer.

This constitutes adding extra layers for the estimator since the model cannot distinguish between types. This structure makes the number of connections sampled, even for a small number of nodes, to grow very fast. For briefness, this section shows only a handful of the specifications mentioned. A complete table of results, including all specifications, is present in the appendix.

Table 3.1 shows the results for the three-layer DGP. For all specifications and layer combinations, the complete set of results is available in the appendix. This selection of models illustrates the results across all specifications. The table shows the results for θ_1, the coefficient for the continuous regressor, and θ_2 for the binary regressor. In all cases, it shows three results: median bias, the standard deviation, and the interquartile range of the empirical distribution. Notice the median bias as the ratio of the median value vs. the true value minus one. The data from the simulation tells us that as the number of nodes sampled increases, the empirical distribution of the estimator centers around the true value of the coefficient. The evidence for this claim is the decreasing bias and dispersion of the empirical distribution. This fact occurs both for binary and continuous cases. The behavior of all other DGP's considered also shows the same behavior. The evidence backs up the theoretical results obtained in section 3.3, which predicted the consistency of the estimator. Notice, however, that the binary outcome does not show a uniform decrease for all specifications; this also seems to occur in the results for four layers. The result could point towards a different choice of instrument, as suggested by Jochmans (2017). Alternatively, the overall decrease in spread and bias, at $N = 200$, means is a cautionary tale about the small sample behavior of this estimator when dealing with binary regressors.

Regarding the normality conjecture, figure 3.2 uses a Q-Q plot to depict the quantiles

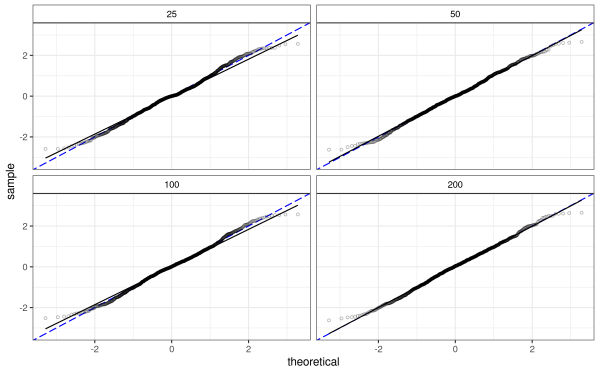
Table 3.1: Simulation Results for a 3 layer DGP

		$\theta_1 = -1$			$\theta_2 = 1$		
	N	Med. Bias	Std. Dev	IQ Range	Med. Bias	Std. Dev	IQ Range
Log Normal							
$\sigma_{ij}^{lk^2} = 1$	25	-0.0018	0.0517	0.0642	0.0088	0.2005	0.2540
	50	0.0007	0.0273	0.0357	0.0025	0.1062	0.1417
	100	-0.0003	0.0148	0.0191	0.0008	0.0558	0.0673
	200	-0.0003	0.0084	0.0113	0.0022	0.0295	0.0386
Negative Binomial							
$\phi = 10$	25	0.0017	0.0367	0.0478	-0.0077	0.1840	0.2076
	50	0.0008	0.0197	0.0250	0.0014	0.0826	0.1046
	100	0.0004	0.0102	0.0124	0.0017	0.0409	0.0516
	200	-0.0002	0.0050	0.0065	-0.0006	0.0191	0.0246
Poisson							
	25	0.0008	0.0362	0.0434	-0.0109	0.1651	0.1839
	50	0.0014	0.0175	0.0216	-0.0023	0.0749	0.0885
	100	0.0003	0.0085	0.0101	0.0000	0.0347	0.0421
	200	-0.0002	0.0043	0.0053	0.0002	0.0169	0.0217

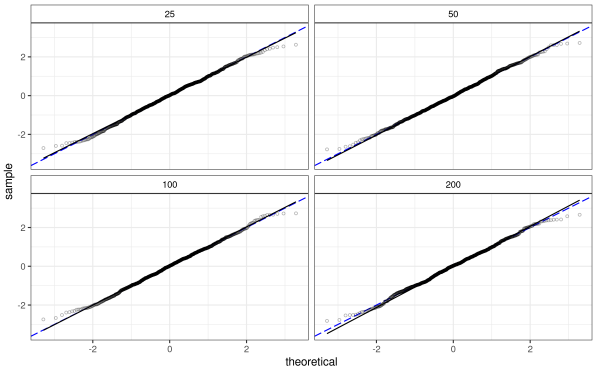
of the estimated distribution against the theoretical quantiles of the normal distribution. Analogously, the figure provides a glimpse of a curated selection of models. All the Q-Q plots for all combinations of specifications and layers are available in the appendix. The black line on each panel represents the forty-five-degree line. The blue dashed line is the fit using the quantiles 25th and 75th. Ideally, a distribution close to a normal would have all the points plotted on top of the black line and the blue line perfectly overlapping the black one. Figure 3.2 tells that as the number of sampled nodes increases, the behavior of the empirical distribution is closer to that of a theoretical normal distribution.

In summary, the simulations show that the GMM estimator is a feasible way to estimate the homophily parameters in an exponential regression model of network formation. Moreover, in the context of a finite number of layers, with the evidence pointing out that the

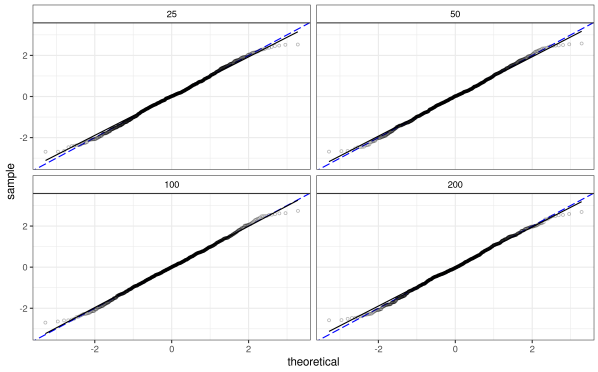
Figure 3.2: Q-Q plot simulation results for 3 layers



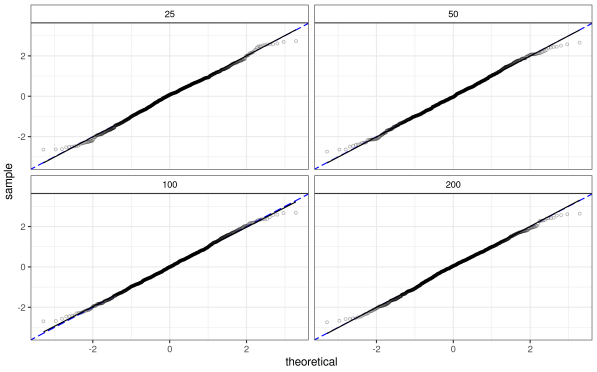
(a) Log-Normal Binary



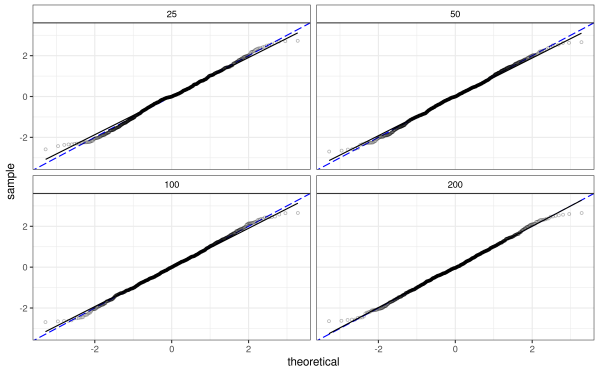
(b) Log-Normal Continuous



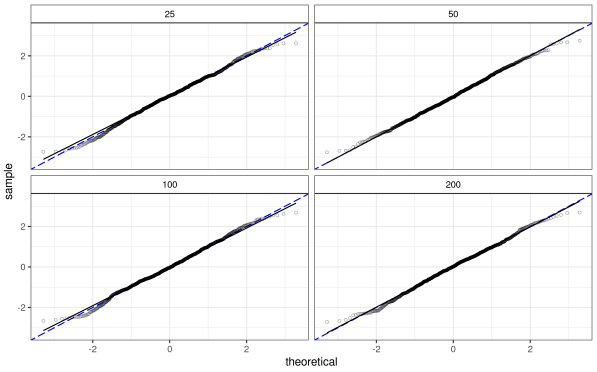
(c) Negative Binomial Binary



(d) Negative Binomial Continuous



(e) Poisson Binary



(f) Poisson Continuous

Note: The results stem from a 1000 simulations. THe specifications for this figure match the ones of table 3.1. Additional specifications available in the appendix.

conjecture about the asymptotic behavior is possibly true, it is also a framework for sound inference. More generally, we know that the performance of the estimator will depend on the functional form of the instrument $\phi(R_{ij};\theta)$ and the specific form of φ.

3.5 Empirical Application

This section aims to illustrate using the method with a real-world example of a multilayer network. The case in question is that of a network of countries. The original gravity model estimates a model using international trade between countries. Countries, however, are connected in several different ways other than trade. The following analysis uses three types of connections: trade, foreign direct investment, and tourism. From a conceptual point of view, the connections in each of these layers occur only within the same layer. For example, one cannot observe a connection between the trade and tourism layers. Hence countries are connected within each aspect or layer and not across them. The monolayer analogous of this network is one where two countries are connected if there is any positive amount of either trade, investment, or tourism between them. As mentioned previously, the advantage gained by reframing the problem in the multilayer model is that it can now observe more complex intricacies of the data that the monolayer version of the graph would not allow. One advantage, for example, is that the estimator allows for each connection to have its own independent idiosyncratic disturbance.

3.5.1 Data

In order to model the trade network, this analysis uses the gravity data set (Head et al., 2010) from the French institute CEPII, which has data for 252 countries of Bilateral trade flows, Geographical distance, Trade facilitation measures, Proxies for cultural proximity, and Macroeconomic indicators. The data set spans from 1948 to 2021 and is an ensemble of measures that borrows from several sources, including the UN, International Monetary Fund, and others. This data set comprises the information for the layer on international trade and the observable characteristics we use for estimating the homophily parameter. The total number of variables in this data set is 179.

The information on the tourism layer comes from the World Tourism Organization UNTWO (2021). The data come from administrative records that each country collects about visitors. UNTWO collects and organizes the information in a data set for their annual report on tourism statistics. Because of the administrative origin of the statistics, there is heterogeneity in the indicators each country reports. For example, some countries report the number of visitors that arrived at their borders, while others report the number of visitors registered at hotels. The analysis's imputation logic is that the number of visitors at the border by country of residence is the best indicator. In case of the absence of this indicator, the variable replaces it with an alternative present on the country's data set. For example, if the number of visitors that arrived at the border by residence is not present, the next best thing is the number of visitors that arrived at the border by nationality. That way, the analysis maximizes the number of countries with information available. Hence, it is crucial to understand that the tourism flow in this data set is not a monolithic measure but

constructed from imperfect sources. After these modifications, the tourism data contains information about 125 countries.

The data set uses the Coordinated Direct Investment survey (IMF, 2015) from the International Monetary Fund for the investment layer. This data set contains direct foreign investment positions reported by countries at the flow's source and destination. Loosely speaking, Foreign direct investment is a type of investment that seeks to establish a lasting investor relationship with a foreign firm. For a deeper look at the different definitions of foreign direct investment, from now on referred to as FDI, see IMF (2015). This analysis uses a foreign direct investment position definition that does not include equity and debt instruments. The data available spans from 2009 until 2020 for 189 countries. A particularity worth mentioning is that FDI can, in some instances, take negative values. The data for this section replaces negative values in the data with zeroes.

For the implementation, the building block of the data is the trade data, which also happens to be the one that reports the largest number of countries. The final data preserves all the countries in the trade data set. Hence, some imputations happen to maintain the structure of the trade layer. For example, some countries report trade data but do not report tourism or foreign direct investment data. Whenever this happens, the assumption that no connection is present, i.e., the connection has value zero. Additionally, the self-loops, connections from a country to itself, are set to zero. The final working data set contains data for 247 countries across three layers from 2015 to 2019. The observable characteristics come from the trade data set. While most observable characteristics are not likely to change from year to year, the results include estimations for all five years to understand the estimator's stability to changes in the outcomes. In total, there are 247^2 connections in each of the three

Table 3.2: Summary Statistics

		¿0 rate	Mean	SD	Median
Dependent Variable	Comercial Trade	0.285	858.149	7667.072	3.121
	IMF trade	0.469	651.909	6552.748	1.998
	FDI origin	0.107	4797.635	36130.394	29.608
	FDI destination	0.081	6747.261	45107.915	81.192
	Tourism	0.133	101.003	850.151	1.006
Observables	Log-distance	1.000	1.574	1.044	1.907
	Contiguous		0.010	0.099	0.000
	Common Language		0.099	0.298	0.000
	Colonial Past		0.143	0.350	0.000
	Trade Agreement		0.144	0.351	0.000

layers. It is worth mentioning that thanks to the matching and imputation effort, any result from this analysis should be taken with a grain of salt. Table 3.2 shows some characteristics of the data. The upper panel contains insights about the different dependent variables considered for the results. Notice that the percentage of connections showing positive values is never higher than 50%. This dispersion increases within the layers where the imputations happened: FDI and tourism. The distributions are scaled to units that make the numbers across different layers comparable in magnitude.

The controls are the same list of regressors used by Jochmans (2017) for comparability. Additional specifications considering transformations of these regressors showed minimal changes in the stability of the estimator. Table 3.2 shows descriptive statistics about the variables used as controls. The selection of regressors is a set of random variables that provide relatedness measures between the countries. The most important measure is distance; however other qualitative measures can also describe relations between different countries. For example, a dummy for a shared official language, a dummy for a similar colonial history, and whether the two countries have a trade agreement in place. The data set has observed

information for the controls for the 61 thousand pairs of countries considered. A noticeable feature is the low percentage of connections with similarities in contiguity, language, colonial past, trade agreements. Of the 61 thousand possible pairs of countries, only 10% are contiguous or have a common language, regarding trade agreement and colonial paste a slightly higher 14%.

3.5.2 Results

Table 3.3 show the results of the estimation where the dependent variables are Comercial Trade, FDI recorded at the destination and Tourism. Unlike Jochmans (2017), the results do not censor the connections without positive values. Since the purpose of this estimator is to model complex networks, ignoring the connections without positive values would effectively bias the model for the formation of links between countries. An additional specification is also part of the results, where the dependent variables are the trade reported by the IMF, the FDI reported at the origin and tourism. The purpose of the second specification and the repeated cross-sections is to shed light on the stability of the estimator performance. Since the results of the second specification tell the same story as the ones from table 3.3, they are available in the appendix.

The results follow what the economic theory and intuition dictate, as they also agree with the empirical evidence from previous work like Jochmans (2017); Silva and Tenreyro (2006). The basic idea of the gravity model, hence its name, is that trade tends to diminish as countries are further apart. Hence one expects a negative sign for the coefficient on distance. Notice that in table 3.3 the coefficients for Log-distance are between -0.34 to -0.37.

Table 3.3: Multilayer gravity estimates for Trade, Tourism, and FDI

	2015	2016	2017	2018	2019
Log-distance	-0.346	-0.349	-0.379	-0.375	-0.357
	(0.012)	(0.013)	(0.012)	(0.011)	(0.011)
Contiguous	0.443	0.441	0.425	0.472	0.59
	(0.024)	(0.024)	(0.023)	(0.023)	(0.028)
Common Lang.	0.636	0.579	0.567	0.647	0.594
	(0.023)	(0.022)	(0.021)	(0.02)	(0.025)
Colonial Past	0.764	0.744	0.67	0.667	0.721
	(0.028)	(0.029)	(0.028)	(0.026)	(0.03)
Trade Agreement	0.654	0.635	0.654	0.624	0.452
	(0.016)	(0.016)	(0.016)	(0.016)	(0.018)

Additionally, all estimates for all are statistically different from zero. Even the magnitude of the estimate, which should be affected by additional layers, resembles that of previous findings.

Next, the intuition behind the similarity indicators is that the more similar two countries are, the more they are likely to connect in any of the aspects. Hence, it is reasonable to expect a positive sign from the coefficients for these indicators. The results from 3.3 show that all similarity indicators have a positive effect on the connections evaluated. They also are statistically different from zero and resemble the magnitude of the coefficients found in earlier literature.

In terms of the estimator's performance, the stability across time and specifications show that the estimates are robust. Also, the estimator for the standard error of each coefficient is stable. This evidence supports, even if weakly, the conjecture about the asymptotic behavior of the estimator. Overall, the estimator provides a feasible way to stably estimate a multilayer gravity model. If the conjecture for the asymptotic distribution is true, then the estimator also provides a sound framework for inference on the parameters. The caveat to

this positive performance is that the estimates could be improved or worsened by choosing another functional form for the instrument and the conditional mean model.

3.6 Conclusions

The renewed interest in network econometrics and the boom in available network data have pushed the field to search for newer ways to deal with data. The availability of multilayer network data is an opportunity to understand better complex relations between agents that can have more than one vehicle or meaning. In this context, this effort takes a shot at extending a well-established model to work in the realm of multilayer network information.

First, we rewrite the model in the context of a multilayer network. Then, I find a set of moment conditions equipped to deal with multilayer data using the framework provided by Jochmans (2017). Moreover, it is shown that for any network with two or more layers, one can always build such moment conditions and difference out individual and layer-specific fixed effects. Next, using the sample analog for the population moment conditions and profiting from symmetry, we can use standard GMM to estimate the parameters for the model. Furthermore, I propose and prove that the estimator achieves consistency and conjecture that it is normally distributed asymptotically. Then, a simulation shows the behavior of the estimator in finite samples. The results show that as the number of nodes grows, the bias and dispersions of the estimator are reduced. Additionally, reasonable supporting evidence for the conjecture of asymptotic normality is provided through visual inspection of Q-Q plots of the empirical distribution of the simulated estimators. Finally, an application to real-world data is showcased. The application uses a data set explicitly constructed for this purpose.

The results show that the estimator is stable, and its estimates are consistent with theory and previous literature.

Appendix

Proofs

Proof of Theorem 1. We follow a similar approach to that of Jochmans (2017). To establish consistency of the estimator we need to show that conditions 1-4 of theorem 2.1 of Newey and McFadden (1994). The first two assumptions imply that conditions 1 through three are satisfied. Condition 4 states that $s(\theta)$ converges in probability to $\bar{s}(\theta)$ uniformly on $\mathcal{S}$ and remains to be shown. By definition we need to show that

$$\lim_{n\to\infty} \Pr\left(\sup_{\theta\in\mathcal{S}} \|s(\theta) - E[s(\theta)]\| > \epsilon\right) = 0$$

for any $\epsilon > 0$.

First define the set Λ as the set of all combinations of layer indexes (l, k, l', k') that satisfy equation (3.2). Define γ_L as the cardinality of the set Λ. For the purpose of this document, the total number of layers L is fixed, hence we interpret γ_L as a constant.

Analogously, define the $\mathcal{I}$ as the set of combinations of node indexes (i, j, i', j') that satisfy equation (3.2). Define ϱ_n as the cardinality of the set $\mathcal{I}$. There is a close expression for ϱ_n:

$$\varrho_n = \left(\frac{n!}{2!(n-2)!}\right)^2 = \frac{n^2(n-1)^2}{4},$$

this expression is of order $O(n^4)$.

For convenience first we need to establish a couple of shorthand notations. First, define

$v(\iota, \lambda, \theta) = \phi(R_{ij}, R_{ij'}, R_{i'j}, R_{i'j'}; \theta)\left(U_{ij}^{lk}(\theta)U_{i'j'}^{l'k'}(\theta) - U_{ij'}^{lk}(\theta)U_{i'j}^{l'k'}(\theta)\right)$. Where the indexes $(i, j, i', j') = \iota \in \mathcal{I}$, and $(l, k, l', k') = \lambda \in \Lambda$. Then, we can rewrite the difference of v and its expectation as: $\bar{v}(\iota, \lambda, \theta) = v(\iota, \lambda, \theta) - E[v(\iota, \lambda, \theta)]$

Using these, we can rewrite the difference of the estimator $s(\theta)$ and its expectation as:

$$s(\theta) - E[s(\theta)] = \varrho_n^{-1}\gamma_L^{-1}\sum_{\lambda\in\Lambda}\sum_{\iota\in\mathcal{I}}\bar{v}(\iota, \lambda, \theta)$$

First determine whether $\bar{v}(\iota, \lambda, \theta)$ is bounded. By Cauchy-Schwarz inequality we can determine that:

$$E\left[\left\|\phi(R_{ij}, R_{ij'}, R_{i'j}, R_{i'j'}; \theta)\, U_{ij}^{lk}(\theta)U_{i'j'}^{l'k'}(\theta)\right\|^2\right]^2$$

is bounded by:

$$E\left[\|\phi(R_{ij}, R_{ij'}, R_{i'j}, R_{i'j'}; \theta)\|^4\right] \times \sqrt{E\left[\left\|U_{ij}^{lk}(\theta)\right\|^8\right]}\sqrt{E\left[\left\|U_{i'j'}^{l'k'}(\theta)\right\|^8\right]}$$

By assumption 4 it must be that each one of these terms is uniformly bound on $\mathcal{S}$ for any i, i', j, j' and λ. Hence, there must exist constant B such that:

$$E\left[\|v(\iota, \lambda, \theta)\|^2\right] < B$$

This, in turn, implies that the variance of $s(\theta)$ is uniformly bounded. Chebychev's inequality yields

$$\Pr(\|s(\theta) - E[s(\theta)]\| > \epsilon) \leq \frac{E\left[\|s(\theta) - E[s(\theta)]\|^2\right]}{\epsilon}$$

Next step in the proof is to show that the variance estimator goes to zero. Notice that, $Var(s(\theta)) = E\|\frac{\varrho^{-1}\gamma^{-1}}{4}\sum_{\lambda}\sum_{\iota}\bar{v}\left(\iota,\lambda,\theta\right)\|^2$. Rearranging terms, and using Minkowski's inequality we can find a bound for the variance:

$$\varrho_n^{-2}\gamma_L^{-2}\sum_{\lambda}\sum_{\iota}\sum_{\lambda'}\sum_{\iota'}E\left[\|\bar{v}\left(\iota,\lambda,\theta\right)\|\|\bar{v}\left(\iota',\lambda',\theta\right)\|\right]$$

Notice that this expression is the sum of the covariance of the terms $\bar{v}\left(\iota,\lambda,\theta\right)$ and $\bar{v}\left(\iota',\lambda',\theta\right)$. We know, by assumption 3, that connections are independent. More over, if none of the indexes in (ι,λ) are present in (ι',λ') we know that their correlation is equal to zero. The asymptotic behavior of this bound is determined by this particularity.

The total number of possible combinations of indexes is the square of the product of the cardinality of the set $\mathcal{I}$ and the cardinality of the set Λ: $(\varrho_n\gamma_L)^2$. This is the denominator of the bound of the variance. The order of magnitude of this term is $O(n^8)\times\gamma_L^2$, however Notice that only ϱ_n is important to determine the asymptotic behavior since γ_L is fixed. Hence, this term is of order $O(n^8)$.

The numerator behavior will be determined by the number of combinations of indexes that show overlapping indexes. We use all possible combinations of λ, λ', ι , and ι' such that:

- at least one index is shared between λ and λ', or
- at least one index is shared between ι and ι'

For fixed λ, and λ' there are $O(n^7)$ combinations of ι and ι' that have overlapping indexes. For all possible combinations of λ, and λ', there are $O(n^7) \times \gamma_L^2$ combinations of ι and ι' that have overlapping indexes. Finally, since we also restrict the combinations to share an index between λ and λ', the number of combinations is of order $O(n^7) \times (\gamma_L^2 - \delta)$ where *delta* is the number of combinations of λ and λ' that do not share an index. Since nor γ_L nor δ depend on n, their behavior is not important to understand the asymptotic behavior of the numerator. Therefore the numerator is of order $O(n^7)$. Hence, we can conclude that uniformly on $\mathcal{S}$:

$$\Pr(\|s(\theta) - E[s(\theta)]\| > \epsilon) = O\left(n^{-1}\right)$$

converging to zero as $n \to \infty$ for any $\epsilon > 0$. Therefore we have shown that condition (4) of Newey and McFadden (1994) is satisfied. In turn, this implies that

$$\theta_n - \theta_0 \xrightarrow{p} 0 \qquad \text{as} \quad n \to \infty$$

■

Implementation of evaluation function

In the more general form the estimation procedure supposes the construction of a sample moment condition that looks like this:

$$s(\theta) = \varrho^{-1} \sum_{l=1}^{m} \sum_{l<l'} \sum_{k=1}^{m} \sum_{k<k'} \sum_{i=1}^{n} \sum_{i<i'} \sum_{j=1}^{n} \sum_{j<j'} \phi\left(R_{ij}, R_{ij'}, R_{i'j}, R_{i'j'}; \theta\right)$$
$$\times \left(U_{l;ij}(\theta) U_{l;i'j'}(\theta) - U_{l;ij'}(\theta) U_{l;i'j}(\theta)\right)$$

The issue that originally Jochmans (2017) found in his contribution was that the high order of the U-statistic. The original estimation procedure had to loop around four summations with n terms. This meant that to evaluate this function, that is finding its critical points, by mere linear search over a grid involved a set of operations $O(n^4)$. Hence as soon as the number of observations went a little over the tenths the evaluation was cumbersome, and time consuming, at best, and impossible at worst. Jochmans (2017), uses the symmetry in the construction of the instrument in order to provide a specific solution.

$$s(\theta) = \varrho^{-1} \sum_{l=1}^{m} \sum_{l<l'} \sum_{k=1}^{m} \sum_{k<k'} \sum_{i=1}^{n} \sum_{i<i'} \sum_{j=1}^{n} \sum_{j<j'} (R_{ij} - R_{ij'}) - (R_{i'j} - R_{i'j'})$$
$$\times \left(U_{ij}^{lk}(\theta) U_{i'j'}^{l'k'}(\theta) - U_{ij'}^{lk}(\theta) U_{i'j}^{l'k'}(\theta)\right)$$

We will define the set of all combinations of l, l', k, k' that comply with the moment construction in section 3.2 as Λ. Hence we can show, using a similar logic as Jochmans (2017), that we can express the previous equation as:

$$s(\theta) = \varrho^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} R_{ij} \sum_{(l,l',k,k') \in \Lambda} \left(U_{ij}^{lk}(\theta) \bar{U}_{\cdot\cdot}^{l'k'}(\theta) - \bar{U}_{i\cdot}^{lk}(\theta) \bar{U}_{\cdot j}^{l'k'}(\theta)\right)$$

Where $\bar{U}_{i\cdot}^{lk}(\theta) = \sum_{j=1}^{n} U_{ij}^{lk}(\theta)$, $\bar{U}_{\cdot j}^{l'k'}(\theta) = \sum_{i=1}^{n} U_{ij}^{l'k'}(\theta)$, and $\bar{U}_{\cdot\cdot}^{l'k'}(\theta) = \sum_{i=1}^{n} \sum_{j=1}^{n} U_{ij}^{l'k'}(\theta)$

This means that instead of doing the $O(n^4)$ operations for the inner loops we can use averages on each one of the layers to make the process considerably more concises. Notice however, that evaluating all possible combinations of (l, l', k, k') that are suitable is also an

$O(m^4)$ process by itself. If we were to evaluate the suitability of each combination, even without executing the mathematical operations, in every evaluation iteration it the whole algorithm becomes unfeasible. However, since the combination of feasible (l, l', k, k') does not depend on the indexes of the nodes, we can easily **pre-calculate** the elements of this set only once before proceeding to evaluate the function. Until the closing of this version, I have not been able to find a closed expression to define the set. However I have devised an algorithm that finds such combinations from the requirements of independence imposed in section 3.2 to build independent moment conditions. This makes that instead of evaluating m^4 possible outcomes we end up only evaluating at most $\frac{1}{4}m^4$. For example in a network with $m = 20$, we would evaluate exactly 36,290 cases instead of 160,000. Since for the scope of this work we suppose that the number of layers is finite and small, the evaluation process becomes feasible. A somewhat similar reasoning can be derived for the second instrument. A brief description of the algorithm is shown below:

```
# m is the number of layers

M<-function(m) {
#define upper bound
upper_bound=round((m^4)/4)

#initialize
r<-0
for l in 1:m {
    for k in l:m {
        veco=1:m
        veco = select elements not in (l,k)
        for l' in veco{
           voco = select elements such that veco >= l'
           for k' in voco {
            p = create combinations (l,k,l',k')

            increase counter r+1
            store comb[r] = p
            }
            }
            }
        }

label comb accordingly
return comb

}
```

Numerical Experiments Extended Results: Consistency evidence

Table 3.4: Simulation Results for a 3 layer DGP

		$\theta_1 = -1$			$\theta_2 = 1$		
	N	Med. Bias	Std. Dev	IQ Range	Med. Bias	Std. Dev	IQ Range
Log Normal							
$\sigma_{ij}^{lk^2} = 1$	25	-0.0018	0.0517	0.0642	0.0088	0.2005	0.2540
	50	0.0007	0.0273	0.0357	0.0025	0.1062	0.1417
	100	-0.0003	0.0148	0.0191	0.0008	0.0558	0.0673
	200	-0.0003	0.0084	0.0113	0.0022	0.0295	0.0386
$\sigma_{ij}^{lk^2} = \lambda_{ij}^{lk^{-1}}$	25	0.0038	0.0336	0.0405	-0.0006	0.1663	0.1747
	50	0.0015	0.0169	0.0194	-0.0007	0.0694	0.0807
	100	0.0007	0.0079	0.0093	-0.0024	0.0342	0.0413
	200	0.0003	0.0040	0.0053	-0.0011	0.0166	0.0203
$\sigma_{ij}^{lk^2} = 1 + \lambda_{ij}^{lk^{-1}}$	25	0.0007	0.0681	0.0747	-0.0025	0.3125	0.3112
	50	0.0010	0.0320	0.0402	-0.0050	0.1229	0.1643
	100	0.0006	0.0171	0.0236	-0.0026	0.0680	0.0835
	200	-0.0004	0.0089	0.0118	0.0009	0.0339	0.0445
$\sigma_{ij}^{lk^2} = \lambda_{ij}^{lk^{-2}}$	25	0.0099	0.0439	0.0410	-0.0075	0.2584	0.2041
	50	0.0038	0.0253	0.0240	-0.0051	0.1336	0.1091
	100	0.0023	0.0147	0.0138	-0.0021	0.0811	0.0615
	200	0.0013	0.0078	0.0078	-0.0006	0.0335	0.0320
Negative Binomial							
$\phi = 1$	25	0.0002	0.0638	0.0813	0.0248	0.2743	0.3249
	50	-0.0002	0.0338	0.0443	0.0003	0.1395	0.1721
	100	0.0011	0.0174	0.0227	0.0002	0.0697	0.0888
	200	0.0001	0.0091	0.0116	0.0007	0.0347	0.0447
$\phi = 10$	25	0.0017	0.0367	0.0478	-0.0077	0.1840	0.2076
	50	0.0008	0.0197	0.0250	0.0014	0.0826	0.1046
	100	0.0004	0.0102	0.0124	0.0017	0.0409	0.0516
	200	-0.0002	0.0050	0.0065	-0.0006	0.0191	0.0246
$\phi = 5$	25	0.0013	0.0399	0.0530	0.0067	0.2045	0.2294
	50	-0.0001	0.0211	0.0272	0.0012	0.0924	0.1139
	100	0.0005	0.0102	0.0135	0.0017	0.0448	0.0573
	200	0.0001	0.0056	0.0071	0.0005	0.0223	0.0291
Poisson							

25	0.0008	0.0362	0.0434	-0.0109	0.1651	0.1839
50	0.0014	0.0175	0.0216	-0.0023	0.0749	0.0885
100	0.0003	0.0085	0.0101	0.0000	0.0347	0.0421
200	-0.0002	0.0043	0.0053	0.0002	0.0169	0.0217

Note:
Bias expressed as median over the true value minus one

Table 3.5: Simulation Results for a 4 layer DGP

		$\theta_1 = -1$			$\theta_2 = 1$		
	N	Med. Bias	Std. Dev	IQ Range	Med. Bias	Std. Dev	IQ Range
Log Normal							
$\sigma_{ij}^{lk^2} = 1$	25	0.0015	0.0399	0.0464	0.0010	0.1629	0.2005
	50	-0.0013	0.0226	0.0283	-0.0008	0.0826	0.1037
	100	-0.0003	0.0122	0.0159	0.0020	0.0450	0.0577
	200	-0.0002	0.0067	0.0085	0.0008	0.0241	0.0299
$\sigma_{ij}^{lk^2} = \lambda_{ij}^{lk^{-1}}$	25	0.0019	0.0230	0.0252	-0.0023	0.1239	0.1239
	50	0.0010	0.0110	0.0136	-0.0037	0.0461	0.0554
	100	0.0000	0.0055	0.0068	0.0003	0.0224	0.0285
	200	-0.0001	0.0027	0.0034	-0.0002	0.0112	0.0137
$\sigma_{ij}^{lk^2} = 1 + \lambda_{ij}^{lk^{-1}}$	25	0.0021	0.0472	0.0611	0.0043	0.1983	0.2457
	50	0.0000	0.0248	0.0326	-0.0021	0.0995	0.1319
	100	-0.0006	0.0132	0.0167	0.0036	0.0494	0.0664
	200	-0.0006	0.0072	0.0096	-0.0011	0.0262	0.0342
$\sigma_{ij}^{lk^2} = \lambda_{ij}^{lk^{-2}}$	25	0.0072	0.0332	0.0294	-0.0045	0.1463	0.1365
	50	0.0026	0.0178	0.0158	-0.0025	0.0860	0.0773
	100	0.0014	0.0094	0.0089	-0.0003	0.0469	0.0441
	200	0.0005	0.0049	0.0052	0.0000	0.0216	0.0231
Negative Binomial							
$\phi = 1$	25	0.0032	0.0501	0.0658	-0.0180	0.1999	0.2575
	50	0.0017	0.0249	0.0342	-0.0013	0.1020	0.1335
	100	0.0004	0.0131	0.0170	-0.0003	0.0503	0.0669
	200	-0.0007	0.0069	0.0090	0.0000	0.0268	0.0346
$\phi = 10$	25	0.0011	0.0266	0.0328	-0.0022	0.1255	0.1478
	50	0.0000	0.0131	0.0165	0.0012	0.0577	0.0750
	100	-0.0003	0.0067	0.0090	-0.0001	0.0282	0.0375
	200	-0.0001	0.0035	0.0046	0.0010	0.0142	0.0195
$\phi = 5$	25	0.0003	0.0306	0.0394	0.0078	0.1407	0.1669
	50	0.0010	0.0156	0.0202	-0.0003	0.0638	0.0808

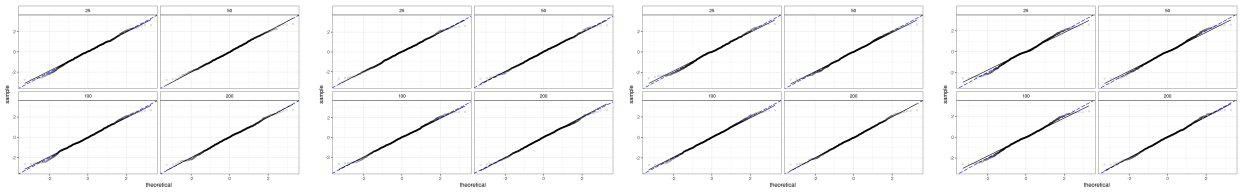
	100	0.0002	0.0077	0.0101	-0.0018	0.0320	0.0416
	200	0.0002	0.0041	0.0052	-0.0005	0.0158	0.0216
Poisson							
	25	-0.0008	0.0241	0.0285	0.0065	0.1652	0.1290
	50	0.0004	0.0114	0.0140	-0.0008	0.0480	0.0596
	100	-0.0001	0.0055	0.0068	0.0014	0.0238	0.0293
	200	-0.0001	0.0026	0.0034	-0.0005	0.0119	0.0144

Note:

Bias expressed as median over the true value minus one

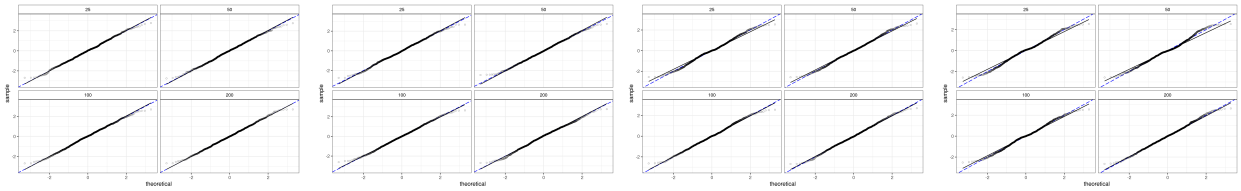
Numerical Experiments Extended Results: Q-Q plots

Figure 3.3: Q-Q plot, simulation results: Poisson Specification



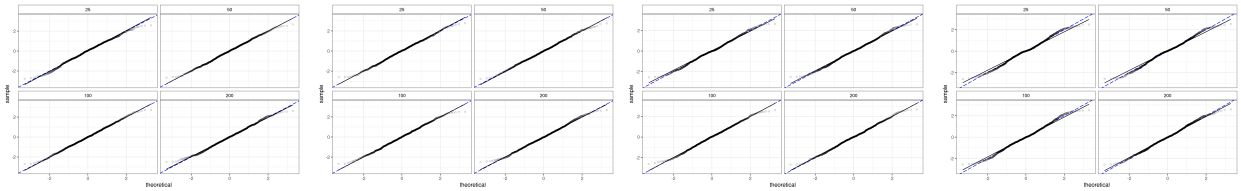
(a) 3 layer Continuous (b) 4 layer Continuous (c) 3 layer Binary (d) 4 layer Binary

Figure 3.4: Q-Q plot, simulation results: Negative Binomial $\phi = 1$



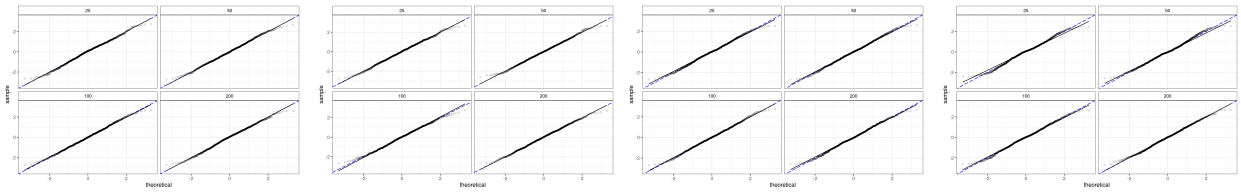
(a) 3 layer Continuous (b) 4 layer Continuous (c) 3 layer Binary (d) 4 layer Binary

Figure 3.5: Q-Q plot, simulation results: Negative Binomial $\phi = 5$



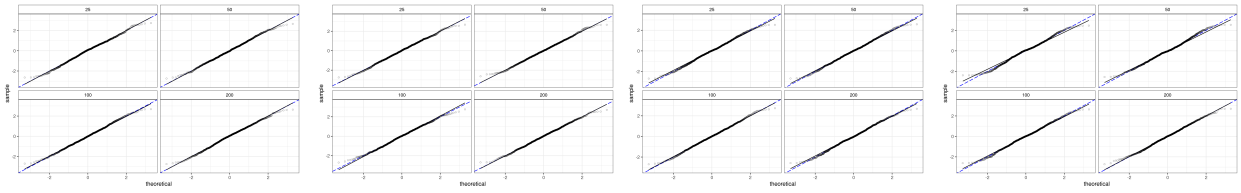
(a) 3 layer Continuous (b) 4 layer Continuous (c) 3 layer Binary (d) 4 layer Binary

Figure 3.6: Q-Q plot, simulation results: Negative Binomial $\phi = 10$



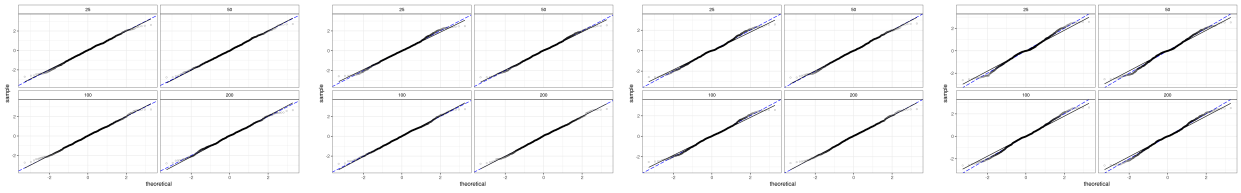
(a) 3 layer Continuous (b) 4 layer Continuous (c) 3 layer Binary (d) 4 layer Binary

Figure 3.7: Q-Q plot, simulation results: Negative Binomial $\phi = 10$



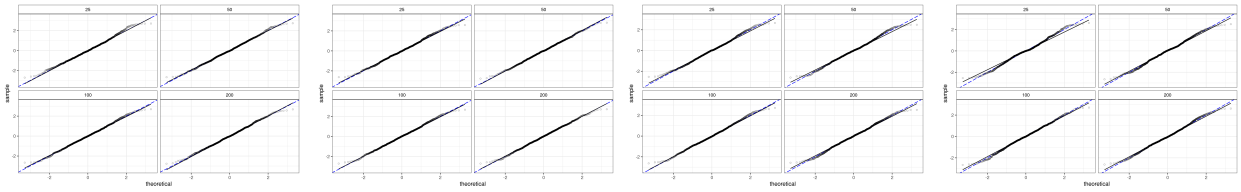
(a) 3 layer Continuous (b) 4 layer Continuous (c) 3 layer Binary (d) 4 layer Binary

Figure 3.8: Q-Q plot, simulation results: Lognormal ${\sigma_{ij}^{lk}}^2 = 1$



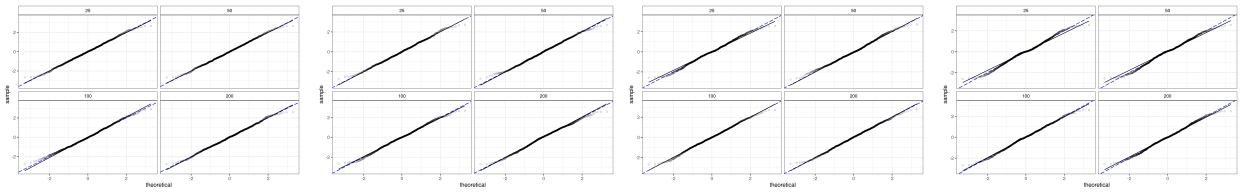
(a) 3 layer Continuous (b) 4 layer Continuous (c) 3 layer Binary (d) 4 layer Binary

Figure 3.9: Q-Q plot, simulation results: Lognormal ${\sigma_{ij}^{lk}}^2 = {\lambda_{ij}^{lk}}^{-1}$



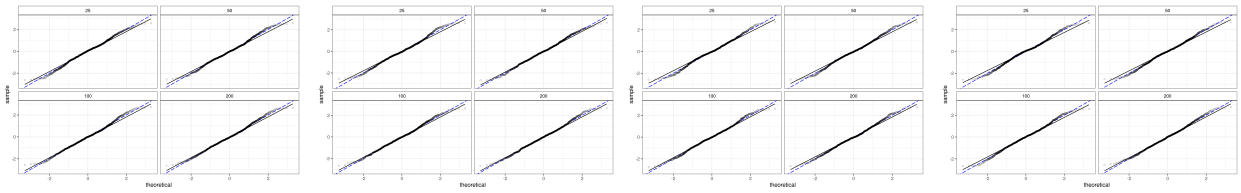
(a) 3 layer Continuous (b) 4 layer Continuous (c) 3 layer Binary (d) 4 layer Binary

Figure 3.10: Q-Q plot, simulation results: Lognormal ${\sigma_{ij}^{lk}}^2 = 1 + {\lambda_{ij}^{lk}}^{-1}$



(a) 3 layer Continuous (b) 4 layer Continuous (c) 3 layer Binary (d) 4 layer Binary

Figure 3.11: Q-Q plot, simulation results: Lognormal ${\sigma_{ij}^{lk}}^2 = {\lambda_{ij}^{lk}}^{-2}$



(a) 3 layer Continuous (b) 4 layer Continuous (c) 3 layer Binary (d) 4 layer Binary

Additional Gravity results

Table 3.6: Multilayer gravity estimates for Trade, Tourism, and FDI (specification 2)

	2015	2016	2017	2018	2019
Log-distance	-0.35	-0.352	-0.396	-0.374	-0.42
	(0.015)	(0.015)	(0.013)	(0.012)	(0.011)
Contiguous	0.409	0.402	0.447	0.498	0.468
	(0.022)	(0.022)	(0.022)	(0.022)	(0.024)
Common Lang.	0.593	0.601	0.562	0.643	0.685
	(0.024)	(0.025)	(0.026)	(0.024)	(0.025)
Colonial Past	1.028	1.059	0.927	0.933	0.732
	(0.03)	(0.031)	(0.029)	(0.028)	(0.027)
Trade Agreement	0.604	0.587	0.539	0.602	0.427
	(0.017)	(0.017)	(0.017)	(0.016)	(0.015)

Bibliography

Alter AL, Oppenheimer DM. 2008. Easy on the mind, easy on the wallet: The roles of familiarity and processing fluency in valuation judgments. *Psychonomic Bulletin and Review* **15**: 985–990. ISSN 10699384.

Anderson JE, van Wincoop E. 2003. Gravity with gravitas: A solution to the border puzzle. *American Economic Review* **93**: 170–192.

Aronow PM, Samii C, Assenova VA. 2017. Cluster–robust variance estimation for dyadic data. *Political Analysis* **23**: 564–577.

Ball GH, Hall DJ. 1965. Isodata, a novel method of data analysis and pattern classification. Technical report, Stanford research inst Menlo Park CA.

Bilkes G. 1997. The new bank note distribution system. *Bank of Canada Review* **1997**: 41–54.

Boccaletti S, Bianconi G, Criado R, Del Genio CI, Gómez-Gardenes J, Romance M, Sendina-Nadal I, Wang Z, Zanin M. 2014. The structure and dynamics of multilayer networks. *Physics reports* **544**: 1–122.

Boeschoten WC, Fase MMG. 1992. The Demand for Large Bank Notes. Technical Report 3.

Bruce V, Gilmore D, Mason L, Mayhew P. 1983. Factors affecting the perceived value of coins. *Journal of Economic Psychology* **4**: 335–347. ISSN 01674870.

Cannon E, Cipriani GP. 2006. Euro-Illusion: A Natural Experiment. Technical Report 5.

Chamberlain G. 1987. Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of econometrics* **34**: 305–334.

Charbonneau KB. 2017. Multiple fixed effects in binary response panel data models. *The Econometrics Journal* **20**: S1–S13.

Charrad M, Ghazzali N, Boiteau V, Niknafs A. 2014. NbClust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software, Articles* **61**: 1–36. ISSN 1548-7660.

Chen H, Huynh KP, Shy O. 2019. Cash versus card: Payment discontinuities and the burden of holding coins. *Journal of Banking & Finance* **99**: 192–201.

Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, Robins J. 2018. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* **21**: C1–C68. ISSN 1368-4221.

Di Muro F, Noseworthy TJ. 2013. Money Isn't Everything, but It Helps If It Doesn't Look Used: How the Physical Appearance of Money Influences Spending. *Journal of Consumer Research* **39**: 1330–1342. ISSN 0093-5301.

Engert W, Fung B, Segendorf B. 2019. A Tale of Two Countries: Cash Demand in Canada and Sweden. Discussion Papers 2019-7, Bank of Canada.

Fehr E, Tyran JR. 2001. Does Money Illusion Matter? Technical Report 5.

Fernández-Val I, Weidner M. 2016. Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics* **192**: 291–312. ISSN 0304-4076.

Gourieroux C, Monfort A, Trognon A. 1984. Pseudo maximum likelihood methods: Applications to poisson models. *Econometrica* **52**: 701–720. ISSN 00129682, 14680262.

Graham BS. 2020a. Chapter 2 - dyadic regression. In Graham B, Áureo de Paula (eds.) *The Econometric Analysis of Network Data*. Academic Press. ISBN 978-0-12-811771-2, 23–40.

Graham BS. 2020b. Chapter 2 - network data. In Durlauf SN, Hansen LP, Heckman JJ, Matzkin RL (eds.) *Handbook of Econometrics, Volume 7A*, volume 7 of *Handbook of Econometrics*. Elsevier, 111–218.

Head K, Mayer T, Ries J. 2010. The erosion of colonial trade linkages after independence. *Journal of International Economics* **81**: 1–14.

Helpman E, Melitz M, Rubinstein Y. 2008. Estimating Trade Flows: Trading Partners and Trading Volumes*. *The Quarterly Journal of Economics* **123**: 441–487. ISSN 0033-5533.

Henry C, Huynh K, Welte A. 2018. 2017 Methods-of-Payment Survey Report. Discussion Papers 2018-17, Bank of Canada.

Hobbes T. 1996. *Hobbes: Leviathan: Revised student edition*. Cambridge Texts in the History of Political Thought. Cambridge, MA: Cambridge University Press.

Huang Z. 1998. Extensions to the K-means algorithm for clustering large data sets with categorical values. *Data mining and knowledge discovery* **2**: 283–304.

Humphrey TM. 1974. The Quantity Theory of Money: Its HIstorical Evolution and Role in Policy Debates. *Economic Review* **1**: 1–19. ISSN 1055-7903.

IMF. 2015. *The Coordinated Direct Investment Survey Guide 2015*. International Monetary Fund.

Jackson MO. 2008. *Social and Economic Networks*. Princeton, NJ, USA: Princeton University Press. ISBN 0691134405, 9780691134406.

Jackson MO, Wolinsky A. 1996. A strategic model of social and economic networks. *Journal of Economic Theory* **71**: 44–74. ISSN 0022-0531.

Jain AK. 2010. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters* **31**: 651 – 666. ISSN 0167-8655. Award winning papers from the 19th International Conference on Pattern Recognition (ICPR).

Jiang JH, Shao E. 2019. The cash paradox. *Review of Economic Dynamics* ISSN 1094-2025.

Jochmans K. 2017. Two-way models for gravity. *Review of Economics and Statistics* **99**: 478–485.

Kahneman D, Tversky A. 1979. Prospect Theory: An Analysis of Decision under Risk. Technical Report 2.

Kalbfleisch JD, Prentice RL. 2002. *The Statistical Analysis of Failure Time Data*. New York, NY: John Wiley & Sons, 2nd edition.

Kalbfleisch JD, Prentice RL. 2011. *The statistical analysis of failure time data*, volume 360. John Wiley & Sons.

Kaplan EL, Meier P. 1958. Nonparametric estimation from incomplete observations. *Journal of the American statistical association* **53**: 457–481.

Kivelä M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA. 2014. Multilayer networks. *Journal of Complex Networks* **2**: 203–271. ISSN 2051-1310.

Klein JP, Moeschberger ML. 2006. *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media.

Lechner M, et al. 2011. The estimation of causal effects by difference-in-difference methods. *Foundations and Trends® in Econometrics* **4**: 165–224.

Lloyd S. 1982. Least squares quantization in PCM. *IEEE transactions on information theory* **28**: 129–137.

MacQueen J. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1. Oakland, CA, USA, 281–297.

Manski CF. 1993. Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies* **60**: 531–542. ISSN 0034-6527.

Menzel K. 2019. Inference with dyadic data: Asymptotic behavior of the dyadic-robust t-statistic. *Journal of Business & Economic Statistics* **37**: 671–680.

Mishra H, Mishra A, Nayakankuppam D. 2006. Money: A Bias for the Whole. *Journal of Consumer Research* **32**: 541–549. ISSN 0093-5301.

Newey WK, McFadden D. 1994. Chapter 36 large sample estimation and hypothesis testing. volume 4 of *Handbook of Econometrics*. Elsevier, 2111–2245.

Noussair CN, Richter G, Tyran JR. 2012. Money illusion and nominal inertia in experimental asset markets. *Journal of Behavioral Finance* **13**: 27–37.

Paskarathas R, Balodis E. 2019. Big Banknote Data. *Keesing Journal of Documents & Identity* : 3–6.

Paskarathas R, Graaskamp L, Balodis E, Garanzotis T. 2017. Polymer versus paper substrate lifespan calculations: the case of Canada. In Montoya I (ed.) *Banknote Management for Central Banks*, chapter 10. Research and Markets, 163–189.

Raghubir P, Capizzani M, Srivastava J. 2017. What's in Your Wallet? Psychophysical Biases in the Estimation of Money. *Journal of the Association for Consumer Research* **2**: 105–122. ISSN 2378-1815.

Raghubir P, Srivastava J. 2002. Effect of Face Value on Product Valuation in Foreign Currencies. *Journal of Consumer Research* : 335–347ISSN 0093-5301.

Raghubir P, Srivastava J. 2008. Monopoly Money: The Effect of Payment Coupling and Form on Spending Behavior. *Journal of Experimental Psychology: Applied* **14**: 213–225. ISSN 1076898X.

Raghubir P, Srivastava J. 2009. The Denomination Effect. *Journal of Consumer Research* **36**: 701–713. ISSN 0093-5301.

Rao BP. 2009. Conditional independence, conditional mixing and conditional association. *Annals of the Institute of Statistical Mathematics* **61**: 441–460.

Rojas D, Estrada J, Huynh KP, Jacho-Chávez DT. 2020. Survival analysis of banknote circulation: Fitness, network structure, and machine learning. *Advances in Econometrics* **42**.

Sant'Anna PH. 2020. Nonparametric tests for treatment effect heterogeneity with duration outcomes. *Journal of Business & Economic Statistics* : 1–17.

Shafir E, Diamond P, Tversky A. 1997. Money Illusion. *QUARTERLY JOURNAL OF ECONOMICS* **CXII**: 341–374.

Silva JS, Tenreyro S. 2006. The log of gravity. *The Review of Economics and statistics* **88**: 641–658.

Steinhaus H. 1956. Sur la division des corp materiels en parties. *Bull. Acad. Polon. Sci* **1**: 801.

Szepannek G. 2016. clustmixtype: K-prototypes clustering for mixed variable-type data. *R package version 0.1-16* .

Szepannek G. 2018. clustmixtype: User-friendly clustering of mixed-type data in R. *The R Journal* **10**: 200–208.

Tabord-Meehan M. 2019. Inference with dyadic data: Asymptotic behavior of the dyadic-robust t-statistic. *Journal of Business & Economic Statistics* **37**: 671–680.

Tinbergen J. 1962. *Shaping the World Economy; Suggestions for an International Economic Policy*. T.

Tschoegl AE. 1997. The Optimal Denomination of Currency. Technical Report 4.

Tversky A, Kahneman D. 1981. The Framing of Decisions and the Psychology of Choice. Technical report.

UNTWO. 2021. *Methodological Notes to the Tourism Statistics Database, 2021 Edition*. World Tourism Organization (UNTWO).

Wallace N, et al. 1998. A dictum for monetary theory. *Federal Reserve Bank of Minneapolis Quarterly Review* **22**: 20–26.

Walsh CE. 2017. *Monetary theory and policy*. MIT press.

Zou H. 2006. The adaptive lasso and its Oracle properties. *Journal of the American statistical association* **101**: 1418–1429.

Zou H, Hastie T. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **67**: 301–320. ISSN 13697412, 14679868.

Zou H, Zhang HH. 2009. On the adaptive elastic-net with a diverging number of parameters. *The Annals of Statistics* **37**: 1733–1751.