Essays in Financial Economics and Econometrics

by

Victor M. Orestes

M.S. Economics São Paulo School of Economics - FGV (2019)
B.A. Economics, University of São Paulo (2016)

Submitted to the Department of Economics and the Statistics and Data Science Center
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN ECONOMICS AND STATISTICS

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2025

Authored by:	Victor M. Orestes Department of Economics September 15, 2025
Certified by:	Stephen Morris Peter A. Diamond Professor of Economics, Thesis Supervisor
Certified by:	Robert Townsend Elizabeth & James Killian Professor of Economics, Thesis Supervisor
Accepted by:	Isaiah Andrews Charles E. and Susan T. Harris Professor of Economics Chairman, Departmental Committee on Graduate Theses

Essays in Financial Economics and Econometrics

by

Victor M. Orestes

Submitted to the Department of Economics and the Statistics and Data Science Center on September 15, 2025 in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN ECONOMICS AND STATISTICS

ABSTRACT

This thesis comprises three essays in finance and econometrics. The first two essays focus on the role of credit access and liquidity in shaping real firm outcomes. The first essay examines the transmission of modern monetary policy through corporate asset markets. Exploiting quasi-experimental variation in the Central Bank of Brazil's collateral framework and implementing a novel dynamic regression discontinuity design, it shows that monetary policy can ease expected future borrowing constraints, reduce firms' precautionary cash holdings, and stimulate employment. The second essay analyzes how receivables financing through factoring helps firms smooth cash flows. Using a shift-share instrument and matched administrative data, it finds that cheaper liquidity leads firms to rely more on permanent labor. The third essay develops a new method for distributional inference—nonparametric quantile mixture models. This framework can be applied to financial settings such as tail risk estimation and density forecasting, as well as to causal inference when the objective is to estimate the distributional effects of interventions. It is used here to quantify the heterogeneous wage effects of a major environmental disaster.

The first chapter (joint with Luis Alvarez and Thiago Christiano Silva) studies how modern monetary policy tools, which increasingly operate through corporate asset markets, affect real firm outcomes. We exploit quasi-experimental variation from the inclusion of specific corporate debt instruments in the Central Bank of Brazil's collateral framework and implement a novel dynamic regression discontinuity design. We find that eligibility increases firms' debt issuance, modestly lowers spreads, and reduces cash holdings, reflecting a decline in precautionary savings. These effects translate into higher employment and greater supply chain liquidity. We interpret the mechanism through the lens of segmented financial markets: by relaxing firms' expected future borrowing constraints, the policy acts as a persistent borrowing subsidy and liquidity injection. This encourages firms to reduce cash hoarding and expand production. Using a semi-structural framework calibrated to our reduced-form estimates, we find that an induced 0.8% borrowing subsidy leads to a 1% increase in debt issuance, a 0.2% reduction in cash holdings, and a 0.4% increase in the wage bill.

The second chapter (joint with Thiago Christiano Silva and Henry Zhang) shows that firms experience large increases in sales and purchases after receiving cheaper liquidity. We focus on factoring, defined as the supplier-initiated sale of receivables. In Brazil, receivables funds (FIDCs) securitize receivables for institutional investors. By assembling a novel transaction-level dataset of factoring with other credit operations for all registered firms and FIDCs, we construct a shift-share instrument for factoring financing supply based on FIDC flows. We then use a novel combination of electronic payments, trade credit, and employer-employee matched data to estimate the impacts. A flow-induced increase in receivables demand reduces firms' factoring interest rate. In response, firms demand more permanent labor and less temporary labor. In our model, these effects arise from factoring's purpose of reducing cash inflow volatility, helping firms match inflows to outflows, which firms otherwise achieve at an efficiency cost through substitution across labor types.

The third chapter (joint with Luis Alvarez) introduces nonparametric quantile mixture models as a computationally convenient and flexible alternative to standard nonparametric density mixtures, which are widely used in Statistics and Econometrics but face significant computational and inferential challenges. We propose a sieve estimator based on a generalized method of L-moments and develop a full inferential theory. In doing so, we contribute to the statistical literature by extending a numerical bootstrap method to high-dimensional settings. As a direct application of our theory, we provide the first inference method for the distributional synthetic controls of Gunsilius (2023), a novel tool for counterfactual analysis that previously lacked formal inference procedures. We apply this method to evaluate the effects of the Brumadinho dam collapse—a large-scale environmental disaster—on the local wage distribution. The results reveal substantial heterogeneity across the distribution, with evidence of displacement effects in which median-paying jobs are replaced by lower-wage contracts.

JEL Codes: C1, E4, E5, G2, G3

Thesis supervisor: Stephen Morris
Title: Peter A. Diamond Professor of Economics

Thesis supervisor: Robert Townsend
Title: Elizabeth & James Killian Professor of Economics

Thesis supervisor: Whitney Newey
Title: Ford Professor of Economics

Acknowledgments

I am deeply grateful to my advisors, Stephen Morris and Robert Townsend, for their guidance, feedback, and unwavering support throughout my PhD. Stephen has shaped my research agenda and the way I think about and produce economics since the very beginning of the program. His extraordinary generosity with his time and his steadfast encouragement have been more valuable than I can fully express, and he serves as a model of advising I aspire to follow with my own students. Rob's vision for the potential of financial innovations was instrumental in defining my research direction and in facilitating my collaboration with the Central Bank of Brazil, which was essential for the development of my papers. His distinctive approach to economic questions has been a constant source of motivation. I am also grateful to Whitney Newey for his encouragement and advice on developing and applying econometric tools to the problems I sought to study. Without their guidance, none of this would have been possible.

I am grateful to the MIT faculty for their feedback and guidance throughout the program. I especially thank Alberto Abadie, Isaiah Andrews, Anna Mikusheva, Victor Chernozhuk, and Dave Donaldson. I am also indebted to my former advisors from the São Paulo School of Economics, Bernardo Guimarães and Ricardo Masini, who continued to support me throughout the PhD program.

I have been fortunate to work with exceptional coauthors and friends—Luis Alvarez, Henry Zhang, and Thiago Silva. I learned a great deal from them, and none of the projects would have been possible without them. Their partnership made the work more engaging and sustained my motivation throughout, and I look forward to many future collaborations. I am also thankful to my classmates and friends for their friendship and support along the way.

I dedicate this thesis to my parents, Léa and José, and to my girlfriend, Debora, whose love, patience, and encouragement have made this journey possible.

Contents

List of Figures

List of Tables

Chapter 1

Corporate Effects of Monetary Policy: Evidence from Central Bank Liquidity Lines

This chapter is jointly authored with Luis Alvarez and Thiago Christiano Silva.[1]

The views expressed in this working paper are those of the authors and do not necessarily reflect those of the Central Bank of Brazil.

1.1 Introduction

Understanding the pass-through channels of monetary policy tools to the real economy is at the core of macroeconomic analysis and is fundamental to the design of effective monetary systems and stabilization frameworks. In recent years, there has been an increasing reliance on unconventional monetary policy (UMP) tools, including interventions in corporate bond markets. These tools were designed to directly influence corporate credit markets—and, by extension, economic activity—when the traditional transmission channels of the baseline interest rate become "clogged" during periods of stress. Major tools include direct purchases of corporate debt and expansion to the collateral framework, the latter being the focus of our paper.

The collateral framework involves the central bank accepting corporate securities as collateral in its lending operations with commercial banks and other financial

[1]We are deeply grateful to Stephen Morris, Robert Townsend, and Whitney Newey for their invaluable guidance and support. We also thank Alberto Abadie, Patrick Adams, Fabio Araujo, Isaiah Andrews, Martin Bereja, Ricardo Caballero, Fernando Chague, Dave Donaldson, Rafael Edreira, Bruno Ferman, Marcelo Fernandes, Sarah Gertler, Bernardo Guimarães, Gustavo Joaquim, Nobuhiro Kyotaki, Ricardo Masini, Anna Mikusheva, Lira Mota, Whitney Newey, Jonathan Parker, Kerry Siani, João Rudge Leite, Tong Liu, David Thesmar, Gabriel Toledo, Maria Tiurina, Robert Townsend, Adrien Verdelhan, Emil Verner, Ivan Werning, Christian Wolf, and Henry Zhang, and the seminar participants at the MIT Macroeconomics Seminar, MIT Sloan Finance Seminar, MIT Econometrics Lunch and the USP Seminar for their valuable comments and suggestions. Any errors are our sole responsibility. The views expressed in this Working Paper are those of the authors and do not necessarily reflect those of the Banco Central do Brasil.

intermediaries. By changing eligibility criteria and haircuts applied to these assets, central banks can affect the liquidity and, thus, the demand for these assets. Despite their widespread use by the major central banks in the world, estimating their real effects on firms and respective transmission mechanisms is challenging because of a combination of both a lack of comprehensive datasets and endogeneity in eligibility criteria, simultaneity with macroeconomic shocks, and general equilibrium spillovers, which complicate causal inference and leave this question largely unanswered. This is important not only to evaluate the effectiveness of past interventions but also to improve the design of monetary policy frameworks going forward, as collateral frameworks are integral to the functioning of monetary systems in many countries.

This paper leverages the inclusion of corporate debt in the Brazilian Central Bank's collateral framework and introduces a novel dynamic regression discontinuity design (RDD) methodology to estimate the causal effects of debt eligibility on firms' financial and real economic outcomes. Additionally, we examine how this eligibility influences liquidity and pricing in both primary and secondary corporate bond markets. The policy we study, the Liquidity Facility Lines (LFL), introduced in November 2021, provides liquidity to financial institutions using corporate bonds as collateral. Each participating financial institution has an LFL limit based on the face value of deposited eligible corporate bonds, which can be used for short- and medium-term credit lines and reducing reserve requirements. This policy marked a shift from the previous reliance on Treasury securities for liquidity provisions, following recommendations made by IMF 2012 to broaden collateral eligibility.[2] Thus the policy directly improves corporate bond liquidity for financial institutions holding eligible debt, potentially affecting asset demand.

Our empirical findings indicate that eligible firms experienced, one year into the policy, a 12% increase in corporate bond debt within a year, alongside a decrease in intermediate bank/fund debt by 5%, and a 20 basis point reduction in cost of capital. However, there were no significant effects on corporate debt maturity or bank/fund interest rates. Additionally, we observed a 6% real increase in the wage bill and a temporary rise in net payment outflows to non-financial firms, driven by higher payment outflows. This increase in expenditure was partially financed by a 12% reduction in liquid cash holdings, suggesting a decreased precautionary savings motive. The policy's impact was amplified by the increased use of corporate bonds as collateral, especially in short-term repo operations, while direct trading decreased, indicating the policy incentivized a "buy-and-hold" strategy by enhancing both directly and indirectly the service flow associated with holding these bonds in the portfolio.

The identification of causal effects leverages the unique eligibility rule based on an

[2]Corporate bonds were chosen due to existing regulations and market infrastructure. The policy had wide coverage, with eligible firms representing a significant portion of the corporate bonds market, including unrated bonds. The LFL policy resulted in liquidity injections in the economy, as institutions quickly approached the limits for reserve requirement reductions and borrowing limits in the facility lines in the few months after the start of the policy.

internal risk score, which uses data from the credit registry system (SCR). Our strategy is based on comparing the outcomes of firms that are close to the eligibility threshold. However, the dynamic nature of eligibility, with firms' treatment status changing over time, presents a significant challenge in isolating the dynamic causal effects accurately for two reasons: (i) - at a given point in time, firms near a threshold may have different treatment exposures and (ii) - treatment today may partially affect treatment status in the future.

To address this challenge, we develop a novel framework for identifying and estimating causal effects in dynamic regression discontinuity designs, where treatment assignment evolves over time based on time-varying scores. Unlike traditional RDDs that assume static treatment assignment, our potential outcomes framework accounts for the influence of treatment duration on outcomes. We establish conditions under which reduced-form RD estimands can be interpreted causally as weighted sums of average marginal effects of exposure to the policy, with non-negative weights, even in the presence of time-varying scores, when the causal impact of the policy may depend on the length of exposure to the policy. Nonetheless, we argue that, even if these conditions are likely to be satisfied in a given empirical setting, the interpretation of the estimand may not be immediate, as weights reflect the margin through which units self-select into treatment on the basis of prior scores. Specifically, if the researcher is interested on the average marginal or cumulative effects of exposure to policy, the reduced-form RD estimand does not recover the parameters of interest. Motivated by this, we outline different strategies to disentangle both the average marginal and average cumulative treatment effects of exposure to policy and introduce an estimation strategy that exploits homogeneity assumptions between populations at different thresholds, while still allowing for nonstationarity and decreasing returns in the exposure to the policy, to construct more efficient estimators of effects. We also propose overidentification and placebo tests to assess the plausibility of our identifying assumptions.

Finally, we develop a semi-structural approach to examine how firm eligibility for the LFL policy affects liquidity injections from financial intermediaries, borrowing costs, and the pass-through to firms' production decisions. The model combines two key features: a segmented financial market and imperfectly elastic demand for firms' debt. We demonstrate that in inelastic financial markets, the reduced-form results imply significant quantity effects of the LFL policy, as opposed to price effects. The quantity effects act like a liquidity injection, in the form of a borrowing subsidy. Using this framework, we estimate the pass-through of liquidity injections to real outcomes, finding that a 1% increase in debt issuance raises the wage bill by 0.4% to 0.75%. This expansion is partially funded by a reduction in internal funds, with cash holdings decreasing by 0.25% to 0.35%, alongside a modest increase in external debt.

Literature Our paper contributes to several strands of the literature. Since the Great Recession, central banks have increasingly turned to operations involving corporate bonds as an additional monetary policy tool. We are the first to estimate the causal effects on real outcomes and firms' capital structure and liquidity management of indirect policies such as the collateral framework.[3] Previous papers have focused mostly on outcomes in the secondary market and a few outcomes in the primary markets of corporate bonds (Fang, Wang, and Wu 2020, Pelizzon et al. 2024). We are able to do so by leveraging a new empirical setting where we can improve upon both the identification and data, both in terms of availability and frequency, and developing a novel econometric methodology appropriate to our setting. Our estimates of significant effects on firms' labor expenditure and liquidity management contribute to filling the gap noted by Nyborg 2016: even though collateral frameworks are at the core of the monetary systems, they are largely overlooked in academic research, and their far-reaching effects on the real economy are poorly understood. It also provides the first causal evidence that haircut policies, such as those advocated by Geanakoplos 2010, can directly influence the production decision of the underlying asset issuers.

We contribute to the literature on safe assets and convenience yield (Krishnamurthy and Vissing-Jorgensen 2012, Brunnermeier, Merkel, and Sannikov 2022, Mota 2023) by providing causal estimates on how (i) monetary policy affects the equilibrium service flow of privately issued securities and (ii) how the issuers respond to this increased service flow. In particular, we estimate large quantity effects (as opposed to price effects) and corresponding large real effects. Additionally, our results are related to the literature of monetary policy implementation through segmented markets and its relationship with the market micro-structure of OTC markets (Lagos and S. Zhang 2020, Eisenschmidt, Y. Ma, and A. L. Zhang 2024).

We provide new empirical evidence on the price elasticity of demand for corporate bonds. Our findings contribute to the literature by offering an alternative to the demand system approach (e.g., Koijen and Yogo 2019), as our estimation strategy is based on the firm's liquidity management decisions. Furthermore, we contribute to the theoretical and empirical literature on the inelasticity of financial markets (e.g., Gabaix and Koijen 2021, Darmouni, Siani, and Xiao 2022), providing a semi-structural approach to quantify the pass-through of interventions in these markets to firms' production decisions.

[3] A branch of literature focuses on direct corporate bond purchase programs such as the ECB's Corporate Sector Purchase Program (CSPP) standing facility lines backed by corporate bonds and the FED's Secondary Market Corporate Credit Facility (SMCCF). A strand of the literature empirically assesses the effects of these policies, pointing out that they boost primary market issuance, lower yields, and create spillover effects that impact both yields and liquidity in the secondary markets (Todorov 2020, Gilchrist et al. 2020, Boyarchenko, Kovner, and Shachar 2022). Few papers focus on the effects at the firm level, as a clear method to compare eligible firms with non-eligible ones is lacking in these settings, and the existing results point to null direct real effects (Darmouni and Siani 2022, Adelino et al. 2023). Our findings contrast with those from this literature.

Finally, from the econometric viewpoint, our paper contributes to the literature on regression discontinuity designs (M. Cattaneo and Titiunik 2022). First, we show that, in dynamic settings where eligibility to a policy is determined by a sequence of running variables (scores) at each period, RD estimands that contrast average outcomes at a given period between individuals just above and below the threshold at an *earlier* period may not have a causal interpretation, in the sense that they may not correctly reflect the *sign* of treatment effects even in an extreme setting where there is no heterogeneity in these in the population (Blandhol et al. 2022). The breakdown in this weak notion of causality stems from the fact that, if prior eligibility to the treatment partly determines future scores, and if agents on the threshold at a given period self-select into treatment *in later periods*, then individuals below and above the threshold at a given period may have different exposures (and treatment effects) in the future, which generates a "forbidden comparison" problem akin to those underlying the breakdown of two-way fixed effects estimators in staggered adoption settings (Roth et al. 2023). Moreover, we argue that even if the conditions that ensure a causal interpretation to the RD estimand hold, it is not possible to interpret the parameter directly as a cumulative or marginal effect. We then show that alternative strategies to recover these marginal or cumulative effects available in the literature, such as those considered in the literature of sequential experiments under imperfect compliance (Sojitra and Syrgkanis 2024) or those specially tailored to specific dynamic RD designs (Cellini, Ferreira, and Rothstein 2010; Hsu and Shen 2021), either impose global independence assumptions that severely limit the dynamic dependence between scores or assume constant marginal returns of exposure to the policy. These assumptions are unlikely to be satisfied in settings such as ours, where scores are tightly linked to past financial decisions of firms and treatment effects reflect firms' production and financing decisions upon exposure to the policy. As an alternative to these approaches, and inspired by a recent literature exploring the identifying power of homogeneity restrictions in cross-sectional RD designs with multi-valued treatments and a single discontinuity (C. Caetano, G. Caetano, and Carlos Escanciano 2023), we propose to recover marginal and cumulative effects by exploring cross-threshold treatment homogeneity restrictions that, while constraining treatment effect heterogeneity, allow for nonstationarity and nonconstant marginal returns of exposure to the policy. We provide applied researchers with a toolkit for efficient estimation under these restrictions, which includes a minimum distance estimator that efficiently explores available overidentifying restrictions, as well as overidentifying restrictions and placebo tests to assess the plausibility of the imposed assumptions.

Outline This paper is organized as follows: Section 1.2 describes the institutional setting and the data, focusing on Brazil's Liquidity Facility Lines (LFL) policy and the corporate bond market. Section 1.3 outlines the identification strategy, using a dynamic regression discontinuity design (RDD) to estimate the causal effects of the LFL policy on firm outcomes.

Section 1.4 presents the empirical results, showing the policy's impact on corporate bond issuance, debt structure, bond costs, and real outcomes such as the wage bill and payment outflows. Section 1.5 presents our semi-structural model. Finally, Section 1.6 concludes by summarizing the key findings and their broader implications for monetary policy and liquidity management. The appendices contain additional empirical results, derivations of the econometric and theoretical results, and a detailed overview of the Brazilian corporate debt market.

1.2 Data and Institutional Setting

1.2.1 Datasets

We have access to data on the BCB's Liquidity Facility Lines (LFL) policy, including daily eligibility, letter grades at the corporate bond and firm levels, and the limits and usage by participating financial institutions. Using the credit operations dataset (SCR), which includes all bank loans, credit lines, collateral-backed loans, and mutual fund-initiated financing (e.g., factoring), we can compute the underlying risk score. This dataset also allows us to construct firm-level outcomes on total outstanding internal and foreign debt, interest rates, and maturity. We exclude operations involving the public development bank (BNDES) from the analysis.

From B3, the central depository, clearinghouse, and trade repository for OTC financial assets in Brazil, we obtain the complete record of electronic transactions involving corporate bonds. This dataset includes all asset movements—transfers, trades in primary and secondary markets, collateral use, and repo operations. It also captures holdings at the final investor and financial institution levels. We supplement this with additional data from the Brazilian Financial and Capital Markets Association (ANBIMA) and B3, providing registry details on corporate bonds, such as issuer, underwriter, issuance spread, volume, and daily outstanding quantity and (nominal) value. Finally, with data from the Brazilian Securities and Exchange Commission (CVM) and the tax registry we map mutual funds and other companies (such as insurance companies) ownership and management structure.

The Annual Report of Social Information (RAIS) is a mandatory annual submission for all formal employers in Brazil, covering all registered employment relationships, including employees with formal contracts. RAIS provides detailed employee-level data, such as employment status, job tenure, and wages. We use this dataset to construct firm-level information on the number of workers and the total wage bill.

The electronic payments dataset includes all payments associated with invoices (boletos), usually via bank transfer, as well as all interbank transfers and instant payments (PIX). In these datasets, we observe identifiers for creditor and debtor, as well as the time stamp of when the transaction is settled.

Finally, we measure firm-level cash-like holdings, which include term deposits, treasury bonds, money market mutual fund shares, and short-term repo operations (with resale obligations) involving public and private assets. For term deposits, we aggregate data from electronic records of the major registration and settlement platforms (B3, CRT4, CERC, and CSDBR), covering nearly all such deposits. Treasury holdings and treasury-backed repo operations data are sourced from the Special System for Settlement and Custody (SELIC), the centralized system for registration, settlement, and custody of federal government bonds. Data on private asset-backed repo operations comes from B3's electronic records.

1.2.2 Liquidity Facility Lines (LFL)

The BCB Liquidity Facility Lines (LFL), launched in November 2021, are tools to provide liquidity to financial institutions using corporate bonds as collateral. From 2000 to 2021, BCB's liquidity provisions were limited to short-term operations involving Treasury securities. The policy followed a recommendation from IMF 2012 to introduce systems for the quick acceptance of collateral beyond government securities, including valuation and haircuts; pre-determine, periodically review, and regularly disseminate haircuts for asset classes to reduce uncertainty. Corporate bonds were selected due to pre-existing regulations and financial market infrastructure requiring centralized registry, custody, and settlement of these securities.[4]

Brazil's corporate bonds (measured by outstanding debt volume) represent roughly 10% of GDP and 20% of total credit to nonfinancial corporations, making them the largest private fixed-income security in the country and surpassing mortgage-backed securities in volume. This contrasts with the U.S., where corporate bonds account for about 40% of GDP and 42% of credit to nonfinancial firms. In Brazil, roughly half of large firms' domestic debt is financed through corporate bonds, and these issuers collectively employ 12% of formal workers while accounting for 15% of payment outflows. The instruments are typically unsecured, with an average spread of the baseline interest rate plus 3% and an average maturity of seven years. Despite their importance, the majority of corporate bonds are neither traded nor externally rated, although they are frequently used as collateral in over-the-counter markets.[5] Holdings are concentrated among the largest commercial banks (44%) and mutual or pension funds (50%). For a more detailed description of the Brazilian corporate bond market, refer to Appendix 1.14.

Each participating financial institution has a LFL limit based on the total amount

[4]See section IV of IMF 2012 and Vote 146/2021-BCB for a detailed exposition of the rationale behind the policy.

[5]The typical use as collateral occurs in short-term repo operations, in which commercial banks sell corporate bonds (with a repurchase obligation) to nonfinancial firms. These transactions provide both liquidity benefits and tax advantages, as they are exempt from the financial transaction tax that applies to term deposits.

of eligible corporate bonds held in a special account in the central depository. These limits can be used for borrowing on short and medium-term lines (up to 45 days and 1 year, respectively), and to reduce reserve requirements in June 2024, the BCB mandates depository institutions to hold 20% of term and demand deposits in a BCB account. The LFL limit can reduce term deposit reserve requirements by up to 15% (3 percentage points). [6] Collateral eligibility is defined and disclosed daily using an internal risk model, we will discuss the eligibility criteria in Section 1.2.3 as it is central to our identification strategy for the causal effects of the LFL on firms.

The policy had broad coverage, with eligible firms representing 9.5% of the formal workers' wage bill and around 40% of long-term corporate bonds (and 50% of short-term bonds) being eligible by the end of 2022 as shown in Figure 1.9. Eligible firms were among the largest corporate bond issuers, with an average of 6,000 workers. Due to internal risk scoring, there was no requirement for bonds to be rated as investment grade by major rating agencies, significantly expanding eligibility since most bonds in Brazil are unrated. This unique feature of the LFL resulted in broader economic coverage compared to typical corporate bond interventions by the Federal Reserve and European Central Bank.

Credit limits are determined by the collateral pool price, discounted by risk and diversification haircuts, and are updated promptly after assets are pre-positioned or withdrawn from the pledge account. Events such as interest payments and amortizations are transferred to a cash pledge account at the BCB, which is accessible by the participating financial institution. We postpone the details on collateral eligibility to the next subsection. [7]

Figure 1.2 shows that financial institutions participating in the policy (those with a positive limit at any point) approached the limit on long-term lines within one year of the policy's implementation. [8] Together, these results indicate a direct liquidity injection into the economy as a result of the policy.

1.2.3 Collateral Eligibility, Risk Score and Causal Effects

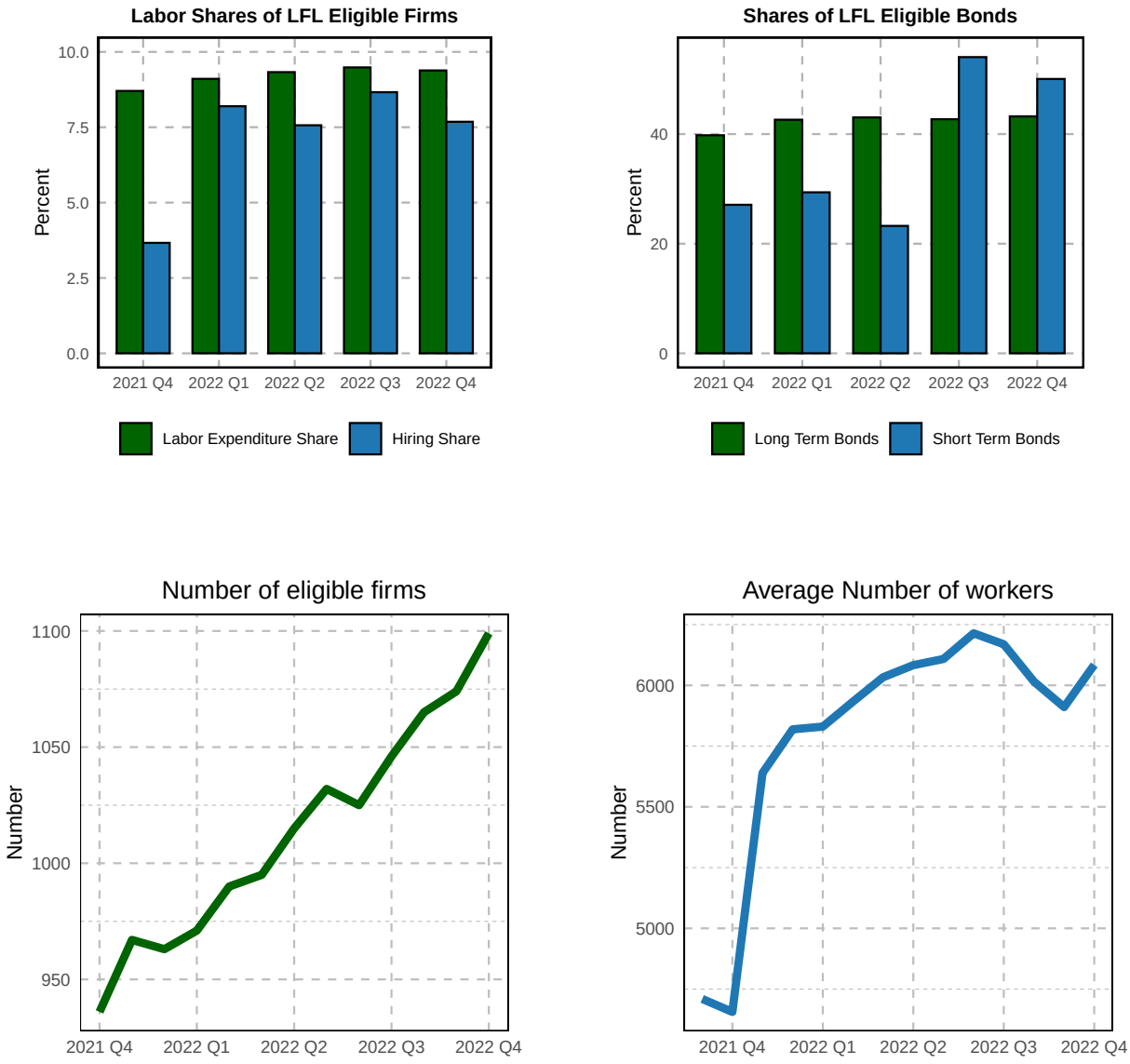
Eligibility requires the issuer to be a non-financial firm with a low-risk score according to an internal risk model. Accepted collateral includes short and long-term corporate bonds issued in local currency by eligible firms. and the payoff structure is not too complex. The first step in determining eligibility is mapping the continuous risk score to a letter grade. Table 1.1 shows the relationship between risk score, letter grade, and eligibility.

[6] Term deposits make up about 80% of total deposits. For further details, see the complete list of requirements.

[7] A complete description of the haircut applied to each corporate bond can be found in Annex I of Vote 146/2021-BCB

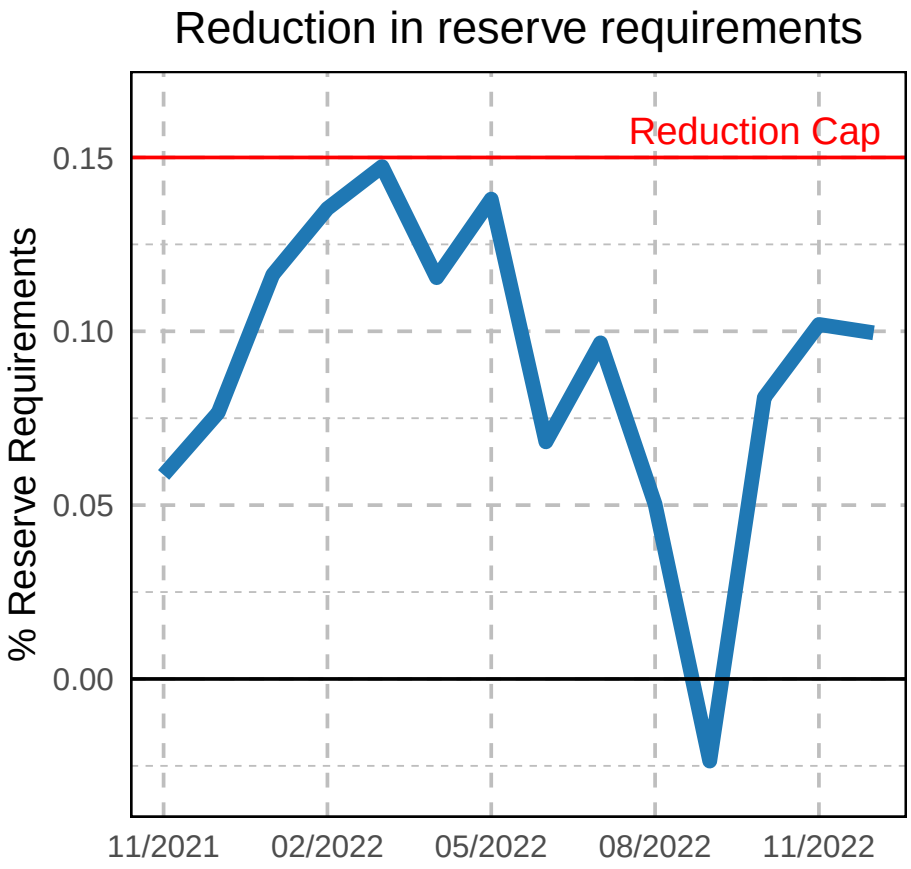
[8] The Liquidity Facility Line (LLT) window has been closed since March 2022, and the operations that were still open were concluded in March 2023.

Figure 1.1: LFL Eligible Firms.



Note: The graph displays four panels showing key metrics for LFL-eligible firms and bonds from Q4 2021 to Q4 2022. The labor share reflects the wage bill for formal workers in the last month of each quarter, while the hiring share represents the wage bill for newly hired workers during the quarter. Bond shares indicate the percentage of eligible corporate bonds by type each quarter. Data sources: B3 (Short Term Bonds), ANBIMA (Long Term Bonds), RAIS (wage bill and number of workers), and BCB - Department of Banking Operations and Payment Systems (LFL Eligibility).

Figure 1.2: LFL Take-up by Participants.



Note: The graph shows participants' institutions' take-up of Liquidity Facility Lines. The figure shows the reduction in reserve requirements as a percentage of the maximum possible reduction, netting out voluntary deposits held in the BCB. Data Sources: BCB - Department of Banking Operations and Payment Systems (LFL data, reserve requirements), BCB - Department of Open Market Operations (Voluntary Deposits).

Notably, bonds issued by firms with a letter grade of B are subject to limits on long-term lines and reduction in reserves but are ineligible for short-term lines and receive a higher haircut. Our identification strategy exploits the sharp discontinuity in eligibility around the corresponding letter grade of B.

The risk score is based on a mechanical rule using exclusive data from credit operations reported by lenders in the credit registry SCR. This score is based on a weighted average of lenders' reported Allowance for Loan and Lease Losses (ALLL) for selected credit categories.[9] Importantly, this score is used exclusively for the LFL policy. The score, the SCR data, and the exact details of its computation are not public information; only an indicator of firm (and asset) eligibility is public information.

The final risk score is calculated daily and is the maximum among the credit risk scores computed using (i) the most recent month's data available in the Credit Information System (SCR), (ii) the average of the last three months' data available in the SCR, or (iii) the average of the last six months' data available in the SCR. Firms without outstanding credit operations in the SCR (including corporate bonds) are not eligible. Based on the BCB's internal score calculation rules and SCR data, we compute the score for the last business day of each month. Therefore, our analysis uses a monthly frequency, with eligibility determined by the status on the last day of each month.

Table 1.1: LFL Eligibility Rule

Letter	**LB** $(>)$	**UB** $(\leqslant)$	**Haircut (%)**
AA	0	0	16%
A	0.00%	0.50%	23%
B	**0.50%**	**1.00%**	**39%**
C	**1.00%**	**3.00%**	**100%**
D	3.00%	10.00%	100%
E	10.00%	30.00%	100%
F	30.00%	50.00%	100%
G	50.00%	70.00%	100%
H	70.00%	100.00%	100%

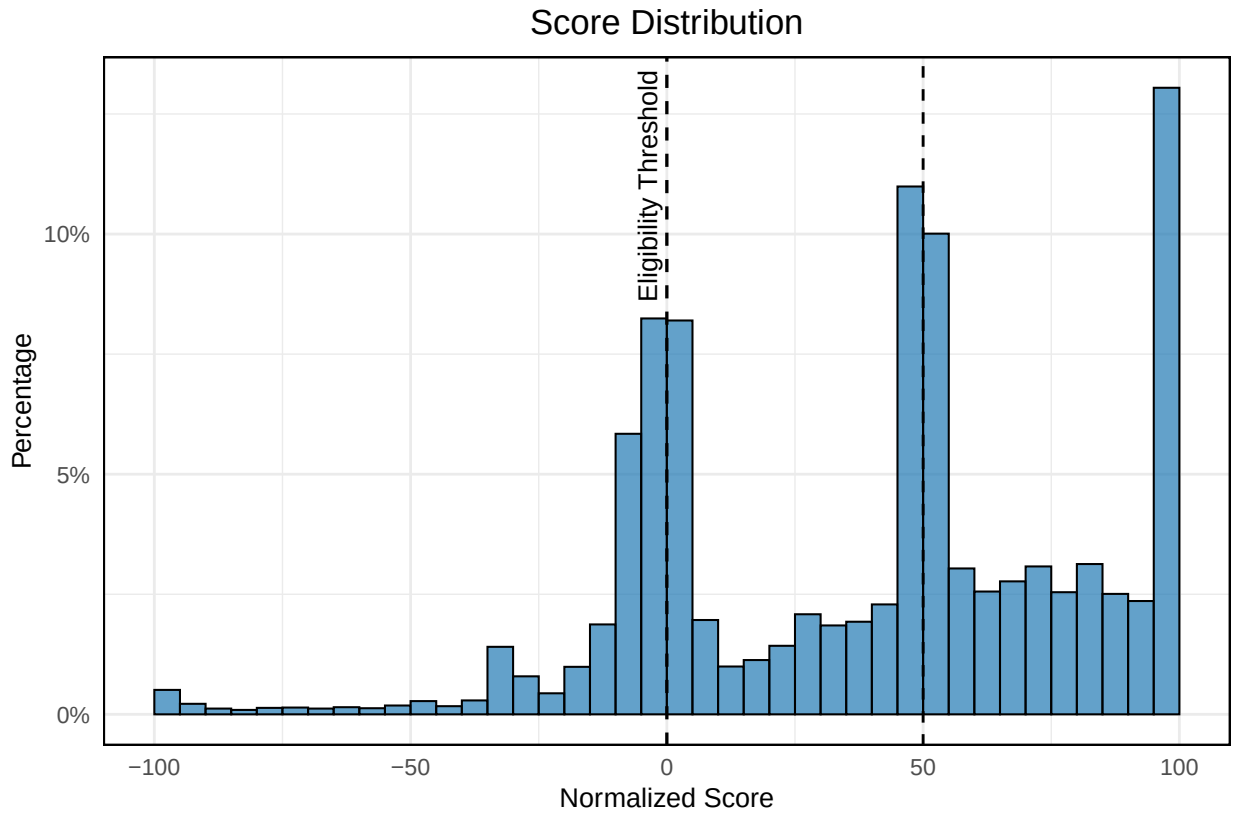
Figure 1.3 shows the distribution of the score shows the score distribution of non-financial firms from November 2021 to December 2022. To visualize the score distribution, we normalize the scores as follows: firms with ratings A and B have scores normalized between 0 and 100, while firms with ratings C and D have scores normalized between -100 and 0, so formalized the normalized score of firm i at time t, Z_{it}, is given by

[9] Certain credit operations are excluded from the risk score calculation. These include the cession of credit rights from financial applications, operations with real estate or vehicles as collateral backed by the Federal Government Treasury, insurance, and third-party coverage.The description of the score computation can be found in section III of Chapter IX of Vote 146/2021-BCB

$$Z_{it} = \begin{cases} 100 \times (1 - \text{Score}_{it}) & \text{if } \text{Score}_{it} \in [0, 1] \\ -100 \times \frac{\text{Score}_{it} - 1}{9} & \text{if } \text{Score}_{it} \in [1, 10] \end{cases}.$$

The key observation from Figure 1.3 is the concentration of firms around the eligibility threshold. This occurs because, for a given credit operation, the score can only be reported in pre-specified discrete numbers that match the letter rating thresholds shown in Table 1.1. This concentration indicates that, aside from the effect of LFL eligibility, firms just above and below the threshold would be comparable. Any manipulation of the score is unlikely, as the score and its underlying data are not public, and financial institutions are mandated to properly document their credit score classification policies and procedures, making them available to the BCB and independent auditors.

Figure 1.3: LFL Score Distribution (2021-2022).



Note: The figure plots the distribution of the normalized score Z_{it} pooling the periods from November 2021 to December 2022. The sample consists of non-financial firms that are considered potential eligible firms in the LFL.

We formally test both claims. Figures 1.4 and 1.5 show summary statistics for pre-policy outcomes for firms near the eligibility threshold when the policy was introduced in November 2021. The plots include fitted lines or curves on either side of the cutoff, showing

no clear discontinuity in any outcome variable. Second, Table 1.3 shows the results of the manipulation test proposed by M. D. Cattaneo, Jansson, and X. Ma 2020. The null hypothesis of no manipulation is rejected in nearly all periods.

The previous results suggest that a Regression Discontinuity Design (RDD) could uncover the causal effects of the LFL policy, as units near the threshold are comparable before the policy starts. However, the dynamic nature of eligibility, with firms' treatment status changing over time, presents a challenge . The next section discusses identification in this dynamic RDD and introduces a methodology to estimate the dynamic causal effects of the LFL policy.

1.3 Identification and Estimation of Causal Effects in a Dynamic Regression Discontinuity Design

1.3.1 Setup

In this section, we introduce a framework to understand the causal content of regression discontinuity estimands, in a setting where treatment assignment is (partially) determined by a sequence of running variables at each time period. In the LFL policy eligibility to the program is determined by a score that is computed every month.

We consider a population of interest, and a policy intervention that begins at period $t = 1$. The sequence of outcomes of a unit in the population of interest is denoted by $(Y_t)_{t=1}^T$. We view $(Y_t)_{t=1}^T$ as a vector of random variables, defined in a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, capturing the fact that paths in observed outcomes are heterogeneous in the population of interest.

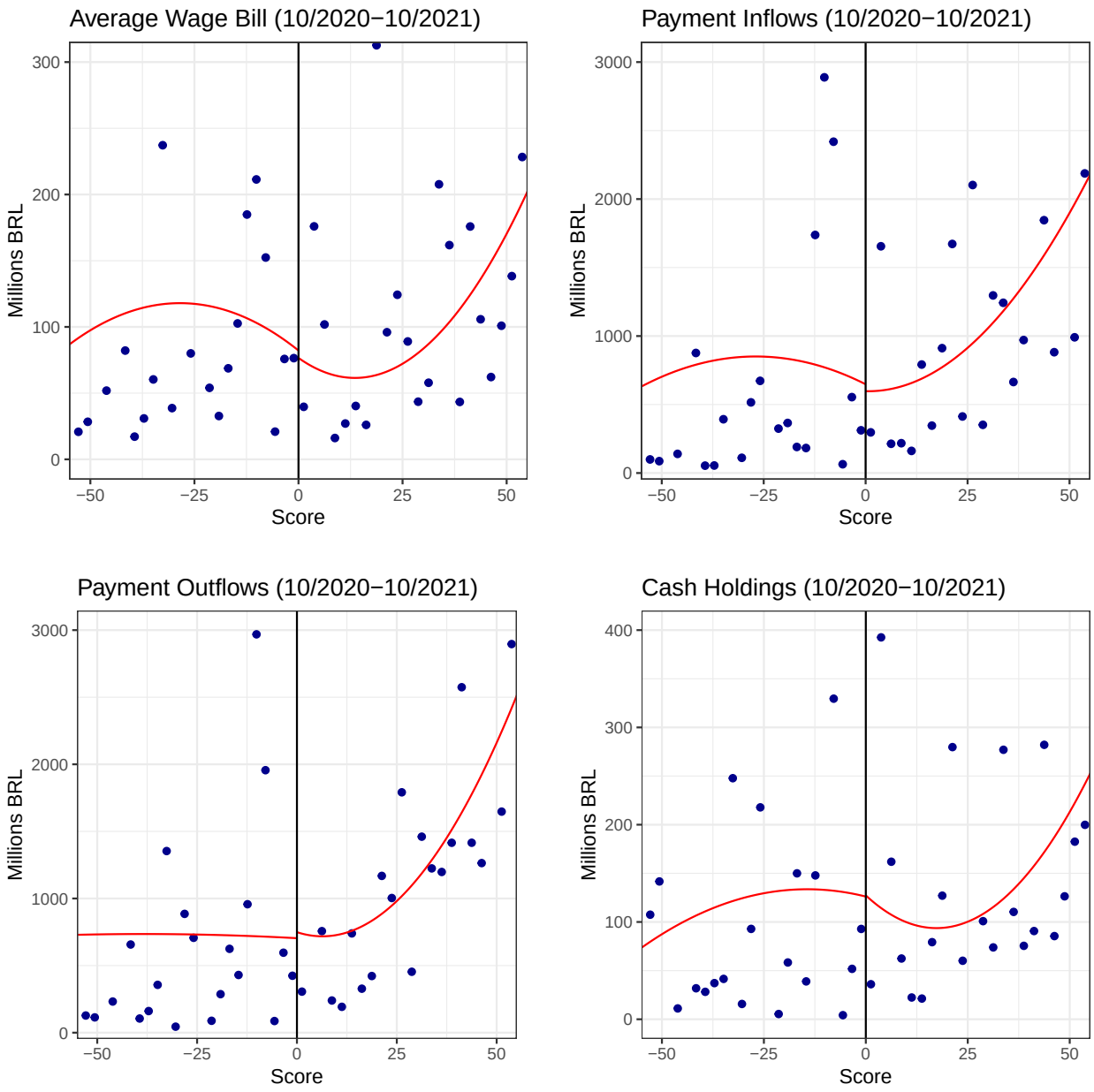
We assume that the length of exposure to the policy may exert a causal effect on the outcome. Denote by D_t the length of exposure of a unit in the population by time period t. Thus, D_t is a random variable taking values on $\{0, 1, \dots, t\}$ and $t \mapsto D_t$ is a nondecreasing random function. We posit that, for every $t = 1, \dots, T$, outcomes Y_t are generated according to:

$$Y_t = Y_t(0) + \sum_{j=1}^{t} \mathbf{1}\{D_t \geqslant j\}(Y_t(j) - Y_{t-1}(j)),$$

where $\{Y_t(j) : j = 0, 1, \dots, t\}$ are the potential outcomes associated with being exposed to the policy for j periods, by period t.

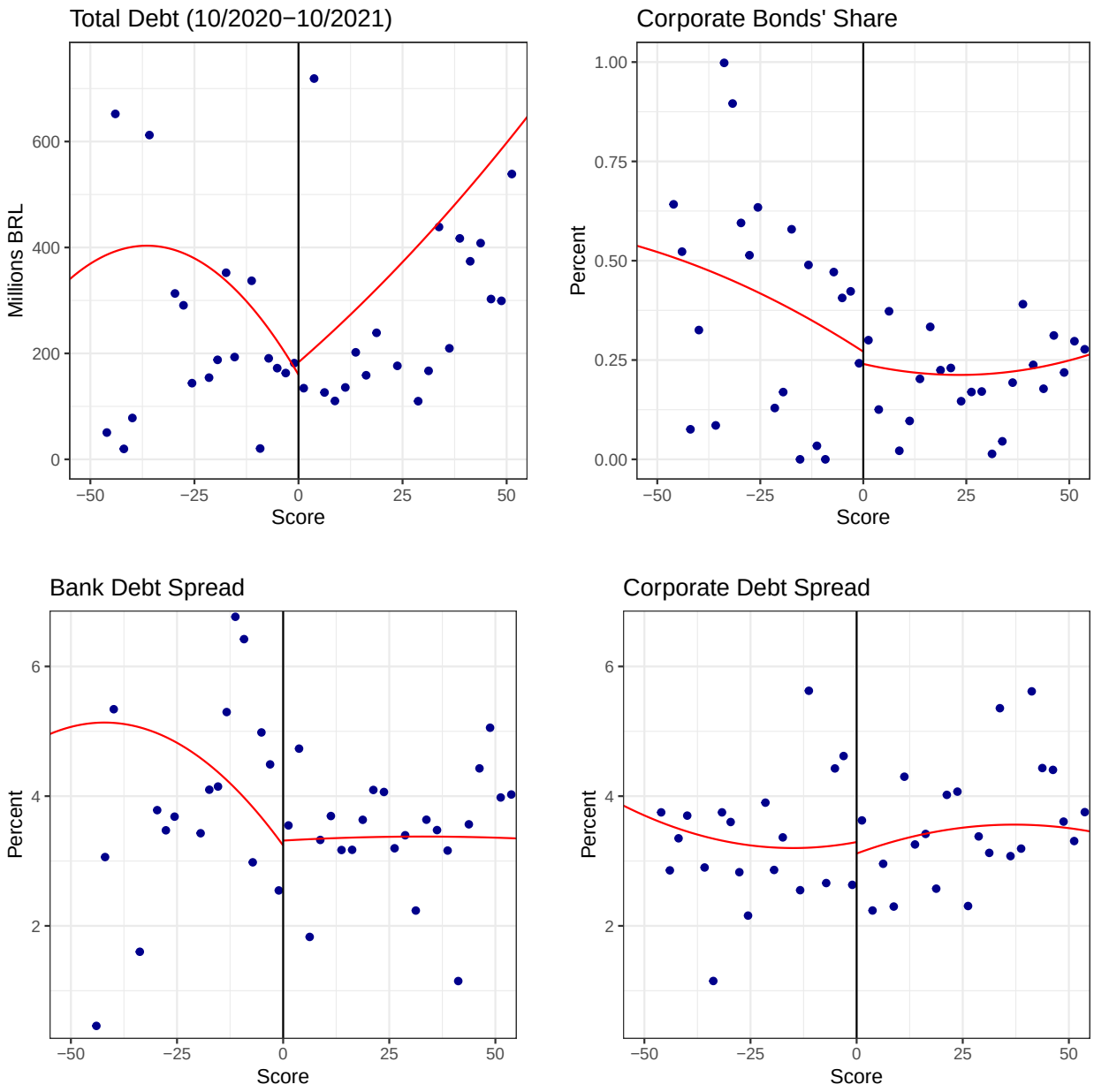
We assume that exposure to the policy is (partly) determined by a sequence of continuous running variables or *scores* Z_t, $t = 1, \dots, T$. Specifically, we assume that $D_t = \boldsymbol{D_t}(Z_1, \dots, Z_t)$, with eligibility being affected by the Z_t crossing a threshold, which we normalize to zero. In the LFL example, the map $\boldsymbol{D_t}$ takes a simple homogeneous and

Figure 1.4: Summary Statistics of firms near the LFL threshold in November 2021.



Note: The graph summarizes firm statistics around the threshold. We fit a second-order polynomial on each side, using 20 equal-sized bins per side.

Figure 1.5: Summary Statistics of firms near the LFL threshold in November 2021.



Note: The graph summarizes firm statistics around the threshold. We fit a second-order polynomial on each side, using 20 equal-sized bins per side.

additive form, $\boldsymbol{D_t}(Z_1,\dots,Z_t) = \sum_{j=1}^{t} \mathbf{1}\{Z_j \geqslant 0\}$, though we consider more general potential treatment rules in our results.

1.3.2 What does RDD identify?

Our first goal is to understand under which interpretable conditions does a "reduced-form" RD estimand of the type:

$$\tau_{t|l} = \lim_{s\downarrow 0} \mathbb{E}[Y_t|Z_l = s] - \lim_{s\uparrow 0} \mathbb{E}[Y_t|Z_l = s]\,, \quad t,s \in \{1,2,\dots,T\} \tag{1.1}$$

have a causal interpretation. Estimand (1.1) is an average contrast of the outcome of interest in period t, between units just below and above the threshold at period l. We assume the following conditions for our dynamic RD design:

Assumption 1 (Existence of limits)**.** *For each* $t,l \in \{1,\dots,T\}$*, let* $\boldsymbol{Z}_{t|l}(z) = (Z_1,\dots,Z_{l-1},z,\dots Z_t)$ *if* $l \leqslant t$ *and* $\boldsymbol{Z}_{t|l} = (Z_1,Z_2,\dots,Z_t)$ *otherwise. We assume that, for each* $t,l,j \in \{1,\dots,T\}$ *and* $j \leqslant t$*, the following limits exist:*

$$\phi^{+}_{t|l}(j) \coloneqq \lim_{s\downarrow 0} \mathbb{E}[Y_t(j)-Y_t(j-1)|D_t \geqslant j, Z_l = s] = \lim_{s\downarrow 0} \mathbb{E}[Y_t(j)-Y_t(j-1)|\boldsymbol{D}_t(\boldsymbol{Z}_{t|l}(s)) \geqslant j, Z_l = s]$$

$$\phi^{-}_{t|l}(j) \coloneqq \lim_{s\uparrow 0} \mathbb{E}[Y_t(j)-Y_t(j-1)|D_t \geqslant j, Z_l = s] = \lim_{s\uparrow 0} \mathbb{E}[Y_t(j)-Y_t(j-1)|\boldsymbol{D}_t(\boldsymbol{Z}_{t|l}(s)) \geqslant j, Z_l = s]$$

$$\omega^{+}_{t|l}(j) \coloneqq \lim_{s\downarrow 0} \mathbb{P}[D_t \geqslant j|Z_l = s] = \lim_{s\downarrow 0} \mathbb{P}[\boldsymbol{D}_t(\boldsymbol{Z}_{t|l}(s)) \geqslant j|Z_l = s]$$

$$\omega^{-}_{t|l}(j) \coloneqq \lim_{s\uparrow 0} \mathbb{P}[D_t \geqslant j|Z_l = s] = \lim_{s\uparrow 0} \mathbb{P}[\boldsymbol{D}_t(\boldsymbol{Z}_{t|l}(s)) \geqslant j|Z_l = s]$$

Assumption 2 (Local monotonicity of treatment rule)**.** *For each* t *and* l*, with* $l \leqslant t$*, there exists* $\check{\epsilon} > 0$ *such that, for every* z,z' *with* $|z| \vee |z| \leqslant \check{\epsilon}$*,* $z \leqslant z' \implies \mathbb{P}[\boldsymbol{D}_t(\boldsymbol{Z}_{t|l}(z)) \leqslant \boldsymbol{D}_t(\boldsymbol{Z}_{t|l}(z'))] = 1$

Assumption 3 (Local experiment around threshold)**.** *For every* $l \in \{1,2,\dots,n\}$*, there exists some* $\bar{\epsilon} > 0$ *such that, on* $|Z_l| \leqslant \bar{\epsilon}$

$$(\{(Y_t(j))_{j=0}^{t}, \boldsymbol{D}_t\}_{t=1}^{T}, (Z_t)_{t=1}^{l-1}) \text{ is independent of } Z_l\,.$$

Assumption 1 is a technical requirement ensuring that (1.1) exists. Assumption 2 is a local monotonicity assumption: it implies that, on a neighborhood of the threshold, externally increasing the score should almost surely not decrease treatment dosage.

Assumption 3 is our main identification condition. It is a local independence condition that ensures a "local experiment" around the threshold.[10] This assumption ensures that, for every period l, predetermined traits – namely, potential outcomes, potential treatment functions, and the scores *prior* to l –are balanced at either side of the threshold at l. The balance condition captures the notion that, at a given period, units around the threshold do not have perfect control over their score (Lee and Lemieux 2010). Note that the Assumption does not impose balance of scores after the reference period l, as it may well be that being above or below the threshold at l partly determines the score after l – in the LFL application, being eligible to the policy at l may increase credit scores in future periods.

Having defined our setup, we are now able to precise in which sense we require $\tau_{t|l}$ to have a causal interpretation. We adapt the notion of "weak causality" of Blandhol et al. 2022 to our setting.

Definition 1 (Weak Causality)**.** *$\tau_{t|l}$ is a weakly causal parameter if, for any vector of random variables defining potential outcomes that satisfy Assumptions 1, 2 and 3 such that $\mathbb{P}[Y_t(j) - Y_t(j-1) \geqslant 0] = 1$ for every $j = 1 \ldots, t$, one has $\tau_{t|l} \geqslant 0$. Similarly, if $\mathbb{P}[Y_t(j) - Y_t(j-1) \leqslant 0] = 1$ for every $j = 1 \ldots, t$, then one has that $\tau_{t|l} \leqslant 0$*

Our notion of weak causality requires the sign of the estimand to correctly reflect a scenario where marginal effects $Y_t(j) - Y_t(j-1)$ share almost-surely the same sign.[11]

The next result characterizes the estimands $\tau_{t|l}$ under our set of assumptions.

Proposition 1. (*LATE Representation*) *Suppose Assumptions 1, 2, and 3 hold. We then have that:*

1. *If $l > t$, $\tau_{t|l} = 0$.*
2. *If $l = t$,*

$$\tau_{t|l} = \sum_{j=1}^{t} \omega_{t|l}(j)\phi_{t|l}(j)\,, \tag{1.2}$$

where

$$\omega_{t|l}(j) = \lim_{\epsilon \downarrow 0} \mathbb{P}[\boldsymbol{D}_t(\boldsymbol{Z}_{t|l}(\epsilon)) \geqslant j, \boldsymbol{D}_t(\boldsymbol{Z}_{t|l}(-\epsilon)) < j || Z_l| \leqslant \epsilon]$$

[10] In the main text, we opt to state the identification assumption in terms of a local independence condition ([)as in, e.g.,][]Hahn2001, instead of relying on a local continuity condition ([)as in, e.g.,][]Dong2018,Arai2022. While the latter is known to be weaker than the former (M. Cattaneo and Titiunik 2022), deriving our representation in Proposition 1 under this form of restriction would require, in our dynamic setup, to introduce additional notation that would burden the interpretation of the identification condition, while keeping the main message of the representation result intact.

[11] Our notion of weak causality is less demanding than Blandhol et al.'s, as they require their parameter under study to accurately capture the sign of *average* marginal effects across predefined complier populations when the signs of average effects in these groups are all the same. In contrast, we only require our target parameter to accurately capture the signs in an extreme scenario where there is not any sign heterogeneity in treatment effects.

$$\phi_{t|l}(j) = \lim_{\epsilon \downarrow 0} \mathbb{E}[Y_t(j) - Y_t(j-1) | \boldsymbol{D}_t(\boldsymbol{Z}_{t|l}(\epsilon)) \geqslant j, \boldsymbol{D}_t(\boldsymbol{Z}_{t|l}(-\epsilon)) < j, |Z_l| \leqslant \epsilon]$$

In this case, the estimand is weakly causal, and the weights $\omega_{t|l}(j)$ *are identified as the first-stage contrasts* $\omega^+_{t|l}(j) - \omega^-_{t|l}(j)$.

3. *If* $t > l$

$$\begin{aligned} \tau_{t|l} = \sum_{j=1}^{t} \omega_{t|l}(j)\phi_{t|l}(j) + \\ (\omega^+_{t|l}(j) - \omega^\downarrow_{t|l}(j))\phi^\downarrow_{t|l}(j) + \omega^\downarrow_{t|l}(j)(\phi^+_{t|l}(j) - \phi^\downarrow_{t|l}(j)) + \\ (\omega^\uparrow_{t|l}(j) - \omega^-_{t|l}(j))\phi^-_{t|l}(j) + \omega^\uparrow_{t|l}(j)(\phi^\uparrow_{t|l}(j) - \phi^-_{t|l}(j)), \end{aligned} \tag{1.3}$$

where $\omega^\downarrow_{t|l}(j) = \lim_{\epsilon \downarrow 0} \mathbb{P}[\boldsymbol{D}_t(\boldsymbol{Z}_{t|l}(\epsilon)) \geqslant j || Z_l| \leqslant \epsilon]$, $\phi^\downarrow_{t|l}(j) = \lim_{\epsilon \downarrow 0} \mathbb{E}[Y_t(j) - Y_t(j-1) | \boldsymbol{D}_t(\boldsymbol{Z}_{t|l}(\epsilon)) \geqslant j, |Z_l| \leqslant \epsilon]$, $\omega^\uparrow_{t|l}(j) = \lim_{\epsilon \uparrow 0} \mathbb{P}[\boldsymbol{D}_t(\boldsymbol{Z}_{t|l}(\epsilon)) \geqslant j || Z_l| \leqslant \epsilon]$ *and* $\phi^\uparrow_{t|l}(j) = \lim_{\epsilon \uparrow 0} \mathbb{E}[Y_t(j) - Y_t(j-1) | \boldsymbol{D}_t(\boldsymbol{Z}_{t|l}(\epsilon)) \geqslant j, |Z_l| \leqslant \epsilon]$. *The estimand is not weakly causal.*

Proof. See Appendix 1.8.1. □

Proposition 1 clarifies under which conditions the parameter $\tau_{t|l}$ has a weakly causal interpretation. There are three main cases. When one compares outcomes determined prior to a given period, for units just below and above that given period's threshold, then, as expected, the RD estimand is zero. This is due to the fact that, in the RD design, predetermined traits are exactly balanced at a given period's threshold.

In the second case, when both the score and the outcome of interest are in the same period, we have a similar result to static fuzzy RD designs. In this case, the reduced-form RD estimand identifies a weighted-average of complier treatment effects, where a unit in the population is defined as a complier with respect to a length of exposure j if her treatment changes from strictly below j to weakly above it when her score is perturbed from a value marginally below the threshold to a value marginally above it. The parameter aggregates each marginal treatment effect $Y_t(j) - Y_t(j-1)$ at each complier subpopulation with respect to each level of exposure j.

Finally, in the third case, when $l < t$, we have a setting where the parameter is not weakly causal. This is due to the dynamic nature of the running variable, and the possibility of being above or below the threshold at l discontinuously affecting future scores. In this case, the comparison underlying the RD estimand need not be causal, as the differences between outcomes just below and above the threshold at l may be effectively comparing units with different exposures to the treatment and that self-select into these exposures according to their expected gains. The effects of differential treatment intensity are captured by the terms $(\omega^+_{t|l}(j) - \omega^\downarrow_{t|l}(j))\phi^\downarrow_{t|l}(j)$ and $(\omega^\uparrow_{t|l}(j) - \omega^-_{t|l}(j))\phi^-_{t|l}(j)$, which measure how the

probability of having an exposure of at least j by period t is affected by being below or above the threshold at l *through its effect over future scores*. The "direct" effect of being above the threshold at l does not enter these quantities, as the score entering the potential treatment function in the comparisons $\omega^{+}_{t|l}(j) - \omega^{\downarrow}_{t|l}(j)$ and $\omega^{+}_{t|l}(j) - \omega^{\downarrow}_{t|l}(j)$ is fixed. Differential exposure to the treatment violates weak causality because the parameter may reflect an individual's *margin of choice over treatment* rather than the actual effects of the policy over outcomes. For example, if there is not any treatment effect heterogeneity, so that there exists constants $\{\delta_j\}_{j=1}^{t}$ such that $Y_t(j) - Y_{t-j}(j-1) = \delta_j$ a.s., for every j, then:

$$\tau_{t|l} = \sum_{j=1}^{t} (\omega^{+}_{t|l}(j) - \omega^{-}_{t|l}(j))\delta_j \,.$$

In this case, if $\omega^{+}_{t|l}(j) < \omega^{-}_{t|l}(j)$ for some j, so that being marginally above the threshold at l generates a "compensation" effect over future scores that decreases the likelihood of having an exposure of at least j by period t – for example, in the LFL example, if the policy induces excessive risk-taking, thus increasing the probability of default over future periods –, then $\tau_{t|l}$ need not accurately reflect the sign of treatment effects, even when there is not any heterogeneity in these.

The effects of self-selection into the gains of treatment are captured by the terms $\omega^{\downarrow}_{t|l}(j)(\phi^{+}_{t|l}(j) - \phi^{\downarrow}_{t|l}(j))$ and $\omega^{\uparrow}_{t|l}(j)(\phi^{\uparrow}_{t|l}(j) - \phi^{-}_{t|l}(j))$. Its impacts over the parameter are similar to the effect of differential treatment intensity. Specifically, if units just above the threshold at l with exposure by period t of at least j have expected effects different than units that would also have exposure of at least j by t if the score at l entering the potential treatment function were externally fixed just above the threshold, but that in practice experimented a score slightly below the threshold, then the difference $\phi^{+}_{t|l}(j) - \phi^{\downarrow}_{t|l}(j)$ may be such that the parameter does not need accurately reflect the sign of treatment effect heterogeneity. Once again, the difference between $\phi^{+}_{t|l}(j)$ and $\phi^{\downarrow}_{t|l}(j)$ arises because being below and above a threshold at l may generate discontinuous effects over future scores, which generates a potential margin of self-selection into the gains of treatment as individuals with higher scores at l may have more or less "options" in the future and may self-select into these according to their expectations regarding $Y_t(j) - Y_t(j-1)$.

The following corollary provides one necessary and one sufficient restriction for weak causality to hold when $l < t$.

Corollary 1. *If $l < t$, a necessary condition for weak causality to hold is that $\omega^{+}_{t|l}(j) \geqslant \omega^{-}_{t|l}(j)$ for every $j = 1, \dots t$. This condition is implied by $\lim_{\epsilon \downarrow 0} Z_{l+1}, Z_{l+2}, \dots, Z_t | Z_l = \epsilon$ dominating $\lim_{\epsilon \uparrow 0} Z_{l+1}, Z_{l+2}, \dots, Z_t | Z_l = \epsilon$ in first-order stochastic dominance. Moreover, a sufficient condition for weak causality is that the environment be such that $\omega^{+}_{t|l}(j) \geqslant \omega^{-}_{t|l}(j)$ and $\phi^{+}_{t|l}(j) = \phi^{-}_{t|l}(j)$ for every j. In this case, we can write:*

$$\tau_{t|l} = \sum_{t=1}^{l} (\omega^{+}_{t|l}(j) - \omega^{-}_{t|l}(j))\phi^{+}_{t|l}(j)\,.$$

The corollary shows that, for $\tau_{t|l}$ to be weakly causal, it must be that the *overall effect* of being slightly above the threshold at l, on the probability of exposure being at least j, must be nonnegative, for every $j = 1, \ldots, t$. This condition is implied by being slightly above the threshold at l increasing the likelihood of observing higher scores in future periods. Finally, if we also restrict treatment effect heterogeneity in the environment, so that the average effects among individuals marginally below the threshold at l that choose an exposure of at least j by t is equal to the average effect among individuals marginally *above* the threshold at period l and that also choose an exposure of at least j by t, then we also have a weakly causal parameter.

1.3.3 Aggregation strategy

The previous section provided conditions for a contrast of average outcomes below and above a given period's threshold to have a weakly causal interpretation. We have seen that, unless $l \geqslant t$, one requires additional conditions restricting differential treatment exposure and self-selection on expected gains in order for the parameter to accurately capture the sign of effects. Weak causality of $\tau_{t|l}$ may not be sufficient for analysing the effects of a policy, though. For example, if one is interested in dynamics, it may be important to identify averages of the *marginal effects*, $Y_t(j) - Y_t(j-1)$, $t = 1, \ldots, t$, $j = 1, \ldots, t$; or a path of averages of the *cummulative* effects $Y_t(t) - Y_t(0)$, $t = 1, \ldots, T$.

We consider how restrictions on treatment effect heterogeneity may enable the identification of average marginal effects (average cummulative effects). Our approach is inspired by C. Caetano, G. Caetano, and Carlos Escanciano 2023, who consider the identifying power of homogeneity restrictions in cross-sectional RD designs with multi-valued treatments and a single discontinuity. We view homogeneity assumptions as the natural type of restriction in our setup, since the interpretation of the "building-block" RD estimand $\tau_{t|l}$ when $l < t$ already necessitates this type restriction. Moreover, existing alternatives in the literature are less appealing in our setting. For example, Sojitra and Syrgkanis 2024 provide identification results for LATEs in a dynamic setting where a vector of instruments is sequentially randomized. It would be difficult to leverage their instruments in our case, because running variables are only *locally* sequentially randomized, and restricting our analysis to the subpopulation that is *always* around the threshold could generate a selection bias, insofar as past scores determine future ones. Even if one were to disregard this type of bias, however, we emphasize that, for estimation purposes, this approach would suffer from a curse of dimensionality, as the number of available observations would rapidly deteriorate as a function of T, for a given bandwidth around

scores.

In Appendix 1.8.2, we also compare our approach to other alternatives especially tailored to dynamic RD settings. Appendix 1.8.2 shows that, if we couple a below-and-above threshold homogeneity condition, $\phi^{+}_{t|l}(j) = \phi^{-}_{t|l}(j)$, with a time-stationarity assumption on marginal effects, it is possible to recursively identify the full path of marginal effects. While appealing, this assumption is not well-suited to settings where there is aggregate trends or volatility in economic conditions, since in this case one would expect the effect of first entering the policy to vary across time. One expects that to be the case in our empirical application, where the effects of the policy should vary with the macroeconomic environment. In Appendix 1.8.2, we show that, by conditioning on subpopulations defined by past-period exposures, it is possible to point-identify, in the subpopulation with prior exposure equal to $t-1$, the average complier marginal effects of increasing exposure from $t-1$ to t *at period* t without additional homogeneity assumptions. However, as we argue in the Appendix, this approach suffers from a curse of dimensionality, becoming considerably more demanding of the data as one entails higher values of t; and would be generally inefficient when interest lies on cummulative effects. Moreover, even if estimation was not a concern, we show that the effects identified by this strategy in our empirical application concern a complier subpopulation that is expected to be the least affected by the policy, which is also unappealing on conceptual grounds. Finally, 1.8.2 discusses alternative approaches to identification available in the literature that were designed with specific dynamic RD settings in mind (Cellini, Ferreira, and Rothstein 2010; Hsu and Shen 2021). We show that, when translated to our setting, the identifying assumptions in these papers either impose a constant marginal return of increasing exposure to the policy, which seems especially unappealing for the types of economic outcomes considered in our empirical application or rely on a conditional independence assumption that strongly limits the association between scores and current period potential outcomes.

In light of the limitations discussed above, we propose instead to resort to a combination of a below-and-above-threshold homogeneity assumption with a between-score homogeneity restriction to identify effects. Formally, we assume that $\phi^{+}_{t|l}(j) = \phi^{-}_{t|l}(j) = \phi_t(j)$ for every t, $l \leqslant t$ and $j \leqslant t$. Notice that, while this assumption limits heterogeneity of effects across subpopulations at different thresholds for given time period t and exposure j, it allows for time-series nonstationarity and non-constant returns in the marginal effects of increasing exposure. Under the homogeneity assumptions, we are able to identify the vector of marginal effects at t, $\phi_t = (\phi_1(j), \phi_2(j), \ldots, \phi_t(j))$, as the unique solution of the linear system:

$$\underbrace{\begin{bmatrix} \omega^+_{t|1}(1) - \omega^-_{t|1}(1) & \omega^+_{t|1}(2) - \omega^-_{t|1}(2) & \dots & \omega^+_{t|1}(t) - \omega^-_{t|1}(t) \\ \omega^+_{t|2}(1) - \omega^-_{t|2}(1) & \omega^+_{t|2}(2) - \omega^-_{t|2}(2) & \dots & \omega^+_{t|2}(t) - \omega^-_{t|2}(t) \\ \vdots & \vdots & \dots & \vdots \\ \omega^+_{t|t}(1) - \omega^-_{t|t}(1) & \omega^+_{t|t}(2) - \omega^-_{t|t}(2) & \dots & \omega^+_{t|t}(t) - \omega^-_{t|t}(t) \end{bmatrix}}_{=\boldsymbol{\Omega}_t} \boldsymbol{\phi}_t = \begin{bmatrix} \tau_{t|1} \\ \tau_{t|2} \\ \vdots \\ \tau_{t|t} \end{bmatrix}, \tag{1.4}$$

under the assumption that the matrix $\boldsymbol{\Omega}_t$ has full rank, which ensures unicity of the solution of the linear system (1.5). The assumption of full rank requires thresholds in different periods to differentially affect exposure at t, which may be seen as a relevance condition on the first-stage contrasts (jumps) $\omega^+_{t|l}(j) - \omega^-_{t|l}(j)$.

Estimation and inference on $\boldsymbol{\phi}_t$ identified through (1.4) is immediate. Specifically, one may estimate $\omega^+_{t|l}(1) - \omega^-_{t|l}(1)$ with the RD estimators of the "first-stage" contrasts $\lim_{s\downarrow 0}\mathbb{E}[\mathbf{1}\{D_t \geqslant j\}|Z_l = s] - \lim_{s\uparrow 0}\mathbb{E}[\mathbf{1}\{D_t \geqslant j\}|Z_l = s]$, and the $\tau_{t|l}$ by the corresponding reduced-form contrasts. One may then estimate $\boldsymbol{\phi}_t$ by numerically solving the sample analog of the system (1.4), i.e. by numerically inverting the estimator $\widehat{\Omega}_t$ of Ω_t. Inference may be performed using the delta-method, under assumptions that ensure asymptotic normality of the building-block RD estimands (Calonico, M. D. Cattaneo, and Titiunik 2014).

While appealing, the plug-in approach outlined in the previous paragraph may perform poorly in practice if $\widehat{\Omega}_t$ is nearly collinear. In this case, the estimator may exhibit substantial biases, and asymptotic approximations may deliver inference methods with severe distortions, as in a weak instrumental variable settings (Andrews, Stock, and Sun 2019).[12] We thus propose to impose additional restrictions on the evolution of the effect $\boldsymbol{\phi}_t$ that may enable the construction of more efficient estimators. Specifically, assuming that the outcome of interest is included in first differences in the specification, we propose to parametrize $\boldsymbol{\psi}_t(j)$, the average cummulative effect of the policy *in levels*, for an individual at period t that has been exposed to the policy continuously over j periods, by:

$$\boldsymbol{\psi}_t(j) = \sum_{l=0}^{j-1}\sum_{s=l}^{j-1}\phi_{t-l}(j-s) = \underbrace{\delta_t}_{\text{calendar time effect}} + \underbrace{\mu_{t-j}}_{\text{cohort effect}} + \underbrace{\gamma(j)}_{\text{exposure effect}}. \tag{1.5}$$

In Appendix 1.8.3, we provide an optimal-minimum distance estimator that efficiently incorporates information from the first-stage matrices $\{\hat{\Omega}_t : t = 1, \dots, T\}$, and the reduced-

[12] Notice that there is a tension between the validity of the homogeneity assumption $\phi^+_{t|l}(j) = \phi^+_{t|l}(j)$ and the full rank of Ω_t. Indeed, one would expect that $\phi^+_{t|l}(j) = \phi^+_{t|l}(j))$ when the score is highly persistent across time, in which case self-selection is limited as $(Z_{l+1}, Z_{l+2}, \dots, Z_t)$ is effectively randomized when the Z_l is in a neighborhood of 0. However, this is also the case where one would not expect many variability across the rows of Ω_t.

form estimators $\{\hat{\tau}_{t|l} : t = 1, \ldots, T, l \leqslant t\}$ to estimate (1.5). We also show how our approach to estimation immediately yields overidentifying restrictions J-test, thus enabling researchers to assess the appropriateness of parametrization (1.5). We further provide computational details of the implementation of our methods, including how to efficiently recover the covariances between the building-block RD estimators in `R`; and propose a placebo test that may be implemented if pre-treatment data is available.

1.4 Empirical Results

1.4.1 Empirical Strategy

Sample: Our unit of observation is the firm. We focus on firms rated by the Central Bank of Brazil, filtering for those that participated in the corporate bond market at any point between 2020 and 2023. Financial firms are excluded, as they are ineligible for the policy. We limit the sample to firms that were active between 2020 and 2022, and further restrict it to those that had at least 100 employees at any point between 2019 and 2022. This filtering aims to exclude firms potentially created solely to issue corporate bonds without employing a workforce or participating in interfirm payments within supply chains. We use the time period 2020-2022, with outcomes at the monthly level. We stop at 2022 because RAIS labor data is only available until this period. This leaves us with 677 unique firms and 24,372 firm $\times$ month observations, as we have a complete panel for most outcomes of interest.

Table 1.2 presents summary statistics for firms in our samples and their respective corporate bonds in the pre-intervention period of October 2021. For firms around the threshold, differences in most outcomes, such as total debt, corporate bond share, payment outflows, wage bill, and the number of workers, are not statistically significant, with p-values well above 0.05. However, when comparing eligible firms around the threshold to all eligible firms, we observe that firms in the full sample tend to have larger total debt and wage bills. The full sample of eligible firms also has more workers on average, borrow more, and at a lower spread, indicating that they tend to be larger and less financially constrained than the marginal firm.

Econometric Implementation: We implement the dynamic RDD estimator from Section 1.3.3 by first differencing all outcomes of interest, ΔY_{it}, and applying a 2% winsorization to mitigate the influence of outliers. We use the normalized score Z_{it} from Section 3.3. Finally, when normalizing the estimator by $\mathbb{E}[\tilde{Y}_0 \mid Z_1 = 0]$, where $\tilde{Y}_0$ is the level of the outcome of interest, we estimate the conditional expectation via local polynomial regression. Numerical implementation details are provided in Appendix 1.8.3.

We parametrize the cumulative treatment effects following the system defined in 1.5. The resulting sequence $\{\psi_t(t)\}_{t=1}^{T}$ represents the cumulative impact of an absorbing

treatment over time, capturing how the treatment effect evolves as the treatment persists for a unit that maintained eligibility status.

We then present the results of the optimal estimator, derived in Appendix 1.8.3, which is designed to minimize the variance of the estimated treatment effects. Additionally, we include the J-test for overidentifying restrictions, which tests the consistency of the imposed parametrization of cumulative treatment effects with the data.

Bandwidth Choice: We select a cohort-specific RDD bandwidth, h_s, ensuring consistency across all time periods and outcomes. Following the bandwidth selection method from Calonico, M. D. Cattaneo, and Farrell 2020, we base this choice on the pre-treatment (October 2021) wage bill. We use two distinct MSE-optimal bandwidth selectors (for firms below and above the cutoff) and apply a triangular kernel in all specifications.

Visualization of Cumulative Treatment Effects The path $\{\psi_t(t)\}_{t=1}^{T}$ represents the cummulative effects of an absorbing state. This cumulative effect is estimated by working with outcomes in the first differences. For a percentage change interpretation, each estimate can be normalized by an estimator of $\mathbb{E}[Y_0 \mid Z_1 = 0]$, where Y_0 is the pre-period outcome in levels. Under our identification assumption, this is valid because units around the threshold are comparable.

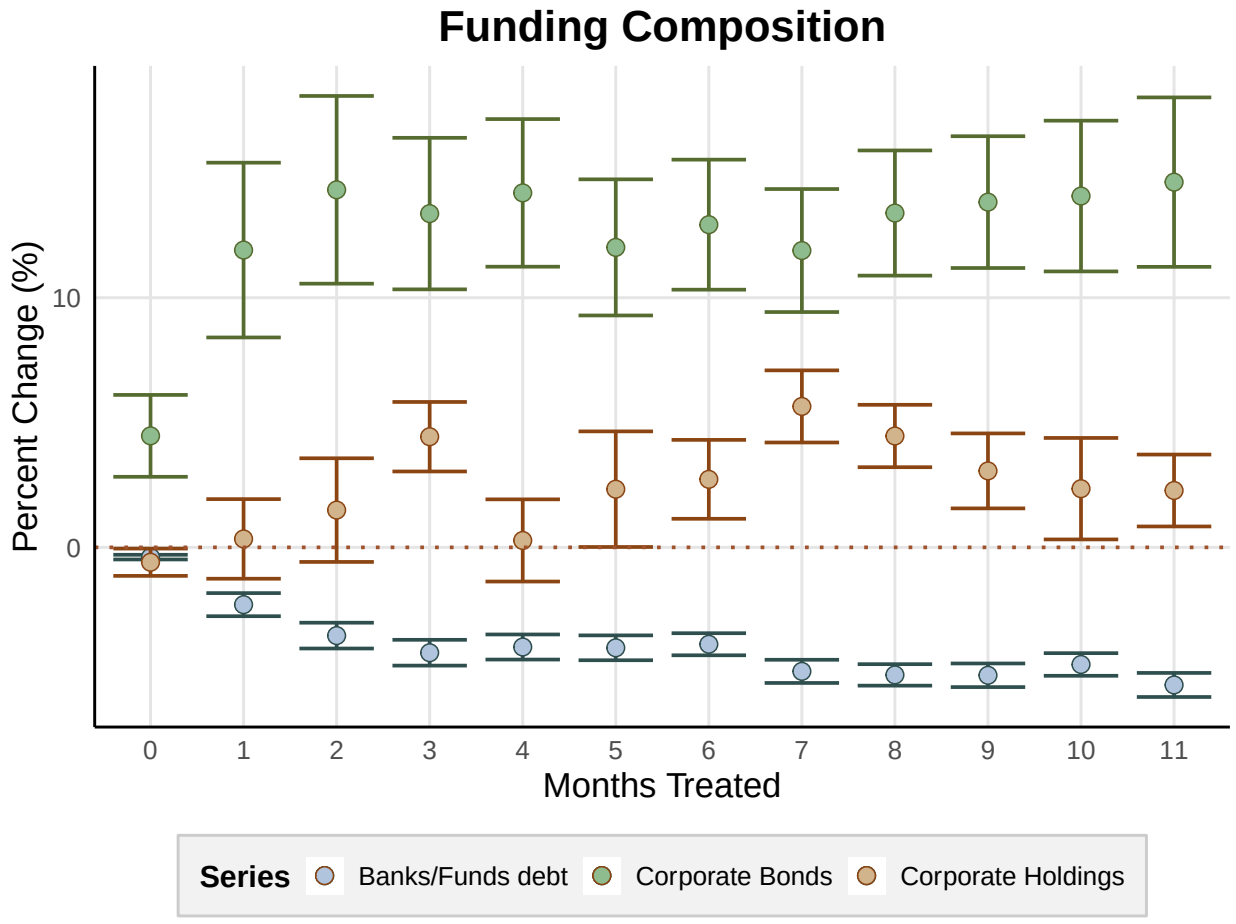
This approach ensures that we track changes in outcomes for the same group of firms. Tracking the same group is important for maintaining consistency and comparability in the analysis. By focusing on stable cohorts, we ensure that observed changes in outcomes are not attributable to mechanical variations in the composition of firms. This reduces the risk of confounding effects from firms entering or exiting the cohort, which could distort treatment effect estimates, and allows for a more precise comparison of within-firm changes across different outcomes.

1.4.2 The effects on primary corporate bonds market

We first estimate the policy's causal effect on primary corporate bond markets. Figure A10 plots the estimates of the dynamic causal effects of eligibility to the LFL policy on firms' outstanding debt. shows the dynamic causal effects of LFL policy eligibility on firms' outstanding debt. There is a significant 12% increase in corporate bond debt after one year in the policy, with a slight rise in foreign debt. However, this is partially offset by a 5% decrease in intermediate bank/fund debt. Overall we observe a decrease in the cost of capital associated with corporate bonds of around 20 basis points. Figure A6 shows that we do not find any statistically significant effect of the policy on corporate debt maturity or bank/fund interest rates charged to eligible firms. Thus we can show that the implied reduction in issuance spread of corporate bonds was around 67 basis points.

We estimate changes in safe asset holdings following eligibility, holdings by around 12%. We conclude that eligibility led firms to increase net total debt, as the rise in corporate bond issuance was only partially offset by reductions in other forms of intermediated financing. Notably, there was a catalyst effect from firms' use of internal funds. In the next subsection, we will link this to a reduction in firms' precautionary savings, likely driven by the expectation of improved future credit access.

Figure 1.6: Dynamic Treatment Effects on Capital Structure.



Note: The figure shows estimates of cumulative treatment effects on outstanding debt volume and safe asset holdings. with error bars representing 95% confidence intervals. Differences in Debt volume are normalized by pre-period total debt.

1.4.3 Effects on bond liquidity in secondary markets

We show how corporate bond liquidity is affected in the secondary markets. To quantify this effect, we measure firm-level turnover liquidity in the over-the-counter (OTC) market through three specific operations: (i) trading in the secondary market, (ii) using corporate bonds as collateral in short-term repurchase (repo) operations, and (iii) utilizing them as collateral in other financial transactions, including those involving the LFL facilities. We define turnover liquidity on a monthly basis at the asset level using three metrics: (i) the total quantity traded in the secondary market divided by the outstanding market quantity, (ii) the total quantity posted as collateral in repo operations on the last business day of the month divided by the outstanding market quantity, and (iii) the total quantity posted as collateral in other operations divided by the outstanding market quantity. Finally, we aggregate these metrics at the firm level using a volume-weighted average based on the outstanding nominal value of each asset.

We estimate an increase in service flow from a "buy-and-hold" corporate bond strategy in Figure 1.9. Initially, there was a mechanical rise in collateral posted under LFL, primarily from commercial banks' holdings already used in short-term repo operations. However, one year after the policy, banks compensate by increasing repo usage relative to initial volumes while trading activity decreases.

1.4.4 The real effects of eligibility

We estimate that the policy had real effects. Figure A10 shows that eligible firms causally increased the wage bill by about 6 % (in real terms) one year after becoming eligible. and temporarily increased net payment outflows to other non-financial firms (Figure 1.8). In the appendix we show that this result is driven by an increase in payment outflows rather than a decrease in payment inflows).

Combining the results from the previous subsection, we conclude that the use of internal finance amplified the real effects of the increased demand for corporate bonds. We interpret this effect as a reduction in precautionary savings, as firms typically accumulate safe assets to self-insure against liquidity shocks. This trend is evident in the time series of firm holdings in our sample in Figure A3, where we observe a significant increase in safe asset holdings during the COVID crisis.

While there are no estimates of the real effects of collateral eligibility on firms, some papers examine the real effects of direct purchases of corporate bonds (Darmouni and Siani 2022, Adelino et al. 2023), and others discuss firm responses to increases in corporate debt convenience yield (Mota 2023). The prevailing wisdom in this literature suggests that such policies do not have direct real effects; instead, firms typically increase cash holdings, repay debt, or distribute payments as dividends. Our results sharply contrast with this

view. Not only do we observe real effects, but these effects are amplified by firms' use of internal funds.

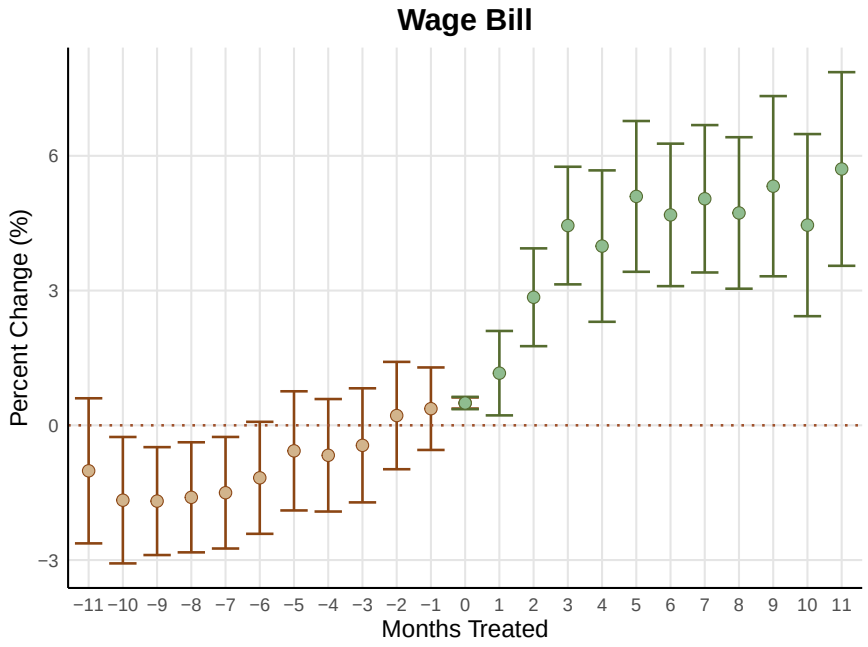


Figure 1.7: Dynamic Treatment Effects on Wage Bill.

Note: The figure shows estimates of cumulative treatment effects on the wage bill, with error bars representing 95% confidence intervals. Wage bill differences are normalized by the pre-period wage bill.

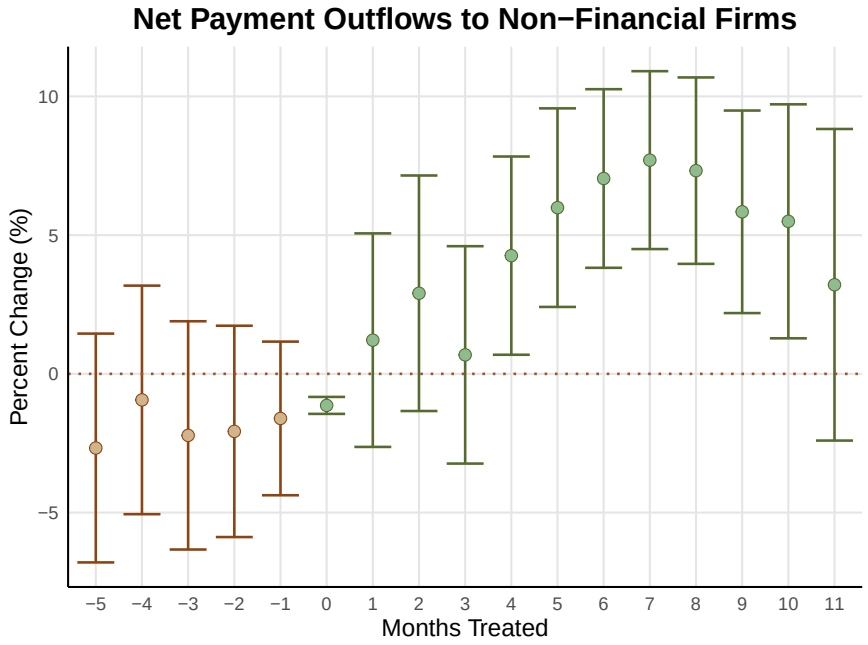


Figure 1.8: Dynamic Treatment Effects on Net Payment Outflows to Non-Financial Firms.

Note: The figure shows estimates of cumulative treatment effects on the net payment outflows to non-financial firms, with error bars representing 95% confidence intervals. Net payment outflows are normalized by total outflows from October 2020 to October 2021.

1.5 Model

In this section, we implement a semi-structural approach that builds on a theoretical model of segmented financial markets to address the following questions: (1) To what extent does the LFL policy induce liquidity injections in eligible firms? (2) How do these injections affect firms' production and financial decisions?

LFL and Liquidity Injections: We analyze the relationship between firm eligibility and the induced liquidity injections from financial intermediaries, using a segmented financial market model.

We develop a model based on the asset-liability management (ALM) practices of commercial banks. In this framework, a set of commercial units handles both loan origination and deposit collection from firms, while a separate ALM sets banks' internal fund transfer prices (FTP). The ALM sets FTP rates that account for liquidity costs and regulatory requirements, thus ensuring that commercial units fully internalize these costs and benefits. When a unit makes a corporate loan or purchases a firm's bond, it "borrows" from the ALM at a defined FTP; conversely, deposits collected from firms are "lent" to the ALM at a set price, aligning incentives across divisions and managing liquidity risk efficiently.[13] In the context of the LFL policy, the ALM determines whether to access central bank facilities and how much corporate bond collateral to post. A key implication is that the transmission mechanism of firm eligibility operates directly through its effect on the FTP charged to lending units, thus shaping their bond demand. If intermediaries' demand for corporate bonds is sufficiently price-inelastic, the reduction in the FTP can be passed through to either bond prices or the quantity borrowed, and, in the latter case, it effectively operates as a liquidity injection that is a subsidy to firm borrowing.[14] See Figure A11 for a diagram.

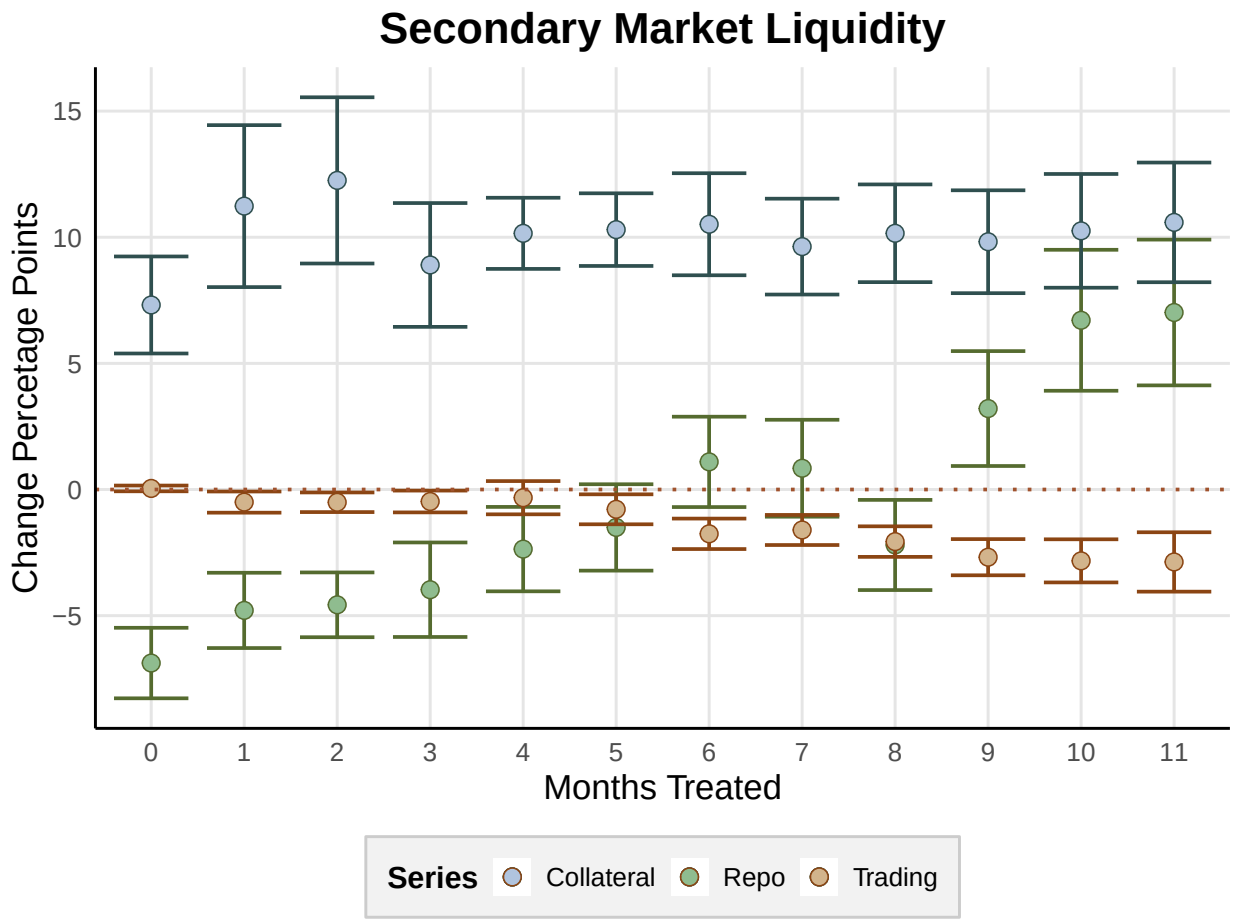
Finally, we develop an approach to estimate this elasticity based on firms' trade-offs between borrowing and holding cash-like assets. We leverage the fact that we do observe from our data, both the firm borrowing cost and return by holding safe assets (from the OTC data). Our results indicate that the demand for corporate bonds is relatively inelastic, with price elasticity of demand between 1.06 and 1.73, magnitudes in the range of those estimated for the U.S. corporate bond market (Darmouni, Siani, and Xiao 2022). While the literature usually follows the demand system approach from Koijen and Yogo 2019, our method is based on price theory arguments. Firms in our sample hold liquid asset reserves while also carrying debt.[15] The trade-off between the marginal cost of holding

[13] See Moody's (2018) for a brief description of the fund transfer pricing methodology of commercial banks.

[14] For example, in Europe, the collateral haircuts imposed by the ECB on corporate bonds are incorporated by banks into the FTP. See, for instance, this report.

[15] That is not specific to our setting; recently, this fact has been well documented in the US and Europe.

Figure 1.9: Dynamic Treatment Effects on OTC markets turnover liquidity.



Note: The figure shows estimates of cumulative treatment effects on turnover liquidity of corporate bonds,with error bars representing 95% confidence intervals. We measure firm-level turnover liquidity in the OTC market through three types of operations: (i) trading in the secondary market, (ii) use as collateral in short-term repo operations, and (iii) use as collateral in other operations (including LFL). Turnover liquidity is defined monthly at the asset level as: (i) the total traded quantity in the secondary market divided by the outstanding market quantity, (ii) the total quantity posted as collateral in repo operations on the last business day of the month divided by the outstanding market quantity, and (iii) the total quantity posted as collateral in other operations divided by the outstanding market quantity. We aggregate this at the firm level using the volume-weighted average based on the outstanding nominal value of each asset. Effective observations are unique firms with positive weights.

cash and the marginal benefit, driven by the firm's liquidity risk, shapes their liquidity management decisions. Rather than holding cash, firms may issue debt to address liquidity needs, though this approach carries risks, particularly due to the potential need for large issuance that could impact market prices. The magnitude of this price impact (which affects the liquidity risk) depends on market elasticity, and we develop our strategy around this observation.

1.5.1 Setup

Asset Generating Unit

The AGU is a short-lived financial intermediary that can invest in firms' debt instruments but needs to raise funds B_t from the ALM. We assume firm debts are imperfect substitutes in the following form. The (ex-post) aggregation of returns to enter the utility function of the intermediary is given by:

$$\tilde{B}_{t+1} = \left(\int \eta_i^{\frac{1}{\varepsilon}} \left(R_{i,t} \omega_{i,t} B_t \right)^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{1.6}$$

where B_t is the total investment in firm debt, $R_{i,t}$ is the ex-post return, and $\omega_{i,t}$ denotes the portfolio weight of firm i's debt, satisfying

$$\int \omega_{i,t} \, di = 1. \tag{1.7}$$

The term $\eta_i > 0$ is a firm-specific demand shifter, and $\varepsilon > 1$ governs the elasticity of substitution between different firms' debts. As $\varepsilon \to \infty$, then we have the perfect substitute case.

To access funds B_t for investment in corporate debt, the AGU obtains financing from the ALM, which provides B_t at a fund transfer price $\mu(B_t)$, where $\mu(\cdot)$ is a well-behaved function. Additionally, the ALM supplements the investment by $(1 + \tilde{\tau}_c) B_{t+1}^{LFL}$, where

$$B_{t+1}^{LFL} = \left(I(i) \int \eta_i^{\frac{1}{\varepsilon}} \left(R_{i,t} \omega_{it} B_t \right)^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{1.8}$$

where $I(i)$ is an indicator function, with value 1 if i is eligible to the LFL policy. Conversely, for each dollar not allocated to eligible firms, the ALM deducts $(1 - \tilde{\tau}_l) B_{t+1}^{NLFL}$, defined analogously:

$$B_{t+1}^{NLFL} = \left((1 - I(i)) \int \eta_i^{\frac{1}{\varepsilon}} \left(R_{i,t} \omega_{it} B_t \right)^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}} \tag{1.9}$$

Finally, the expected utility of the AGU is:

$$\mathbb{E}_t \left[\eta_b^{\frac{1}{\varepsilon}} \frac{\tilde{B}_{t+1}^{1-\gamma}}{1-\gamma} + \tilde{\tau_c} B_t^{LFL} - \tilde{\tau_l} B_t^{NLFL} - \mu(B_t) B_t \right], \tag{1.10}$$

where we set $\gamma = \frac{1}{\varepsilon}$. Taken the distribution of returns $R_{i,t}$, the eligibility status $I(i)$, $\mu(\cdot)$, $\tilde{\tau_{c,t}}$ and $\tilde{\tau_{l,t}}$ as given, the AGU maximizes (1.10), subject to (1.6) - (1.9).

One key modeling assumption first is that the AGU is short-lived or exhibits myopic decision horizons. This assumption is analogous to the assumption in Consumer Asset Pricing Models of Gabaix and Koijen 2021 but within the context of Intermediary Asset Pricing Models (e.g., He and Krishnamurthy 2013). However, the interpretation differs: while the latter emphasizes behavioral finance aspects of household portfolio decisions (e.g., the "narrow framing" in Barberis, Huang, and Thaler 2006), the AGU in our model may be driven by short-term performance incentives prevalent in financial markets (e.g., Morris and Shin 2015; Morris, Shim, and Shin 2017; and, specifically for corporate bond markets, Darmouni, Siani, and Xiao 2022). A critical implication of this distinction is that the AGU's risk aversion need not align with that of households.

We show in the appendix that we can decompose the relative demand between eligible and non-eligible firms' debt in this setting as:

$$\frac{b_{i,t}^{LFL}}{b_{j,t}^{NLFL}} = \left(\frac{\eta_i}{\eta_n}\right) \left(\frac{S_{i,t}}{S_{j,t}}\right)^{\varepsilon-1} \left(\frac{1+\tau_{c,i,t}}{1-\tau_{l,j,t}}\right)^{\varepsilon}, \tag{1.11}$$

where $\left(\frac{1+\tau_{c,i,t}}{1-\tau_{l,j,t}}\right) := \left(\frac{1-\mathbb{P}_{i,t}}{1-\mathbb{P}_{j,t}}\right)\left(\frac{1+\bar{\tau}_{c,t}}{1-\bar{\tau}_{l,t}}\right)$ has an interpretation of a (risk-adjusted) relative transfer between eligible and noneligible firms. The implied liquidity injection (as a % of total debt) for the marginal firms is given by:

$$L_{i,t}^{LFL} = 1 - \left(\frac{1+\tau_{c,i,t}}{1-\tau_{l,i,t}}\right)^{-\varepsilon}.$$

Remark: perfectly elastic markets We briefly discuss what would happen if the market were perfectly elastics, i.e., $\varepsilon \to \infty$, which is a common assumption in many asset pricing and macro-finance models. If we restrict to firms with the same default risk and η_i (and for, simplicity, assuming the LFL does not affect risk), as will be the case for firms close the eligibility threshold, we can rearrange (1.11):

$$\left(\frac{b_{i,t}^{LFL}}{b_{j,t}^{NLFL}}\right)^{\frac{1}{\varepsilon}} = \left(\frac{S_{i,t}}{S_{j,t}}\right)^{\frac{\varepsilon-1}{\varepsilon}} \left(\frac{1+\tau_{c,t}}{1-\tau_{l,t}}\right).$$

Taking the limit as $\varepsilon \to \infty$ we will have that $\left(\frac{1+\tau_{c,t}}{1-\tau_{l,t}}\right) = \left(\frac{S_{j,t}}{S_{i,t}}\right)$, so changes in prices will directly reflect changes in implied subsidies. In cases where effects on cost of capital are

small, assuming perfectly elastic markets would imply small subsidies.

Liability Generating Unit

This unit determines $R^f_{i,t}$, the interest rate on cash-like assets that the firm i will receive at time t. For our analysis, it is sufficient that there exists cross-sectional variation in $R^f_{i,t}$. Such variation may arise from high turnover in a firm's cash holdings, as banks typically offer lower remuneration to firms with high turnover due to the associated liquidity risk. This could be implemented by offering term deposit contracts that penalize early withdrawals (in terms of return) or by providing lower returns in checking accounts.

In the next subsection, we discuss how the firm's borrowing cost is linked to the return on cash-like assets, forming the basis for our estimation strategy of the structural parameter ε.

1.5.2 Firms

For this part, we only require that the following first-order condition holds for the firm,

$$\frac{1}{R^f_{it}} = \frac{\varepsilon - 1}{\varepsilon - 2}\mathbb{E}_t\left[\Lambda_{t+1} s_{i,t+1}\right], \tag{1.12}$$

where R^f_{it} represent the interest rate on cash-like assets that firm i faces, $s_{i,t+1}$ denote the spread between the cost of capital and the savings rate, and Λ_{t+1} be the firm's stochastic discount factor (SDF). The left-hand side of the equation captures the marginal cost of holding one unit of cash for the next period, while the right-hand side represents the marginal benefit, which reflects the firm's liquidity risk. Rather than holding cash, a firm might choose to issue debt when liquidity is needed, but the uncertainty around interest rates and the potential necessity to issue a large volume introduces risk. In the appendix, we present one model where this condition holds.[16]

It's worth noting a logical implication of market elasticity for firms' liquidity management decisions. If markets are not perfectly elastic, then firms internalize the price impact of a potential large debt issuance in case there is a liquidity shock or an immediate need for external finance, which can lead to a larger liquidity preference.[17]

[16] Our approach builds on the assumption that firms simultaneously borrow and save, creating a trade-off in liquidity management. The three largest financial institutions in our Brazilian empirical setting referenced this in a recent proposal (September 2024), suggesting a Central Bank Digital Currency use case that would enable firms to use term deposits as collateral for short-term working capital loans. They argued that this would impact both the return on deposits and the expected spread, influencing the trade-off we emphasize and enhancing risk-sharing between banks and firms.

[17] A simple case where this would hold is if the deposit rate is fixed and the marginal benefit decreases with cash holdings. With cash buffers, firms borrow less during liquidity shocks. Given $\frac{\varepsilon-1}{\varepsilon-2} > 1$, the marginal benefit curve flattens, making firms hold more cash. Generally, this requires the marginal benefit curve to be steeper than the marginal cost curve in the relevant region.

1.5.3 Identification, estimation, and results

We require estimates of some crucial parameters to answer our original questions. First, in Section 1.5.3, we outline a strategy to identify and estimate the elasticity of substitution in the demand for bonds ϵ. Having recovered ϵ, Section 1.5.3 proposes a strategy to recover the relative transfer rate between eligible and non-eligible firms implied by the policy. Section 1.5.3 then uses these estimates to compute the effect of LFL-implied subsidies.

Identifying and estimating the elasticity parameter ε

Rearranging equation (1.12), we have

$$\mathbb{E}_t[s_{i,t+1}] = \mathbb{E}_t[(1-\Lambda_{t+1})s_{i,t+1}] + \underbrace{\frac{\varepsilon-2}{\varepsilon-1}}_{\beta}\frac{1}{R^f_{it}}\,. \tag{1.13}$$

Using (1.13), we generate an estimating equation, which we take to the data, as follows. First, we consider a two-way fixed effects approximation to $\mathbb{E}_t[(1-\Lambda_{t+1})s_{i,t+1}]$, i.e. we replace it with $\alpha_i + \gamma_t$ in (1.13). This produces the following linear specification:

$$s_{i,t+1} = \alpha_i + \gamma_t + \beta\frac{1}{R^f_{i,t}} + e_{i,t+1}\,, \tag{1.14}$$

where the error term satisfies the sequential exogeneity assumption:

$$\mathbb{E}[e_{i,t+1}|R_{i,t}, R_{i,t-1}, \ldots] = 0\,.$$

Consistent estimation of β in short-panels may then be performed by first taking the first-difference of (1.14):

$$\Delta s_{i,t+1} = \delta_t + \beta\Delta\frac{1}{R^f_{i,t}} + \Delta e_{i,t+1}\,, \tag{1.15}$$

and then estimating (1.15) through a generalized method of moments estimator, using $\frac{1}{R^f_{i,t-c}}$, $c \leqslant 1$, as instruments to $\Delta\frac{1}{R^f_{i,t}}$, and including a full set of time dummies as controls, in analogy to Arellano and Bond 1991.

The validity of our proposed estimation strategy hinges on two requirements. First, estimation of (1.15) presupposes the existence of both time-series as well as cross-sectional variation in the firm-level savings rate $R^f_{i,t}$. This is indeed the case in our empirical setting, where we observe substantial variability in the return to firms' cash-like holdings both between firms as well as over time (see Figures A7 and A8 in the Appendix).

Secondly, the validity of our approach requires the two-way fixed effects structure to be indeed a good approximation to the conditional expectation $\mathbb{E}_t[(1-\Lambda_{t+1})s_{i,t+1}]$, i.e.

that $\mathbb{E}_t[(1-\Lambda_{t+1})s_{i,t+1}] \approx \alpha_i + \gamma_t$. Our option for a two-way fixed effects approximation rests on a "double" justification for this method. To see this, notice that, if the economy were on an aggregate steady-state, then Λ_{t+1} would be constant over time, and the only source of variability in $\mathbb{E}_t[(1-\Lambda_{t+1})s_{i,t+1}]$ would rest on cross-sectional heterogeneity in the expectations over $s_{i,t+1}$. In this case, $\mathbb{E}_t[(1-\Lambda_{t+1})s_{i,t+1}] \approx \alpha_i$, and a one-way fixed effects strategy leveraging time series variation in the savings rate would be sufficient to identify effects. Conversely, if the economy was not in a steady-state, but there was no cross-sectional heterogeneity in firms' assessments regarding $s_{i,t+1}$, then $\mathbb{E}_t[(1-\Lambda_{t+1})s_{i,t+1}] \approx \gamma_t$, and a one-way fixed effects approximation that leverages cross-sectional variation in the savings-rate would suffice. Therefore, when we employ a two-way fixed effects approximation, we are able to construct a "doubly-robust" estimator, in the sense that its validity hinges on either one of the two cases being valid.

Empirical Strategy and Results: We estimate specification (1.15) using the Arellano-Bond type estimator described in the section (1.5.3). We measure the return on cash-like assets $R^f_{i,t}$ as follows. The majority of firms' cash-like holdings have returns indexed to the baseline interest rate, including term deposits, portions of treasury holdings, money market fund shares, and returns on repo operations. At the end of each month, we compute $R^f_{i,t} := R^f_t S^f_{i,t}$, where R^f_t is the baseline rate and $S^f_{i,t}$ represents the volume-weighted average spread over the risk-free rate for all of the firm's cash-like instruments, with adjustments for any instruments with predetermined returns.[18]. We measure the borrowing spread $s_{i,t}$ as the difference between the weighted average interest rate on all observed domestic credit operations—excluding foreign debt, for which interest rates are unavailable in our dataset—and the corporate bond rate in month t, relative to the saving interest rate $R^f_{i,t}$. We estimate using monthly data from January 2020 to December 2023, restricting firms as in the empirical analysis. Results are robust to restricting to the pre-LFL treatment period, using quarterly differences, and including all firms that issue corporate bonds.

Table 1.4 reports our estimation results. We find $\hat{\varepsilon} = 2.31$, corresponding to a price elasticity of demand of approximately 1.31. This estimate aligns with the inelastic market hypothesis proposed by Gabaix and Koijen 2021, which posits that financial markets exhibit significantly lower elasticity than commonly assumed in standard asset pricing and macro-finance models. Additionally, our results are consistent with findings from the U.S. corporate bond market; for example, Darmouni, Siani, and Xiao 2022 estimate elasticities between 0.7 and 1.9, depending on the type of financial institution.[19] However, the interpretation of our estimates differs slightly: while this literature focuses on security-

[18]For instance if a firm has a return of 90% of the baseline interest rate, and the (annualized) baseline interest rate is 3%, then the (annualized) return on cash-like instruments is $R^f_{i,t} = 1.027$.

[19]Darmouni, Siani, and Xiao 2022 find lower elasticities than Bretscher et al. 2022, who find elasticities for individual bonds ranging from one to 14 and argue that the discrepancy likely stems from the use of less granular security groupings than Bretscher et al. 2022.

level or institution-specific elasticities, our estimates capture the aggregate elasticity of firm-level debt. This distinction is important, as we note that the aggregate debt elasticity is important for firms' liquidity management decisions and the subsequent pass-through effects from the financial to production decisions. Thus, our estimation strategy departs from the demand-system approach of Koijen and Yogo 2019 by incorporating the interaction between demand and debt supply, mediated by liquidity risk, as captured in (1.12). Furthermore, our estimates imply a risk aversion parameter of $\hat{\gamma} = 0.43$, indicating that the AGU is risk-averse but less so than a typical household with log utility function over consumption.

In the next subsection, we combine our elasticity estimates with reduced-form dynamic RDD results to quantify the implicit debt subsidy induced by the LFL policy for marginal firms (those near the eligibility threshold).

Identifying and estimating the relative subsidy rate implied by the LFL policy

Having backed-out estimates for the elasticity parameter ϵ, we are able to identify the relative transfer rate between eligible and non-eligible firms implied by the LFL policy as follows. Let $b_{i,t}(d)$ and $S_{i,t}(d)$, $d \in \{0,1\}$, denote the potential bond levels and spreads associated with being externally assigned eligibility to the LFL policy continuously over t periods since its implementation ($d = 1$) or never being assigned to it during this period ($d = 0$). The first-order conditions of the AGU problem imply that the relative debt of a firm i eligible to the LFL for t periods, $b_{i,t} = b_{i,t}(1)$, and a firm j never eligible to the policy over this period, $b_{j,t} = b_{j,t}(0)$, satisfy:

$$\frac{b_{i,t}(1)}{b_{j,t}(0)} = \left(\frac{S_{i,t}(1)}{S_{j,t}(0)}\right)^{\varepsilon-1} \left(\frac{\eta_i}{\eta_j}\right)^{\varepsilon} \left(\frac{1+\tau_{c,i,t}(1)}{1-\tau_{l,i,t}(0)}\right)^{\varepsilon},$$

or, taking logs

$$[\log(b_{i,t}(1)) - \log(b_{j,t}(0))] = (\varepsilon-1)[\log(S_{i,t}(1)) - \log(S_{j,t}(0))] + \varepsilon[\log(\eta_i) - \log(\eta_j)] + \varepsilon \log\left(\frac{1+\tau_{c,i,t}(1)}{1-\tau_{l,i,t}(0)}\right),$$

where η_l is the share parameter for the bonds b_l of firm l in the CES aggregator. This parameter reflects an idiosyncratic preference for a firm's bonds – beyond those captured by the asset return — due to unmodeled factors such as complementarity or agency concerns. Importantly, we assumed this parameter to be invariant to a firm's participation in the LFL policy. Using this fact, we may average the relation across eligible firms i and non-legible firms j on the threshold of eligibility in the first-period to obtain:

$$\begin{aligned}&\mathbb{E}[\log(b_{i,t}(1))|\,\text{Score}_{i,1}=0]-\mathbb{E}[\log(b_{j,t}(0))|\,\text{Score}_{j,1}=0]= \\ &(\epsilon-1)(\mathbb{E}[\log(S_{i,t}(1))|\,\text{Score}_{i,1}=0]-\mathbb{E}[\log(S_{j,t}(0))|\,\text{Score}_{j,1}=0]) \\ &+\epsilon\left(\mathbb{E}[\log(1-\tau_{c,i,t}(1))|\,\text{Score}_{i,1}=0]-\mathbb{E}[\log(1-\tau_{l,i,t}(0))|\,\text{Score}_{j,1}=0]\right),\end{aligned} \tag{1.16}$$

where we have used the RD assumption that individuals around the threshold of eligibility should be balanced with respect to predetermined traits (in this case, the shifters η_l). Now, our RD strategy identifies the differences in conditional expectations – the cumulative effects of exposure to the policy (the $\psi_t(t)$ from our empirical setting) on bonds and spreads – in the formula above. Consequently, having obtained an estimate of ϵ, we back out an estimate of the relative risk-adjusted transfer at the marginal firm.

Empirical Strategy and Results: We first note that up to a first-order approximation, we can recast (1.16) as:

$$\psi_t^b(t) \approx (\varepsilon-1)\psi_t^S(t)+\varepsilon\psi_t^\tau(t), \tag{1.17}$$

where $\psi_t^b(t)$ and $\psi_t^S(t)$ represent the reduced-form cumulative effect estimands for total debt and total cost of capital. $\{\psi_t^\tau(t)\}_{t=0}^T$ has the interpretation of a sequence of implied subsidies on the marginal firm. We thus follow a plug-in approach to get the point estimates

$$\widehat{\psi_t^\tau(t)}=\left(\frac{\widehat{\psi_t^b(t)}-(\hat{\varepsilon}-1)\widehat{\psi_t^S(t)}}{\hat{\varepsilon}}\right),$$

where $\widehat{\psi_t^b(t)}$, $\widehat{\psi_t^S(t)}$ are the estimates from the empirical section and $\hat{\varepsilon}$ is the elasticity estimate from the previous subsection. Figure A9 shows, of 4%, which implies, on average, the empirical results. We estimate a stable induced subsidy, on average of 4%, which implies on average a liquidity injection (as a percentage of total debt) of 8.5%. Given that the average haircut on marginal firms' corporate bonds in the LFL is 39%, a back-of-the-envelope calculation suggests that each 1% reduction in the haircut by the BCB induced a 0.14% direct liquidity injection by the financial sector to the marginal firms. In the next subsection we discuss the pass-through for firms' production and liquidity management decisions.

Recovering the corporate effects of LFL-implied transfers

Let $P_{it}(a,s,\{\bar{\tau}_j\}_{j=1}^t)$ be a policy function chosen by the firm and $\psi_t(t)$ the reduced form dynamic RDD estimand associated with the outcome $P_{i,t}$. This policy function depends crucially on three key parts: on an aggregate state vector A_t^d, which may depend on the

policy being in place since period 1 ($d = 1$), or not ($d = 0$); on idiosyncratic state variables S_{it} not affected by the policy, and on the sequence of exposures to the policy from its implementation, which in our model translate into a sequence of subsidy rates $\{\bar{\tau}_j\}_{j=1}^t$.

In this section, we are interested in computing the effects of LFL-implied transfers on a firm that is always eligible, i.e. we are interested in computing, for an alternative path of subsidies $\{\tilde{\tau}_j\}_{j=1}^t$, the expected effect:

$$\mathbb{E}[P_{it}(A_t^1, S_{it}, \{\tilde{\tau}_j\}_{j=1}^t) - P_{it}(A_t^1, S_{it}, \mathbf{0}_{t\times 1})|\text{score}_{i,1} = 0]\,. \tag{1.18}$$

In Appendix 1.10.2, we show that, under high-level assumptions on monotonicity and differentiability of the policy functions, we can bound, up to a first-order approximation,

$$\min_{j=1,\dots,t} \psi_t(t)\frac{\tilde{\tau}_j}{\psi_j^\tau(j)} \leqslant \mathbb{E}[P_{it}(A_t^1, S_{it}, \{\tilde{\tau}_j\}_{j=1}^t) - P_{it}(A_t^1, S_{it}, \mathbf{0}_{t\times 1})|\text{score}_{i,1} = 0] \leqslant \max_{j=1,\dots,t} \psi_t(t)\frac{\tilde{\tau}_j}{\psi_j^\tau(j)}\,. \tag{1.19}$$

With this representation, we *partially* identify the counterfactual outcomes from a given subsidy path. We quantify the policy's effects by assessing changes in wage bills, cash holdings, and supplier payments from a (permanent) subsidy that induces a 1% increase in debt issuance and determine the required subsidy amount.

Empirical Results: Figure 1.10 presents the results. A 1% induced increase in debt issuance is estimated to raise the wage bill by 0.4–0.75%, partially funded through internal funds, with cash holdings decreasing by 0.2–0.35%. The results also show a small increase in external debt and a temporary rise in net payment outflows to suppliers. Achieving this would require a (permanent) subsidy equivalent to 0.8% of total debt issuance. The implied effect on issuance spreads is small. These findings relate to recent estimates on the direct effects of quantitative easing on firms, primarily through indirect mechanisms such as the portfolio rebalancing channel. Evidence suggests that quantitative easing impacts firms' capital structure—specifically increasing cash holdings and investment in a lesser magnitude—while having a near-zero, albeit statistically significant, effect on corporate bond prices (see Selgrad 2023).[20]

Our results directly address the Central Bank's capacity to "target" the real economy by employing additional policy tools that enhance firms' borrowing opportunities, particularly when financial markets are illiquid. The inclusion of corporate bonds in the Collateral Framework under the LFL policy can be interpreted as a form of haircut policy, as advocated by Geanakoplos 2010. Our estimates indicate that the Central Bank can directly influence private agents' (net) leverage by affecting both the demand for their debt instruments and their liquidity management policies. This dual effect impacts both the supply and demand for funds, ultimately influencing employment and generating spillovers throughout the supply chain. Importantly, our estimates are estimated in

[20] More broadly to the empirical literature on the real effects of financial shocks such as Chodorow-Reich 2014 and Huber 2018, where the effects are expected to come from credit rationing in the crisis period. In the Brazilian context, similar large quantity and small price effects are observed following working capital supply shocks, which induce permanent increases in sales and employment (Orestes, Silva, and H. Zhang 2024).

recovery periods rather than crisis periods, indicating that the Central Bank's collateral policies can effectively stimulate economic activity outside financial crises.

1.6 Conclusion

We estimated that the LFL policy significantly influenced firms' financial and real outcomes. By allowing eligible corporate bonds to be used as collateral, the policy facilitated increased corporate bond issuance, reduced reliance on intermediate bank/fund debt, and lowered bond costs. This, in turn, led to a real impact on firms' operations, including a rise in the wage bill and net payment outflows, financed partially by a reduction in liquid cash holdings. The policy's influence extended beyond direct financial metrics, encouraging a "buy-and-hold" strategy among institutions, as evidenced by the decreased trading activity and increased use of bonds as collateral in repo operations.

These findings highlight the broader implications of expanding collateral eligibility to include corporate bonds within monetary policy frameworks. By affecting firms' and banks liquidity management strategies, the LFL policy contributed to easing borrowing constraints and promoting economic activity, even for relatively large firms that have access to capital markets.

References

Adelino, Manuel, Miguel A Ferreira, Mariassunta Giannetti, and Pedro Pires (2023). "Trade Credit And The Transmission of Unconventional Monetary Policy". In: *The Review of Financial Studies* 36.2, pp. 775–813.

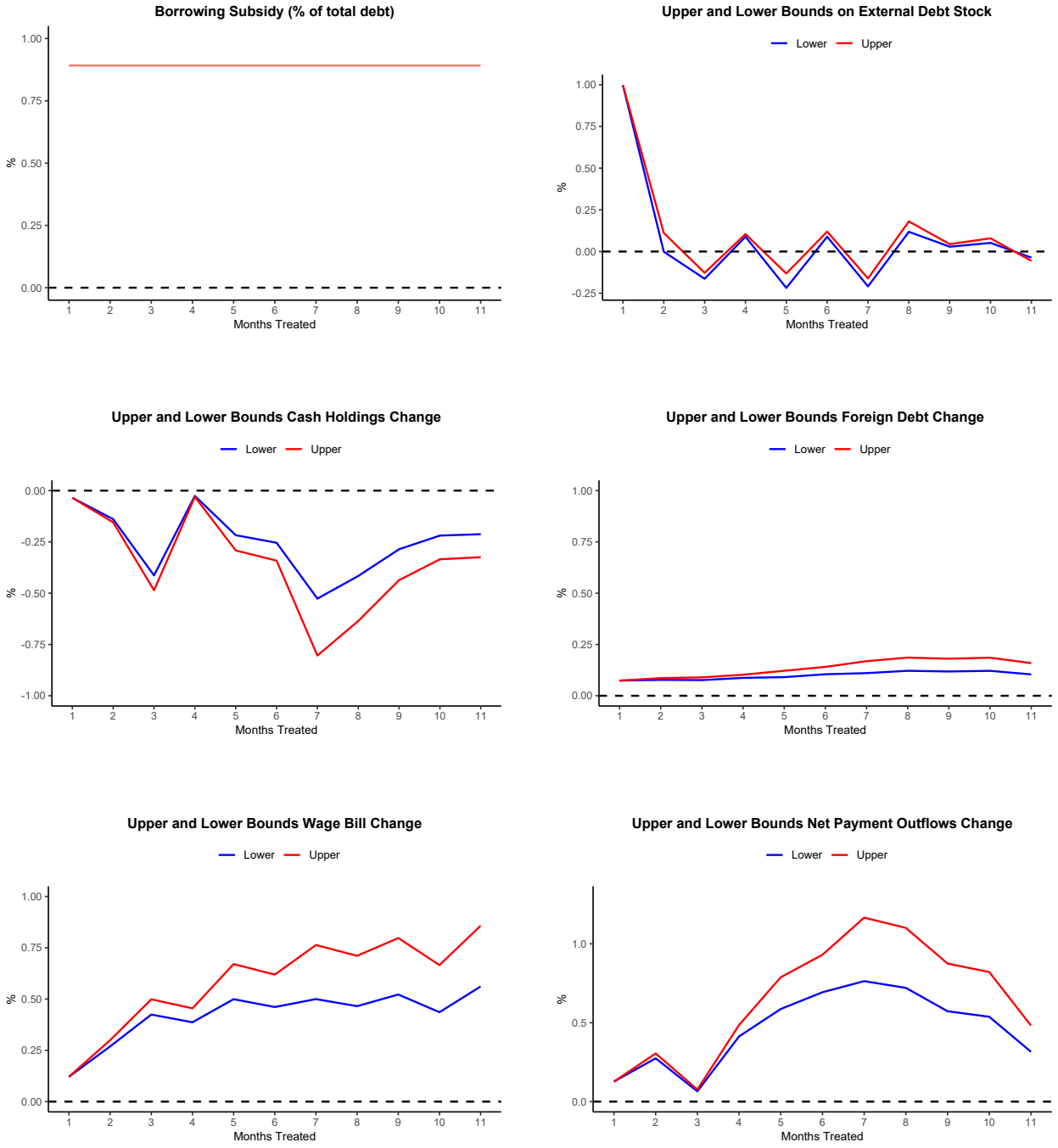
Andrews, Isaiah, James H Stock, and Liyang Sun (2019). "Weak instruments in instrumental variables regression: Theory and practice". In: *Annual Review of Economics* 11.1, pp. 727–753.

Arellano, Manuel and Stephen Bond (Apr. 1991). "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations". In: *The Review of Economic Studies* 58.2, pp. 277–297. ISSN: 0034-6527. DOI: 10.2307/2297968. eprint: https://academic.oup.com/restud/article-pdf/58/2/277/4454458/58-2-277.pdf. URL: https://doi.org/10.2307/2297968.

Barberis, Nicholas, Ming Huang, and Richard H Thaler (2006). "Individual preferences, monetary gambles, and stock market participation: A case for narrow framing". In: *American economic review* 96.4, pp. 1069–1090.

Blandhol, Christine, John Bonney, Magne Mogstad, and Alexander Torgovitsky (Jan. 2022). *When is TSLS Actually LATE?* Working Paper 29709. National Bureau of Economic Research. DOI: 10.3386/w29709. URL: http://www.nber.org/papers/w29709.

Figure 1.10: Responses to an 0.8% increase in borrowing subsidy



Boyarchenko, Nina, Anna Kovner, and Or Shachar (2022). "It's what you say and what you buy: A holistic evaluation of the corporate credit facilities". In: *Journal of Financial Economics* 144.3, pp. 695–731.

Bretscher, Lorenzo, Lukas Schmid, Ishita Sen, and Varun Sharma (2022). "Institutional corporate bond pricing". In: *Swiss Finance Institute Research Paper* 21-07.

Brunnermeier, Markus K, Sebastian A Merkel, and Yuliy Sannikov (2022). *Debt as safe asset*. Tech. rep. National Bureau of Economic Research.

Caetano, Carolina, Gregorio Caetano, and Juan Carlos Escanciano (2023). "Regression discontinuity design with multivalued treatments". In: *Journal of Applied Econometrics* 38.6, pp. 840–856. DOI: https://doi.org/10.1002/jae.2982. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jae.2982. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2982.

Calonico, Sebastian, Matias D Cattaneo, and Max H Farrell (2020). "Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs". In: *The Econometrics Journal* 23.2, pp. 192–210.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik (2014). "Robust nonparametric confidence intervals for regression-discontinuity designs". In: *Econometrica* 82.6, pp. 2295–2326.

Cattaneo, Matias D, Michael Jansson, and Xinwei Ma (2020). "Simple local polynomial density estimators". In: *Journal of the American Statistical Association* 115.531, pp. 1449–1455.

Cattaneo, Matias D. and Rocio Titiunik (2022). "Regression Discontinuity Designs". English (US). In: *Annual Review of Economics* 14. Publisher Copyright: © 2022 Annual Reviews Inc.. All rights reserved., pp. 821–851. ISSN: 1941-1383. DOI: 10.1146/annurev-economics-051520-021409.

Cellini, Stephanie Riegg, Fernando Ferreira, and Jesse Rothstein (2010). "The value of school facility investments: Evidence from a dynamic regression discontinuity design". In: *The Quarterly Journal of Economics* 125.1, pp. 215–261.

Chodorow-Reich, Gabriel (2014). "The employment effects of credit market disruptions: Firm-level evidence from the 2008–9 financial crisis". In: *The Quarterly Journal of Economics* 129.1, pp. 1–59.

Darmouni, Olivier and Kerry Siani (2022). "Bond market stimulus: Firm-level evidence". In: *Bond Market Stimulus: Firm-Level Evidence: Darmouni, Olivier| uSiani, Kerry*. [Sl]: SSRN.

Darmouni, Olivier, Kerry Siani, and Kairong Xiao (2022). "Nonbank fragility in credit markets: Evidence from a two-layer asset demand system". In: *Available at SSRN 4288695*.

Eisenschmidt, Jens, Yiming Ma, and Anthony Lee Zhang (2024). "Monetary policy transmission in segmented markets". In: *Journal of Financial Economics* 151, p. 103738.

Fang, Hanming, Yongqin Wang, and Xian Wu (2020). *The collateral channel of monetary policy: Evidence from China*. Tech. rep. National Bureau of Economic Research.

Gabaix, Xavier and Ralph SJ Koijen (2021). *In search of the origins of financial fluctuations: The inelastic markets hypothesis*. Tech. rep. National Bureau of Economic Research.

Geanakoplos, John (2010). "The leverage cycle". In: *NBER macroeconomics annual* 24.1, pp. 1–66.

Gilchrist, Simon, Bin Wei, Vivian Z Yue, and Egon Zakrajšek (2020). *The Fed takes on corporate credit risk: An analysis of the efficacy of the SMCCF*. Tech. rep. National Bureau of Economic Research.

He, Zhiguo and Arvind Krishnamurthy (2013). "Intermediary asset pricing". In: *American Economic Review* 103.2, pp. 732–770.

Hsu, Yu-Chin and Shu Shen (2021). "Dynamic regression discontinuity under treatment effect heterogeneity". In: *Work. Pap., Univ. Calif., Davis*.

Huber, Kilian (2018). "Disentangling the effects of a banking crisis: Evidence from German firms and counties". In: *American Economic Review* 108.3, pp. 868–898.

IMF (2012). "Brazil: Financial System Stability Assessment". In: *IMF*.

Koijen, Ralph SJ and Motohiro Yogo (2019). "A demand system approach to asset pricing". In: *Journal of Political Economy* 127.4, pp. 1475–1515.

Krishnamurthy, Arvind and Annette Vissing-Jorgensen (2012). "The aggregate demand for treasury debt". In: *Journal of Political Economy* 120.2, pp. 233–267.

Lagos, Ricardo and Shengxing Zhang (2020). "Turnover liquidity and the transmission of monetary policy". In: *American Economic Review* 110.6, pp. 1635–1672.

Lee, David S. and Thomas Lemieux (June 2010). "Regression Discontinuity Designs in Economics". In: *Journal of Economic Literature* 48.2, pp. 281–355. DOI: 10.1257/jel.48.2.281. URL: https://www.aeaweb.org/articles?id=10.1257/jel.48.2.281.

Morris, Stephen, Ilhyock Shim, and Hyun Song Shin (2017). "Redemption risk and cash hoarding by asset managers". In: *Journal of Monetary Economics* 89, pp. 71–87.

Morris, Stephen and Hyun Song Shin (2015). "Risk premium shifts and monetary policy: A coordination approach". In: *Princeton University William S. Dietrich II Economic Theory Center Research Paper* 075_2016.

Mota, Lira (2023). "The corporate supply of (quasi) safe assets". In: *Available at SSRN 3732444*.

Nyborg, Kjell G (2016). *Collateral frameworks: The open secret of central banks*. Cambridge University Press.

Orestes, Victor, Thiago Christiano Silva, and Henry Zhang (2024). *Firm-Level and Aggregate Effects of Cheaper Liquidity: Evidence from Factoring*. Tech. rep.

Pelizzon, Loriana, Max Riedel, Zorka Simon, and Marti G Subrahmanyam (2024). "Collateral eligibility of corporate debt in the Eurosystem". In: *Journal of financial economics* 153, p. 103777.

Roth, Jonathan, Pedro H.C. Sant'Anna, Alyssa Bilinski, and John Poe (2023). "What's trending in difference-in-differences? A synthesis of the recent econometrics literature".

In: *Journal of Econometrics* 235.2, pp. 2218–2244. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2023.03.008. URL: https://www.sciencedirect.com/science/article/pii/S0304407623001318.

Selgrad, Julia (2023). *Testing the portfolio rebalancing channel of quantitative easing*. Tech. rep. Working Paper.

Sojitra, Ravi B. and Vasilis Syrgkanis (2024). *Dynamic Local Average Treatment Effects*. arXiv: 2405.01463 [econ.EM]. URL: https://arxiv.org/abs/2405.01463.

Todorov, Karamfil (2020). "Quantify the quantitative easing: Impact on bonds and corporate debt issuance". In: *Journal of Financial Economics* 135.2, pp. 340–358.

Appendix A

1.7 Data Appendix

1.7.1 Sample Summary Statistics

Table 1.2: Summary statistics for the sample used in the analysis

	Around the Threshold			Eligible (All)
Firms (Outcomes 10/2021)	Non-eligible	Eligible	p-value (RDD)	
Total Debt (USD Mil.)	42.4	40.6	0.53	121
Corporate Bonds Share	34%	32%	0.86	45%
Payment Outflows (USD Mil.)	140	152	0.35	606
Wage Bill (USD Mil.)	17.8	16.5	0.38	56
Number of Workers	1619	1552	0.46	4464
Number of Firms	**115**	**112**	-	**680**
Assets (Outcomes 10/2021)	Non-eligible	Eligible	p-value (RDD)	
Issuance Spread (% over baseline)	4.78%	4.67%	0.72	3.80%
Outstanding Volume (BRL Mil.)	261	258	0.93	362
Maturity (Years)	7.6	6.9	**0.02**	7.1
Turnover Liquidity Trading (% monthly)	2.30%	2.10%	0.71	2.62%
Premium (% face value)	99.50%	100.50%	0.18	100.50%
Number of Assets	**164**	**294**	-	**1451**

1.8 Econometric Appendix

1.8.1 Proof of Proposition 1

We first begin noticing that, by Assumption 3, $\lim_{\epsilon\downarrow 0}\mathbb{E}[Y_t(0)|Z_l=\epsilon]-\lim_{\epsilon\uparrow 0}\mathbb{E}[Y_t(0)|Z_l=\epsilon]=0$ for any t or l. We then consider each of the three cases separately:

(a) *Case* $t<l$. Recall that:

$$Y_t = Y_t(0)+\sum_{j=1}^{t}\mathbf{1}\{D_t\geqslant j\}(Y_t(j)-Y_{t-1}(j)) \tag{1.20}$$

But, then, for any $|\epsilon|<\bar{\epsilon}$ and $j\in\{1,\ldots,t\}$:

$$\begin{aligned}
\mathbb{E}[\mathbf{1}\{D_t\geqslant j\}(Y_t(j)-Y_{t-1}(j))|Z_l=\epsilon] = \\
\mathbb{E}[\mathbf{1}\{D_t(Z_1,\ldots,Z_t)\geqslant j\}(Y_t(j)-Y_{t-1}(j))|Z_l=\epsilon] = \\
\mathbb{E}[\mathbf{1}\{D_t(Z_1,\ldots,Z_t)\geqslant j\}(Y_t(j)-Y_{t-1}(j))|Z_l=\epsilon,|Z_l|\leqslant\bar{\epsilon}] = \\
\mathbb{E}[\mathbf{1}\{D_t(Z_1,\ldots,Z_t)\geqslant j\}(Y_t(j)-Y_{t-1}(j))||Z_l|\leqslant\bar{\epsilon}],
\end{aligned}$$

where the last equality uses the fact that the random variable whose conditional expectation is being taken, $\psi_t(j)$, is a measurable function of $\boldsymbol{D}_t$, $Y_t(j)$, $Y_t(j-1)$ and $\{Z_s\}_{s=1}^{t}$, with $t<s$. Consequently, it follows from Assumption 3 that this random variable is independent from Z_l, conditional on $|Z_l|\leqslant\bar{\epsilon}$, and therefore $\mathbb{E}[\psi_t(j)|Z_l=\epsilon,|Z_l|\leqslant\bar{\epsilon}]=\mathbb{E}[\psi_t(j)||Z_l|\leqslant\bar{\epsilon}]$. The desired conclusion then follows from the additivity of conditional expectations, since then $\lim_{\epsilon\downarrow 0}\mathbb{E}[Y_t|Z_l=\epsilon]-\lim_{\epsilon\uparrow 0}\mathbb{E}[Y_t|Z_l=\epsilon]=\lim_{\epsilon\downarrow 0}\mathbb{E}[Y_t(0)|Z_l=\epsilon]-\lim_{\epsilon\uparrow 0}\mathbb{E}[Y_t(0)|Z_l=\epsilon]+\sum_{j=1}^{t}(\lim_{\epsilon\downarrow 0}\mathbb{E}[\psi_t(j)|Z_l=\epsilon]-\lim_{\epsilon\uparrow 0}\mathbb{E}[\psi_t(j)|Z_l=\epsilon])=0$.

(b) *Case* $t=l$. Take $|\epsilon|<\bar{\epsilon}$ and $j\in\{1,\ldots,t\}$ and observe that:

$$\begin{aligned}
\mathbb{E}[\mathbf{1}\{D_t\geqslant j\}(Y_t(j)-Y_{t-1}(j))|Z_t=\epsilon] = \\
\mathbb{E}[\mathbf{1}\{D_t(Z_1,\ldots,\epsilon)\geqslant j\}(Y_t(j)-Y_{t-1}(j))|Z_t=\epsilon] = \\
\mathbb{E}[\mathbf{1}\{D_t(Z_1,\ldots,\epsilon)\geqslant j\}(Y_t(j)-Y_{t-1}(j))|Z_r=\epsilon,|Z_t|\leqslant|\epsilon|] = \\
\mathbb{E}[\mathbf{1}\{D_t(Z_1,\ldots,\epsilon)\geqslant j\}(Y_t(j)-Y_{t-1}(j))||Z_t|\leqslant|\epsilon|],
\end{aligned}$$

where the last equality follows from Assumption 3. Consequently, taking $\epsilon\in(0,\bar{\epsilon}\wedge\check{\epsilon})$, we have:

$$\begin{aligned}
\mathbb{E}[\mathbf{1}\{D_t\geqslant j\}(Y_t(j)-Y_{t-1}(j))|Z_t=\epsilon]-\mathbb{E}[\mathbf{1}\{D_t\geqslant j\}(Y_t(j)-Y_{t-1}(j))|Z_t=-\epsilon] = \\
\mathbb{E}[(\mathbf{1}\{D_t(Z_1,\ldots,\epsilon)\geqslant j\}-\mathbf{1}\{D_t(Z_1,\ldots,-\epsilon)\geqslant j\})(Y_t(j)-Y_{t-1}(j))||Z_t|\leqslant\epsilon] = \\
\mathbb{E}[\mathbf{1}\{D_t(Z_1,\ldots,\epsilon)\geqslant j, D_t(Z_1,\ldots,-\epsilon)<j\}(Y_t(j)-Y_{t-1}(j))||Z_t|\leqslant\epsilon],
\end{aligned}$$

where the last equality follows from local monotonicity (Assumption 2). The desired conclusion then follows from $\mathbb{E}[\mathbf{1}\{D_t(Z_1,\ldots,\epsilon)\geqslant j, D_t(Z_1,\ldots,-\epsilon)<j\}(Y_t(j)-Y_{t-1}(j))||Z_t|\leqslant$

$\epsilon] = \mathbb{P}[D_t(Z_1,\ldots,\epsilon) \geqslant j, D_t(Z_1,\ldots,-\epsilon) < j | |Z_t| \leqslant \epsilon]\mathbb{E}[(Y_t(j) - Y_{t-1}(j))||Z_t| \leqslant \epsilon, D_t(Z_1,\ldots,\epsilon) \geqslant j, D_t(Z_1,\ldots,-\epsilon) < j]$ and the Assumption on the existence of limits.

(c) *Case* $t > l$*:* the representation in this case follows immediately from the definition of conditional expectations and the existence of limits. To see why the representation is not weakly causal, it suffices to provide an example of a setting where potential outcomes and selection into treatment satisfy Assumptions 1, 2, and 3; treatment effects have almost-surely the same sign, and the estimand has the opposite sign. We consider $3 = t > l = 1$, though similar arguments can be made for other choices of (t, l) with $t > l$. We assume that $Z_1 = \mathcal{U}[-1, 1]$ and that $Z_j = \mathbf{1}\{Z_1 > 0\}\mathcal{U}[-1, 0] + \mathbf{1}\{Z_{1<0}\}\mathcal{U}[-1, 0]$ for $j \in \{2, 3\}$. We consider $\boldsymbol{D}_3(\boldsymbol{z}) = \sum_{j=1}^{3} \mathbf{1}\{z_j \geqslant 0\}$, and assume that $Y_t(0) = 0$ a.s., and $Y_t(j) - Y_t(j-1) = \delta_j$ a.s. for $j \in \{1, 2, 3\}$. Observe that this is a setting that satisfies Assumptions 1, 2, and 3. Specifically, it exhibits an extreme version of the "compensation effect" discussed in the main text. Using the representation in the statement of the theorem, we obtain that:

$$\tau_{3|1} = -\delta_2\,,$$

thus showing that the estimand is not weakly causal.

1.8.2 Alternative identification strategies

In this Appendix, we discuss alternative identification strategies to the one adopted in the main text.

Identification based on time-series stationarity restrictions

Suppose that, for each t, average marginal effects below and above the threshold at t are homogeneous and that these marginal effects are stationary, i.e. $\phi^+_{t|t}(j) = \phi^-_{t|t}(j) = \phi(j)$ for every $t = 1,\ldots,T$ and $j \in \{1,\ldots,t\}$. In this case, we are able to write the following system:

$$\begin{bmatrix} \omega^+_{1|1}(1) - \omega^-_{1|1}(1) & 0 & 0 & \ldots & 0 \\ \omega^+_{2|2}(1) - \omega^-_{2|2}(1) & \omega^+_{2|2}(2) - \omega^-_{2|2}(2) & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ \omega^+_{T|T}(1) - \omega^-_{T|T}(1) & \omega^+_{T|T}(2) - \omega^-_{T|T}(2) & \omega^+_{T|T}(3) - \omega^-_{T|T}(3) & \ldots & \omega^+_{T|T}(T) - \omega^-_{T|T}(t) \end{bmatrix} \begin{bmatrix} \phi(1) \\ \phi(2) \\ \vdots \\ \phi(T) \end{bmatrix} = \begin{bmatrix} \tau_{1|1} \\ \tau_{2|2} \\ \vdots \\ \tau_{T|T} \end{bmatrix}.$$

If the identified lower-triangular matrix of first-stage contrasts is invertible, which will be the case under the relevance condition $\min_{t=1,\ldots,T} |\omega^+_{t|t}(t) - \omega^-_{t|t}(t)| > 0$, then it is possible to identify the vector $(\phi(1),\ldots,\phi(T))$. Indeed, observe that, under this Assumption, marginal effects may be identified recursively. First, one identifies $\phi(1)$ by:

$$\phi(1) = \frac{\tau_{1|1}}{\omega^+_{1|1}(1) - \omega^-_{1|1}(1)}\,,$$

then, one may identify $\phi(2)$ by:

$$\phi(2) = \frac{\tau_{2|2} - (\omega^+_{2|2}(1) - \omega^-_{2|2}(1))\phi(1)}{\omega^+_{2|2}(2) - \omega^-_{2|2}(2)},$$

and so on.

The recursive approach discussed in this section is especially suited to settings where the marginal effects of a given exposure remain stable over time. This does not seem the case in our setting, which encompasses a recovery from a sanitary emergency that rattled financial and real markets. In this case, it does not seem plausible to assume that the effect of being exposed to the LFL for one period to be the same at the beginning of our time window as well as in the end of it.

Remark 1. *An alternative stationarity condition that can be used for identification is to assume that, for a fixed threshold* l, $\phi^+_{t|l}(j) = \phi^-_{t|l}(j) = \phi(j)$ *for every* $t = 1, \dots, T$. *Under this assumption, below-and-above marginal effects around a fixed threshold are homogeneous and stationary across-time. Using this assumption, one obtains, the following triangular system:*

$$\begin{bmatrix} \omega^+_{1|l}(1) - \omega^-_{1|l}(1) & 0 & 0 & \dots & 0 \\ \omega^+_{2|l}(1) - \omega^-_{2|l}(1) & \omega^+_{2|l}(2) - \omega^-_{2|l}(2) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \omega^+_{T|l}(1) - \omega^-_{T|l}(1) & \omega^+_{T|l}(2) - \omega^-_{T|l}(2) & \omega^+_{T|l}(3) - \omega^-_{T|l}(3) & \dots & \omega^+_{T|l}(t) - \omega^-_{T|l}(t) \end{bmatrix} \begin{bmatrix} \phi(1) \\ \phi(2) \\ \vdots \\ \phi(T) \end{bmatrix} = \begin{bmatrix} \tau_{1|l} \\ \tau_{2|l} \\ \vdots \\ \tau_{T|l} \end{bmatrix},$$

which is identified under a relevance condition on the main diagonal of the first-stage matrix. One advantage of this assumption is that it only requires knowledge of the score at a single period (for example, the first period score if $l = 1$).

Identification by conditioning on past exposures

Yet another alternative identification strategy consists in conditioning on the realisation of prior exposures, relying on the fact that $x \mapsto D_x$ is nondecreasing. Specifically, consider the following RDD estimand:

$$\tilde{\phi}_{t|t} = \frac{\lim_{\epsilon \downarrow 0} \mathbb{E}[Y_t | D_{t-1} = t-1, Z_t = \epsilon] - \lim_{\epsilon \downarrow 0} \mathbb{E}[Y_t | D_{t-1} = t-1, Z_t = \epsilon]}{\lim_{\epsilon \downarrow 0} \mathbb{E}[D_t | D_{t-1} = t-1, Z_t = \epsilon] - \lim_{\epsilon \downarrow 0} \mathbb{E}[D_t | D_{t-1} = t-1, Z_t = \epsilon]}$$

Under Assumptions 1, 2, and 3, this estimand identifies the LATE:

$$\tilde{\phi}_{t|t} = \mathbb{E}[Y_t(t) - Y_t(t-1) | \boldsymbol{D}_t(\boldsymbol{Z}_{t|t}(\epsilon)) = t, \boldsymbol{D}_t(\boldsymbol{Z}_{t|t}(-\epsilon)) = t-1, D_{t-1} = t-1],$$

i.e. we identify the average marginal effect of changing the exposure from $t-1$ to t, among the individuals around the threshold at t who had exposure prior exposure of $t-1$, and whose exposure is changed if their scores were perturbed to be marginally above the threshold (scores).

While this approach does not require homogeneity assumptions to identify marginal treatment effects, we note that it may be extremely demanding of the data for larger values of t, as we

effectively restrict ourselves to the subpopulation that is at the threshold around t, but that has been always treated prior to that. If treatment adoption substantially affects future scores, moving them away from a vicinity of the threshold, we expect this subpopulation to be rather small, for a given bandwidth around the score, which would severely hinder estimation. In addition, the proposed approach only identifies marginal effects without homogeneity assumptions. If one were to identify cummulative effects, then homogeneity assumptions that allowed comparing different $\tilde{\phi}_{t|t}$ would be required. In this case, and if there are overidentifying restrictions, we believe that estimation of cummulative effects that *impose* the homogeneity assumptions – such as the ones in the main text – would be preferable on efficiency grounds.

Remark 2 (On the compliers captured by $\tilde{\phi}_{t|t}$)**.** *In our empirical application, $\tilde{\phi}_{t|t}$ captures the average marginal effect among those firms that were exposed to the policy for $t-1$ periods, but remain around the threshold at t. Given that we expect the policy to operate through the relaxation of borrowing constraints, and that our running variable measures default risk, one expects this subpopulation to be precisely the one that is least affected by the policy. Therefore, from a policy perspective, this does not appear to be an interesting parameter, even if estimation risk was not a concern.*

Comparison with other identification strategies in the literature

Cellini, Ferreira, and Rothstein 2010 propose a recursive identification strategy of average marginal effects in a dynamic RD design. When translated to our setting, their key identifying condition requires constant marginal effects of increasing exposure, i.e. that $Y_t(j)-Y_t(j-1)$ does not depend on j.[21] This assumption seems especially restrictive for economic outcomes over a longer time window, where one would expect "decreasing returns" to policy exposure as agents actions converge to their "steady-state" policy choice. As an alternative to Cellini, Ferreira, and Rothstein, Hsu and Shen 2021 propose identifying marginal effects in dynamic RDDs under an assumption that, when translated to our setting with potential treatment function $\boldsymbol{D}_t(\boldsymbol{z}) = \sum_{j=1}^{t} \mathbf{1}\{z_j \geqslant 0\}$, would require that, for $z \in (-\epsilon, 0)$:

$$Y_2(0) \perp Z_2 | Z_1 = z\,.$$

Notice that this Assumption limits association between the second-period score and the unexposed second-period potential outcome, for units just below the first-period threshold. It is unlikely to be satisfied in our empirical setting, where the current period credit score is expected to correlate with the unexposed potential outcome, even after controlling for the first-period score, insofar as scores exhibit a contemporaneous association with outcomes beyond those mediate by the treatment.

[21] Their approach also requires an homogeneity around the threshold. See Lemma 2.1 of Hsu and Shen 2021 foir details.

1.8.3 Identification and optimal-estimation under (1.5)

Point Estimates and Var/Cov Matrix:

Consider the stacked model:

$$I_{o,j,t,s} = \alpha_{o,t,s} + \beta^{+}_{1,o,t,s}\mathbf{1}\{Z_{j,s} \geqslant c\}(Z_{j,s} - c) + \beta^{+}_{1,o,t,s}\mathbf{1}\{Z_{j,s} < c\}(Z_{j,s} - c) + \gamma_{o,t,s}\mathbf{1}\{Z_{j,s} \geqslant c\} + \epsilon_{j,s}\,,$$

where $I_{o,j,t,s}$ is an input $o \in \{D, Y\}$ at time t for unit j, using the jump at s. We estimate the regression coefficients by constructing a dataset stacked over o, j, t, and s, and then applying OLS with weights $K\left(\frac{Z_{j,s}-c}{h_s}\right)$. By clustering standard errors at the unit level, we account for serial correlation and obtain robust variance estimates. The estimated $\gamma_{o,t,s}$ corresponds to the jumps in both the reduced form and the first stage. The resulting joint covariance matrix is appropriate for the Delta Method. Here, h_s is a cohort-specific bandwidth, and we adjust the standard errors following Calonico, Cattaneo, and Titiunik 2014.

In our empirical strategy, we set a fixed bandwidth, h_s, for each cohort across all time periods and outcomes. This approach ensures comparability of the same group of units across all outcomes (and time). However, the methodology can be easily extended to a more general case, where the bandwidth is allowed to vary with j, t, and s.

Optimal weighting scheme and variance estimator

Recall that from our identification results, the system in (1.4) given by:

$$\Omega_t \phi_t = \boldsymbol{\tau_t}\,,$$

Note that there exists a matrix R such that the restrictions implied by (1.5) can be recast as:

$$\boldsymbol{a}^* - \boldsymbol{V}^* R\beta = \mathbf{0}_{T(T+1)/2}$$

where $\boldsymbol{a}^{*'} = (\boldsymbol{\tau}_1^{'}, \boldsymbol{\tau}_2^{'}, \ldots, \boldsymbol{\tau}_T^{'})'$, and:

$$\boldsymbol{V}^* = \begin{bmatrix} \Omega_1 & \mathbf{0}_{1\times 2} & \ldots & \mathbf{0}_{1\times T} \\ \mathbf{0}_{2\times 1} & \Omega_2 & \ldots & \mathbf{0}_{2\times T} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{T\times 1} & \mathbf{0}_{T\times 2} & \ldots & \Omega_T \end{bmatrix}$$

and β is the matrix with D parameters encoding the cumulative exposure effects, calendar time effects and cohort effects, with R being a $T(T+1)/2 \times D$ matrix appropriately piecing these cumulative effects in terms of average marginal effects.

The weighted minimum distance estimator can now be written as:

$$\min_b (\widehat{\boldsymbol{a}^*} - \widehat{\boldsymbol{V}^*} Rb)' W (\widehat{\boldsymbol{a}^*} - \widehat{\boldsymbol{V}^*} Rb)\,,$$

with W a $T(T+1)/2 \times T(T+1)/2$ weighting matrix. This implies that:

$$\hat{b} = (R'\widehat{\boldsymbol{V}^*}'W\widehat{\boldsymbol{V}^*}R)^{-1}R'\widehat{\boldsymbol{V}^*}'W\widehat{\boldsymbol{a}^*}$$

It follows that the optimal choice of weights is given by:

$$W^* = \mathbb{V}[\widehat{\boldsymbol{a}^*} - \widehat{\boldsymbol{V}^*}R\beta]^{-1}$$

But, by the properties of row-wise vectorization:

$$\widehat{\boldsymbol{V}^*}R\beta = (\beta'R' \otimes \mathbb{I}_{T(T+1)/2})\,\mathrm{vec}(\widehat{\boldsymbol{V}^*})$$

Implying that

$$(W^*)^{-1} = \mathbb{V}[\widehat{\boldsymbol{a}^*}] + (\beta'R' \otimes \mathbb{I}_{T(T+1)/2})\mathbb{V}\left[\mathrm{vec}(\widehat{\boldsymbol{V}^*})\right](\beta'R' \otimes \mathbb{I}_{T(T+1)/2})' - \mathrm{cov}(\widehat{\boldsymbol{a}^*}, \widehat{\boldsymbol{V}^*}R\beta) - \mathrm{cov}(\widehat{\boldsymbol{V}^*}R\beta, \widehat{\boldsymbol{a}^*})$$

But

$$\mathrm{cov}(\widehat{\boldsymbol{a}^*}, \widehat{\boldsymbol{V}^*}R\beta) = \mathrm{cov}(\widehat{\boldsymbol{V}^*}R\beta, \widehat{\boldsymbol{a}^*})' = \mathrm{cov}(\widehat{\boldsymbol{a}^*}, \mathrm{vec}(\widehat{\boldsymbol{V}^*}))\left((R\beta)' \otimes \mathbb{I}_{T(T+1)/2}\right)',$$

implying that we may estimate W^* using a preliminary estimator for β (e.g. by implementing the minimum distance estimator using identity weights). In the numerical implementation, we consider a multi-step procedure to estimate W^*. Specifically, we start by solving the minimum distance problem using the identity weights and compute a first-step estimate $\hat{W}_1$ of W^*. In the following steps, we sequentially re-estimate W^* by re-estimating β using the minimum-distance estimator that uses the previous-step estimate of the weighting matrix. We stop this procedure after 10 iterations.

Asymptotic variance formula Suppose that, for a sequence $c_n \to \infty$:

$$c_n \begin{bmatrix} (\mathrm{vec}(\widehat{\boldsymbol{V}^*}) - \mathrm{vec}(\boldsymbol{V}^*)) \\ (\widehat{a^*} - a^*) \end{bmatrix} \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \mathbb{V})$$

It is then possible to see that:

$$c_n(\hat{b} - \beta) = (R'\boldsymbol{V}^{*\prime}W\boldsymbol{V}^*R)^{-1}[R'\boldsymbol{V}^{*\prime}Wc_n(\widehat{\boldsymbol{a}^*} - \boldsymbol{a}^*) + R'c_n(\widehat{\boldsymbol{V}^*} - \boldsymbol{V}^*)'W\boldsymbol{a}^*] + o_p(1)\,.$$

i.e. we can discard the uncertainty in the inverse. But then, by noting that:

$$R'c_n(\widehat{\boldsymbol{V}^*} - \boldsymbol{V}^*)'W\boldsymbol{a}^* = ((W\boldsymbol{a}^*)' \otimes R')c_n(\mathrm{vec}(\widehat{\boldsymbol{V}^*}) - \mathrm{vec}(\boldsymbol{V}^*))$$

we can calculate the variance formula as

$$\mathbb{V}\left[c_n(\hat{b}-\beta)\right] = (R'\boldsymbol{V}^{*\prime}W\boldsymbol{V}^{*}R)^{-1}\Big[R'\boldsymbol{V}^{*\prime}W\mathbb{V}\left[c_n(\widehat{\boldsymbol{a}^{*}}-\boldsymbol{a}^{*})\right](R'\boldsymbol{V}^{*\prime}W)'+ ((W\boldsymbol{a}^{*})'\otimes R')\mathbb{V}\left[c_n(\mathrm{vec}(\widehat{\boldsymbol{V}^{*}})-\mathrm{vec}(\boldsymbol{V}^{*}))\right]((W\boldsymbol{a}^{*})'\otimes R')'+ S+S'\Big](R'\boldsymbol{V}^{*\prime}W\boldsymbol{V}^{*}R)^{-1'} \tag{1.21}$$

where

$$S = R'\boldsymbol{V}^{*\prime}W\,\mathrm{cov}(c_n(\widehat{\boldsymbol{a}^{*}}-\boldsymbol{a}^{*}), c_n(\mathrm{vec}(\widehat{\boldsymbol{V}^{*}})-\mathrm{vec}(\boldsymbol{V}^{*})))((W\boldsymbol{a}^{*})'\otimes R')'$$

J-test

Under the null of correct-model specification, and, when using the plug-in estimator of the optimal weights W^{*}, we have:

$$(\widehat{\boldsymbol{a}^{*}}-\widehat{\boldsymbol{V}^{*}}R\hat{b})'\widehat{W^{*}}(\widehat{\boldsymbol{a}^{*}}-\widehat{\boldsymbol{V}^{*}}R\hat{b}) \sim \chi^{2}_{T(T+1)/2-D},$$

which may be used in specification testing.

Placebo Test

To assess the robustness of our results, we conduct a placebo test by applying the same dynamic regression discontinuity design (RDD) to a pre-policy period, during which the Liquidity Facility Lines (LFL) policy had not yet been implemented. Specifically, we shift the timeline backward and consider a placebo treatment assignment based on the same running variable but prior to the actual policy rollout.

Let $\{Y_1, Y_2, \ldots, Y_T\}$ denote the observed outcomes in the post-policy period. We construct a placebo dataset by using pre-policy outcomes $\{Y_0, Y_{-1}, \ldots, Y_{-T+1}\}$ and the same running variable Z_t, while maintaining the original threshold for treatment eligibility. The placebo regression discontinuity (RD) estimand is defined as:

$$\tau^{\text{placebo}}_{t|l} = \lim_{s\downarrow 0}\mathbb{E}[Y^{\text{pre}}_{t}|Z_l = s] - \lim_{s\uparrow 0}\mathbb{E}[Y^{\text{pre}}_{t}|Z_l = s] \tag{1.22}$$

Under the null hypothesis of no treatment effect before the policy implementation, we expect that $\tau^{\text{placebo}}_{t|l} = 0$ for all t. If the estimated placebo effects are statistically indistinguishable from zero, this would support the validity of our identifying assumptions and confirm that the observed effects in the main analysis are not driven by pre-existing trends or spurious correlations.

For each pre-policy period t, we estimate the placebo treatment effect using the same specification as in the main analysis, controlling for the same covariates. We then test whether the placebo effects differ from zero by constructing confidence intervals around the estimates and checking whether zero lies within these intervals. In particular that implies that all cumulative treatment effects should be 0.

Formally, we test the hypothesis:

$$H_0 : \psi_t^{\text{placebo}} = 0 \quad \text{for all } t \in \{1, 2, \ldots, T\} \tag{1.23}$$

Failure to reject the null hypothesis is consistent with policy effects observed in the post-policy period not being confounded with pre-existing dynamics, thus making it more credible to attribute estimated effects to the actual policy intervention.

1.9 Additional Empirical Results

1.9.1 Density Manipulation Tests

Table

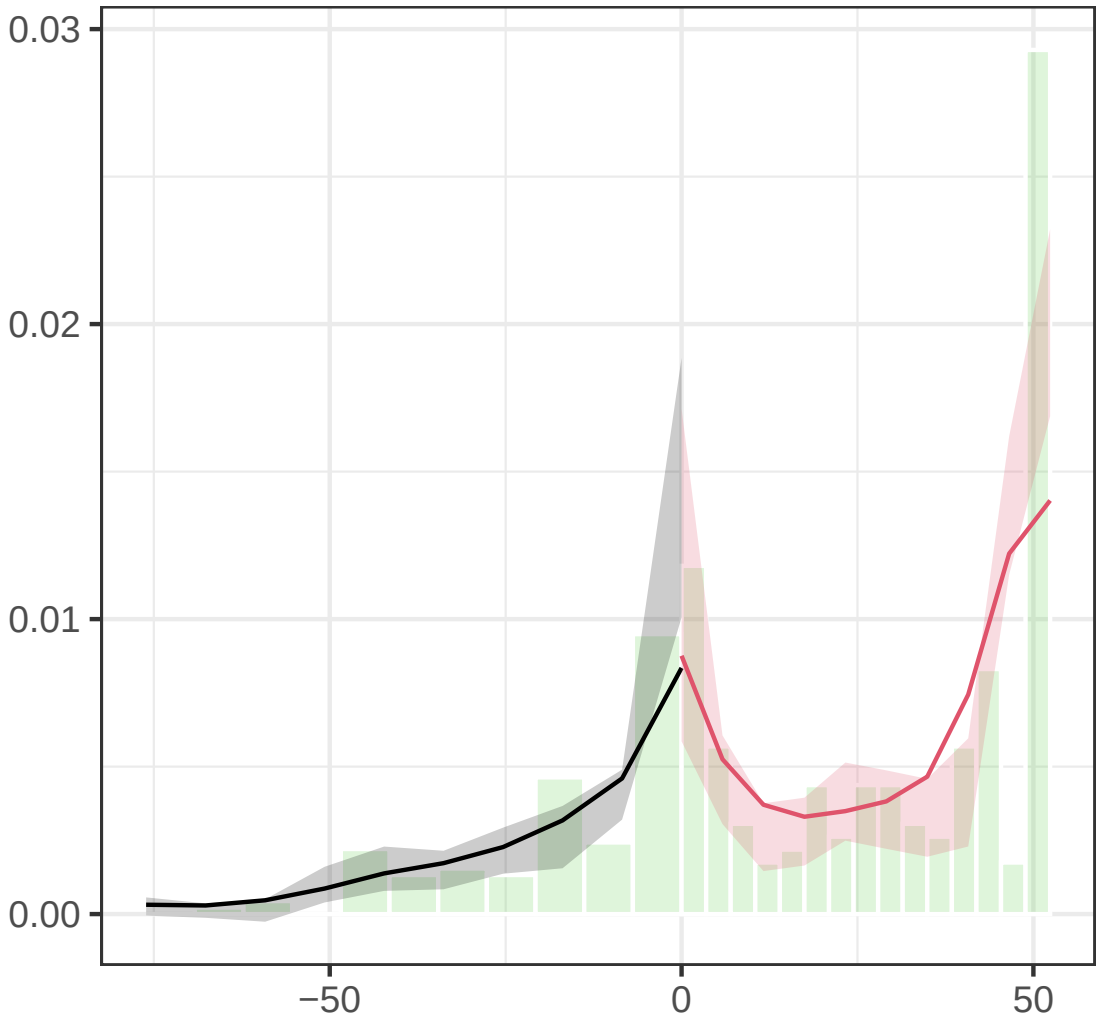
Table 1.3: Manipulation Test for Z_{it}

Time	Test Statistic	p-value	Eff. N Left	Eff. N Right	h_L	h_R
Nov 2021	-0.5825	0.56	115	112	25.3640	17.2677
Dec 2021	0.0910	0.93	108	126	20.8068	17.8393
Jan 2022	-0.0023	0.99	102	134	21.8986	20.4582
Feb 2022	-0.1634	0.87	106	138	22.2285	21.4362
Mar 2022	-0.4009	0.69	108	128	22.9810	19.8630
Apr 2022	-0.2123	0.83	103	130	22.3442	19.7005
May 2022	-0.4902	0.62	107	130	20.9249	20.2777
Jun 2022	-0.6521	0.51	107	128	22.5062	20.9442
Jul 2022	-0.5157	0.61	107	122	19.2556	18.2163
Aug 2022	-1.6286	0.10	120	122	20.7579	19.4576
Sep 2022	-0.3994	0.69	115	116	20.2911	17.7385
Oct 2022	-2.1302	**0.03**	115	126	18.9888	17.8679
Nov 2022	-1.6646	0.10	112	130	21.4372	20.2427
Dec 2022	-0.6888	0.49	103	126	21.4340	19.5926

Note: This table presents the manipulation test results following Cattaneo, Jansson, and Ma 2020. Eff. N denotes the effective number of observations (those with positive weights) on each side of the threshold. Bandwidth (h) refers to the optimally chosen bandwidth for each side of the threshold. h_L and h_R are, respectively, the left and right bandwidths.

Manipulation Test Plot

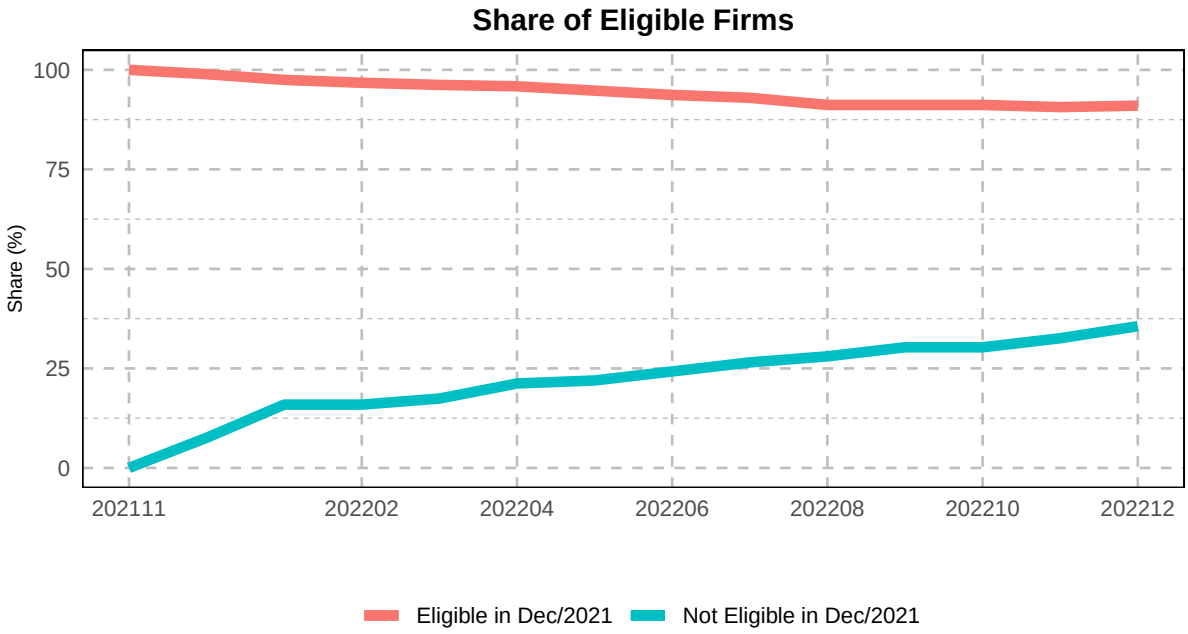
Figure A1: Time Series of Safe Asset Holdings



Note: The figure presents presents the manipulation test results following Cattaneo, Jansson, and Ma 2020 for November 2021, when the policy started.

1.9.2 Evolution of eligibility over time

Figure A2: Time Series of Safe Asset Holdings



Note: The figure plots the evolution of eligibility status over time given the treatment status in November 2021, when the policy started.

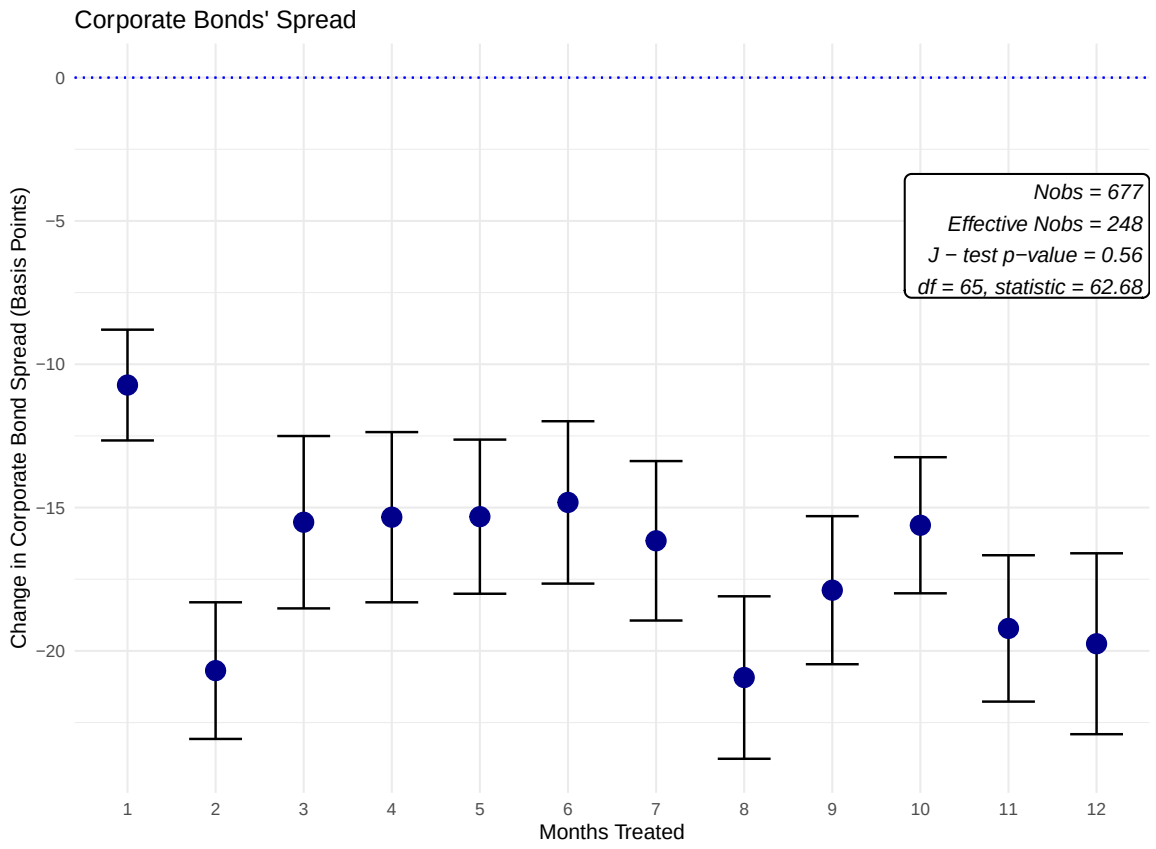
1.9.3 Time Series of Safe Asset Holdings

Figure A3: Time Series of Safe Asset Holdings

Note: The figure presents the total cash holdings of firms in our sample, with values normalized to 100 as of January 2020. The solid vertical line marks the month preceding the onset of Covid-19.

1.9.4 Effects on Cost of Capital

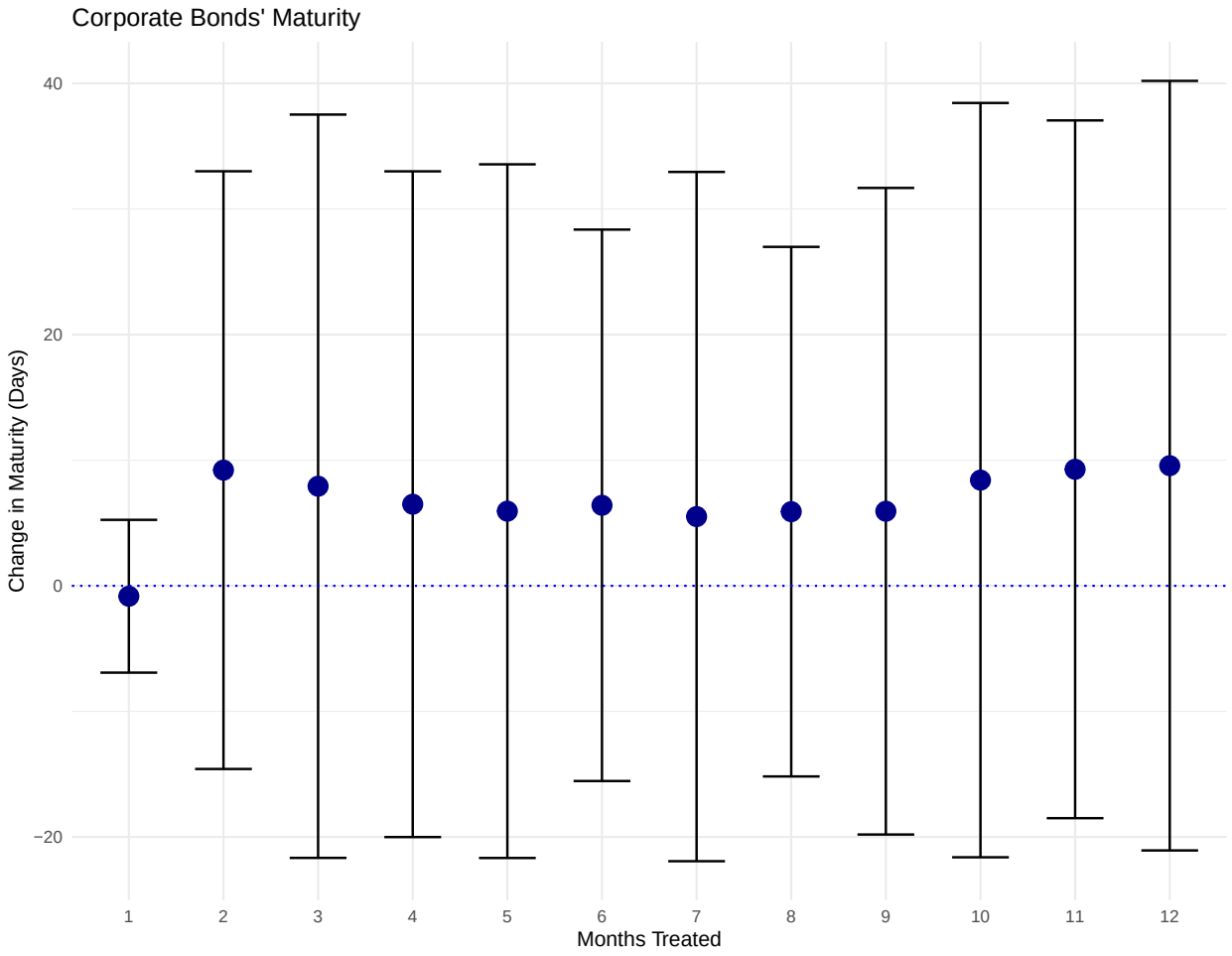
Figure A4: Dynamic Treatment Effects on Corporate Bonds Spread.



Note: The figure shows estimates of cumulative treatment effects on outstanding in the bond spread over the risk-free rate, with error bars representing 95% confidence intervals. The spread is the outstanding-volume-weighted average of credit spreads for firms' active corporate bonds.

1.9.5 Effects on Debt Maturity

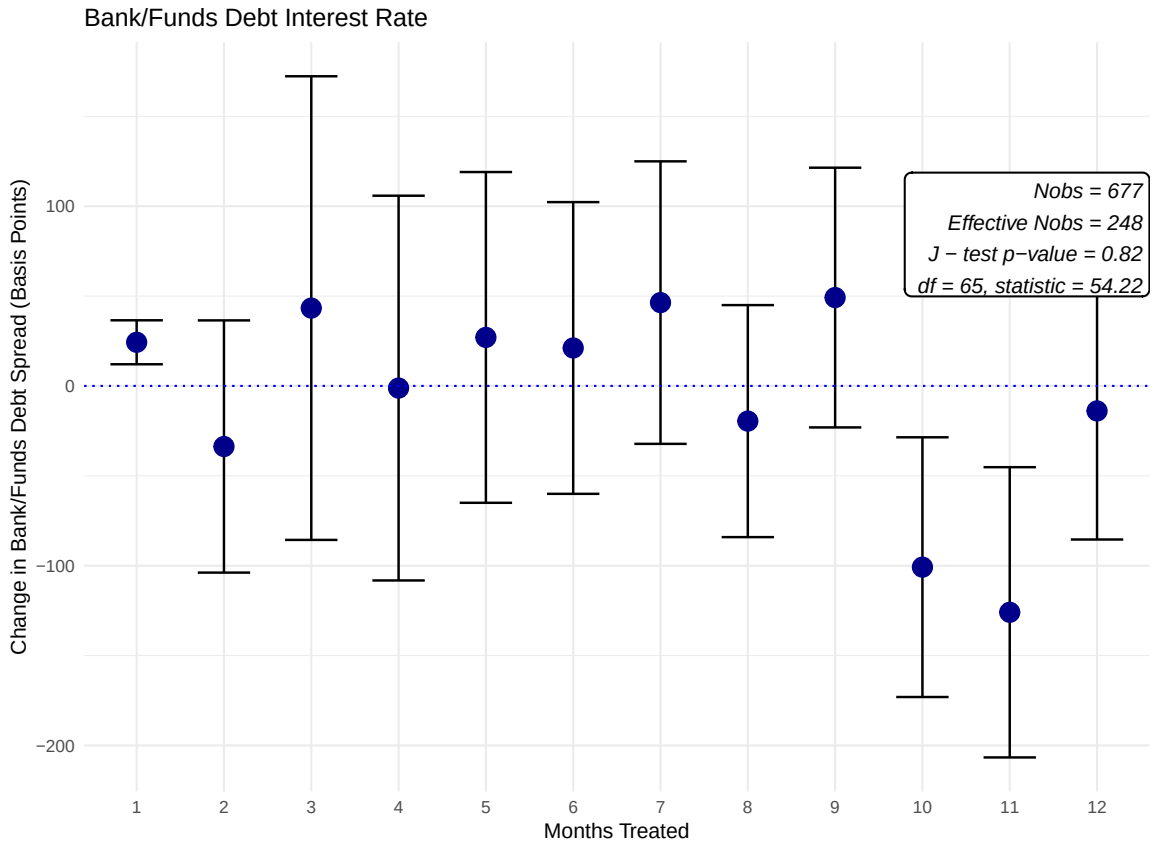
Figure A5: Dynamic Treatment Effects on Maturity



Note: The figure shows estimates of cumulative treatment effects on outstanding corporate bond debt maturity.

1.9.6 Effect on Interest Rates

Figure A6: Dynamic Treatment Effects on Maturity



Note: The figure shows estimates of cumulative treatment effects on outstanding bank debt (excluding corporate bonds) cost of capital.

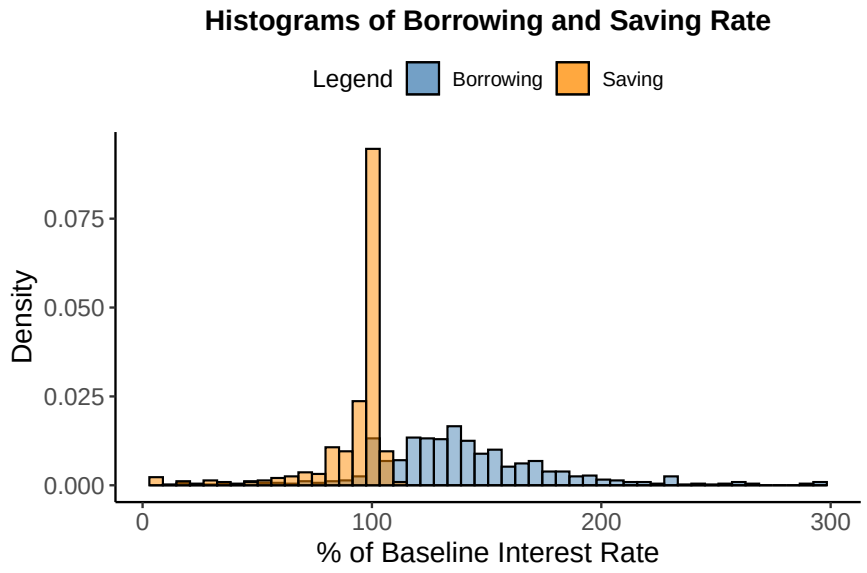


Figure A7: Dispersion of cost of capital and savings rate.

Note: The figure shows the dispersion of borrowing and savings interest rates for firms in our sample in October 2021. Borrowing rate is define as the volume-weighted average cost of capital on all intermediate debt. Savings rate is define as the total return on cash holdings. Returns are expressed in % of the (annualized) baseline interest rate.

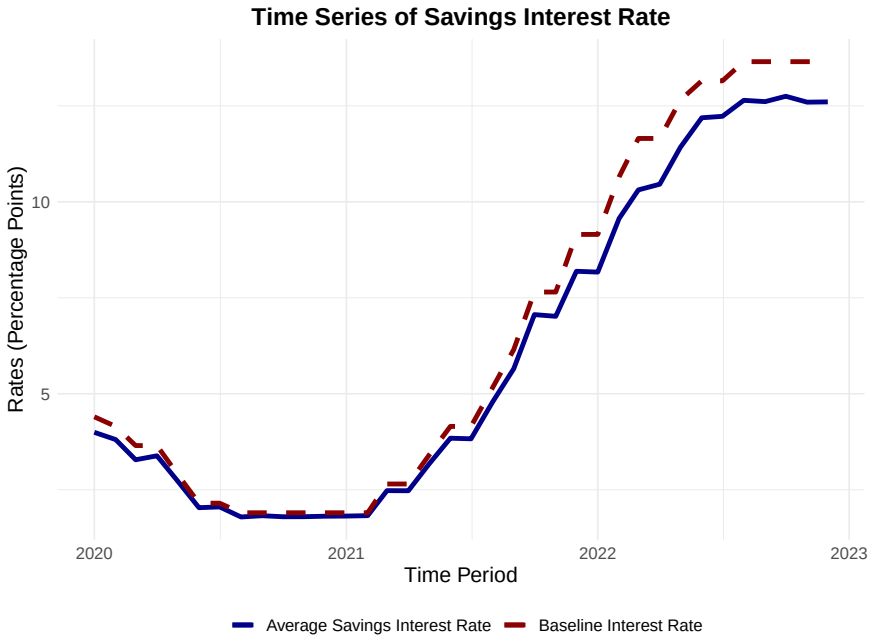


Figure A8: Time series of savings rate.

Note: The figure shows the the time series of the (volume weighted) average savings interest rates for firms in our sample. Savings rate is define as the total return on cash holdings. Returns are annualized and expressed in percentage points.

1.9.7 Elasticities Estimates

Table 1.4: Elasticity and Coefficient Estimates (Monthly Difference)

	GMM		FE-IV		FE		OLS	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β	**0.24****	0.07	0.31***	0.24*	−0.40***	−0.51***	0.21***	0.42***
	(0.061)	(0.13)	(0.050)	(0.118)	(0.025)	(0.084)	(0.007)	(0.033)
ε	**2.31*** **	2.07***	2.46***	2.32***	1.71***	1.66***	2.27***	2.74***
	(0.117)	(0.208)	(0.121)	(0.175)	(0.012)	(0.036)	(0.011)	(0.101)
FE	YES	YES	YES	YES	YES	YES	NO	NO
N	24,204	9,405	28,201	16,872	28,201	12,875	29,083	13,626
T	ALL	PRE	ALL	PRE	ALL	PRE	ALL	PRE
J-test (p-value)	0.39	0.72	-	-	-	-	-	-

Notes: Robust standard errors in parentheses. ***p < 0.001; **p < 0.01; *p < 0.05.

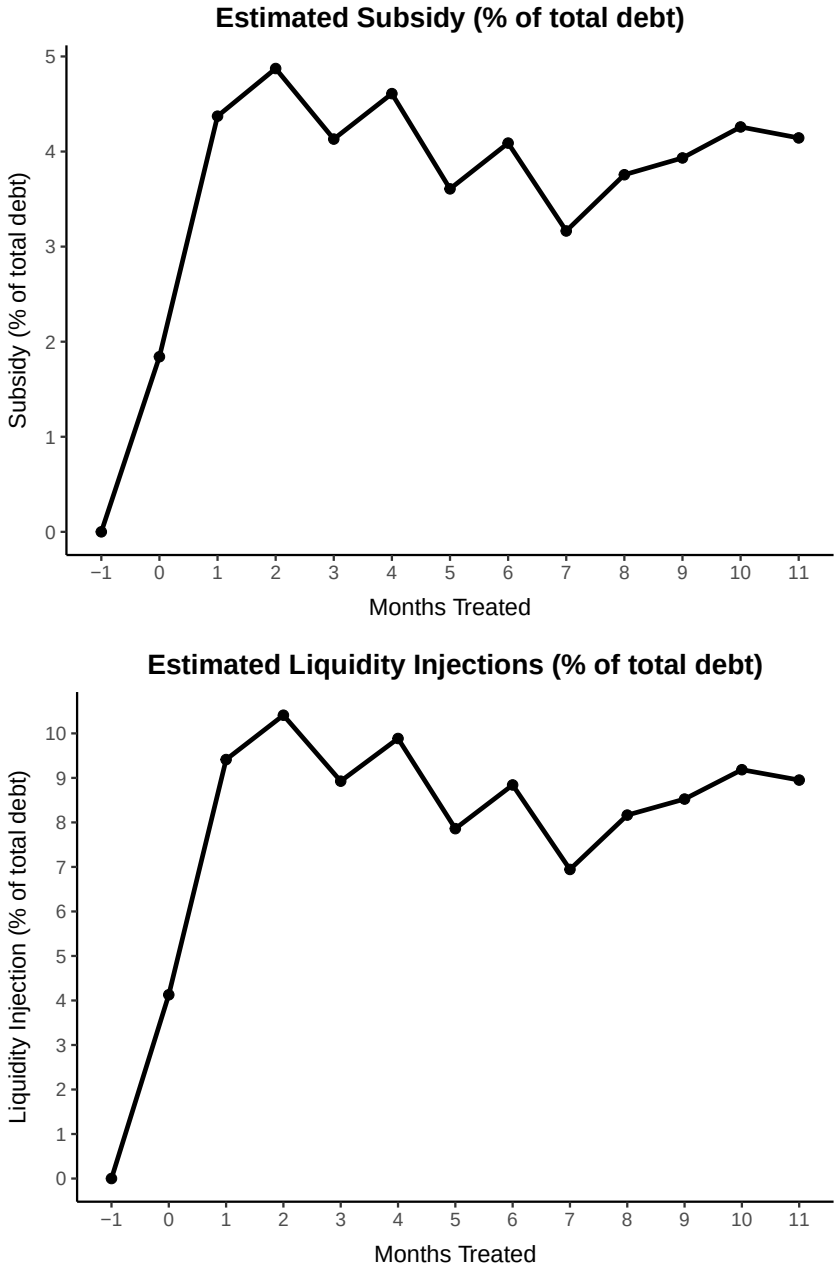
Table 1.5: Elasticity and Coefficient Estimates (Quarterly Difference)

	GMM	IV-FE	FE	OLS
	(1)	(2)	(3)	(4)
β	**0.23*****	0.29***	−0.39***	0.18***
	(0.061)	(0.06)	(0.02)	(0.007)
ε	**2.30*****	2.41***	1.71***	2.22***
	(0.105)	(0.124)	(0.010)	(0.010)
FE	YES	YES	YES	NO
N	21,469	25,289	25,289	27,588
T	ALL	ALL	ALL	ALL
J-test (p-value)	0.33	-	-	-

Notes: Robust standard errors in parentheses. ***p < 0.001; **p < 0.01; *p < 0.05.

1.9.8 Implicit Transfers Estimates

Figure A9



1.10 Proofs and derivation - Model Section

1.10.1 Derivation of equation (1.11)

We can write the AGU problem as:

$$\max_{B_t,\{\omega_{it}\}} \mathbb{E}_t \left[\eta_b^{\frac{1}{\varepsilon}} \frac{\tilde{B}_{t+1}^{1-\gamma}}{1-\gamma} + \tau_{c,t}^{\sim} B_{t+1}^{LFL} - \tau_{l,t}^{\sim} B_{t+1}^{NLFL} - \mu(B_t) B_t \right].$$

subject to

$$\int \omega_{it}\, di = 1, \tag{1.24}$$

$$\tilde{B}_{t+1} = \left(\int \eta_i^{\frac{1}{\varepsilon}} \left(R_{i,t} \omega_{it} B_t \right)^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{1.25}$$

$$B_{t+1}^{LFL} = \left(I(i) \int \eta_i^{\frac{1}{\varepsilon}} \left(R_{i,t} \omega_{it} B_t \right)^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{1.26}$$

$$B_{t+1}^{NLFL} = \left((1 - I(i)) \int \eta_i^{\frac{1}{\varepsilon}} \left(R_{i,t} \omega_{it} B_t \right)^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{1.27}$$

where $I(i)$ is an indicator function, with value 1 if i is eligible to the policy. We can write the first-order condition with respect to ω_{it}.

$$\mathbb{E}_t \left[\left(\frac{\omega_{it}}{\eta_i} \right)^{-\frac{1}{\varepsilon}} R_{i,t}^{\frac{\varepsilon-1}{\varepsilon}} \left\{ \tilde{B}_{t+1}^{\frac{1}{\varepsilon}-\gamma} + I(i) \tau_{c,t}^{\sim} \left(\frac{B_{t+1}^{LFL}}{\eta_b} \right)^{\frac{1}{\varepsilon}} - (1 - I(i)) \tau_{l,t}^{\sim} \left(\frac{B_{t+1}^{NLFL}}{\eta_b} \right)^{\frac{1}{\varepsilon}} \right\} \right] = \frac{\lambda_t}{\eta_b^{\frac{1}{\varepsilon}} B_t^{\frac{\varepsilon-1}{\varepsilon}}}$$

Now we use the following definitions,

1. $\mathbf{1}\{ND_{i,t}\}$ in an indicator whether firm i did not default.
2. $\mathbb{P}_{i,t}$ is the default probability of firm i,
3. $S_{i,t}$ as the nominal (gross) spread over the risk-free rate R_t^f,
4. $V_{i,t}^W := \mathbb{E}_t \left[\tilde{B}_{t+1}^{\frac{1}{\varepsilon}-\gamma} \middle| ND_{i,t} \right]$,
5. $\tau_{c,t}^{-} := \tau_{c,t}^{\sim} \mathbb{E}_t \left[\left(\frac{B_{t+1}^{LFL}}{\eta_b} \right)^{\frac{1}{\varepsilon}} \middle| ND_{i,t} \right]$, analogously defining $\tau_{l,t}^{-}$,

combining with the property that if Y is a binary random variable, $\mathbb{E}[XY] = \mathbb{E}[X|Y=1]\mathbb{P}(Y=1)$, we can simplify the first order condition to:

$$\left(\frac{\omega_{it}}{\eta_i} \right)^{-\frac{1}{\varepsilon}} S_{i,t}^{\frac{\varepsilon-1}{\varepsilon}} \left(1 - \mathbb{P}_{i,t} \right) \left(V_{i,t}^W + I(i) \tau_{c,t}^{-} - (1 - I(i)) \tau_{l,t}^{-} \right) = \frac{\lambda_t}{R_t^f \eta_b^{\frac{1}{\varepsilon}} B_t^{\frac{\varepsilon-1}{\varepsilon}}}$$

Comparing the choice between investing in an eligible versus ineligible firm, and using the definition $\omega_{i,t} = \frac{b_{i,t}}{B_t}$:

$$\frac{b_{i,t}^{LFL}}{b_{j,t}^{NLFL}} = \left(\frac{\eta_i}{\eta_j}\right)\left(\frac{S_{i,t}}{S_{j,t}}\right)^{\varepsilon-1}\left\{\left(\frac{1-\mathbb{P}_{i,t}}{1-\mathbb{P}_{j,t}}\right)\left(\frac{V_{i,t}^{W}+\bar{\tau}_{c,t}}{V_{j,t}^{W}-\bar{\tau}_{l,t}}\right)\right\}^{\varepsilon}$$

Using the assumption $\gamma = \frac{1}{\varepsilon}$, we get the desired result by defining the risk-adjusted transfers

$$\left(\frac{1+\tau_{c,i,t}}{1-\tau_{l,i,t}}\right) := \left(\frac{1-\mathbb{P}_{i,t}}{1-\mathbb{P}_{j,t}}\right)\left(\frac{1+\bar{\tau}_{c,t}}{1-\bar{\tau}_{l,t}}\right)$$

1.10.2 Derivation of bound (1.19)

First, define

$$\mathbb{P}_{l,t} := \mathbb{E}\left[\mathbb{P}_{i,t}\left(A_t^1, S_{it}, \{\bar{\tau}_{c,j}\}_{j=1}^{t}\right)|\text{score}_{i,1}=0\right],$$

$$\mathbb{P}_{c,t} := \mathbb{E}\left[\mathbb{P}_{i,t}(A_t^1, S_{it}, \{-\bar{\tau}_{l,j}\}_{j=1}^{t})|\text{score}_{i,1}=0\right],$$

where the first is the (expected) probability of default of a marginal firm that is always eligible, and the second term is the counterfactual probability of default in case this firm were not to be eligible.

Consider the RDD estimand $\psi_t(t)$ in our empirical application, which captures the effect of being continuously exposed to the policy since its introduction in period $t = 1$. Through the lenses of our model, this estimand identifies:

$$\begin{aligned}\psi_t(t) = \mathbb{E}[P_{it}(A_t^1, S_{it}, \{\bar{\tau}_{c,j}\}_{j=1}^{t}) - P_{it}(A_t^1, S_{it}, \{-\bar{\tau}_{c,j}\}_{j=1}^{t})|\text{score}_{i,1}=0] \approx \\ \sum_{j=1}^{t}\mathbb{E}\left[\frac{\partial P_{it}(A_t^1, S_{it}, \mathbf{0}_{t\times 1}))}{\partial \bar{\tau}_j}\Big|\,\text{score}_{i,1}=0\right](\bar{\tau}_{c,j}+\bar{\tau}_{l,j}) = \\ \sum_{j=1}^{t}\mathbb{E}\left[\frac{\partial P_{it}(A_t^1, S_{it}, \mathbf{0}_{t\times 1}))}{\partial \bar{\tau}_j}\Big|\,\text{score}_{i,1}=0\right](\bar{\tau}_{c,j}-\mathbb{P}_{c,t}+\bar{\tau}_{l,j}+\mathbb{P}_{l,t}+\mathbb{P}_{c,t}-\mathbb{P}_{l,t}) \approx \\ \left(\sum_{j=1}^{t}\mathbb{E}\left[\frac{\partial P_{it}(A_t^1, S_{it}, \mathbf{0}_{t\times 1}))}{\partial \bar{\tau}_j}\Big|\,\text{score}_{i,1}=0\right]\psi_j^{\tau}(j)\right)+\mathbb{R}_t\end{aligned} \tag{1.28}$$

where we rely on first-order Taylor expansions combined with the definition of $\psi_t^{\tau}(t)$ to achieve the expressions, and the remainder term $\mathbb{R}_t$ can be written as:

$$\begin{aligned}\mathbb{R}_t := \sum_{j=1}^{t}\mathbb{E}\left[\frac{\partial P_{it}(A_t^1, S_{it}, \mathbf{0}_{t\times 1}))}{\partial \bar{\tau}_j}\Big|\,\text{score}_{i,1}=0\right](\mathbb{P}_{c,t}-\mathbb{P}_{l,t}) \approx \\ \sum_{j=1}^{t}\sum_{k=1}^{t}\mathbb{E}\left[\frac{\partial P_{it}(A_t^1, S_{it}, \mathbf{0}_{t\times 1}))}{\partial \bar{\tau}_j}\Big|\,\text{score}_{i,1}=0\right]\mathbb{E}\left[\frac{\partial \mathbb{P}_{it}(A_t^1, S_{it}, \mathbf{0}_{t\times 1}))}{\partial \bar{\tau}_k}\Big|\,\text{score}_{i,1}=0\right](\bar{\tau}_{c,k}+\bar{\tau}_{l,k}) \approx 0.\end{aligned}$$

Where we used the fact that the probability of default is itself a endogenous choice for the firm, and

that for all $1 \leqslant j, k \leqslant t$, $\mathbb{E}\left[\frac{\partial P_{it}(A_t^1, S_{it}, \mathbf{0}_{t\times 1}))}{\partial \bar{\tau}_j}\Big| \text{score}_{i,1} = 0\right] \mathbb{E}\left[\frac{\partial \mathbb{P}_{it}(A_t^1, S_{it}, \mathbf{0}_{t\times 1}))}{\partial \bar{\tau}_k}\Big| \text{score}_{i,1} = 0\right] \approx 0$, i.e., these terms are second order. We then conclude:

$$\psi_t(t) \approx \sum_{j=1}^{t} \underbrace{\mathbb{E}\left[\frac{\partial P_{it}(A_t^1, S_{it}, \mathbf{0}_{t\times 1}))}{\partial \bar{\tau}_j}\Big| \text{score}_{i,1} = 0\right]}_{\Delta_j^*} \psi_j^{\tau}(j). \tag{1.29}$$

The expression above shows that, in a first-order approximation, the estimand $\psi_t(t)$ recovers the effect, for those units around the threshold at the first period, of being continuously exposed to a sequence of "effective subsidies" $\{\bar{\tau}_{c,j} + \bar{\tau}_{l,j}\}_{j=1}^t$, in an aggregate equilibrium where the policy is in place. In other words, the RD estimand captures the effects of subsidizing bond issuance in the new aggregate equilibrium induced by the policy – general equilibrium effects are not captured by the estimator since units marginally below and above the threshold are equally subject to aggregate state variables. Recall that we can identify the sequence of $\psi_t^S(t)$.

As a consequence, for an alternative path of subsidies $\{\tilde{\tau}_j\}_{j=1}^t$, the expected effect:

$$\mathbb{E}[P_{it}(A_t^1, S_{it}, \{\tilde{\tau}_j\}_{j=1}^t) - P_{it}(A_t^1, S_{it}, \mathbf{0}_{t\times 1})|\text{score}_{i,1} = 0] \approx \sum_{j=1}^{t} \Delta_{j,t}^* \tilde{\tau}_j \,. \tag{1.30}$$

The above expression shows that, to compute multipliers, it is necessary, in a first-order approximation, to identify the partial derivatives $\left\{\Delta_{j,t}^*\right\}_{t=1}^t$. While it would be possible to leverage the homogeneity assumption underlying our RDD estimator to identify these partial derivatives by relying on the path of $\psi_t^{\tau}(t)$ and the path of estimands $\{\psi_t(j)\}_{j=1}^t$,[22] this approach will result in noisy estimates if some of the $\psi_t(j)$ are noisily estimated.[23] As an alternative, we thus propose an approach based on bounding the treatment path by solely relying on $\psi_t(t)$. Suppose that the policy function is monotone increasing in the subsidy rate τ_j, for every $j = 1, \ldots, t$. Consider any alternative policy path of transfers, $\{\tilde{\tau}_j\}_{j=1}^t$. In this case, the effect of this alternative path can be bounded above, in a first-order approximation, by:

[22] Indeed, consider the estimand $\psi_t(s)$ that captures the effect of being exposed to the policy for the s last periods, by period t. Under the cross-threshold homogeneity condition underlying our RD strategy, and through the lenses of our model, this estimand identifies:

$$\psi_t(s) = \mathbb{E}[P_{it}(A_t^1, S_{it}, \{-\bar{\tau}_{l,j}\}_{j=1}^{t-s} \cup \{\bar{\tau}_{c,j}\}_{j=t-s+1}^t) - P_{it}(A_t^1, S_{it}, \{-\tau_{l,j}\}_{j=1}^t)|\text{score}_{i,1} = 0] \approx$$
$$\sum_{j=t-s+1}^{t} \mathbb{E}\left[\frac{\partial P_{it}(A_t^1, S_{it}, \mathbf{0}_{t\times 1})}{\partial \tau_j}\Big| \text{score}_{i,1} = 0\right] \log\left(\frac{1+\tau_{c,j}}{1-\tau_{l,j}}\right).$$

By computing this formula for $s = 1, \ldots t$, and using the fact that the path of implied subsidies $\left\{\log\left(\frac{1+\tau_{c,j}}{1-\tau_{l,j}}\right)\right\}_{j=1}^t$ is identified, we can recover the partial derivatives recursively $\Delta_{j,t}^*$.

[23] This will be the case if the subpopulation (cohort) that starts treatment at period $t - j + 1$ are small.

$$\mathbb{E}[P_{it}(A_t^1, S_{it}, \{\tilde{\tau}_j\}_{j=1}^t) - P_{it}(A_t^1, S_{it}, \mathbf{0}_{t\times 1})|\text{score}_{i,1} = 0] \leqslant$$
$$\sup_{\Delta\in\mathbb{R}_+^t:\psi_t(t)=\sum_{j=1}^t \Delta_j \psi_j^\tau(j)} \sum_{j=1}^t \Delta_j\tilde{\tau}_j = \max_{j=1,\dots,t} \frac{\psi_t(t)\tilde{\tau}_j}{\psi_j^\tau(j)}, \tag{1.31}$$

and similarly, we may bound this quantity below by:

$$\mathbb{E}[P_{it}(A_t^1, S_{it}, \{\tilde{\tau}_j\}_{j=1}^t) - P_{it}(A_t^1, S_{it}, \mathbf{0}_{t\times 1})|\text{score}_{i,1} = 0] \geqslant$$
$$\inf_{\Delta\in\mathbb{R}_+^t:\psi_t(t)=\sum_{j=1}^t \Delta_j \psi_j^\tau(j)} \sum_{j=1}^t \Delta_j\tilde{\tau}_j = \min_{j=1,\dots,t} \frac{\psi_t(t)\tilde{\tau}_j}{\psi_j^\tau(j)}. \tag{1.32}$$

1.10.3 Firm Model

Here, we specify a firm model where the condition (1.12) holds.

Setup

Intermediary goods producers and liquidity risk Intermediary firms are indexed by $i \in [0,1]$, and are owned by a representative household. Each firm consists of a production function given by

$$y_{it} = a_{it} k_{it}^\alpha n_{it}^{1-\alpha},$$

where a_{it}, k_{it} and n_{it} are productivity, capital stock and labor, respectively. For simplicity, we assume that labor is hired at wage w_t. Firms can access a risk-free asset with (gross) return R_{it}^f, and m_{it} is the firms' holdings of the risk-free asset. Firms can issue debt $q_{it}b_{it}$, and firms distribute (real) dividends to the representative household d_{it}. For simplicity, we assume firms can commit to repay debt.

Firms are subject to liquidity risk in the following sense. At period $t-1$, firms hire both next period labor $l_{i,t}$ at wage w_t, invest in capital $k_{i,t}$, and have to repay b_{it}. At t, firms are subject to a productivity shock, i.e., $a_{i,t}$ is stochastic. We define the firm working capital as

$$W_{it} = p_{it}y_{it} + m_{it} - w_t l_{it} - b_{it}.$$

We model liquidity risk in analogy to the "buffer-stock" idea, where if firms do not have enough cash on hand to pay obligations, they have to liquidate capital stock. We model that in reduced form by assuming after the productivity shock is realized, production takes place and firm decides on borrowing, the remaining capital is given by $\gamma\left(W_{it} + q_{it}b_{it+1}\right)k_{it}$, where $\gamma(x)$ is a continuously differentiable function such that:

$$\gamma'(x) > 0, \quad \gamma(0) = (1-\delta), \quad \lim_{x\to-\infty}\gamma(x) = 0, \quad \lim_{x\to+\infty}\gamma(x) = (1-\bar{\delta}), \quad 0 < \delta < \bar{\delta} < 1.$$

Final Goods Producer Final goods producer operates in a competitive market with technology given by:

$$Y_t = \int \left(y_{it}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} di.$$

Intermediate goods demand is given by,

$$y_{it} = p_{it}^{-\sigma},$$

where the price index $P_t := \left(\int p_{it}^{1-\sigma} di \right)^{\frac{1}{1-\sigma}}$, is normalized to 1.

Asset Demand We follow the financial block from Section 1.5.1

$$q_{i,t} = b_{it+1}^{-\frac{1}{\varepsilon-1}} (1+\bar{\tau_{i,t}})^{\frac{\varepsilon}{\varepsilon-1}} \frac{B_t^{\frac{1}{\varepsilon-1}}}{R_t^B},$$

where B_t is a measure of aggregate corporate debt, and R_t^B is an aggregate (gross) return of corporate debt.

Representative Household From the household block, we assume only the existence of a well-defined stochastic discount factor, Λ_{t+1}, which the firm takes as given in its valuation.

Firm Maximization Problem

The firm maximization problem is given by (where we drop the i subscript),

$$v_t(k_t, a_t, l_t, W_{it}, X_t) = \max_{l_{t+1}, k_{t+1}, b_{t+1}, m_{t+1}} d_t + \mathbb{E}_t \left[\Lambda_{t+1} v_{t+1}(k_{t+1}, l_{t+1}, a_{t+1}, W_{it+1}, X_{t+1}) \right]$$

subject to

$$W_{t+1} = p_{t+1} y_{it+1} + m_{t+1} - w_{t+1} l_{t+1} - b_{t+1}$$

$$\frac{y_t}{Y_t} = p_t^{-\sigma},$$

$$y_t = a_t k_t^{\alpha} l_t^{1-\alpha},$$

$$q_t = b_{t+1}^{-\frac{1}{\varepsilon-1}} (1+\tau_t)^{(\frac{\varepsilon}{\varepsilon-1})} \frac{B_t^{\frac{1}{\varepsilon-1}}}{R_t^B},$$

$$d_t \leqslant W_t + q_t b_{t+1} + \gamma \left(W_t + q_t b_{t+1} \right) k_t - k_{t+1} - \frac{m_{t+1}}{R_t^f},$$

$$b_{t+1} \geqslant 0, \quad m_{t+1} \geqslant 0, \quad k_{t+1} \geqslant 0$$

where X_t captures the exogenous (from the perspective of the firms) aggregate state variables, $X_t := (B_t, R_t^B, R_t^f, w_t)$. Condition (1.12) can then be derived by combining the first order conditions with respect to debt b_{t+1}, cash-holdings m_{t+1} and the envelope theorem and will hold with equality

when firms choose an interior solution for both variables, i.e., they have positive debt and savings at a given point in time.

1.11 Simple Model - Borrowing constraints and Cash Holdings

This section presents a simple model exploring how firms manage liquidity risk under borrowing constraints. Firms risk default if they fail to produce enough intermediate output to cover costs. A reduction in borrowing constraints enables capacity expansion while decreasing cash holdings. The model suggests that empirical results are consistent with easing these constraints, with real effects amplified by the endogenous decision of firms to reduce cash buffers due to a lower need for precautionary savings.

1.11.1 Setup

We consider a three-period model, $t = 0, 1, 2$. At period $t = 0$ firms decide on a project consisting of inputs k that produces a total amount of output $y = k^{\alpha}$, $0 < \alpha < 1$. There is a risk-free asset (cash). Firms start without cash, $m_0 = 0$, and cannot issue equity. Input unit cost is q and requires a fraction δ of payment upfront, rest must be paid at period 1.

Firms need external finance to conduct the project, all credit is due at period 2. At period 0 firms can issue "long-term" corporate bonds, at a price q^b, subject to the following borrowing constraint:

$$b \leqslant \bar{b},$$

where $\bar{b} > 0$ is an exogenous borrowing limit.

Firms face liquidity risk in the sense that receive only θy of the project total output in period 1, where $\theta \sim \mathcal{U}[0, 1]$ is realized in period 1. The firm receives the remaining $R(1-\theta)y$ in period 2, where we assume the return is interest rate adjusted to make the shock a pure liquidity shock. After the realization of θ firms can borrow to cover working capital (wages) expenses and are subject to a borrowing limit:

$$Rl \leqslant \lambda(1-\theta)Ry_2,$$

where $\lambda \in [0, 1]$. If the firm does not have enough funds to pay inputs, $y_2 = 0$ and $y_2 = y$ otherwise.[24] Firms can hold cash, and the rate of return is $R > 1$.

Debt is priced competitively by risk-neutral lenders, such that $q^b = \mathbb{P}(b, k)$ and $q^l = 1$, where $\mathbb{P}(\cdot)$ is the probability of repaying debt at period 2. The firm is risk-neutral and maximizes expected returns in period 2.

[24] In this model, y_2 is potentially a function of l, which may lead to multiple equilibria. When this is the case we assume the best equilibrium is played (the maximum possible y_2).

1.11.2 Solution

We start by analyzing firm behavior in period 1 to solve the model. It is enough to restrict to the case where the borrowing constraint binds at period 1 since the borrowing and saving interest rates are the same. Conditional on the choices from period 0 and the realization of θ, firm default if and only if:

$$\theta k^{\alpha} + m_1 + \lambda(1-\theta)k^{\alpha} - q(1-\delta)k \leqslant 0,$$

thus we can define the default threshold, firm default if and only if

$$\theta \leqslant \bar{\theta}(k, m_1) := \frac{q(1-\delta)k - \lambda k^{\alpha} - m_1}{(1-\lambda)k^{\alpha}}.$$

With this characterization, we can simplify the firm's maximization problem, summarized in Lemma 1. Cash holdings after period 1 are given by $m_1 = \mathbb{P}(\theta \geqslant \bar{\theta})b - \delta q k$.

Lemma 1. *1. Firm's choice is characterized by the solution to the optimization problem:*

$$\begin{aligned}
&\max_{b,k,\bar{\theta}} A(\bar{\theta})k^{\alpha} - qk - S(\bar{\theta})b \\
&\textit{s.t. } b \leqslant \bar{b}, \\
&A(\bar{\theta}) = \mathbb{P}(\theta \geqslant \bar{\theta}) + (1 - \mathbb{P}(\theta \geqslant \bar{\theta}))\mathbb{E}[\theta|\theta \leqslant \bar{\theta}], \\
&S(\bar{\theta}) = 1 - \mathbb{P}(\theta \geqslant \bar{\theta}), \\
&\bar{\theta} = \frac{qk - \lambda k^{\alpha} - \mathbb{P}(\theta \geqslant \bar{\theta})b}{(1-\lambda)k^{\alpha}}.
\end{aligned}$$

2. If $\bar{\theta}^* = 0$, then the optimal choice of capital and cash holdings are defined implicitly by:

$$qk^* - \lambda {k^*}^{\alpha} = \bar{b}.$$

$$m_1^* = (1-\delta)qk^* - \lambda {k^*}^{\alpha}$$

1.11.3 Comparative Statics

From Lemma 1 we can derive the following comparative statics:

Lemma 2. *1.*

$$\frac{dk^*}{d\lambda} > 0, \quad \frac{dm_1^*}{d\lambda} < 0, \quad \frac{d\bar{\theta}}{d\lambda} = 0$$

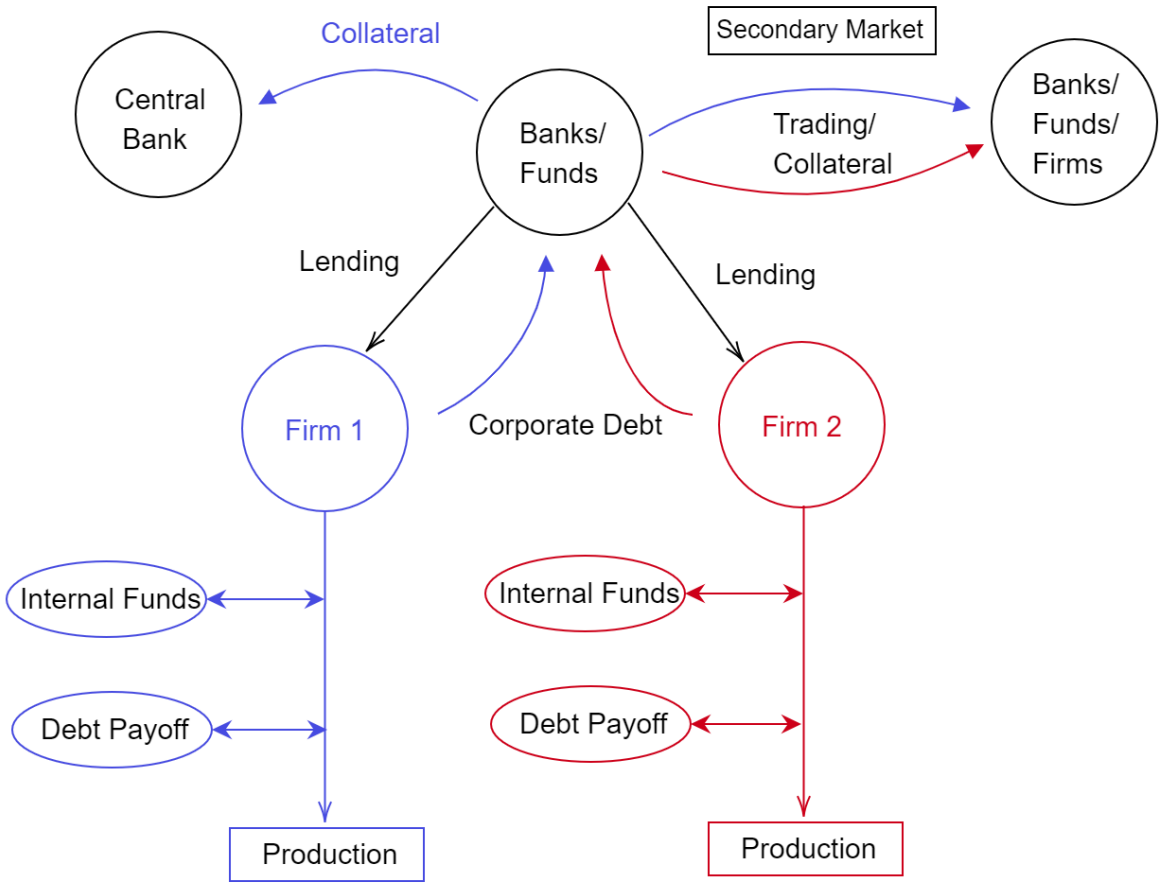
2.

$$\frac{dk^*}{d\bar{b}} > 0, \quad \frac{dm_1^*}{d\bar{b}} > 0, \quad \frac{d\bar{\theta}}{d\bar{b}} = 0$$

The model illustrates how firms manage liquidity risk under borrowing constraints. Firms face default if they lack sufficient intermediate output to cover costs. Higher liquidity allows for increased capital investment but reduces cash holdings.

1.12 Diagram of the empirical analysis

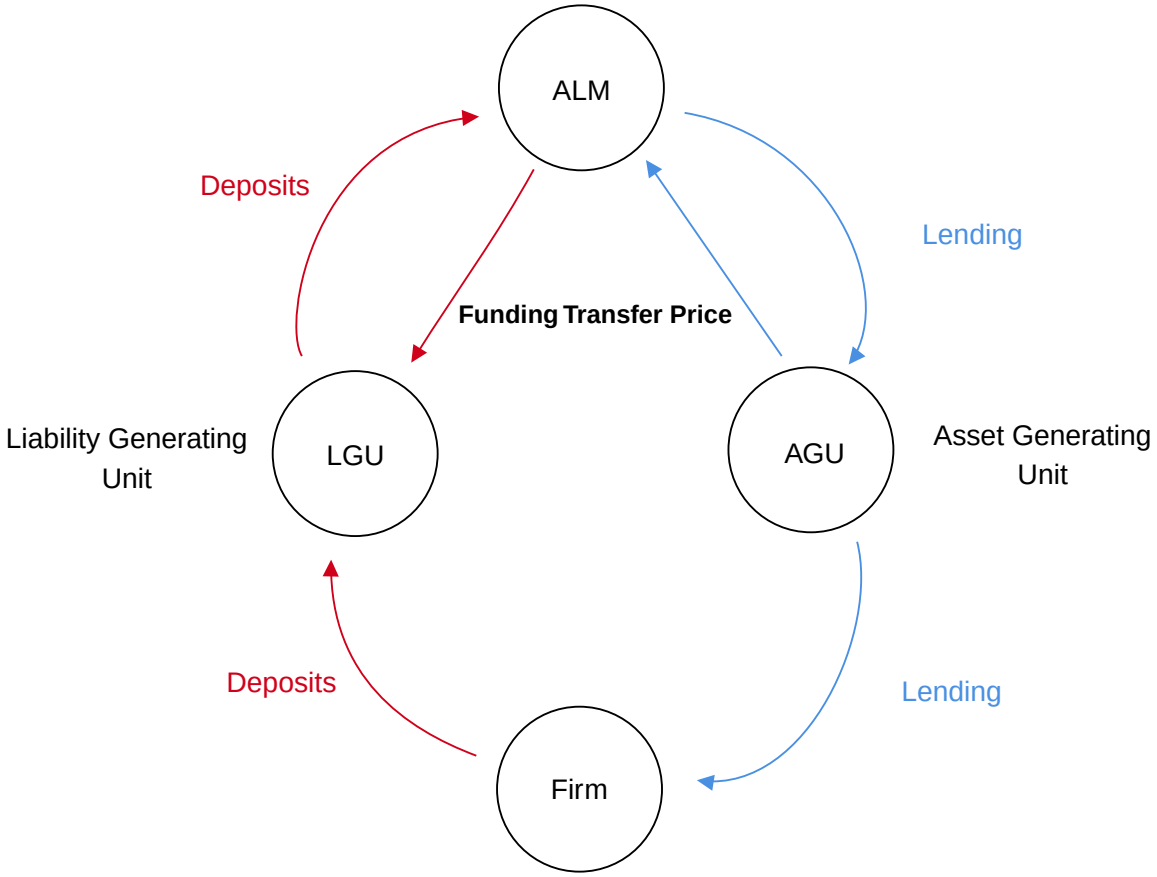
Figure A10: Diagram Empirical Analysis



Note: This diagram illustrates in a simplified way the identification and empirical strategy, as well as the connection between the empirical results.

1.13 Diagram of theoretical model

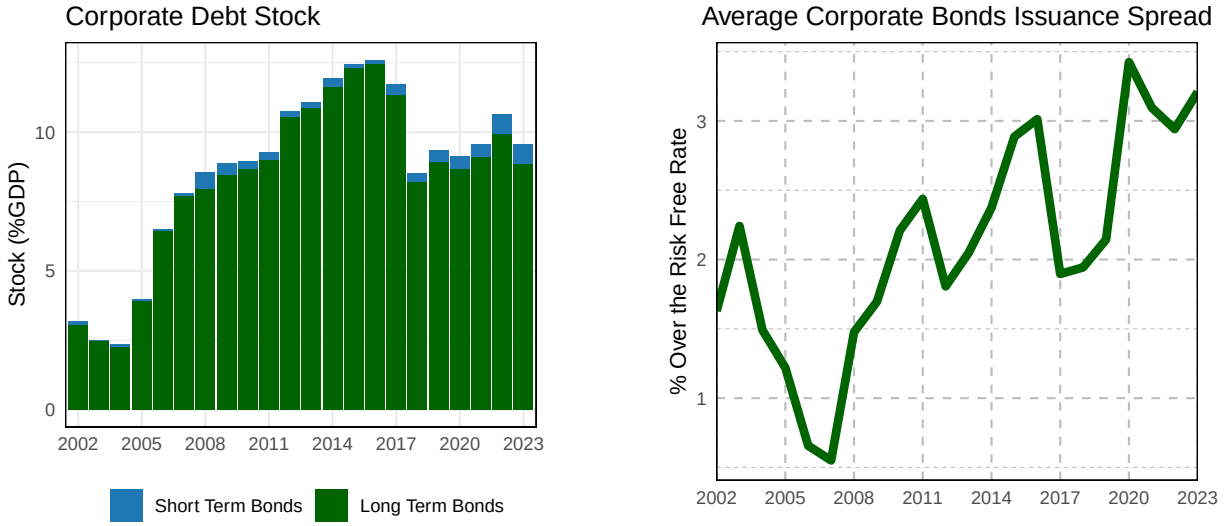
Figure A11: Model Diagram



1.14 Corporate Bonds Market In Brazil

As of December 2023, Brazil's corporate bonds accounted for about 10% of GDP (20% of total credit to nonfinancial corporations). In comparison, the US corporate bonds market accounted for approximately 40% of GDP and 42% of total credit. In Brazil, there are two types of corporate bonds: long-term (*Debêntures*) and short-term (*Notas Comerciais*). The left panel on Figure A12 shows the time series evolution of the outstanding corporate debt (as a share of GDP). The debt stock is primarily in long-term bonds, but the past two years have shown an increase in short-term bonds following a regulatory reform in late 2021. The low share of GDP for Brazil's corporate bonds market is due to the overall low share of credit to GDP, yet the market remains relatively important, especially for larger firms.

Figure A12: Evolution of Corporate Bonds Market.



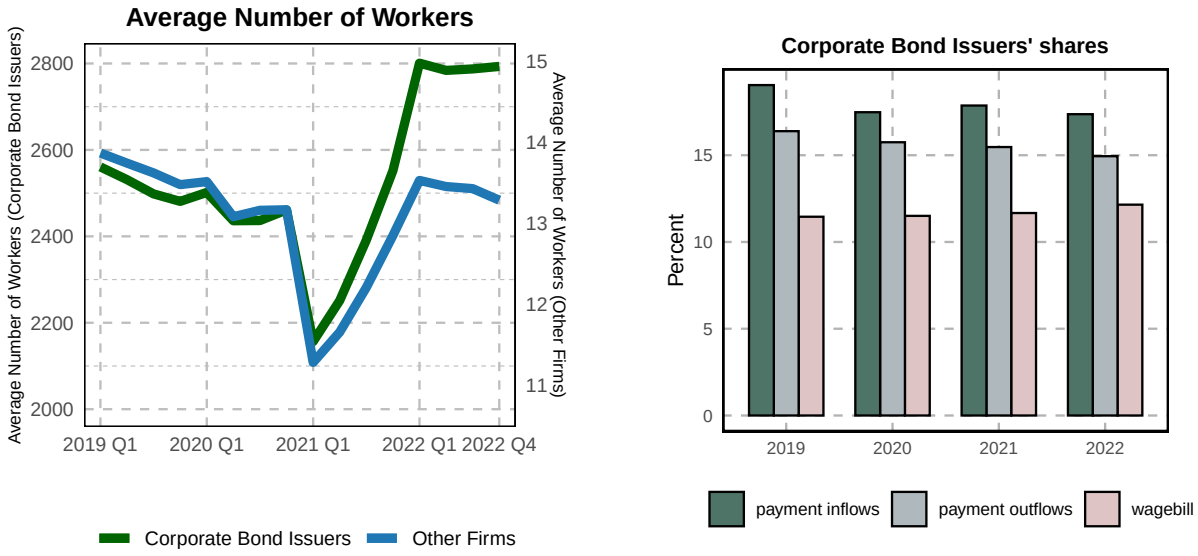
Note: The graph shows the evolution of corporate bond stock and spreads over time. Stocks are calculated on the last business day of each year. Issuance spreads are the (simple) average of spreads on assets linked to the baseline interest rate, which represents about 70% of the issuance volume and number in the sample period. Data sources: B3 (Short Term Bonds) and ANBIMA (Long Term Bonds).

The average spread over the baseline interest rate (risk-free rate in the inter-bank market) on corporate bonds is notably low, around 3.2% in 2023 for interest-rate indexed bonds, compared to an aggregate spread of 7.5% on all corporate credit. The right panel of Figure A12 shows the time series evolution of these bond spreads. Unlike the US, where most corporate bonds are nominal liabilities (see Gomes, Jermann, and Schmid 2016), 70% of Brazil's corporate bonds are indexed to the baseline interest rate, and 23% to the inflation rate.

Corporate bond issuers are usually the largest firms in the economy, with more access to credit and significant roles in labor markets and the supply chain. By the end of 2022, these firms employed an average of 2,800 workers, compared to 13 in the rest of the economy. They represented 12% of the formal labor share and 15% of payment flows between non-financial firms, which measures

their importance in intermediary goods and services transactions. Figure A13 shows the time series of firm size, labor share, and payment share. Notably, corporate bond issuers experienced a smaller decline in labor force during COVID and a significantly stronger recovery compared to other firms in the economy.

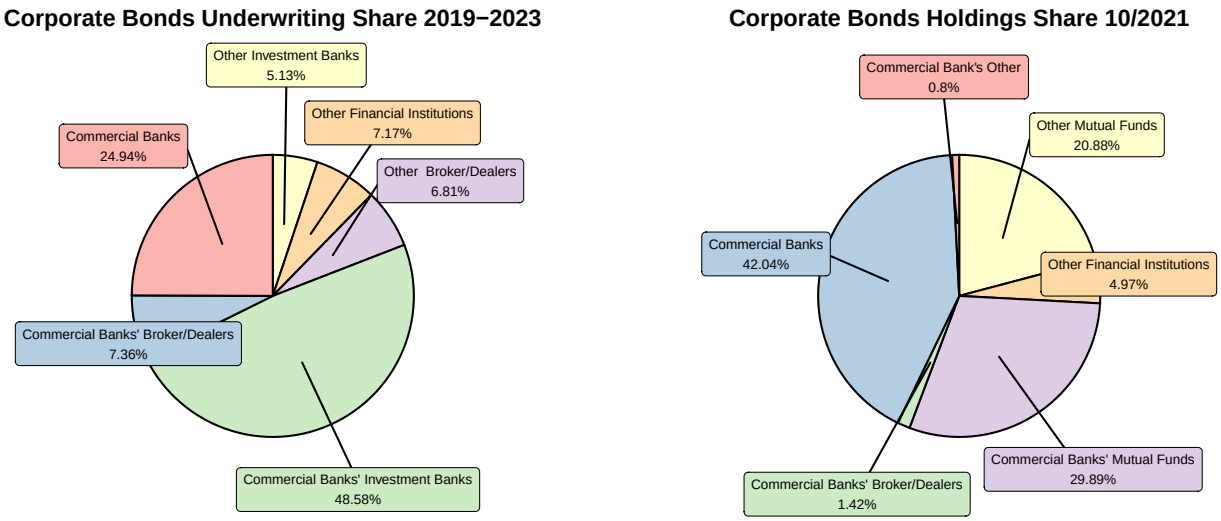
Figure A13: Corporate Bond Issuers Size.



Note: The graph shows the average number of workers, the labor share, and the payment share in inter-firm transactions by corporate bond issuers. The labor share represents the share of the wage bill for formal workers in the last quarter of the respective year. The payment share refers to the share of total electronic payments (invoices, bank transfers, and PIX) between non-financial firms in the respective year. Data Sources: B3 and ANBIMA (Corporate Bond Issuers), RAIS (wage bill and number of workers), and BCB - Department of Banking Operations and Payment Systems (electronic payments).

Corporate bonds can be traded in the secondary market, but liquidity is low, with most bonds rarely or never traded (Carvalho and Marques 2020). The market is highly concentrated, and dominated by commercial banks. Figure A14 shows that the 7 largest commercial banks and their subsidiaries account for 81% of corporate bond underwriting and 75% of holdings (in October 2021). These banks often follow a "buy-and-hold" strategy, allocating bonds to their funds or own books, which likely contributes to the low liquidity in secondary markets. As a result of this market structure, over half of commercial bonds are not even rated by rating agencies (Carvalho and Marques 2020).

Figure A14: Corporate Bonds Market Structure.



Note: The graph displays the underwriting and holdings share of local currency-denominated corporate bonds for Brazil's 7 largest commercial banks, their subsidiaries, and other financial institutions. Underwriting shares are based on total issuance value from January 2019 to December 2023, while holdings shares are based on the reported book value of each corporate bond on the last business day of October 2021. A fund is considered a subsidiary of a commercial bank if managed by one of its companies. Data sources: ANBIMA (Underwriting), B3 (Holdings), IRS and CVM (Ownership and Fund Management).

Commercial banks have incentives to hold corporate bonds in their portfolios because a significant portion of their deposits (68% as of December 2023) are term deposits with returns indexed to the baseline interest rate. This creates a natural hedge for their deposit franchise, similar to the US market where banks lend long-term on fixed rates as discussed by Drechsler, Savov, and Schnabl 2021. Additionally, banks can generate service flow from these bonds by using them in repo operations with noncommercial firms, which provides both liquidity benefits and tax advantages, as these transactions are exempt from the financial transaction tax that applies to term deposits.[25]

[25]Holdings microdata shows that repo operations are exclusively conducted with large companies whose deposits exceed insurance limits. Backing these operations with unliquidated long-term assets can substitute for deposit insurance by reducing the incentive for runs Payne and Weiss 2023, providing additional liquidity benefits. Moreover, repo operations are not classified as deposits and face fewer regulatory constraints. In the Brazilian market, almost all corporate bond holdings by nonfinancial institutions are short-term with resale obligations, making them essentially cash-like assets for firms. This contrasts with the practices of large firms in the U.S. market. discussed by Darmouni and Mota 2023, where cross-border tax incentives drive corporate holdings of bonds.

Chapter 2

Firm-Level and Aggregate Effects of Cheaper Liquidity: Evidence from Factoring

This chapter is jointly authored with Thiago Christiano Silva and Henry Zhang.[1]

The views expressed in this paper are those of the authors and do not necessarily reflect those of the Central Bank of Brazil.

2.1 Introduction

How important is the working capital financing that firms use to bridge the gap in cash flows between paying upfront for labor and waiting for future payments from customers? How responsive are firms' production decisions to working capital financing terms? These questions are relevant to central banks and most firms around the world, especially firms in sectors with long production timelines or long payment clearing cycles. Small and medium enterprises (SMEs) tend to rely more heavily than large firms on working capital financing, due to limited cash and access to corporate treasury operations, but SMEs also tend to face challenges acquiring financing due to limited collateral and credit history (BIS 2023). This paper studies how a change in the price of a specific type of working capital financing affects firms' outcomes, across the entire distribution of firms.

The most common form of working capital financing worldwide is trade credit, defined as suppliers allowing buyers to pay at a later time than the time of transaction. Formally, a trade credit contract consists of an upfront transaction, where the supplier receives a receivable note in return for the good or service, and a payment clearing transaction, where the buyer gives the supplier the

[1]We are deeply grateful to Dave Donaldson, Tong Liu, Stephen Morris, and Rob Townsend for their invaluable guidance and support. We also thank Patrick Adams, Taha Choukhmane, Sarah Gertler, Brice Green, Jack Li, Daniel O'Connor, Ben Olken, Christopher Palmer, Jonathan Parker, Hong Ru, Kerry Siani, David Thesmar, Maria Tiurina, Adrien Verdelhan, Emil Verner, Christian Wolf, Jiaheng Yu, and participants at the MIT Macroeconomics Seminar and Finance Lunch for their valuable comments and suggestions. For fianncial support, we thank the Shultz Fund, and Henry thanks NSF GRFP grant 1745302. Any mistakes are our own.

promised payment in return for retiring the receivable. Suppliers who want cash before payment clearing can sell the receivable to a financial intermediary called a factor, and this type of sale is known as factoring. The worldwide share of receivables that are factored has increased from 7% in 1997 to 24% in 2017 (Boissay, Patel, and Shin 2020). In Brazil, the empirical setting for this paper, factoring is the highest volume type of intermediated working capital financing. Few papers have estimated the impact of working capital financing terms on firms' outcomes, and none uses a dataset encompassing all firms in an economy.

This paper studies the impact of a shift in the factoring interest rate – or equivalently the price of receivables – on firms' financing, labor demand, supply chain relationship terms, and sales. An important feature of the Brazilian setting is the presence of specialized mutual funds called FIDCs, which purchase and securitize firms' receivables, selling debt tranches to institutional investors in a similar mechanism to a mortgage-backed security. FIDCs' share of all factoring in Brazil has grown from 7% in 2015 to 32% in 2023, following trends of greater use of financial technology in factoring and greater asset demand from investors, and FIDCs' share is likely to continue to increase as ongoing reforms broaden the investor base and reduce the informational advantage of banks over FIDCs in factoring. FIDCs' regulatory structure mandates them to primarily purchase receivables, along with short-term Brazilian Treasury notes for liquidity management. FIDCs' main investors are broad market mutual funds and pension funds who face constraints on asset allocation, so there is a liquidity-driven component of flows to FIDCs. In turn, there are sticky relationships between FIDCs and firms due to costly screening: the risk of a firm's receivables depends on the firm's customers in addition to the firm's own creditworthiness, and firms must establish and manage escrow accounts in FIDCs. We use FIDCs' past receivables purchases and current flows to instrument for firms' factoring interest rate, and we are the first to estimate causal responses of firms' decisions and outcomes to the factoring interest rate. Our instrument is valid because institutional investors cannot observe the identities of the numerous firms whose receivables are held by FIDCs, the flows to FIDCs have a liquidity-driven rebalancing component, and FIDCs' current demand for receivables must respond to flows to FIDCs due to FIDCs' capital allocation constraints.

We construct a new dataset using several restricted-access databases at the Central Bank of Brazil (BCB). The dataset contains factoring transactions from banks and FIDCs for all formally registered firms in Brazil that have ever sold their receivables. We merge the factoring data, at the firm by month level, to the employer-employee matched dataset (RAIS), the universe of electronic payments, and boleto contracts that specify trade credit payment terms. There are three main takeaways from our summary statistics: firms have greater cash inflow volatility than cash outflow volatility, firms that are smaller and riskier tend to factor a larger share of receivables, and FIDCs tend to purchase receivables from riskier firms. Our interpretations are as follows. The differences in cash flow volatility generates firms' demand for short-term liquidity through working capital financing. Firms that face tighter financing constraints rely more on factoring, since factoring is inherently available to any firm that offers trade credit, and most firm-to-firm transactions in Brazil feature trade credit.

In light of these facts, the results from our regressions can be interpreted as a local average treatment effect for firms that factor as their marginal source of liquidity. In the first stage, we use

local projections to show that a firm's exposure to net fund inflows leads to a decrease in the factoring interest rate, of similar magnitude in the same month as the flow and the next month, and then the effect on the interest rate decays to zero after four months. In the IV regressions, we show that a flow-induced one percentage point decrease in the interest rate leads to large contemporaneous increases in factoring volume (16.2%), revenue (6.1%), intermediate input expenditure (3.6%), and permanent contract labor (1.1%), as well as a contemporaneous decrease in temporary contract labor (2.1%). The IV local projections show that factoring volume reverts to the previous level after several months, following the reversion of the interest rate, but revenue and intermediate input expenditure persist 1 to 3 percentage points above the previous level. The increase in permanent labor persists at around 1%, as expected. However, the impact on temporary labor reverts after two months; cheaper factoring leads to greater use of temporary labor as well in the long term. We explain the contemporaneous effect through the lens of cash flow volatility, in which hiring labor with flexibility is an imperfect substitute for factoring, and the long term effect through the lens of growth with loosened financial frictions.

We build a model of factoring to explain how the cash flow volatility affects firms' outcomes, how cheaper factoring can have real consequences, including on aggregate output in general equilibrium. Firms' cash inflow volatility arises from the timing of customers' payments and uncertain demand. Factoring directly eliminates volatility from the timing of payments, while other forms of financing do not. When factoring is more expensive, firms not only substitute towards other financing, but also demand more temporary labor to match the fluctuation in cash outflows to cash inflows. In the presence of fixed capacity costs or efficiency costs of temporary versus permanent labor, it is more efficient for firms to factor rather than fluctuate production. We calibrate the model using moments in the data, and apply it to two counterfactuals of how firms' outcomes and aggregate output would respond to a decrease in the factoring interest rate. The first counterfactual is analogous to partial equilibrium, where we only change factoring interest rate for an infinitesimal subset of firms, holding constant aggregate prices and allocations. The second counterfactual is analogous to general equilibrium, where we reduce the factoring interest rate for all firms and allow aggregate prices and allocations to adjust. We motivate the general equilibrium counterfactual through three broad trends: the increased use of financial technology in factoring to reduce transaction costs, the introduction of receivables registries to reduce search and verification costs, and regulatory changes that increase demand for receivables through expanding investor access to FIDCs. We find that the partial equilibrium response is similar in magnitude to the regression results, with elasticities around 3.1, while the general equilibrium response is an order of magnitude smaller, with elasticities of 0.3 to 0.5. General equilibrium dampening arises from higher prices, particularly for permanent contract wages.

This paper relates to several strands of literature in finance and macroeconomics. The literature on the real effects of credit supply shocks has shown large cross-sectional impacts on employment and output, primarily using data on large firms. Our regression estimates have similar magnitudes to those of Chodorow-Reich (2014), where working capital loans comprise a large share of the credit supply shock. Our estimates of the long term impacts on wage bill are comparable to those of Huber (2018), and our aggregate elasticities are comparable to those of Herreno (2023). Our

contribution is threefold: first, we provide the first causal estimates of elasticities of real outcomes to factoring, an important and under-studied type of financing; second, we are the first to show heterogeneity across the entire distribution of firm size and other attributes; and third, we show that there are small effects on financial outcomes, both trade credit and other financing.

In the literature on working capital financing, Lian and Ma (2021) show that one channel for real effects of working capital financing terms are borrowing constraints based on cash flows. Several papers have shown that the debt portfolios of small and medium enterprises (SMEs), and SMEs' responses to financing shocks, are systematically different from those of large firms (Bahaj et al. 2022; Chodorow-Reich et al. 2022; Custódio, Ferreira, and Laureano 2013), with an important role for collateral (Luck and Santos 2019). Caglio, Darst, and Kalemli-Ozcan (2022) show that while large public firms primarily rely on unsecured credit lines, all other types of firms rely heavily on accounts receivable backed financing. In our setting, factoring is the main form of working capital financing for all but the largest firms, and small firms factor a larger share of receivables, yet all firms have significant responses in real outcomes to the factoring interest rate. Hahn et al. (2024) show that a decrease in the flexibility of temporary labor reduces firms' cash holdings; we show the converse, that cheaper financing through factoring lead sto substitution away from temporary labor. Almeida, Carvalho, and Kim (2024) compare the investment and output responses to working capital financing shocks across more versus less constrained firms; our contribution is that we instrument the financing shock, we demonstrate labor substitution patterns that are key to understanding the mechanism for working capital financing terms to have real effects, and we have rich microdata that allow us to study the impact of firm-level shocks with firm-to-firm trade credit, rather than industry-level shocks on firms' total trade credit exposure.

The long literature on trade credit discusses how firms use trade credit for risk sharing (Yang and Birge 2018) and reserve liquidity (Amberg, Jacobson, Von Schedvin, et al. 2021), with high substitutability for bank loans (Restrepo, Cardona-Sosa, and Strahan 2019). There is a higher demand for trade credit in environments with weaker creditor protection due to the information advantage of suppliers versus other creditors (Fabbri and Menichini 2010), and trade credit eases bank credit constraints (Adelino et al. 2023; Garcia-Martin, Justel, and Schmidt-Eisenlohr 2023; Skrastins 2021). Through the interplay of trade credit and bank credit, trade credit can amplify or dampen aggregate fluctuations (Altinoglu 2021; Bocola and Bornstein 2023; Reischer 2024), both through production linkages and through default risk (Jacobson and Von Schedvin 2015; Mateos-Planas and Seccia 2021). However, the empirical literature on the financial intermediation of trade credit is nascent, comprising of three papers that have crucial differences with this paper. The closest paper is Bottazzi, Gopalakrishna, and Claudio (2023), who use a one-time shock to the supply of factoring services and show that factoring alleviates financial constraints. In comparison, our dataset contains the factoring interest rate and leverages quasi-exogenous variation in the interest rate, allowing us to estimate financing semi-elasticities, while Bottazzi, Gopalakrishna, and Claudio (2023) only measure the total effect of the shock on factoring share and firms' outcomes. In addition, our dataset is much larger, with 1.03 million firms in our dataset versus 2,663 firms in Bottazzi, Gopalakrishna, and Claudio (2023), and our dataset includes the numerous small firms that rely most heavily on factoring. By comparison, most papers in the literature on trade credit

and working capital financing only have data on large firms. Amberg, Jacobson, and Qi (2024) study supply-chain finance (SCF), which is buyer-initiated unlike supplier-initiated factoring, and similarly to Bottazzi, Gopalakrishna, and Claudio (2023) study the total effects of SCF enrollment, rather than marginal changes to interest rates. In our setting in Brazil, factoring volume is far higher than SCF volume. Yu (2023) study accounts receivable backed lending in the US, which is the primary form of working capital financing for the set of 695 sellers and 527 buyers in his sample of publicly traded firms, and focuses on the moral hazard motivation of trade credit. As Caglio, Darst, and Kalemli-Ozcan (2022) show, large publicly traded firms differ from most other firms in their composition of working capital financing, primarily comprising of lines of credit, loans, and bonds rather than trade credit and factoring. In relation to the trade credit and factoring literature, our contribution is to show empirical facts that motivate the importance of factoring for trade credit, estimate the heterogeneous impact of factoring on trade credit terms, and model how factoring enhances trade credit by alleviating the cash flow variation that arises from extending financing through trade credit. We are the first to estimate the causal impact of a change in the factoring interest rate on firms' outcomes, and the first to document the heterogeneous impacts across the distribution of firms.

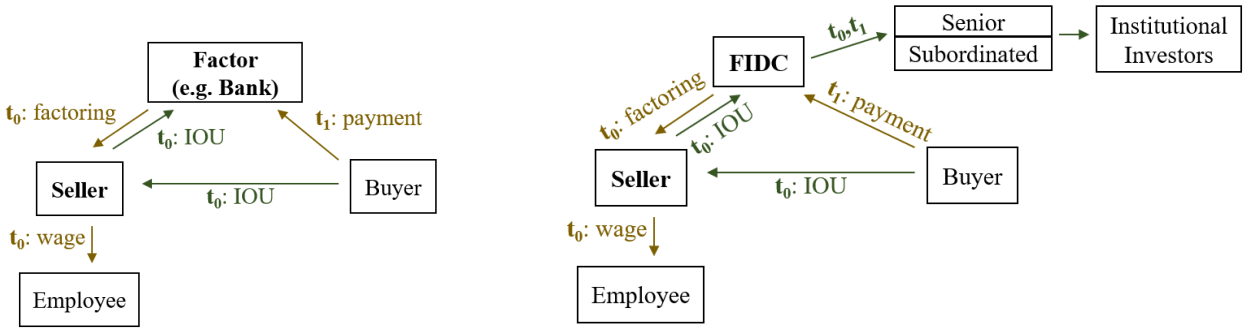
The importance of FIDCs to factoring, through investors' asset demand for receivables, relates to the literature on non-bank credit supply, specifically through funds. Our instrument is inspired by flow-induced trading, namely that firms buy or sell assets in proportion to their holdings, rather than in order of liquidity or in the same allocation as the market portfolio (Coval and Stafford 2007; Darmouni, Siani, and Xiao 2022; Dou, Kogan, and Wu 2022; Edmans, Goldstein, and Jiang 2012; Van der Beck 2022; Wardlaw 2020). While the literature focuses on equity and bond funds, we adopt flow-induced trading to the factoring setting, featuring short maturity and a high rate of recurring purchases, by using past issuance rather than lagged holdings as the measure of a firm's exposure to a fund. The primary justification for the exclusion restriction, that fund flows only affect firms' outcomes through funds' asset purchases, is that flows arise from liquidity or asset class rebalancing motives, rather than expectations of firm-level future returns, for instance from productivity shocks. In our setting, almost all factoring goes to firms whose assets are otherwise not exposed to rebalancing because they are not publicly traded, they do not issue bonds, and there are no analogous funds to FIDCs for long-term debt. Also, most FIDCs purchase receivables from thousands of firms, and neither the firms seeking factoring nor their buyers with payment obligations are reported to investors, so it is unlikely that FIDC flows respond to specific productivity shocks. For these reasons, we believe that the exclusion restriction is more plausible in our setting than in the literature.

We organize the remainder of the paper as follows. Section 2 describes the institutional setting of factoring and FIDCs. Section 3 describes the data and the key facts. Section 4 introduces the methodology and discusses the regression results. Section 5 interprets the results in the lens of a model of factoring and discusses the results from the counterfactuals. Section 6 concludes.

2.2 Factoring and FIDCs

When firms transact, they choose payment terms along with the price and quantity of the good or service. These payment terms stipulate the payment date or dates, often in relation to the contract date. When the payment date t_1 is later than the contract date t_0, the seller has offered trade credit to the buyer. We represent this in Figure 2.1a with an arrow from the buyer to the seller labeled with "IOU." The contract that the buyer signs, stipulating the payment terms, serves as proof of accounts receivable, which is an asset on the seller's balance sheet whose maturity equals the difference between the payment date t_1 and the contract date t_0. The discrepancy between cash inflows at t_1 and cash outflows at t_0, as well as greater volatility for cash inflows relative to outflows, generates firms' demand for working capital financing.

Figure 2.1: Diagram of the Financing Operations: Trade Credit, Factoring, and FIDCs



(a) Typical Factoring

(b) Factoring to a FIDC

Notes: These diagrams show the sequence of transactions for factoring and FIDCs in our setting. On the left is the typical environment for factoring around the world, where the seller extends trade credit to the buyer upon the transaction at time t_0, and the buyer repays at time t_1. If the seller wants cash before t_1, the seller can sell the receivable "IOU" to the factor, receiving the discounted value of the receivable. The discount can be converted into the factoring interest rate. The buyer directly repays the factor through an escrow account set up by the seller. The diagram on the right shows the institutional setting in Brazil when the factor is a FIDC: there is an additional step of securitization where the buyers' payments flow to the standardized share classes held by institutional investors.

After providing trade credit, the seller can choose whether to retain the receivable on its balance sheet or sell the receivable to a financial intermediary for a discount to its face value. Factoring is defined as the sale of the receivable, and the factor is the financial intermediary who purchases the receivable. By factoring the receivable, the seller obtains cash upfront that the seller can use to pay its employees, while still providing trade credit to its buyer for various reasons, such as alleviating moral hazard.

Both factoring and trade credit are widely used around the world (Boissay, Patel, and Shin 2020). A unique feature about the Brazilian institutional setting is the receivables fund, written in Portuguese as the Fundo de Investimento em Direitos Creditórios (FIDC). The regulation that enabled the creation of FIDCs was Instruction 356 from the Securities and Exchange Commission of Brazil (CVM), passed in December 2001. In the subsequent two decades, FIDCs have steadily

grown to become a major asset class. Similarly to other types of funds, net asset value (NAV) is the primary metric of fund size. From January 2013 to January 2024, total FIDC NAV grew from 46 billion US dollars (USD) to 111 billion USD,[2] while the total number of FIDCs grew from 396 to 2,551. All FIDCs must have at least 50% of their NAV invested in receivables at all times. However, a large share of the growth in the number of FIDCs has come from funds that purchase distressed debt, much of which is consumer debt, including credit card receivables. This paper focuses on the 762 FIDCs that primarily purchase firms' receivables, specifically recourse factoring.[3] These FIDCs have combined NAV of 17.7 billion USD, purchasing 4.1 billion USD of firms' receivables per month.

The purpose of FIDCs is to securitize receivables for institutional investors, who wish to have exposure to short-term corporate debt for a wide cross-section of firms in a standardized asset.[4] As the debtors repay the FIDCs, the payments first go to senior shareholders, then to subordinated shareholders, with greater risk but also higher returns for subordinated shares. See Figure 2.1b for a visual depiction of the process.

The primary types of factoring are recourse, where the seller retains residual liability to the factor, and non-recourse, where only the buyer is liable. This paper focuses on recourse factoring, both because of data quality and because FIDCs primarily purchase receivables via recourse factoring.

2.3 Data and Summary Statistics

We use a novel combination of transaction-level datasets from the BCB. These datasets cover the universe of trade credit, electronic payments, and intermediated credit operations in Brazil. We measure trade credit payment terms using boletos, the standardized form of invoices in Brazil. Almost all firm-to-firm transactions use boletos, which can be settled using bank transfers, cash, and other payment rails. The supplier and buyer both observe and stamp the boleto, which the supplier's bank registers with a notary and reports to the BCB. In the boletos data, we observe identifiers for buyer and seller, date of invoice, the due date of payment, the actual date of payment, the amount due, the amount paid, and the reason for delay. The electronic payments dataset includes all payments associated with boletos, usually via bank transfer or cash, as well as all interbank transfers and instant payments. Using firm identifiers, we have the location and sector from the tax registry, and labor variables at the monthly level from the matched employer-employee dataset (RAIS).

The final component of our dataset are credit operations (SCR). SCR includes all debt financing for firms whose total debt since June 2016 exceeds 200 Brazilian reais (BRL), equal to around 40 USD at current exchange rates. In SCR, we observe numerous small firms, including firms with

[2]All monetary figures in this paper are expressed in current (September 2024) US dollars.

[3]See Figure B5 for the distribution of FIDC size.

[4]There are 251 thousand firms in Brazil whose receivables were purchased by FIDCs during our study period. By comparison, there are around 1,500 firms that issued corporate bonds, and fewer than 400 firms that are publicly traded, i.e. have easily accessible equity exposure.

annual revenue under 10 thousand USD per year and firms without full-time employees, so we are not concerned about the coverage of the 200 BRL threshold. We are the first to recognize the unique structure of the factoring data in the SCR and to correctly use the factoring data for all firms in SCR. We are also the first to construct the dataset of FIDC operations in SCR, combining direct purchases of receivables from firms with bulk purchases of receivables from other financial intermediaries. The SCR also contains transactions and snapshots for lines of credit and loans backed by accounts receivable as the main alternative forms of working capital financing.

Our final dataset consists of almost all formally registered firms in Brazil from November 2018 to March 2024. While there are over 6 million registered firms in Brazil, a large majority do not appear active in any given month, where we define activity by sending or receiving payment, or initiating any financing. There are 1.03 million firms in Brazil whose receivables are ever purchased by a FIDC. Under the definition of the instrument, this is a superset of the firms for the firm-by-month level dataset in the regressions, since firms who never have receivables purchased by FIDCs have instrument value equal to zero and are dropped from the regression via the firm fixed effects. These 1.03 million firms receive an average of $68.2 billion USD of trade credit per month, of which $6.9 billion (10%) is factored with recourse, and another $10.5 billion (15%) is financed through other means. By comparison, total issuance of all other working capital financing, comprising of credit lines, short-term loans, and short-term bonds, sums to $4.7 billion USD per month.

Table 2.1 shows summary statistics about the comprehensive contract-level data for trade credit, factoring, and other short-term financing in Brazil. There are two main takeaways from Table 2.1: trade credit is the largest form of short-term lending in Brazil, and factoring is the majority of intermediated capital financing in Brazil.

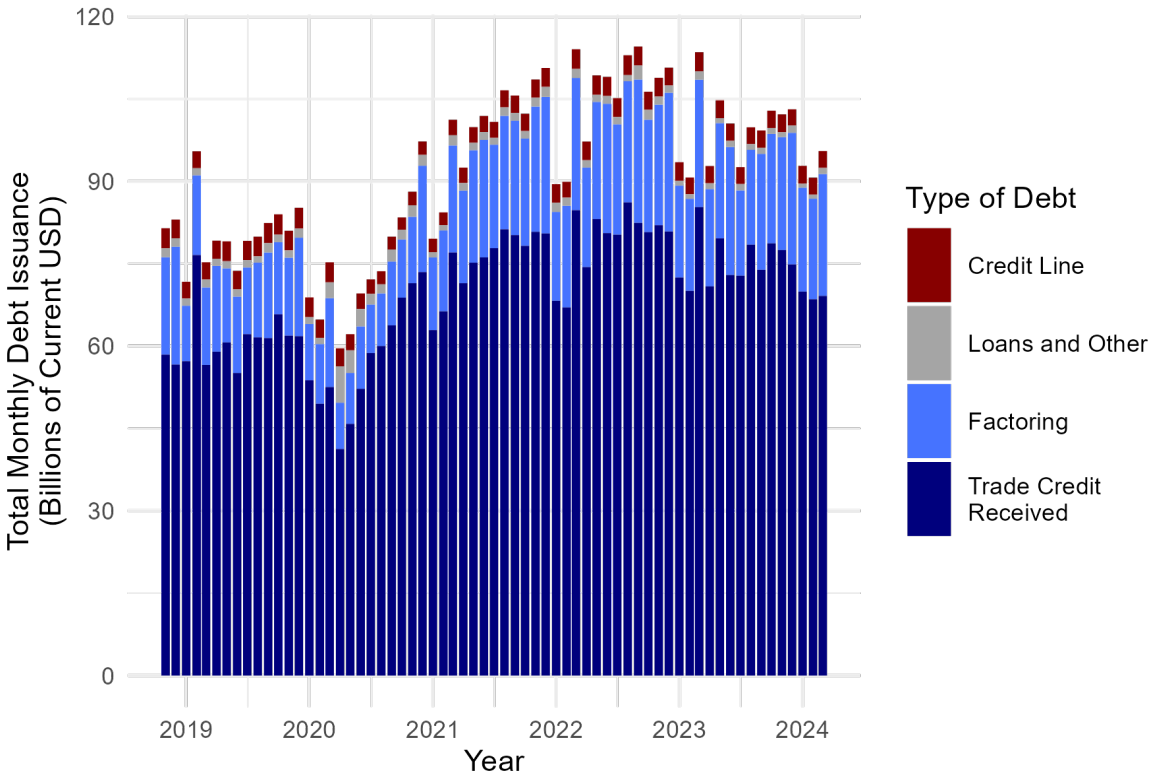
Table 2.1: Annual Means of Trade Credit, Factoring, and Other Short-Term Debt

	Mean Overall	Mean Small Firms	Mean Medium Firms	Mean Large Firms
Panel A: Trade Credit Received, by Seller				
Volume (Million USD)	1.32	0.42	6.17	44.32
Maturity (Days)	30.36	33.99	29.67	28.22
Panel B: Recourse Factoring, by Seller				
Volume (Million USD)	0.13	0.08	0.56	1.97
Interest Rate (%)	19.42	18.35	21.58	19.52
Maturity (Days)	121.19	138.49	95.77	105.06
Panel C: Non-Recourse Factoring, by Buyer				
Volume (Million USD)	0.15	0.01	0.50	9.12
Interest Rate (%)	13.04	15.54	12.50	13.07
Maturity (Days)	79.86	94.81	86.48	77.17
Panel D: Secured Credit Lines				
Volume (Million USD)	0.04	0.02	0.26	0.72
Interest Rate (%)	30.63	45.13	23.24	18.67
Maturity (Days)	66.10	57.22	72.86	69.14
Panel E: Unsecured Credit Lines				
Volume (Million USD)	0.02	0.01	0.07	0.17
Interest Rate (%)	332.07	356.74	316.49	206.32
Maturity (Days)	42.20	41.01	44.82	43.89
Panel F: Other Short-Term Debt (Maturity Under 1 year)				
Volume (Million USD)	0.03	0.00	0.13	1.20
Interest Rate (%)	7.04	22.24	10.09	6.65
Maturity (Days)	187.61	181.17	176.71	188.82

Notes: The data are from the Central Bank of Brazil There are 40,790,100 firm-by-month observations. We define the mean number of employees at the firm level, averaging across months, then classify firms as small (0 up to 50 mean employees), medium (50 up to 500 mean employees), or large (500 or more mean employees). In Panel A, the trade credit means are at the *debtor* firm level. We measure trade credit using boletos, and we measure revenue using the sum of boletos and interfirm electronic payments. In our sample, there are 1.69 billion trade credit transactions per year, an average of 2,687 per debtor per year. In Panel B, for recourse factoring, the debtor is the seller, who initiates the factoring. In our sample, there are 39.3 million recourse factoring transactions per year. In Panel C, for non-recourse factoring, the debtor is the buyer, and we do not always observe the seller who initiates the factoring. In our sample, there are 17.3 million non-recourse factoring transactions per year. In Panel D, there are 1.72 million drawdowns of secured credit lines per year. In Panel E, there are 7.64 million drawdowns of secured credit lines per year. In Panel F, there are 3.57 million other working capital financing transactions per year. For comparison of interest rates, the mean federal funds rate in Brazil (SELIC) during the sample period was 7.83%.

Figure 2.2 shows that factoring has consistently been the primary type of intermediated working capital financing over time. Note that trade credit is not intermediated. Of the total monthly working capital volume that varied from a trough of $60 billion in April 2020 to a peak of $114 billion in March 2022, 70% consisted of trade credit that was not factored, i.e. receivables that suppliers held on their balance sheets. Another 24% was trade credit that was factored, 3% was secured credit lines, and 1% was unsecured credit lines, and the remaining 2% were short-term loans and bonds.[5]

Figure 2.2: Time Series of Working Capital Financing in Brazil

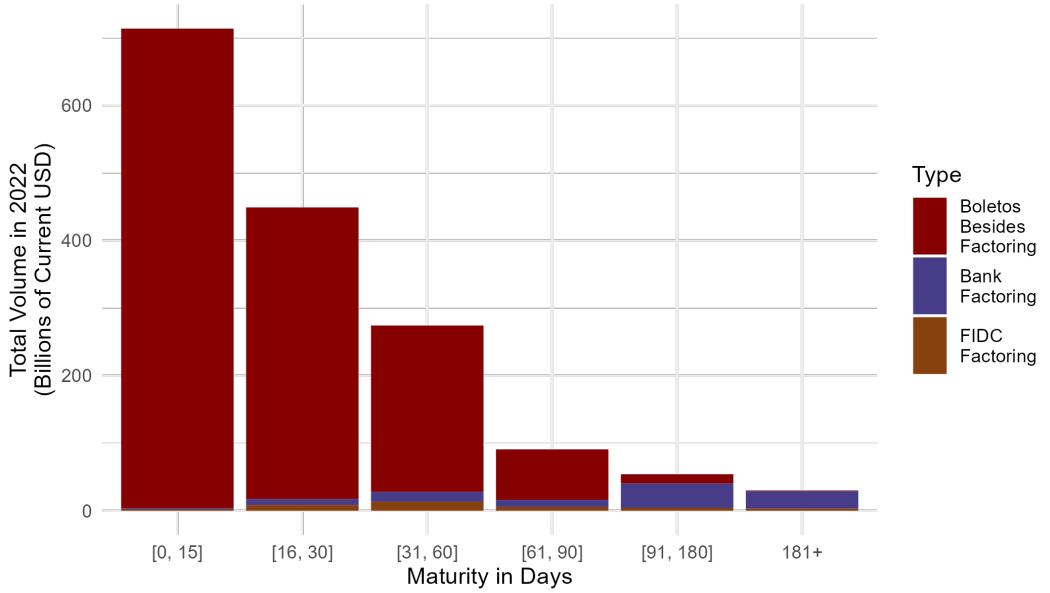


Notes: The data are from the Central Bank of Brazil. This figure shows the composition of firms' short term financing, with maturity under 1 year, among the 1 million firms in our sample. In dark blue is the value of trade credit that a firm receives from its suppliers. In light blue is factoring, the sale of receivables from the trade credit that a firm offers its customers. In red are credit lines, which generally require firms to post collateral. In gray are working capital loans and bonds.

Figure 2.3 shows that the factoring share increases over the maturity of the receivable. Most receivables have short maturity, and the seller retains most short-maturity receivables. By comparison, almost all long maturity receivables are factored.

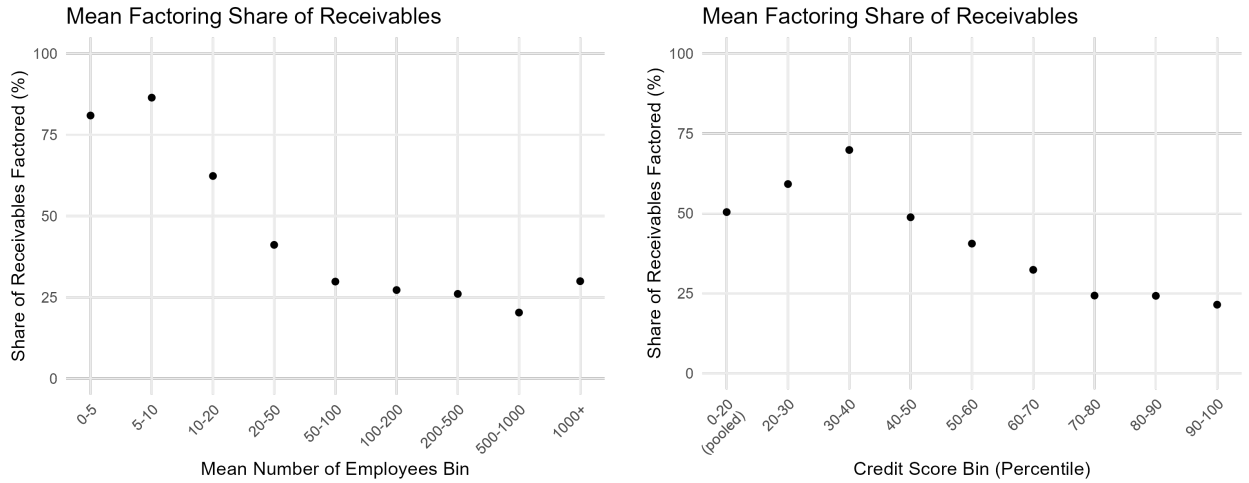
[5]Figure B1 shows that the relative importance of each type of working capital financing is invariant across the distribution of firms' credit score, although the exact magnitudes vary across the distribution.

Figure 2.3: Distribution of Maturity for Trade Credit and Factoring



Notes: The data are from the Central Bank of Brazil. This figure partitions factoring and trade credit at the contract level by the maturity bin, then sums the contract value by bin. The red bars are receivables that the sellers retain and do not factor. The purple bars are factored to banks, and the brown bars are factored to FIDCs, who largely purchase receivables in the 31 to 60 day maturity bin.

Figure 2.4: Factoring Share of Payment Inflows



(a) Factoring Share by Number of Employees (b) Factoring Share by Credit Score

Notes: The data are from the Central Bank of Brazil. The denominator is the issuance volume of all receivables, and the numerator is the volume of receivables that the sellers factor. Each subfigure partitions firms into bins along the horizontal axis. On the left, we classify firms by the mean number of employees across all months, where each bin includes the lower bound and excludes the upper bound. On the right, we classify firms by deciles of credit score in June 2023, the only month with available data. The bottom 19% of firms have a credit score of 0, generally signaling a lack of any credit history, so we pool together the bottom two deciles.

Figure 2.4 is a bin-scatter plot that shows that the share of receivables that are factored is greater for firms with few employees (left) and low credit score (right). Each dot represents the revenue-weighted share of receivables that are factored for a bin of firms, by the mean number of employees across all months on the left, and by credit score decile on the right. The credit scores are a one-time snapshot in 2023 from the largest corporate credit scoring agency in Brazil.

Receivables funds (FIDCs), who purchase and securitize receivables by offering shares at different seniorities, are one reason for the high rate of factoring in Brazil. The CVM Instruction 356 in December 2001 defined the FIDC and set common standards. The number of FIDCs grew steadily over the 2000s and 2010s. Figure B4 in the appendix shows that FIDCs now comprise over 30% of all recourse factoring, a share that has increased over time. Flows to funds explain some of the variation in factoring prices across firms and across time. FIDC flows have weak autocorrelation of 0.21 from month to month. Table 2.2 shows summary statistics on FIDCs:

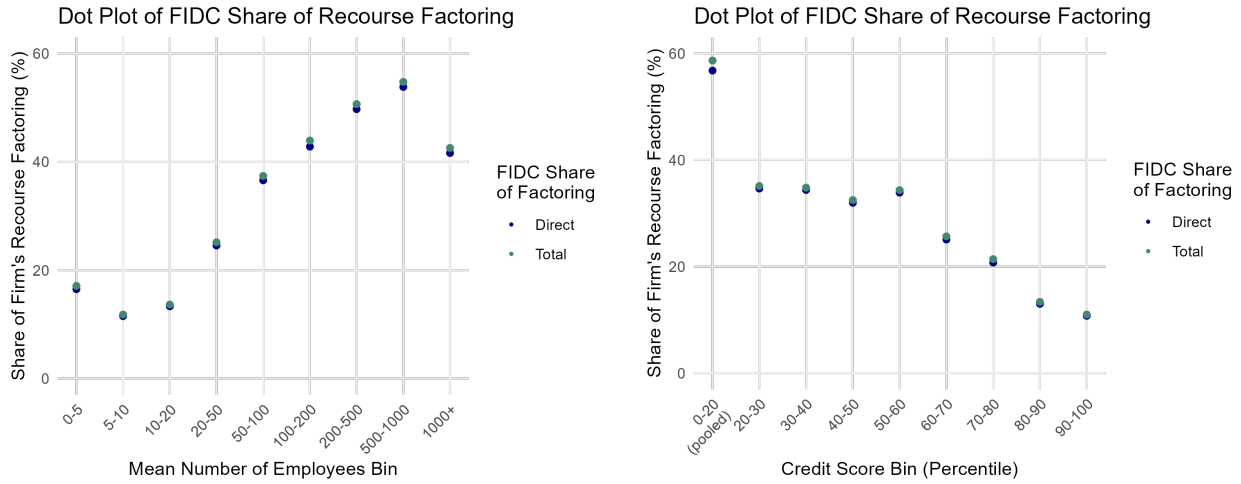
Table 2.2: Summary Statistics on FIDCs at the FIDC by Month Level

	Mean	Std. Dev.	10th Percentile	Median	90th Percentile
Net asset value	23.59	51.36	1.14	7.78	51.22
Monthly recourse factoring	4.48	11.45	0.00	0.94	11.86
Annualized net return (%)	13.83	22.66	-4.04	11.88	34.37
IR recourse factoring (%)	32.75	61.22	10.87	36.97	76.10
Monthly net flow | positive	1.27	6.28	0.00	0.12	2.53
Monthly net flow | negative	-1.09	5.58	-2.04	-0.07	-0.00

Notes: The data are from the Securities and Exchange Commission of Brazil (CVM). The net asset value, monthly recourse factoring, and flow variables are expressed in millions of USD. The interest rate (IR) is an issuance volume weighted average. The net flow is defined to be the difference between current net asset value NAV_t and net return adjusted previous month net asset value $NAV_{t-1}(1+\tilde{r})$, where $\tilde{r}$ is the net return. "Monthly net flow | positive" is the subset of FIDC by month observations with positive net flow, while "Monthly net flow | negative" is the subset of FIDC by month observations with negative net flow. The difference between the factoring interest rate and the net return is explained by funds' holdings of low yield Brazilian Treasury bills for liquidity, performance fees, and default, usually in the form of delayed payment. The net asset value, recourse factoring purchase, and net flow are reported in millions of USD.

Figure 2.5 is a bin-scatter plot that shows that FIDCs tend to purchase receivables from firms with many employees (left) yet with low credit score (right). Although larger firms tend to have higher credit scores, each decile of credit score features firms across most of the size distribution. In each plot, we categorize each firm into a bin, then compute the mean share of factored receivable volume that is purchased by FIDCs rather than banks.

Figure 2.5: FIDC Share of Factoring



(a) FIDC Share by Number of Employees

(b) FIDC Share by Credit Score

Notes: The data are from the Central Bank of Brazil. Each subfigure partitions firms into bins along the horizontal axis. On the left, we classify firms by the mean number of employees across all months, where each bin includes the lower bound and excludes the upper bound. On the right, we classify firms by deciles of credit score in June 2023, the only month with available data. The bottom 19% of firms have a credit score of 0, generally signaling a lack of any credit history, so we pool together the bottom two deciles. In both plots, the denominator is all recourse factoring volume, and the numerator is recourse factoring volume in which the factor is the FIDC. The lower dots are FIDC purchases of receivables directly from firms, while the upper dots are total FIDC purchases of receivables, including from banks.

The main takeaways about the setting for working capital financing in Brazil are that trade credit is by far the most common type of firm-to-firm borrowing in Brazil, small and low credit score firms factor a large share of trade credit, and factoring volume is much higher than other types of intermediated working capital financing. FIDCs purchase a larger share of receivables for low credit score and large firms than for other firms.

2.4 Empirical Analysis

The ideal experiment would be to randomize the price offered to each receivable, across banks and funds (asset demand) and across firms (asset supply). This would allow us to trace out the financing demand curve of factoring, and for each marginal increment along the curve, compare firms' input demand, revenues, trade credit and other firm-to-firm decisions to compute elasticities with respect to the factoring interest rate. However, such an experiment is logistically difficult, and infeasible at the scale of the Brazilian economy, so we use an instrumental variable strategy instead.

2.4.1 First Stage

The first stage regression has factoring interest rate $r^{Fac}_{j,t}$ on the left hand side and firm-level expected flow-driven "exposure" (net purchases of receivables) on the right hand side, constructed as follows: $x^{Fac}_{j\to f,t}$ is fund f's exposure to firm j's recourse factoring, either directly purchasing receivable from firm j or from a bank/fintech, in month t. $X_{f,t}$ is fund f's total purchases of assets. $F_{f,t}$ is the net inflow to fund f, based on net asset value (NAV) following the literature: $F^{NAV}_{f,t} := \frac{NAV_{f,t} - NAV_{f,t-1}R_{f,t}}{NAV_{f,t-1}}$. We then define fund by factoring type by month exposure $e^{Fac}_{j,t}$ to funds' flows is the firm's share of 3-month lagged fund receivables purchases, scaled by the flow to the fund

$$e^{Fac}_{j,t} := \sum_{f} \frac{x_{j\to f,\underline{t}}}{X_{f,\underline{t}}} F_{f,t},$$

We normalize exposure $e^{Fac}_{j,t}$ so that its units are standard deviations from the mean. Then the first stage regression is

$$r^{Fac}_{j,t} = \alpha_j + \alpha_t + \gamma_1 e^{Fac}_{j,t} + \varepsilon_{j,t}. \tag{2.1}$$

The outcome variable of the first stage is the overall interest rate (IR) on factoring, shown in column 1 of Table 2.3. The interpretation is that a one standard deviation increase in expected fund purchases of receivables, due to net fund inflows, leads to a 0.12 percentage point decrease in the firm's factoring interest rate. The first-stage F-statistic is 91.1.

Table 2.3: First Stage Regression and its Decomposition into Bank vs FIDC

	First Stage	Decomposition	
	IR Factoring Issuance (All)	IR Factoring Issuance (Funds)	IR Factoring Issuance (Banks)
	(1)	(2)	(3)
$e_{j,t}$	$-\mathbf{0.1212}^{***}$	-0.1957^{***}	-0.0530^{***}
	(0.0127)	(0.0172)	(0.0068)
Num. Obs.	4,146,540	1,734,458	2,424,888
Num. Firms	511,896	251,391	313,390
Num. Months	65	65	65

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\cdot}p < 0.1$

Notes: These regressions use data from the Central Bank of Brazil. The dataset is at the firm by month level, with firm and month fixed effects, and *standard errors are clustered at the firm level and shown in parentheses*. **The first stage coefficient is in column 1**. If a firm does not factor in a given month, then the interest rate is undefined, and the observation is dropped from the regression.

The column 1 interest rate is a value-weighted average of the interest rates on receivables purchased directly by funds (column 2) and by banks (column 3). Banks retain most of the receivables that

they purchase; banks only re-sell 1.04% of the face value of receivables to funds, usually on the same day. We interpret column 3, the bank interest rate, as an equilibrium object; under our hypothesis, banks only change the interest rate because of competition with funds or the prospect of re-selling to funds.

The exclusion restriction states that fund flows only affect firms' revenues, expenditures, and trade credit decisions through the factoring interest rate, and are not a sign of expectations of firms' creditworthiness nor higher returns conditional on fixed effects. Because we focus on FIDCs, who are mandated to hold the majority of asset value in receivables and purchase negligible amounts of other corporate debt, we do not believe that there is contamination through other corporate interest rates. FIDCs' asset values and flows are a tiny share of total fund asset values and flows,[6] so we believe that FIDC flows are not large enough to affect monetary policy.

We argue that reverse causality, arising from flows chasing firms with high expected returns, is not a concern for three reasons. First, we do not believe that investors select FIDCs based on specific exposure. There are 762 FIDCs in our sample, and there are 251 thousand non-financial firms who sell receivables to FIDCs, with the majority of firms selling receivables to multiple FIDCs, so most FIDCs purchase receivables from thousands of firms. FIDCs do not report the identities of the firms, so investors cannot select FIDCs based on exposure to specific firms. We also believe that sectoral selection is not strong, as FIDCs tend to be diversified (see Figure B6). We show in Section 2.7.2 that our results are robust to inclusion of sector-time and location-time fixed effects that capture many of the potential sources of fundamental shocks that could threaten identification. Of the 762 FIDCs, 651 of them purchase receivables from multiple high-level sectors,[7] and 430 have at least 20% AUM share in multiple sectors. Second, we believe that flows to FIDC are largely driven by institutional investors' portfolio allocation constraints across broad asset classes, like receivables versus equity and fixed income. Third, we show in Table 2.9 that firms' characteristics are balanced across firm-by-month observations with positive flows, negative flows, and flows near zero.

2.4.2 IV Regression

The structural regressions estimate the contemporaneous impact of the recourse factoring interest rate, fitted on the FIDC flows, on a variety of response variables:

$$y_{j,t} = \alpha_j + \alpha_t + \beta_1 r^{Fac}_{j,t} + \varepsilon_{j,t} \tag{2.2}$$

The main response variables are shown in Table 2.4, which shows that a one percentage point increase in the factoring interest rate causes a large decrease in firms' contemporaneous revenue of 6.1 log points from column 1, a moderate decrease in intermediate input expenditure of 3.6 log points in column 2, and small decrease in expenditure on labor of 0.56 log points in column 3.[8]

[6] In March 2024, there are 18 billion USD of assets under management for FIDCs that primarily purchase firms' receivables, the focus of this paper. In comparison, there are 129 billion USD of assets under management for fixed income funds, and the total stock market capitalization is 972 billion USD.

[7] We define the high-level sectors to be manufacturing, retail, wholesale, transport, professional services, and other sectors.

[8] We proxy for revenue and intermediate input expenditure using all boleto transactions, including

We decompose the decrease in labor expenditure into a decrease in labor demand for permanent employees and an increase in labor demand for temporary employees, with small increases in wages that likely reflect composition effects. Column 4 shows that the number of hours worked by permanent employees decreases by 1.1 log points, while column 5 shows that the number of hours worked by temporary employees increases by 2.1 log points. The average number of permanent employees per firm is 46.6, while the average number of temporary employees is 4.3; this difference explains why employment and the wage bill decrease despite a larger percentage change in temporary labor than permanent labor. The mean number of weekly working hours for permanent employees is 41.8 and for temporary employees is 37.4, including overtime work, so the employment results vary little when measured by number of employees instead of hours. See Table 2.6 and Table 2.10 for the decomposition of the impact on labor market variables.

Table 2.4: IV Regressions of the Main Outcomes on the Factoring Interest Rate

	(1)	(2)	(3)	(4)	(5)
	Log Revenue	Log Expenditure	Log Wage Bill	Labor Demand (Log Hours, Permanent)	Labor Demand (Log Hours, Temporary)
$r^{Fac}_{j,t}$	-0.0614^{***}	-0.0357^{***}	-0.0056^{*}	-0.0110^{***}	0.0212^{***}
	(0.0093)	(0.0056)	(0.0023)	(0.0023)	(0.0064)
Num. Obs.	2,668,026	4,076,721	2,543,940	2,545,958	607,036
Num. Firms	217,956	476,418	287,108	287,201	93,156
Num. Months	65	65	50	50	50

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\cdot}p < 0.1$

Notes: All regressions use data from the Central Bank of Brazil. All regressions use firm and month fixed effects, with standard errors clustered at the firm level in parentheses. The predictor variable is the firm-level interest rate on factoring in percentage points, instrumented by the expected change in receivables purchases driven by fund flows. The response variables are the log revenue proxied by payment inflows, log intermediate input expenditure proxied by payment outflows to firms, log wage bill, log labor demand for permanent workers, and log labor demand for temporary workers. There are only 50 months of data for labor variables because the labor data has only been published through December 2022.

transactions where the counterparty is a consumer. Since boleto transactions can be settled by cash, bank transfer, PIX, or other means, we believe that our proxy is representative of firms' real outcomes.

Table 2.5: IV Regressions of Trade Credit Outcomes on the Factoring Interest Rate

	(1)	(2)	(4)	(5)
	Maturity Offer (Days)	Percentage Offer (%)	Maturity Receive (Days)	Percentage Receive (%)
$r_{j,t}^{Fac}$	−0.0354	−0.3885**	0.6737***	0.0272
	(0.0996)	(0.1274)	(0.1358)	(0.0379)
Num. Obs.	4,146,540	4,146,540	4,146,540	4,146,540
Num. Firms	511,896	511,896	511,896	511,896
Num. Months	65	65	65	65

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\cdot}p < 0.1$

Notes: All regressions use data from the Central Bank of Brazil. All regressions use firm and month fixed effects, with standard errors clustered at the firm level. The predictor variable is the firm-level interest rate on factoring in percentage points. The instrumental variable is the expected change in receivables purchases driven by fund flows. The response variables are the firm by month level mean maturity of trade credit, offered and received, as well as the share of receivables with at least 15 days maturity, the effective lower bound for factoring.

Table 2.5 shows that there are small spillovers through the firm-to-firm trade credit network from the change in the factoring interest rate, which we later show is the shadow cost of trade credit, to trade credit terms that firms offer and receive. Column 1 shows that firms that face a one percentage point higher factoring interest rate do not change the maturity of the trade credit that they extend, but column 2 shows that affected firms are -0.39 percentage points less likely to offer any trade credit on the extensive margin.[9] Column 3 shows that affected firms receive slightly longer trade credit terms by 0.67 days, compared to the baseline mean of 22.8 days. Column 4 shows that the proportion that receive trade credit is a precisely estimated zero.[10]

Labor Outcomes

Now we decompose the wage bill coefficient of 0.56% from Table 2.4 into the hourly wage and hours worked in Table 2.6. Column 1 of Table 2.6 shows that the hourly wage rises slightly, which we interpret as firms choosing a higher marginal revenue product of labor in response to the higher marginal revenue product of capital (which equals the composite interest rate in an efficient equilibrium). Columns 2 and 3 show the total reduction in hours worked by each type of employee, with a larger decrease of 1.4% for new hires in Column 2 than the 0.6% decrease for existing employees in Column 3. Table 2.10 in Section 2.7.1 shows that the results from Table 2.4 and Table 2.6 are similar when using the number of employees rather than the total number of hours worked.

[9] This leads affected firms to offer 5.7% less contemporaneous trade credit, compared to their 6.0% decrease in revenue from Table 2.4.

[10] The reduction in trade credit receipt of 3.1% is similar to the 3.6% reduction in expenditure.

Table 2.6: IV Regressions of Hours Employed Outcomes on Factoring Interest Rate

	(1)	(2)	(3)
	Log Wage (Hourly)	Log Employment (Hours Worked by New Hires)	Log Employment (Hours Worked by Existing Employees)
$r^{Fac}_{j,t}$	0.0037$^{\cdot}$	−0.0135**	−0.0057**
	(0.0020)	(0.0045)	(0.0020)
Num. Obs.	2,543,608	1,124,594	2,526,986
Num. Firms	287,082	183,930	284,845
Num. Months	50	50	50

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\cdot}p < 0.1$

Notes: All regressions use data from the Central Bank of Brazil. All regressions use firm and month fixed effects, with standard errors clustered at the firm level. The predictor variable is the firm-level interest rate on factoring in percentage points. The instrumental variable is the expected change in receivables purchases driven by fund flows. The response variables come from restricted access month-level RAIS data. An employee is defined as new if the employee began working at the firm that month.

Altogether, we explain the labor impacts of the factoring interest rate using a cash flow mismatch story. Firms' sales are volatile month to month, and most firms have limited pricing power, so revenue is both volatile and not perfectly forecasted. Factoring allows firms to smooth their cash inflows; in months where firms have less than typical revenue, they factor more. On the other hand, labor laws impose constraints on cash outflows. For the majority of firms, labor is the largest expense, and firms must commit in advance to pay permanent employees each month an amount that varies little from month to month. However, firms can adjust total labor expenditure on the margin through hiring more temporary employees and fewer permanent employees. When the factoring interest rate is high, meaning that it is expensive to smooth cash inflows, firms use the labor margin of adjustment to match cash outflows to cash inflows. We codify this explanation in the model in Section 2.5

Financial Outcomes

Table 2.7 shows that factoring volume is highly responsive to the factoring interest rate, with little substitution to quantities of other types of financing. Note that the response variable in column 3, the logarithm of actual factoring issuance at the firm level, differs conceptually from the instrumental variable, the expected change in firm-level factoring relative to a baseline of zero FIDC-level flows, based on FIDC-level flows and past factoring. Table 2.11 in the appendix shows that the interest rate on unsecured credit lines responds to the change in the factoring price, primarily through banks' factoring rates. Unsecured credit lines have a high baseline mean interest rate of 333% and high variance across firms, with standard deviation of 85%. See Tables 2.11 and 2.12 in the appendix for additional results on financial outcomes.

Table 2.7: IV Regressions of Debt Issuance Outcomes on Factoring Interest Rate

	(1)	(2)	(3)	(4)	(5)
	Log Debt Issuance (Debt Under 1 Year)	Log Debt Issuance (Debt Over 1 Year)	Log Debt Issuance Factoring All Issuance	Log Debt Issuance Credit Line (Unsecured)	Log Debt Issuance Credit Line (Secured)
$r^{Fac}_{j,t}$	−0.1627***	0.0163	−0.1692***	0.0167	−0.0319
	(0.0174)	(0.0259)	(0.0180)	(0.0170)	(0.0654)
Num. Obs.	4,146,540	508,179	4,146,540	829,816	410,208
Num. Firms	511,896	130,522	511,896	123,370	57,997
Num. Months	65	65	65	65	65

$^{***}p < 0.001; {}^{**}p < 0.01; {}^{*}p < 0.05; {}^{\cdot}p < 0.1$

Notes: All regressions use data from the Central Bank of Brazil. All regressions use firm and month fixed effects, with standard errors clustered at the firm level. The predictor variable is the firm-level interest rate on factoring in percentage points. The instrumental variable is the expected change in receivables purchases driven by fund flows. The response variables are log debt issuance by category of debt. Column 1 is the subset with maturity of up to 365 days. Column 5 is the subset with maturity of over 365 days. Column 3 is factoring. Column 4 and 5 are unsecured and secured credit lines, respectively, where issuance is defined as any drawdown of the credit line, not a change in the credit limit. Across all firms in Brazil, loans are the highest issuance form of long-term debt, and bonds are second highest.

2.4.3 Dynamic Effects

In this section, we use local projections, following Jordà (2005) and Plagborg-Møller and Wolf (2021), to estimate the dynamic effects of FIDC flows on the factoring interest rate and firms' outcomes.

Local Projection First Stage

Funds ought to allocate the majority of their asset holdings to receivables. Given the short maturity of receivables, we believe that funds quickly purchase receivables, with short-term transmission to the interest rate that decays over time. The following local projection generalizes the first stage regression (2.1) over horizons $h \geqslant 0$:

$$r^{Fac}_{j,t+h} = \alpha_{j,h} + \alpha_{t,h} + \beta_h e_{j,t} + \varepsilon_{j,t+h}. \tag{2.3}$$

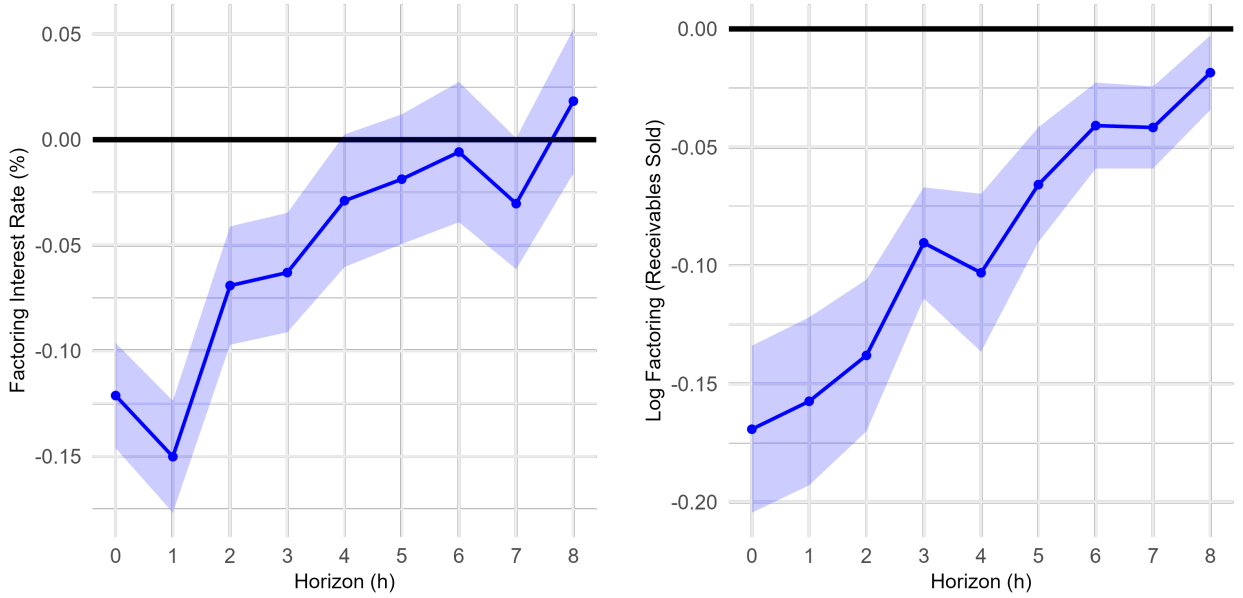
Similarly, to see how the timing of firms' outcomes y change with respect to fund flows, we run the panel data IV local projection (IV-LP):

$$y_{j,t+h} = \alpha_{j,h} + \alpha_{t,h} + \beta_h r^{Fac}_{j,t} + \varepsilon_{j,t+h}, \tag{2.4}$$

Figure 2.6 shows the coefficients β_h over the horizons from 0 to 8 months, with the first stage (2.3) on the left in Figure 2.6a, and the factoring volume outcome from the IV-LP (2.4) on the

right in Figure 2.6b. The interpretation of Figure 2.6a is that one standard deviation exposure to funds' net inflows results in a decrease in the interest rate of 12 basis points in the current month, corresponding to column 1 of Table 2.3, 15 basis points in the next month, seven basis points in the subsequent month, and then decays to zero by the fourth month after the shock. The interpretation of Figure 2.6b is that a fund flow causing a one percentage point contemporaneous increase in the interest rate, which would be an eight standard deviation exposure to funds' net outflows, causes a 17 log point decrease in current factoring volume, corresponding to column 3 of Table 2.7, and only gradually decays to 2 log points by the eighth month after the shock.

Figure 2.6: Local Projection of Factoring Interest Rate and IV-LP of Factoring Volume

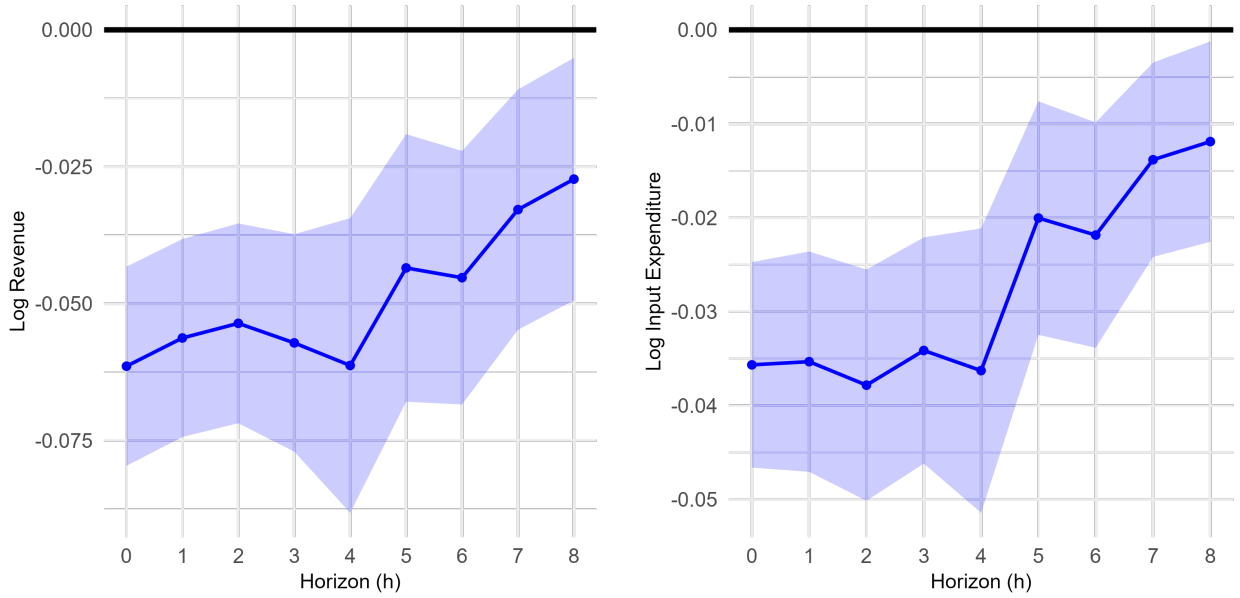


(a) Factoring Interest Rate

(b) Factoring Volume

Notes: The data are from the Central Bank of Brazil. Figure 2.6a on the left corresponds to the local projection in equation (2.3) with the factoring interest rate as the outcome variable. Figure 2.6a on the left corresponds to the panel IV local projection in equation (2.4) with log factoring issuance volume as the outcome variable.

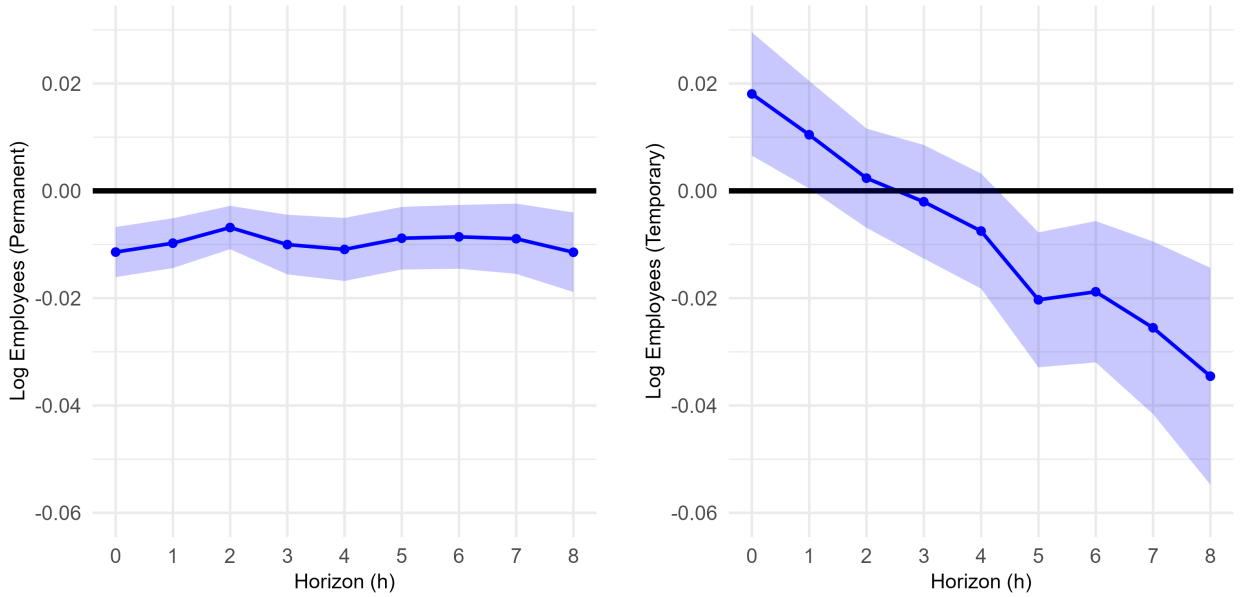
Figure 2.7: IV-LP of Revenue and Input Expenditure



(a) Log Revenue (b) Log Input Expenditure

Notes: The data are from the Central Bank of Brazil. Both plots correspond to the panel IV local projection in equation (2.4); on the left is log revenue as measured through boleto transactions (seller), and on the right is log input expenditure as measured through boleto transactions (buyer).

Figure 2.8: IV-LP of Permanent and Temporary Contract Labor Demand



(a) Log Permanent Labor Demand (b) Log Temporary Labor Demand

Notes: The data are from the Central Bank of Brazil. Both plots correspond to the panel IV local projection in equation (2.4); on the left is log permanent labor demand, and on the right is log temporary labor demand, both measured in log number of employees.

Figure 2.7 shows that the impacts on revenue and input expenditure persist with only gradual decay. Our explanation for the long lasting effects on revenue and expenditure is in Figure 2.8a, which shows that the change in permanent labor demand, relative to baseline, persists at the same level as the contemporaneous effect. Our explanation is that the adjustment is on the dimension of hiring; once a firm adjusts its hiring decision, the firm must commit to it because of the high cost of firing a permanent employee. The benefit is that permanent contract employees gain more job-specific human capital, and we find that wage growth is faster for permanent employees than temporary employees, demeaned on firm and employee fixed effects. Also of interest is the reversal of temporary labor demand in Figure 2.8b. In the short run, temporary labor demand moves in the opposite direction of permanent labor demand because of the cash flow volatility mechanism, where labor flexibility is an imperfect substitute for factoring in matching cash outflows to cash inflows. In the long run, the firm does not need the liquidity from labor flexibility, and instead chooses a level of temporary employment to match the marginal revenue productivity to that of other inputs.

2.4.4 Heterogeneity

In this section, we show two categories of regressions that illustrate the heterogeneity across firms in the impacts of the cost of factoring. The first category is an interaction of firm type with the factoring interest rate. The second category is quantile regression.

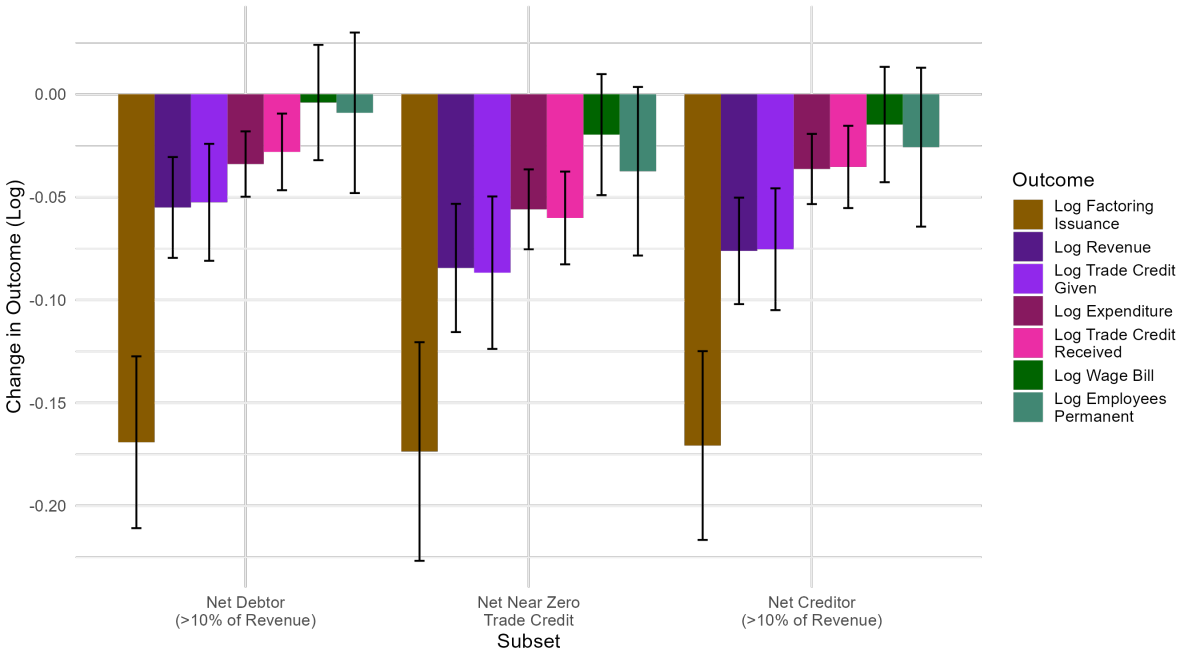
Heterogeneity by Firm Type

Consider regressions with interactions of the factoring interest rate $r_{j,t}^{Fac}$ with bins of firm heterogeneity γ_j:

$$y_{j,t} = \alpha_j + \alpha_t + \beta_1 r_{j,t}^{Fac} \gamma_j + \varepsilon_{j,t} \tag{2.5}$$

We classify each firm as a net creditor, net debtor, or neither for trade credit by taking the difference of total trade credit extended and total trade credit received over the sample period, and dividing by the firm's total revenue. If the ratio is greater than 0.1, then we consider firm to be a net creditor. If the ratio is less than -0.1, then we consider the firm a net debtor. We expect net creditors to have larger responses to the factoring price because they factor more receivables, so they receive more cash on hand for inframarginal factoring, and they may be more sensitive to factoring the marginal receivable, akin to the shadow price of trade credit. Figure 2.9 shows that net debtors indeed have smaller responses to the factoring price, with statistically significant differences for revenue, input expenditure, trade credit received, and trade credit extended, but not for labor outcomes. While the error bars overlap, the t-stats on the differences range from 2 to 3.

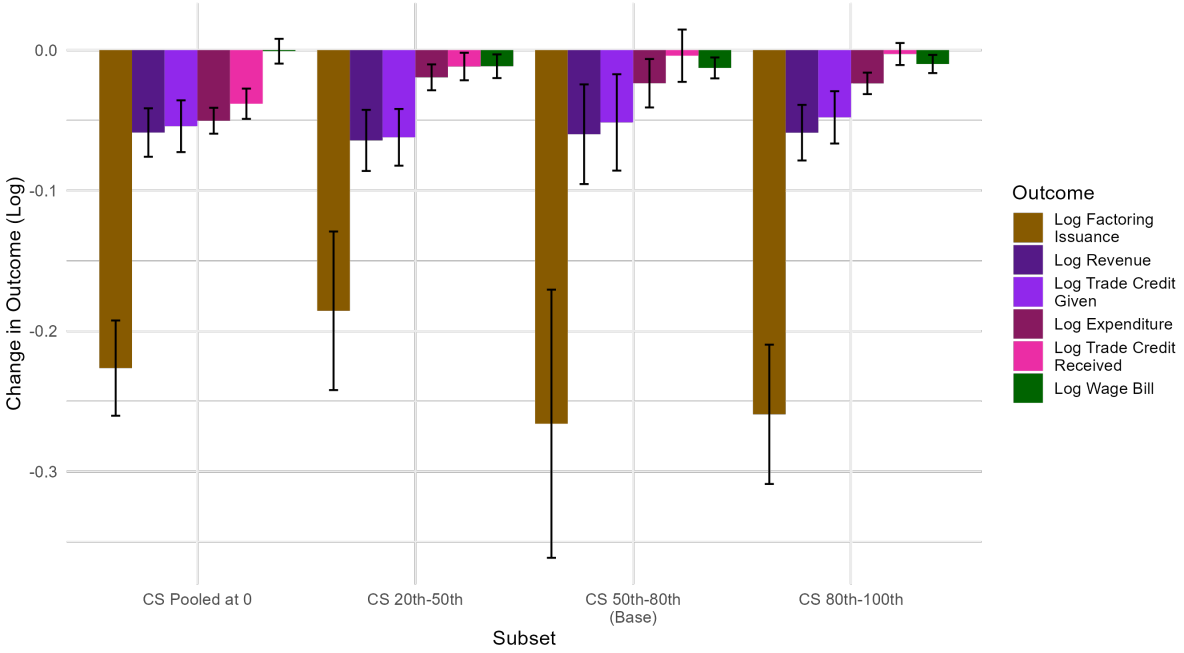
Figure 2.9: Heterogeneous Effects of the Factoring Interest Rate by Net Trade Credit



Notes: The data are from the Central Bank of Brazil. Each color corresponds to the heterogeneity interaction regression (2.5) with a different outcome variable, with bins defined by whether the firm is a net trade creditor with average net lending exceeding 10% of revenue, a net trade debtor with average net borrowing exceeding 10% of revenue, or neither. The error bars show 95% confidence intervals.

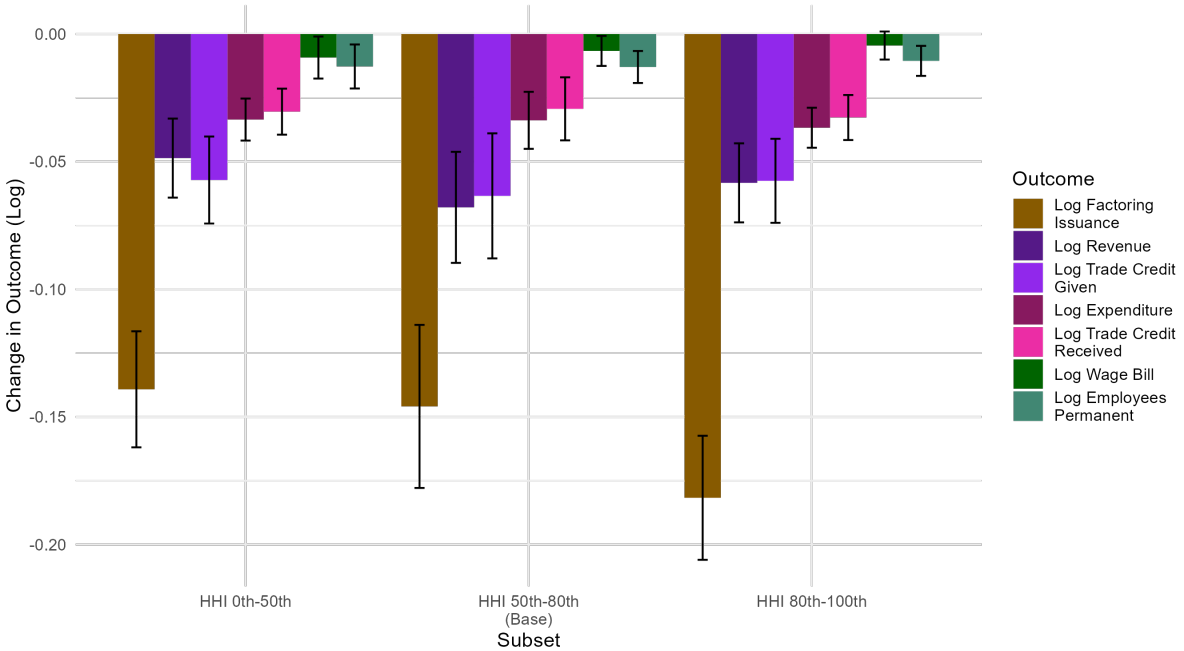
Figure 2.10 shows that the main results in Table 2.4 are true across the distribution of firms, not specific to firms with low credit scores that are more financially constrained. The base category is firms with credit scores between the 50th and 80th percentiles. Firms with the lowest credit scores, pooled at the minimum score, decrease intermediate input expenditure more and do not reduce expenditure on labor. Otherwise, the interactions of credit score bins with the factoring interest rate are similar across bins, despite the greater financial constraints faced by low credit score firms. Likewise, Figure 2.11 shows that there is minimal heterogeneity by the Herfindahl-Hirschman Index (HHI) of the firm's sector defined at the 7-digit CNAE level, which roughly corresponds to 6-digit HS code. HHI proxies for market concentration, and Dass, Kale, and Nanda (2015), Fabbri and Klapper (2016), and Giannetti, Serrano-Velarde, and Tarantino (2021) suggest that trade credit varies with firms' bargaining power and market power, but this does not pass through to factoring.

Figure 2.10: Heterogeneous Effects of the Factoring Interest Rate by Credit Score



Notes: The data are from the Central Bank of Brazil. Each color corresponds to the heterogeneity interaction regression (2.5) with a different outcome variable, with bins defined by the quantile of the firm's credit score in June 2023 from the main corporate credit bureau in Brazil. The error bars show 95% confidence intervals.

Figure 2.11: Heterogeneous Effects of the Factoring Interest Rate by HHI



Notes: The data are from the Central Bank of Brazil. Each color corresponds to the heterogeneity interaction regression (2.5) with a different outcome variable, with bins defined by the tercile of the firm's HHI. The error bars show 95% confidence intervals.

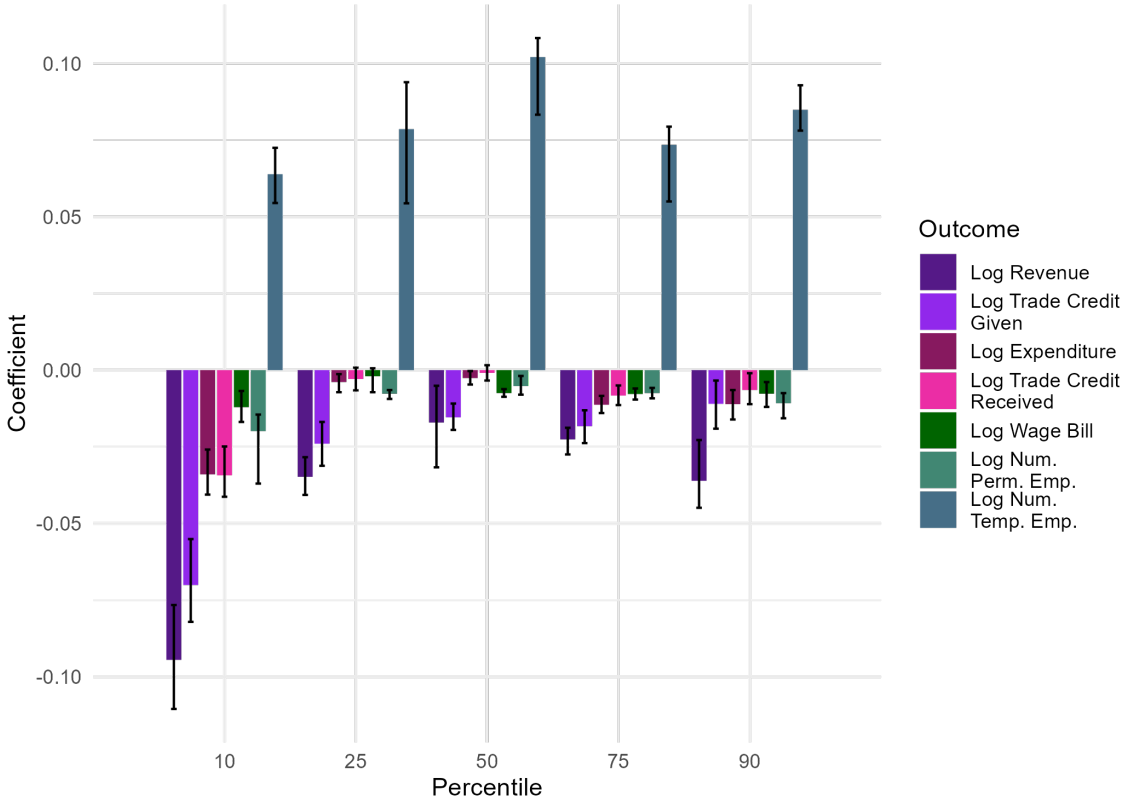
Heterogeneity by Quantile

We also estimate quantile regressions for each of the main outcomes, to assess how firms at different points in the outcome distribution respond to the factoring interest rate. For each quantile τ and outcome y, we run the IV quantile regression based on Canay (2011):

$$Q_{\tilde{y}_{j,t}}(\tau \mid \widetilde{\hat{r}}^{Fac}_{j,t}) = \beta(\tau)\widetilde{\hat{r}}^{Fac}_{j,t} + \varepsilon_{j,t}(\tau), \tag{2.6}$$

where the outcome $\tilde{y}$ and the factoring interest rate $\widetilde{r}$ are de-meaned on firm and month fixed effects, and the factoring interest rate $\widetilde{\hat{r}}$ is fitted from the first stage (2.1).

Figure 2.12: Quantile Treatment Effects of the Factoring Interest Rate



Notes: The data are from the Central Bank of Brazil. Each bar represents a quantile regression using the fund flow instrument corresponding to equation (2.6). Each color represents a different outcome variable, while each set of bars is a given quantile: the 5th, 25th, 50th, 75th, and 95th percentiles. Standard errors are calculated using rank inversion.

The interpretation of each coefficient in Figure 2.12 is the IV treatment effect at the given quantile of the distribution of the outcome. Note that the simple average of the quantile treatment effects does not reconstruct the average treatment effect in the earlier tables due to differences in methodology with fixed effects. Rather, the quantile regression coefficients are useful for comparison across quantiles. The dark purple bar represents log revenue. The impact of factoring interest rate on revenue is largest for the firms with least revenue. The same pattern holds for trade credit offered,

in light purple, as well as expenditure, in maroon, and trade credit received, in pink. The impact on wage bill is a bit larger for the smallest firms, and otherwise close to zero. This is a weighted average of a decrease in permanent employees across the entire distribution, and an increase in temporary employees.

In conclusion, the heterogeneity results by revenue, trade credit, and expenditure suggest that credit constraints may amplify the effects for the smallest firms, while the labor results suggest that the cash flow volatility motivation of factoring is important for firms across the distribution of size and creditworthiness.

2.5 Model and Counterfactual

2.5.1 Conceptual Overview

Our regression results show that factoring volume is the most responsive to the interest rate; revenue is highly responsive, more so for small firms that factor a lot; intermediate input purchases are responsive but not as much as revenue; labor demand decreases, with a short-run increase in temporary workers and a persistent decrease in permanent workers, mainly through reduced hiring. We interpret the results as micro-elasticities: micro both in the sense of at the firm level, with firm and month fixed effects, and in the sense of temporary changes, idiosyncratic at the firm level.

The purpose of the model is to rationalize the micro-elasticities in the empirical results, and also estimate "macro-elasticities:" how do aggregate output and factoring volume respond to the factoring spread. The only other empirical factoring papers, Bottazzi et al (2023) and Amberg et al (2023), estimate the total impact of introducing factoring to firms, rather than elasticities using fine variation in interest rates. However, factoring already exists in many countries, and the main challenge for policymakers is reducing the underlying frictions that keep factoring expensive. These frictions include verification costs that a receivable has not been double-pledged, screening costs of the creditworthiness of the buyer (in addition to the seller as typical with financing), and search costs for a firm to find the factor offering the best price.

2.5.2 Model Setup

There is a unit continuum of producer firms with identical baseline productivity who produce differentiated goods. There are two time periods, morning (period 0) and afternoon (period 1). Firms produce in both the morning and afternoon. Firm j produces its good with a Cobb-Douglas production function over labor ℓ and intermediate inputs x with constant labor share α:

$$y_{jt} = \ell_{jt}^{\alpha} x_{jt}^{1-\alpha}.$$

Firms sell to a representative aggregator firm who bundles the differentiated goods into a final good with elasticity of substitution $s > 1$:

$$Y_t = \left(\int_{j=0}^{1} y_{jt}^{\frac{s-1}{s}} \right)^{\frac{s}{s-1}}, \quad P = \left(\int_{j=0}^{1} p_j^{-(s-1)} \right)^{-\frac{1}{s-1}} \equiv 1.$$

The aggregator firm sells the final good to producer firms and to households. In the morning, producer firms offer trade credit on all sales, meaning that producer firms pay in the afternoon for intermediate inputs used in the morning.[11] All afternoon sales are paid upfront. Firms must pay their employees at the end of each period.

Firms can hire two types of labor: permanent labor ℓ^P, whose wage and quantity must be the same in the morning and afternoon, and temporary labor ℓ_t^T, which the firm can freely adjust. In the morning, permanent labor has the same relative productivity $\psi_0 = 1$ as temporary labor, while in the afternoon, permanent labor is more productive: $\psi_1 = \psi > 1$. Normalize the price of the final good to 1 in each period, so the real wages are w^P for permanent and w_t^T for temporary labor. There is constant elasticity of substitution $\sigma > 1$ between permanent and temporary labor, with share ω on permanent labor:

$$\ell_{jt} = \left(\omega \left(\psi_t \ell_j^P \right)^{\frac{\sigma-1}{\sigma}} + (1-\omega) \left(\ell_{j1}^T \right)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}. \tag{2.7}$$

Firm j chooses ℓ_j^P at the beginning of period 0 and chooses ℓ_{jt}^T at the beginning of period t.

The only dimensions of firm heterogeneity are the realization and distribution of the liquidity shock. Let $\epsilon_j \in [0,1]$ be the share of receivables y_{j0} promised in the morning that fail to materialize in the afternoon. Let $\zeta_j = \mathbb{E}_0 \epsilon_j$ be its mean. Let G_ζ denote the CDF of ζ_j and let $G_\epsilon(\epsilon \mid \zeta_j)$ denote the conditional CDF of ϵ. Heterogeneity in ζ_j represents the ex ante differences across firms in their buyers' creditworthiness, due to differences in sectoral volatility and firm-to-firm matching, without the complication of explicitly modeling the firm network. Firms observe the shock before choosing temporary labor and inputs in the afternoon, but must continue to pay permanent employees the contracted wage. The proceeds of the liquidity shock are rebated lump sum to consumers in the second period. Due to the timing of the liquidity shock, firms behave as if they were choosing their morning and afternoon production allocations in the morning, together with their financing decisions.

In the baseline model, the only type of financing available to the firm is factoring. The firm borrows at the beginning of the morning and repays at the end of the afternoon. Firm j borrows B_j^F, up to the face value of the morning accounts receivable $p_j y_{j0}$ discounted by gross interest rate R_j^F. Because firms begin with zero cash on hand, firms must factor at least the morning wage bill $w^P \ell_j^P + w_0^T \ell_{j0}^T$. Factoring services are imperfectly competitive with constant spread (markup)

[11] We observe a high trade credit share across the firm distribution. To micro-found always offering trade credit, assume that there are two quality levels of the final good. Low quality is worthless. Firms incur a small cost to produce high quality with certainty. Quality is verifiable by the customer and any third party in the afternoon, and the customer can refuse to pay if she discovers the quality to be low.

$\mu^F \geqslant 1$. The main counterfactual for the model is how outcomes change when the factoring spread μ^F decreases. Then the factoring interest rate is based on the share of inflows ζ_j that are shocked:

$$R_j^F = \frac{\beta^{-1}\mu^F}{1-\zeta_j}.$$

Firms begin with zero cash on hand. Firms do not earn a return on cash, but firms can retain cash between the morning and afternoon. Firms cannot default to suppliers in the afternoon, nor to labor in either period, because payments are made upfront.

The producer firm's objective in the morning is to maximize expected profits at the end of the afternoon, by choosing intermediate inputs x_{jt}, permanent labor ℓ_j^P, temporary labor ℓ_{jt}^T, and factoring B_j^F, taking as given wages $\{w^P, w_t^T\}$, factoring interest rate R_j^F, shock e_j, and model parameters. The firm faces a cost of default η, applied to negative profits, which occur when the firm does not have enough cash in the afternoon to repay suppliers with whom it contracted in the morning.

$$\max_{\{y_{jt}, x_{jt}, \ell_j^P, \ell_{jt}^T, B_j^F\}} \pi_j := \beta \mathbb{E}_0 \left[\pi_{j1} + \eta \pi_{j1} \mathbb{1}\{\pi_{j1} < 0\}\right] + m_{j0}, \tag{2.8}$$

$$\text{s.t. } B_j^F \leqslant \frac{p_j y_{j0}}{R_j^F}, \tag{2.9}$$

$$0 \leqslant m_{j0} \equiv B_j^F - \ell_j^P w^P - \ell_{j0}^T w_0^T, \tag{2.10}$$

$$\pi_{j1} := p_j y_{j1} - \ell_j^P w^P - \ell_{j1}^T w_1^T - P x_{j1} + \tilde{m}_{j1},$$

$$\tilde{m}_{j1} = (1-\epsilon_j)\left(p_j y_{j0} - R_j^F B_j^F\right) - P x_{j0}.$$

The key feature of factoring is that the upper bound $\frac{p_j y_{j0}}{R_j^F}$ in (2.9) is inherently endogenous to the firm's output choice y_{j0}. The lower bound $\ell_j^P w^P + \ell_{j0}^T w_0^T$ in (2.10) is also endogenous to the firm's decision.

The producer firm's objective in the afternoon is to maximize marginal profits, taking as given the choices made in the morning: $\{y_{j0}, x_{j0}, \ell_j^P, \ell_{j0}^T, p_j, B_j^F\}$ and the realization of the shock ϵ_j. Since there is no residual uncertainty, the firm's objective is deterministic:

$$\begin{aligned}
&\max_{\{x_{j1}, \ell_{j1}^T\}} \pi_{j1} := p_j y_{j1} - \ell_j^P w^P - \ell_{j1}^T w_1^T - P x_{j1} + \tilde{m}_{j1}, \\
&\quad \text{s.t. } y_{j1} = \tilde{\ell}_{j1}^{\alpha} x_{j1}^{1-\alpha}, \\
&\qquad \tilde{\ell}_{j1} = \left(\omega\left(\psi \ell_j^P\right)^{\frac{\sigma-1}{\sigma}} + (1-\omega)\left(\ell_{j1}^T\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}, \\
&\quad \ell_{j1}^T, x_{j1} \geqslant 0.
\end{aligned}$$

The aggregator firm's objective in each period is standard: choose purchases y_{jt} to minimize

expenditure $\int_0^1 p_j y_{jt} dj$ subject to $Y_t = \left(\int_{j=0}^1 y_{jt}^{\frac{s-1}{s}} \right)^{\frac{s}{s-1}}$. The first-order condition implies

$$\frac{p_j}{P} = \left(\frac{y_{jt}}{Y_t} \right)^{-\frac{1}{s}}. \tag{2.11}$$

There is a representative household. The household's utility is logarithmic over consumption.[12] The household has exponential disutility $\xi > 1$ from labor in each period, i.e. the household prefers to supply similar labor in the morning and afternoon. The household has relative preference ν for permanent versus temporary labor.[13]

$$u_t(c_t, \ell_t^T) = \log(c_t) - \sum_{t=0}^{1} \left[\frac{1}{\xi} \left(\ell^P + \ell_t^T \right)^{\xi} - \nu(\ell^P - \ell_t^T) \right].$$

The household receives its pay in each period, owns the financiers who lend to the firms, and pays for its consumption in the afternoon. Because of the timing of its income and expenditure, the household never demands to borrow. The household begins with zero cash. The household's optimization problem is to choose ℓ^P and ℓ_t^T to maximize discounted utility, given real wages w^P for permanent and w_t^T for temporary labor, subject to its budget constraint.

$$\begin{aligned} &\max_{\{c_0, c_1, \ell^P, \ell_0^T, \ell_1^T\}} \log(c_0) + \beta \log(c_1) - \sum_{t=0}^{1} \left[\left(\ell^P + \ell_t^T \right)^{\xi} + \nu \left(\ell^P - \ell_t^T \right) \right], \\ &\text{s.t. } c_0 + c_1 = 2\ell^P w^P + \sum_t \ell_t^T w_t^T. \end{aligned} \tag{2.12}$$

Equilibrium

Given model parameters, firms optimize (2.19), households optimize (2.18), and markets clear in each period:

$$Y_t = \left(\int_{j=0}^1 y_{jt}^{\frac{s-1}{s}} \right)^{\frac{s}{s-1}} = c_t + \int_{j=0}^1 x_{jt} \, dj, \tag{2.13}$$

$$\int_{j=0}^1 \ell_j^P \, dj = \ell^P, \tag{2.14}$$

$$\int_{j=0}^1 \ell_{jt}^T \, dj = \ell_t^T. \tag{2.15}$$

[12] In this model, the shape of household utility over consumption is unimportant because there is only effectively one period, the afternoon, when the household pays for its consumption, and because there is no heterogeneity among households. With linear or CARA or CRRA utility, the results are qualitatively unchanged.

[13] In an extension, we generalize this to heterogeneous worker types, and we use the mix as a reduced form way to aggregate over this heterogeneity. e.g. older workers who prefer permanent, vs young inexperienced workers who prefer temporary because the search costs are too high for them to receive permanent offers.

See Section 2.8 in the appendix for the method that we use to solve the model.

2.5.3 Model Calibration

We calibrate the model primarily using moments in the data that are implied by the model structure, summarized in Table 2.8. The parameters α, β, ψ, and μ^F are calculated using aggregate moments. The Cobb-Douglas parameter α is expenditure share on labor. The discount rate β scales the overnight interst rate by the mean maturity of factored receivables; the value is almost exactly the same when using 3-month T-bills instead. The relative slope of the hourly wage to experience curve for permanent contract versus temporary contract employees is the gain to experience ψ for the afternoon versus the morning. The difference between FIDC factoring interest rates and the total cost of capital, as a weighted average across all factoring transactions, is the aggregate factoring spread μ^F. For the labor type elasticity of substitution σ, we regress $\log \ell_{j1}^T$ on $\log w_1^T$, net of firm and month fixed effects, then we use σ and equation (2.7) to calibrate ω. We calibrate ν at the indifference point in household FOCs between supplying an additional unit of permanent versus temporary labor. For the Frisch elasticity $\frac{1}{\xi-1}$ and the goods elasticity of substitution, we use the values in the BCB's calibration of its DSGE model. Finally, because we cannot observe the cost of default, we use the value 25% from Glover (2016). See Section 2.8.2 in the appendix for more details.

Table 2.8: Summary of Model Calibration

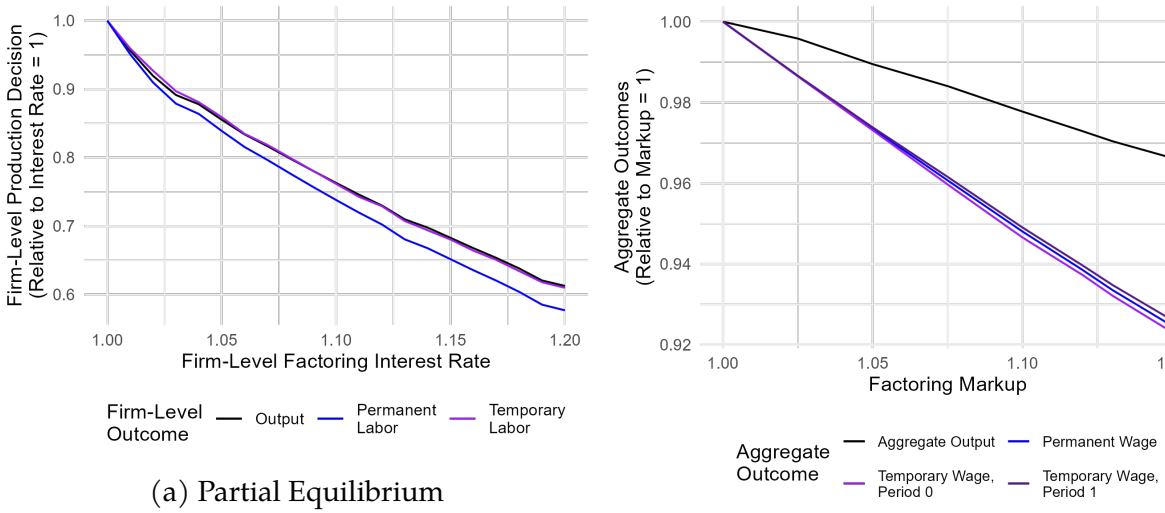
Parameter	Value	Description	Method
α	0.43	Cobb-Douglas labor	Data: Expenditure share
ψ	1.31	Gain to experience	Data: Ratio of existing to new hire wage for permanent vs temporary
μ^F	1.13	Factoring spread	Data: IR minus (default rate + 3-month T-bill)
σ	1.80	EoS permanent vs temporary labor	Data: Regression
ω	0.89	CES share parameter on permanent employees	Data: σ and model-derived moment
ν	0.009	Relative labor preference term	Data: model-derived moment
ξ	5.48	Exponential disutility of labor supply, equiv. to a Frisch elasticity of 0.22	BCB SAMBA DSGE
s	11	EoS across differentiated goods	BCB SAMBA DSGE
β	0.979	Discount rate between morning and afternoon	Data: 3-month T bill
η	0.25	Cost of default	Glover (JFE 2016).

Notes: This table shows the calibration of each parameter in the model. For more details on the calibration methods, see Section 2.8.2 in the appendix.

2.5.4 Counterfactuals

There are two main counterfactuals that we consider. The first counterfactual is the partial equilibrium equivalent of the regressions, shown in Figure 2.13a, where we decrease the interest rate R_j^F for a specific firm, holding fixed the equilibrium values $\{Y_j, w_{jt}^T, w^P\}$. As the firm's factoring risk ζ_j increases, permanent labor demand decreases faster than temporary labor demand and output. The second counterfactual, shown in Figure 2.13b, is the general equilibrium hypothetical, where we decrease the factoring spread μ^F, which decreases R_j^F for all firms. Across equilibria, as the factoring spread decreases, there is a larger increase in wage than in output due to the inelasticity of labor supply. See Section 2.8.3 in the appendix for additional model results.

Figure 2.13: Model-Implied Counterfactuals



(a) Partial Equilibrium

(b) General Equilibrium

Notes: These figures show two counterfactuals from solving the model under different conditions. On the left, for the partial equilibrium counterfactual, we decrease the interest rate R_j^F for a specific firm, holding fixed the equilibrium values $\{Y_j, w_{jt}^T, w^P\}$, as well as all parameters. On the right, for the general equilibrium counterfactual, we decrease the factoring spread μ^F, which decreases R_j^F for all firms, holding all other parameters fixed.

The main takeaway from Figure 2.13 is that the partial equilibrium elasticities with respect to the factoring interest rate, which are around -3.1, are similar in magnitude to the regression results from Section 2.4.2, while the general equilibrium elasticities are -0.3 for output and -0.5 for the wage bill, an order of magnitude smaller than the partial equilibrium effects. The general equilibrium dampening is due to the change in wage and the inelasticity of labor supply when firms collectively reallocate labor demand from permanent to temporary contracts.

2.6 Conclusion

This paper contributes new evidence that explain why factoring is the main form of working capital financing in Brazil and increasingly important worldwide, especially for small firms with low creditworthiness. We are the first to estimate the causal impact of factoring interest rate on firms' production decisions, trade credit, and firm outcomes, using a novel dataset that covers trade credit terms, factoring, other lending, payments, and employment for almost all formally registered firms in Brazil. We show that a decrease in the factoring interest rate leads to a large contemporaneous increase in firms' sales and input purchases, with an increase in permanent employment and a decrease in temporary employment. While the sales and input purchase effects mostly dissipate after several months, the increase in permanent employment is persistent, and temporary employment also increases in the long run as the firm grows and the short-term liquidity motive for labor substitution is no longer pertinent. These results highlight the dual function of

factoring in mitigating cash flow volatility and enabling firms to offer trade credit in the face of other financial constraints. Overall, this paper provides evidence for the importance of factoring as a form of financing for liquidity management, particularly for small and credit-constrained firms.

Our model provides a framework to understand how firms' demand for factoring is driven by cash flow volatility and non-payment risk from customers. Unlike other forms of financing, factoring directly decreases cash inflow volatility through shifting the non-payment risk to a financial intermediary. Factoring also differs from other financing because its borrowing constraint features bounds that are directly endogenous to firms' output decision, which amplifies the partial equilibrium response of factoring volume and output to the factoring interest rate. These effects are dampened in general equilibrium due to adjustments in wages and in firms' responses to changes in aggregate output.

Future research will examine how specific policy reforms to receivables registries, tokenization in the supply chain, expanded access to FIDCs, and fintech market approval affect the factoring interest rate through reducing transaction costs and increasing competition between banks and FIDCs in the supply of factoring.

References

Adelino, Manuel, Miguel A Ferreira, Mariassunta Giannetti, and Pedro Pires (2023). "Trade Credit And The Transmission of Unconventional Monetary Policy". In: *The Review of Financial Studies* 36.2, pp. 775–813.

Almeida, Heitor, Daniel Carvalho, and Taehyun Kim (2024). "The Working Capital Credit Multiplier". In: *The Journal of Finance* Forthcoming.

Altinoglu, Levent (2021). "The Origins of Aggregate Fluctuations in a Credit Network Economy". In: *Journal of Monetary Economics* 117, pp. 316–334.

Amberg, Niklas, Tor Jacobson, and Yingjie Qi (2024). *Supply-Chain Finance: An Empirical Evaluation of Supplier Outcomes*. Tech. rep.

Amberg, Niklas, Tor Jacobson, Erik Von Schedvin, and Robert Townsend (2021). "Curbing Shocks to Corporate Liquidity: The Role of Trade Credit". In: *Journal of Political Economy* 129.1, pp. 182–242.

Bahaj, Saleem, Angus Foulis, Gabor Pinter, and Paolo Surico (2022). "Employment And The Residential Collateral Channel of Monetary Policy". In: *Journal of Monetary Economics* 131, pp. 26–44.

BIS (2023). "Project Dynamo: Catalysing Innovation for SME Growth". In: URL: https://www.bis.org/publ/othp68.pdf.

Bocola, Luigi and Gideon Bornstein (2023). *Macroeconomics of Trade Credit*. Tech. rep.

Boissay, Frederic, Nikhil Patel, and Hyun Song Shin (2020). "Trade Credit, Trade Finance, And The Covid-19 Crisis". In: *Trade Finance, And The COVID-19 Crisis (June 19, 2020)*.

Bottazzi, Laura, Goutham Gopalakrishna, and Tebaldi Claudio (2023). *Supply Chain Finance And Firm Capital Structure*. Tech. rep.

Caglio, Cecilia, R Matthew Darst, and Sebnem Kalemli-Ozcan (2022). *Collateral Heterogeneity And Monetary Policy Transmission: Evidence from Loans to SMEs And Large Firms*. Tech. rep. Working Paper.

Canay, Ivan A (2011). "A Simple Approach to Quantile Regression for Panel Data". In: *The Econometrics Journal* 14.3, pp. 368–386.

Chodorow-Reich, Gabriel (2014). "The Employment Effects of Credit Market Disruptions: Firm-Level Evidence from the 2008–9 Financial Crisis". In: *The Quarterly Journal of Economics* 129.1, pp. 1–59.

Chodorow-Reich, Gabriel, Olivier Darmouni, Stephan Luck, and Matthew Plosser (2022). "Bank Liquidity Provision Across the Firm Size Distribution". In: *Journal of Financial Economics* 144.3, pp. 908–932.

Coval, Joshua and Erik Stafford (2007). "Asset Fire Sales (and Purchases) in Equity Markets". In: *Journal of Financial Economics* 86.2, pp. 479–512.

Custódio, Cláudia, Miguel A Ferreira, and Luís Laureano (2013). "Why Are US Firms Using More Short-Term Debt?" In: *Journal of Financial Economics* 108.1, pp. 182–212.

Darmouni, Olivier, Kerry Siani, and Kairong Xiao (2022). "Nonbank Fragility in Credit Markets: Evidence from a Two-Layer Asset Demand System". In: *Available at SSRN 4288695*.

Dass, Nishant, Jayant R Kale, and Vikram Nanda (2015). "Trade Credit, Relationship-Specific Investment, And Product Market Power". In: *Review of Finance* 19.5, pp. 1867–1923.

Dou, Winston Wei, Leonid Kogan, and Wei Wu (2022). *Common Fund Flows: Flow Hedging And Factor Pricing*. Tech. rep. National Bureau of Economic Research.

Edmans, Alex, Itay Goldstein, and Wei Jiang (2012). "The Real Effects of Financial Markets: The Impact of Prices on Takeovers". In: *The Journal of Finance* 67.3, pp. 933–971.

Fabbri, Daniela and Leora F Klapper (2016). "Bargaining Power And Trade Credit". In: *Journal of Corporate Finance* 41, pp. 66–80.

Fabbri, Daniela and Anna Maria C Menichini (2010). "Trade Credit, Collateral Liquidation, And Borrowing Constraints". In: *Journal of Financial Economics* 96.3, pp. 413–432.

Garcia-Martin, Alvaro, Santiago Justel, and Tim Schmidt-Eisenlohr (2023). *Trade Credit, Markups, And Relationships*. Tech. rep.

Giannetti, Mariassunta, Nicolas Serrano-Velarde, and Emanuele Tarantino (2021). "Cheap Trade Credit And Competition in Downstream Markets". In: *Journal of Political Economy* 129.6, pp. 1744–1796.

Glover, Brent (2016). "The Expected Cost of Default". In: *Journal of Financial Economics* 119.2, pp. 284–299.

Hahn, Jaehoon, Sandy Klasa, Hyuksoon Lim, and S Katie Moon (2024). "Temporary Workers And Corporate Liquidity Management Decisions". In: *Available at SSRN 3675086*.

Herreno, Juan (2023). *Aggregating the Effect of Bank Credit Supply Shocks on Firms*. Tech. rep. mimeo.

Huber, Kilian (2018). "Disentangling the Effects of a Banking Crisis: Evidence from German Firms and Counties". In: *American Economic Review* 108.3, pp. 868–898.

Jacobson, Tor and Erik Von Schedvin (2015). "Trade Credit And The Propagation of Corporate Failure: An Empirical Analysis". In: *Econometrica* 83.4, pp. 1315–1371.

Jordà, Òscar (2005). "Estimation And Inference of Impulse Responses by Local Projections". In: *American Economic Review* 95.1, pp. 161–182.

Lian, Chen and Yueran Ma (2021). "Anatomy of Corporate Borrowing Constraints". In: *The Quarterly Journal of Economics* 136.1, pp. 229–291.

Luck, Stephan and Joao AC Santos (2019). "The Valuation of Collateral in Bank Lending". In: *Journal of Financial and Quantitative Analysis*, pp. 1–30.

Mateos-Planas, Xavier and Giulio Seccia (2021). *Trade Credit Default*. Tech. rep.

Plagborg-Møller, Mikkel and Christian K Wolf (2021). "Local Projections And VARs Estimate the Same Impulse Responses". In: *Econometrica* 89.2, pp. 955–980.

Reischer, Margit (2024). *Trade Credit Origins of Aggregate Fluctuations*. Tech. rep.

Restrepo, Felipe, Lina Cardona-Sosa, and Philip E Strahan (2019). "Funding Liquidity without Banks: Evidence from a Shock to the Cost of Very Short-Term Debt". In: *The Journal of Finance* 74.6, pp. 2875–2914.

Skrastins, Janis (2021). "Barter Credit: Warehouses as a Contracting Technology". In: *Journal of Finance*.

Van der Beck, Philippe (2022). *On the Estimation of Demand-Based Asset Pricing Models*. Tech. rep. 22-67.

Wardlaw, Malcolm (2020). "Measuring Mutual Fund Flow Pressure as Shock to Stock Returns". In: *The Journal of Finance* 75.6, pp. 3221–3243.

Yang, S Alex and John R Birge (2018). "Trade Credit, Risk Sharing, And Inventory Financing Portfolios". In: *Management Science* 64.8, pp. 3667–3689.

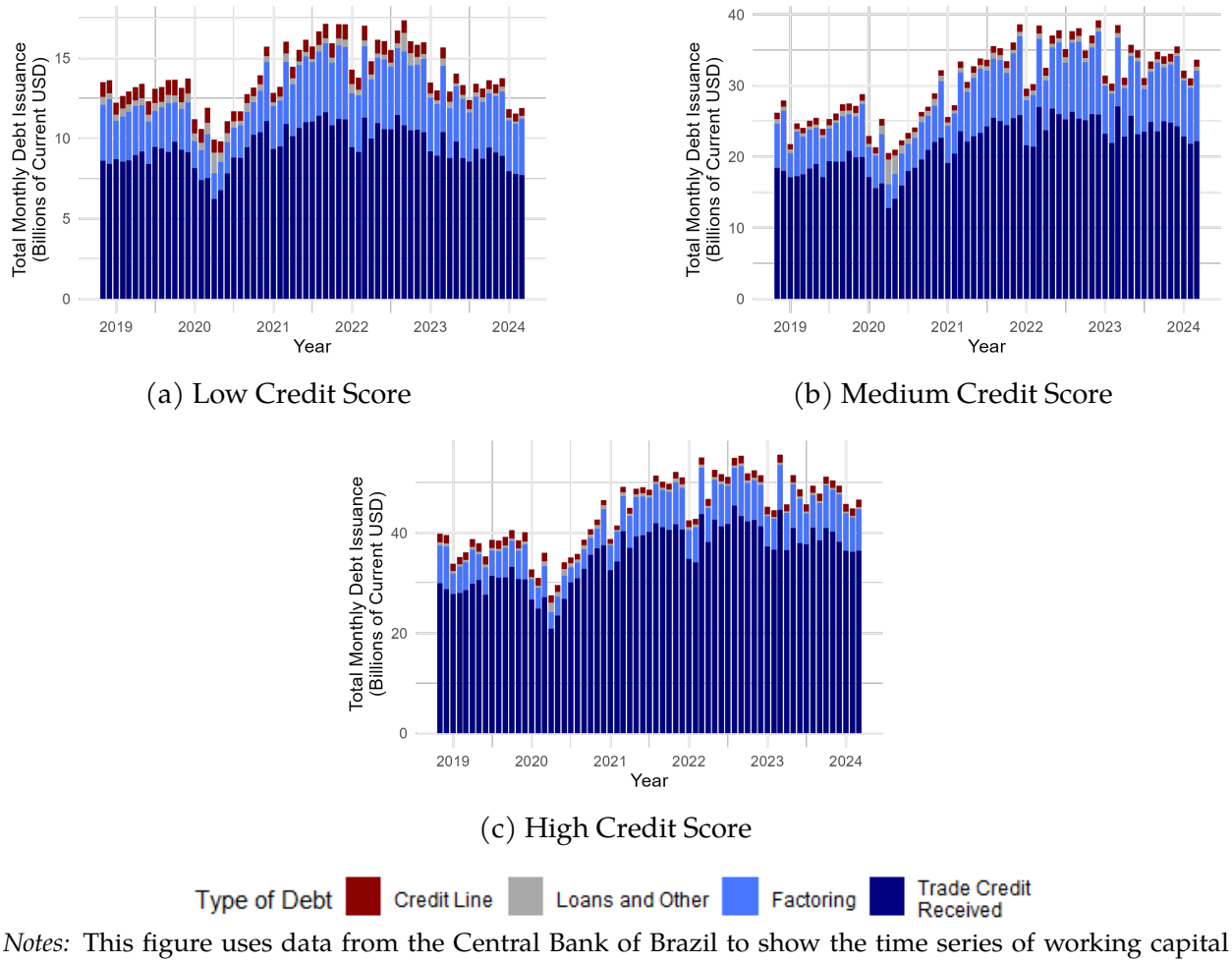
Yu, Jiaheng (2023). *Getting the Banks on Board: Accounts Receivable Financing in the US*. Tech. rep.

Appendix B

2.7 Empirical Appendix

Figure B1 shows that low credit score firms use factoring as a greater share of working capital financing than high credit score firms.

Figure B1: Working Capital Financing Composition by Firms' Credit Score (Right)



(a) Low Credit Score

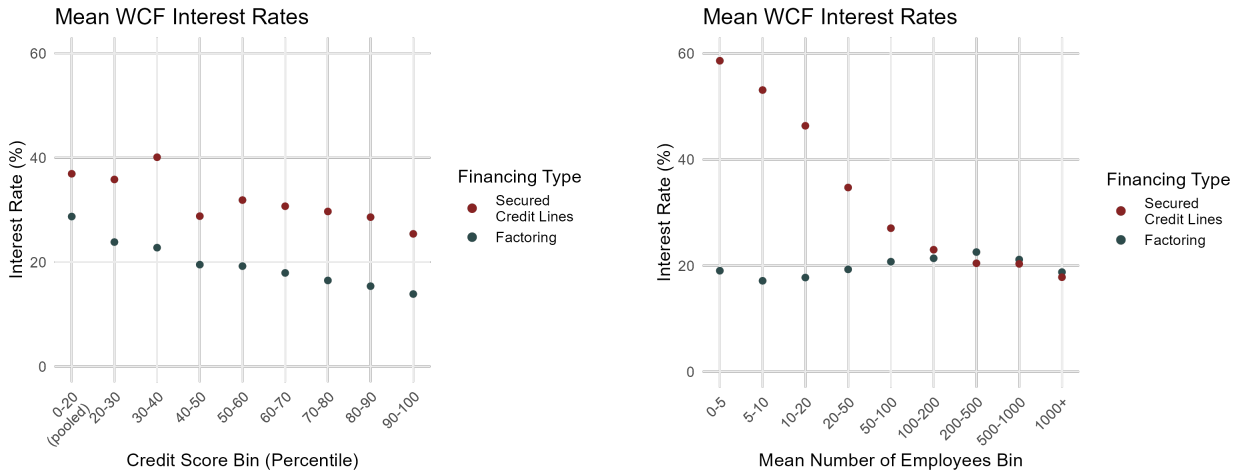
(b) Medium Credit Score

(c) High Credit Score

Notes: This figure uses data from the Central Bank of Brazil to show the time series of working capital financing, partitioned by the tercile of the credit score of the firms as of June 2023, the only month with available data. In dark blue is the value of trade credit that a firm receives from its suppliers. In light blue is factoring, the sale of receivables from the trade credit that a firm offers its customers. In red are credit lines, which generally require firms to post collateral. In gray are working capital loans and bonds.

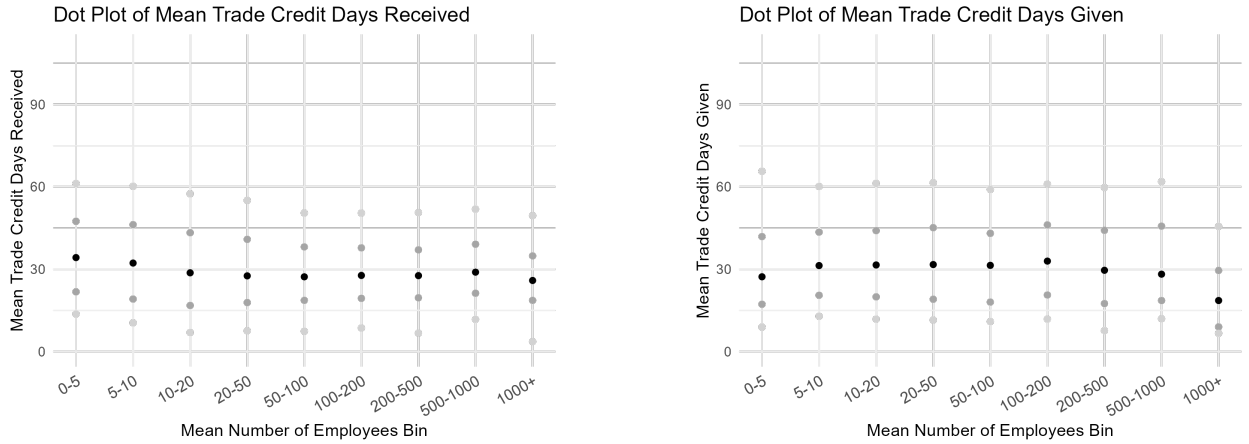
Figure B2 shows that the interest rates of secured credit lines and recourse factoring are both increasing in the risk of the debtor (decreasing in its credit score), but there is a stronger relationship for credit lines than factoring over the distribution of firm size, due to the factoring risk depending in part on the risk profile of the customers. This is one explanation for why small firms factor receivables at a higher rate than large firms.

Figure B2: Mean Interest Rates of Factoring and Credit Lines across the Firm Distribution



Notes: This figure uses data from the Central Bank of Brazil. Each subfigure partitions firms into bins along the horizontal axis and computes the issuance volume weighted mean interest rate within each bin for each type of working capital financing (WCF). On the left, we classify each firm by its mean number of employees across time. On the right, we classify firms by deciles of credit score in June 2023, the only month with available data. The bottom 19% of firms have a credit score of 0, generally signaling a lack of any credit history, so we pool together the bottom two deciles.

Figure B3: Trade Credit Maturity Quantiles across the Firm Size Distribution

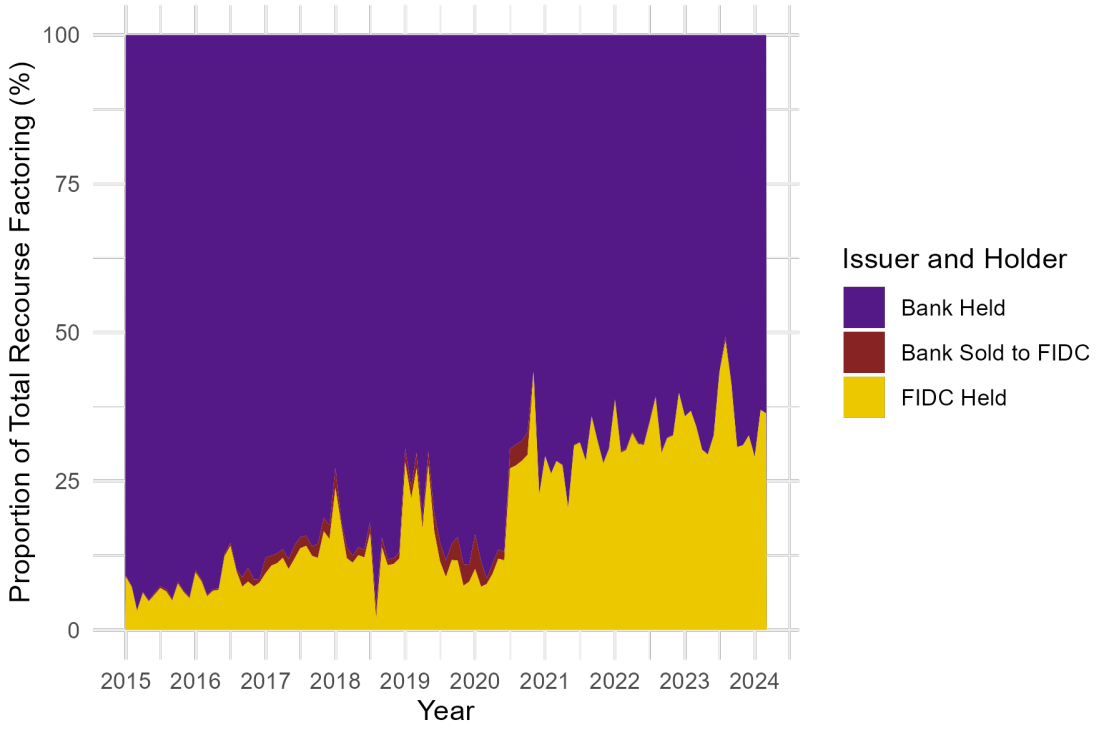


Notes: This figure uses data from the Central Bank of Brazil. Each subfigure partitions firms into bins of credit score as of June 2023, the only month with available data. The bottom 19% of firms have a credit score pooled near 0, generally signaling a lack of any credit history, so we pool together the bottom two deciles. The figure on the left shows the maturity distribution of trade credit received from suppliers, while the figure on the right shows the maturity distribution of trade credit given to customers. The black dots show the mean maturity, the inner medium gray dots show the 25th and 75th percentile, and the outer light gray dots show the 10th and 90th percentile, among the set of firms within each bin.

Figure B3 shows that there is also no strong relationship between a firm's credit score and the maturity of the trade credit that it gives or receives. However, the very largest firms, with over 2,000 employees and/or annual revenue of $1 billion USD, roughly corresponding to the set of publicly traded firms in Brazil, tend to pay upfront for their purchases from suppliers, and offer fewer days of trade credit to customers as well.

Figure B4 shows that the FIDC share of factoring has increased from 7% in 2015 to 32% in 2023.

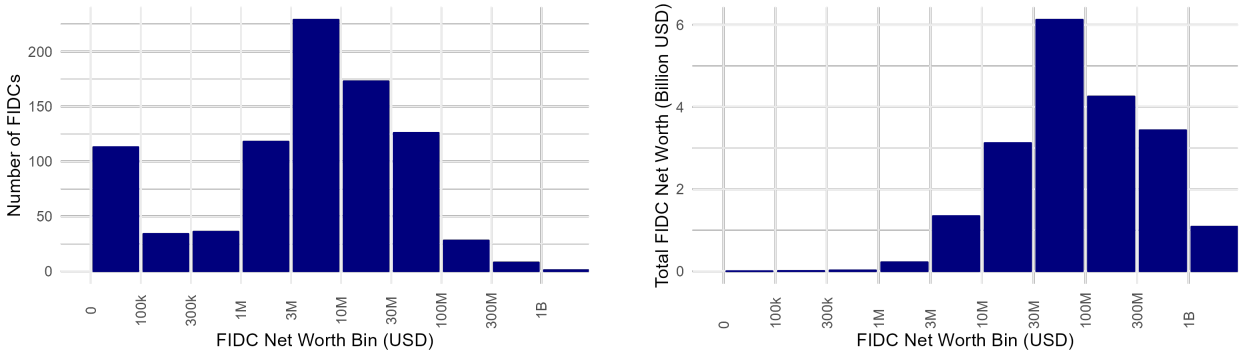
Figure B4: Time Series of FIDC Share of Factoring



Notes: This figure uses data from the Central Bank of Brazil to show the time series of the share of all receivables that are purchased by FIDCs, shown in gold, versus banks, shown in purple. A small proportion of receivables are purchased by banks and then resold to FIDCs; this is shown separately in red.

Figure B5 shows the distribution of FIDC size (left) and the distribution of FIDC net asset value (NAV) across bins of FIDC size (right) at the end of the sample period, March 2024. FIDCs are small compared to other asset classes of mutual funds. Many FIDCs have net asset value between $1 million USD and $100 million USD.

Figure B5: Distribution of FIDC Size



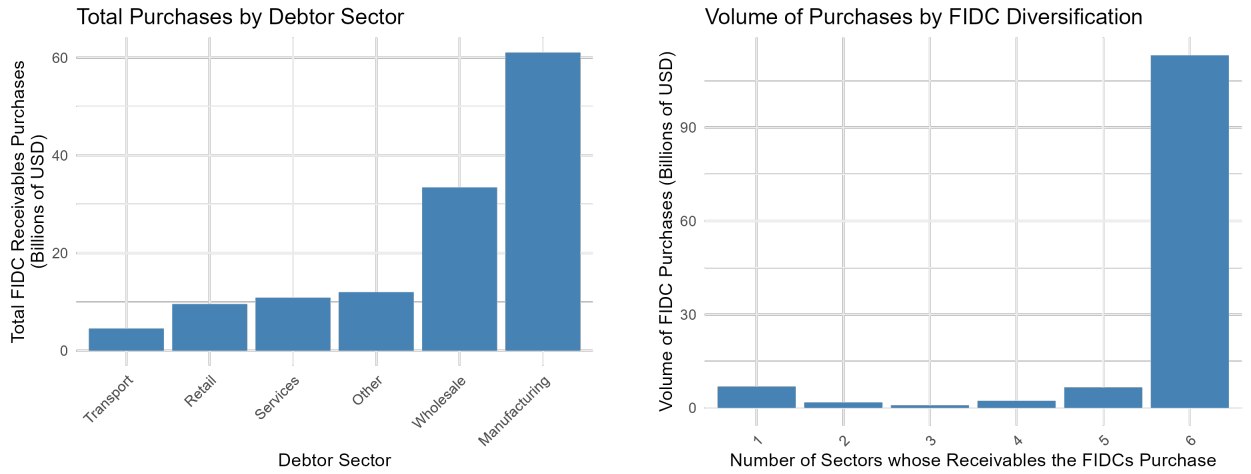
(a) Count of FIDCs by NAV Bin

(b) Total FIDC NAV by NAV Bin

Notes: This figure uses monthly fund report data from the CVM to show the distribution of FIDC size, by the count of the number of FIDCs in each size bin (left), and by the total NAV among FIDCs in each size bin (right). The count of FIDCs includes FIDCs that were inactive at the time of the snapshot in March 2024, but not formally registered as shut down.

Figure B6 shows the diversification of FIDCs across major sectors, as measured by the firm-level CNAE code, and total receivables purchases during the 65 months in the sample from November 2018 to March 2024. On the right, most receivables purchases by FIDCs is specifically by FIDCs that purchase receivables from all major sectors.

Figure B6: FIDC Diversification across Major Sectors



(a) FIDCs' Total Receivables Purchases by Major Sector

(b) FIDCs' Total Receivables Purchases by the Number of Major Sectors

Notes: This figure uses data on FIDCs' receivables purchases from the Central Bank of Brazil, merged to seller firms' CNAE code. On the left is the total receivables purchases by FIDCs by major sector, using the first two digits of the CNAE code. On the right, each FIDC is classified by the number of major sectors from which the FIDC purchases receivables, and then computes the sum of receivables purchase volume. The "6" bar shows that most FIDCs' receivables purchases are by FIDCs that purchase from all five major sectors, as well as other sectors like agriculture.

Table 2.9 shows that firm-month characteristics are similar across positive and negative flows.

Table 2.9: Average Firm Characteristics by Magnitude of Flow

	Negative Flow	Near Zero Flow	Positive Flow
Credit Score (0 to 1000)	517	501	500
Factoring Interest Rate (%)	30.6	30.7	30.3
Monthly Factoring (Thousand USD)	112.6	120.4	98.1
Monthly Other Debt (Thousand USD)	24.6	23.8	14.6
Trade Credit Maturity Offered (Days)	21.6	21.9	27.8
Trade Credit Maturity Received (Days)	37.1	36.9	41.0

Notes: We construct the flows from data from the Securities and Exchange Commission of Brazil (CVM), and firms' characteristics are from the Central Bank of Brazil. We define a flow to be "near zero" if it is between -0.1 and 0.1 standard deviations of zero, where the standard deviation is defined over nonzero values. 26% of firm-month observations have zero flow, mostly because the FIDCs that purchase the given firm's receivables did not receive any inflow nor outflow in the given month. The third and fourth rows refer to the amount of financing issuance.

2.7.1 Additional Regression Results

Additional Outcome Variables

Column 1 of Table 2.10 shows that a 1 percentage point increase in the factoring interest rate causes a 0.47% decrease in the number of employees, which is slightly smaller in magnitude than the 0.56% decrease in wage bill from Table 2.4. This can be further decomposed into a 1.1% decrease in the number of permanent employees in column 3, and a 1.8% increase in the number of temporary employees in column 4, which are comparable in magnitude to the labor demand results in terms of hours of work from Table 2.4. The mean number of permanent employees is 46.6, and the mean number of temporary employees is 4.3, so columns 3 and 4 of Table 2.10 correspond to a decrease of 0.61 permanent employees and an increase of 0.08 temporary employees, respectively. The mean number of permanent employees is 46.6, and the mean number of temporary employees is 4.3, so columns 3 and 4 of Table 2.10 correspond to a decrease of 0.61 permanent employees and an increase of 0.08 temporary employees, respectively.

Table 2.10: IV Regressions of Number of Employee Outcomes on Factoring Interest Rate

	(1)	(2)	(3)	(4)
	Log Number of Employees (Total)	Log Number of Employees (New Hire)	Log Number of Employees (Permanent)	Log Number of Employees (Temporary)
$r^{Fac}_{j,t}$	-0.0047^{*}	-0.0141^{**}	-0.0114^{***}	0.0181^{**}
	(0.0021)	(0.0045)	(0.0024)	(0.0059)
Num. Obs.	2,556,738	1,126,587	2,548,410	608,088
Num. Firms	288,507	184,070	287,381	93,219
Num. Months	50	50	50	50

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\cdot}p < 0.1$

Notes: All regressions use data from the Central Bank of Brazil. All regressions use firm and month fixed effects, with standard errors clustered at the firm level. The predictor variable is the firm-level interest rate on factoring in percentage points. The instrumental variable is the expected change in receivables purchases driven by fund flows. The response variables come from restricted access month-level RAIS data. An employee is defined as new if the employee began working at the firm that month.

Table 2.11 shows that the interest rate on unsecured credit lines highly responds to the change in the factoring price, primarily through banks' factoring rates. Unsecured credit lines have a high baseline mean interest rate of 333% and high variance across firms, with standard deviation of 85%.

Table 2.11: IV Regressions of Interest Rate Outcomes on Factoring Interest Rate

	(1)	(2)	(3)	(4)	(5)
	IR (Debt Under 1 Year)	IR (Debt Over 1 Year)	IR Credit Line (Unsecured)	IR Credit Line (Secured)	IR (Loans Over 1 Year)
$r_{j,t}^{Fac}$	1.5021***	−2.0535*	6.9139**	−3.5990	−0.0787
	(0.2217)	(0.9247)	(2.2363)	(3.9875)	(0.1800)
Num. Obs.	4,146,540	508,179	829,816	410,208	438,844
Num. Firms	511,896	130,522	123,370	57,997	123,553
Num. Months	65	65	65	65	65

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\cdot}p < 0.1$

Notes: All regressions use data from the Central Bank of Brazil. All regressions use firm and month fixed effects, with standard errors clustered at the firm level. The predictor variable is the firm-level interest rate on factoring in percentage points. The instrumental variable is the expected change in receivables purchases driven by fund flows. The response variables are the interest rates by category of debt. Column 1 is the subset with maturity of up to 365 days. Columns 2 and 3 are unsecured and secured credit lines, where issuance is defined as any drawdown of the credit line, not a change in the credit limit. Column 4 is loans with maturity of over 365 days.

Table 2.12 shows that default rates on factoring increase substantially with the factoring interest rate, but the default rate on other debt is unchanged. A one percentage point higher factoring interest rate causes a 0.27 percentage point higher default rate to banks, from a baseline of 0.40%, and a 1.68 percentage point higher default rate to FIDCs, from a baseline of 10.3%. We believe that the much higher default rate to FIDCs corresponds to a weaker threat of exclusion in response to default. While FIDCs only provide factoring services to firms, banks provide a wide range of financial services, and there is far more concentration among banks in most financial services compared to concentration in factoring.

Table 2.12: IV Regressions of Default Rate Outcomes on Factoring Interest Rate

	(1)	(2)	(3)	(4)
	Default Rate Rec. Factoring (to Banks, %)	Default Rate Other (to Banks, %)	Default Rate Rec. Factoring (to FIDCs, %)	Default Rate Other (to FIDCs, %)
$r^{Fac}_{j,t}$	0.2772*	−0.1945	1.6804**	0.0098
	(0.1174)	(0.2567)	(0.5329)	(0.0558)
Num. Obs.	2,739,575	2,739,575	1,435,934	1,435,934
Num. Firms	234,524	234,524	243,683	243,683
Num. Months	65	65	64	64

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\cdot}p < 0.1$

Notes: All regressions use data from the Central Bank of Brazil. All regressions use firm and month fixed effects, with standard errors clustered at the firm level. The predictor variable is the firm-level interest rate on factoring in percentage points. The instrumental variable is the expected change in receivables purchases driven by fund flows. The response variables are the default rates on recourse factoring and other debt issued by banks and FIDCs. The default rate is defined to be the percentage of debt not paid on its due date. This is lower than the percentage of debt that the creditor eventually collects. The issuance-weighted default rate for recourse factoring to banks is 0.40% and the issuance-weighted default rate for recourse factoring to FIDCs is 10.3%.

2.7.2 Robustness: Fixed Effects

In this section, we show that the results in Table 2.4 are robust to alternative specifications of fixed effects in (2.16) that interact aggregated firm characteristics k with month t to capture time-varying shocks.

$$r^{Fac}_{j,k,t} = \alpha_j + \alpha_{k,t} + \gamma_1 e^{Fac}_{j,k,t} + \varepsilon_{j,k,t}. \tag{2.16}$$

The base regression (column 1) uses firm and month fixed effects. Column 2 uses firm and state-by-month fixed effects, for the 26 states and one federal district in Brazil. Column 3 uses firm and sector-by-month fixed effects, for the 285 three-digit CNAE codes. Column 4 uses firm, state-by-month, and sector-by-month fixed effects. This controls for any time-varying shocks across locations and sectors, and the residual variation uses individual firm exposure to FIDC flows demeaned on any shocks in those dimensions. Table 2.13 compares the first stage regression in equation (2.1), of the factoring interest rate $r^{Fac}_{j,t}$ on exposure $e^{Fac}_{j,t}$ to fund flows, across the aforementioned fixed effects:

Table 2.13: First Stage Regressions Across Fixed Effect Specifications

	(1)	(2)	(3)	(4)
$e^{Fac}_{j,t}$	-0.1212^{***}	-0.1137^{***}	-0.1256^{***}	-0.1188^{***}
	(0.0127)	(0.0135)	(0.0128)	(0.0136)
Fixed Effects:				
Firm	X	X	X	X
Month	X	X	X	X
State-Month			X	X
Sector-Month		X		X

$^{***}p < 0.001; {}^{**}p < 0.01; {}^{*}p < 0.05; {}^{\cdot}p < 0.1$

Notes: These regressions use data from the Central Bank of Brazil. The dataset is at the firm by month level. Standard errors are clustered at the firm level and shown in parentheses. If a firm does not factor in a given month, then the interest rate is undefined, and the observation is dropped from the regression.

Table 2.14 compares the reduced form regression in (2.17) across the following fixed effect interactions of firms' aggregated characteristics k with month t.

$$y_{j,k,t} = \alpha_j + \alpha_{k,t} + \beta_1 r^{Fac}_{j,k,t} + \varepsilon_{j,k,t} \tag{2.17}$$

Table 2.14: IV Regressions of Main Outcomes Across Fixed Effect Specifications

	(1)	(2)	(3)	(4)
Panel A: Log Revenue				
$r^{Fac}_{j,t}$	-0.0614^{***}	-0.0678^{***}	-0.0584^{***}	-0.0648^{***}
	(0.0093)	(0.0112)	(0.0092)	(0.0110)
Panel B: Log Input Expenditure				
$r^{Fac}_{j,t}$	-0.0357^{***}	-0.0520^{***}	-0.0355^{***}	-0.0501^{***}
	(0.0056)	(0.0076)	(0.0054)	(0.0071)
Panel C: Log Wage Bill				
$r^{Fac}_{j,t}$	-0.0056^{*}	$-0.0051^{\cdot}$	-0.0060^{**}	-0.0056^{*}
	(0.0023)	(0.0027)	(0.0023)	(0.0028)
Panel D: Log Permanent Labor Demand				
$r^{Fac}_{j,t}$	-0.0114^{***}	-0.0112^{***}	-0.0124^{***}	-0.0121^{***}
	(0.0024)	(0.0028)	(0.0025)	(0.0030)
Panel E: Log Temporary Labor Demand				
$r^{Fac}_{j,t}$	0.0181^{**}	0.0160^{*}	0.0166^{**}	$0.0139^{\cdot}$
	(0.0059)	(0.0075)	(0.0056)	(0.0071)
Fixed Effects:				
Firm	X	X	X	X
Month	X	X	X	X
State-Month			X	X
Sector-Month		X		X

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\cdot}p < 0.1$

Notes: All regressions use data from the Central Bank of Brazil. All regressions use firm and month fixed effects, with standard errors clustered at the firm level in parentheses. The predictor variable is the firm-level interest rate on factoring in percentage points, instrumented by the expected change in receivables purchases driven by fund flows. The response variables are the log revenue proxied by payment inflows, log intermediate input expenditure proxied by payment outflows to firms, log wage bill, log labor demand for permanent workers, and log labor demand for temporary workers.

2.8 Model Appendix

2.8.1 Solving the Model

In this section, we describe how we solve the model, beginning with the firms' objective in (2.8) and the household's objective in (2.12).

Household Block

We solve the labor supply block by taking the FOCs of the Lagrangian in (2.18)

$$\begin{aligned}\max_{\{c_0,c_1,\ell^P,\ell_0^T,\ell_1^T\}} \mathcal{L}^{HH} &= \log(c_0) + \beta\log(c_1) - \sum_{t=0}^{1}\left[\frac{1}{\xi}\left(\ell^P + \nu\ell_t^T\right)^{\xi} - \nu(\ell^P - \ell_t^T)\right] \\ &\quad - \lambda^H \cdot \left(c_0 + c_1 - 2\ell^P w^P - \sum_t \ell_t^T w_t^T\right), \qquad (2.18)\\ \text{s.t. } c_0 + c_1 &= 2\ell^P w^P + \sum_t \ell_t^T w_t^T.\end{aligned}$$

The consumption FOCs give $c_0 = \beta c_1$ and $\lambda^H = \frac{1}{c_0}$. The labor FOCs are

$$\begin{aligned}\frac{\partial\mathcal{L}^{HH}}{\partial\ell^P} &= -\sum_{t=0}^{1}\left(\ell^P + \ell_t^T\right)^{\xi-1} + 2\nu + 2\lambda_1 - 2\lambda^H w^P = 0,\\ \frac{\partial\mathcal{L}^{HH}}{\partial\ell_t^T} &= -\left(\ell^P + \ell_t^T\right)^{\xi-1} - \nu + \lambda_1 - \lambda^H w_t^T = 0.\end{aligned}$$

We solve numerically. There are 6 unknowns $\{c_0, c_1, \ell^P, \ell_0^T, \ell_1^T, \lambda^H\}$ with 6 equations: the 2 consumption FOCs, the 3 labor FOCs, and the two budget constraints. In practice, it is easier to solve a reduced system of 3 unknowns $\{\ell^P, \ell_0^T, \ell_1^T\}$ in 3 equations by substituting out λ^H.

$$\lambda^H(w_t^T - w^P) = \frac{1}{2}\sum_{t'=0}^{1}\left(\ell^P + \ell_{t'}^T\right)^{\xi-1} - \left(\ell^P + \ell_t^T\right)^{\xi-1} - 2\nu.$$

After solving for $\{\ell^P, \ell_0^T, \ell_1^T\}$, we can use the consumption FOC and the budget constraint to solve for $\{c_0, c_1\}$.

Firm Block

Firms must factor at least what is paid to labor in the morning. Firms may decide to factor more to avoid the risk of high ϵ_j, which affects the otherwise risk-neutral firm's objective function through the penalty for ending the afternoon with negative profit. Firms generally do not hit the factoring borrowing constraint in (2.10), and there is no return on cash nor motive for precautionary (excess) borrowing. Work backwards by first solving for the constrained optimal x_{j1}, ℓ_{j1}^T in the afternoon.

The firm takes as given the choices made in the morning: $\{y_{j0}, x_{j0}, \ell_j^P, \ell_{j0}^T, p_j, B_j^F\}$ and the realization of the shock ϵ_j. Since there is no residual uncertainty, the firm's objective is to maximize marginal profits

$$\begin{aligned}
\max_{\{x_{j1}, \ell_{j1}^T\}} \ & \pi_{j1} := p_j y_{j1} - \ell_j^P w^P - \ell_{j1}^T w_1^T - P x_{j1}, \\
\text{s.t. } & y_{j1} = \tilde{\ell}_{j1}^{\alpha} x_{j1}^{1-\alpha}, \\
& \tilde{\ell}_{j1} = \left(\omega \left(\psi \ell_j^P \right)^{\frac{\sigma-1}{\sigma}} + (1-\omega) \left(\ell_{j1}^T \right)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \\
& \frac{p_j}{P} = \left(\frac{y_{j1}}{Y_1} \right)^{-\frac{1}{s}}, \\
& \ell_{j1}^T, x_{j1} \geqslant 0.
\end{aligned}$$

The unconstrained FOCs are

$$\begin{aligned}
\frac{\partial \pi_{j1}}{\partial x_{j1}} &= \frac{(s-1)(1-\alpha)}{s} P Y_1^{\frac{1}{s}} \tilde{\ell}_{j1}^{\frac{(s-1)\alpha}{s}} x_{j1}^{-\frac{1+(s-1)\alpha}{s}} - P = 0, \\
\frac{\partial \pi_{j1}}{\partial \ell_{j1}^T} &= \frac{(s-1)\alpha}{s} P Y_1^{\frac{1}{s}} \tilde{\ell}_{j1}^{-\frac{\alpha+s(1-\alpha)}{s}} x_{j1}^{\frac{(s-1)(1-\alpha)}{s}} \frac{\partial \tilde{\ell}_{j1}}{\partial \ell_{j1}^T} - w_1^T = 0,
\end{aligned}$$

where $\tilde{C}_1^T \equiv (1-\omega) \frac{(s-1)\alpha}{s} \left(\frac{(s-1)(1-\alpha)}{s} \right)^{\frac{(s-1)(1-\alpha)}{1+(s-1)\alpha}}$ is a constant.

$$\begin{aligned}
x_{j1} &= \left(\frac{(s-1)(1-\alpha)}{s} \right)^{\frac{s}{1+(s-1)\alpha}} Y_1^{\frac{1}{1+(s-1)\alpha}} \tilde{\ell}_{j1}^{\frac{(s-1)\alpha}{1+(s-1)\alpha}}, \\
w_1^T &= \tilde{C}_1 Y_1^{\frac{1}{1+(s-1)\alpha}} \tilde{\ell}_{j1}^{\frac{1}{\sigma} - \frac{1}{1+(s-1)\alpha}} \left(\ell_{j1}^T \right)^{-\frac{1}{\sigma}}.
\end{aligned}$$

In the morning, choose $\{y_{j0}, x_{j0}, \ell_j^P, \ell_{j0}^T, B_j^F\}$ to maximize expected profit, given the choices of $\ell_{j1}^T, x_{j1}, \pi_{j1}$ in the afternoon, and given wages w^P, w_0^T and parameters:

$$\max_{\{y_{j0}, x_{j0}, \ell_j^P, \ell_{j0}^T, B_j^F\}} \beta \mathbb{E}\left[\pi_{j1} + \eta \mathbb{1}\{\pi_{j1} < 0\} \right] + B_j^F - \left(\ell_j^P w^P + \ell_{j0}^T w_0^T \right), \tag{2.19}$$

$$\text{s.t. } B_j^F \leqslant \frac{p_j y_{j0}}{R_j^F}, \tag{2.20}$$

$$0 \leqslant B_j^F - \ell_j^P w^P - \ell_{j0}^T w_0^T, \tag{2.21}$$

$$\begin{aligned}
\pi_{j1} &= p_j y_{j1} - \ell_j^P w^P - \ell_{j1}^T w_1^T - P x_{j1} + \tilde{m}_{j1}, \\
\tilde{m}_{j1} &= (1-\epsilon_j) \left(p_j y_{j0} - R_j^F B_j^F \right) - P x_{j0}.
\end{aligned}$$

B_j^F is endogenous subject to the lower bound (2.21) and upper bound (2.20). The penalty for default introduces a non-convexity that requires tedious case work for an analytical solution. For each firm type j, corresponding to a cash flow risk value ζ_j, create a grid of $\{B_j^F, \ell_j^P\}$. Conditional on

B_j^F and ℓ_j^P, the morning objective function $\mathbb{E}_0\pi_j$ has no residual uncertainty over $\{\ell_{j0}^T, x_{j0}\}$, and the afternoon objective function π_{j1} has no uncertainty over $\{\ell_{j1}^T, x_{j1}\}$. We calculate the argmax of $\mathbb{E}_0\pi_j$ over $\{B_j^F, \ell_j^P\}$.

$$R_j^F w_0^T = \tilde{C}_0^T Y_0^{\frac{1}{1+(s-1)\alpha}} \tilde{\ell}_{j0}^{\frac{1}{\sigma}-\frac{1}{1+(s-1)\alpha}} \left(\ell_{j0}^T\right)^{-\frac{1}{\sigma}},$$

where $\tilde{C}_0^T \equiv (1-\omega)\frac{(s-1)\alpha}{s}\left((1-\zeta_j)\frac{(s-1)(1-\alpha)}{s}\right)^{\frac{(s-1)(1-\alpha)}{1+(s-1)\alpha}}$ is a constant. Note that if the problem were convex, we could directly solve for ℓ_j^P as shown in the appendix:

$$\begin{aligned} 0 = {} & \tilde{C}_0^P Y_1^{\frac{1}{1+(s-1)\alpha}} \tilde{\ell}_{j1}^{\frac{1}{\sigma}-\frac{1}{1+(s-1)\alpha}} \left(\ell_j^P\right)^{-\frac{1}{\sigma}} - w^P \\ & + (1-\zeta_j)\left[\tilde{C}_1^P Y_0^{\frac{1}{1+(s-1)\alpha}} \tilde{\ell}_{j0}^{\frac{1}{\sigma}-\frac{1}{1+(s-1)\alpha}} \left(\ell_j^P\right)^{-\frac{1}{\sigma}} - R_j^F w^P\right], \end{aligned} \tag{2.22}$$

where $\tilde{C}_0^P \equiv \omega\psi^{\frac{\sigma-1}{\sigma}}\frac{(s-1)\alpha}{s}\left(\frac{(s-1)(1-\alpha)}{s}\right)^{\frac{(s-1)(1-\alpha)}{1+(s-1)\alpha}}$ and $\tilde{C}_1^P \equiv \omega\frac{(s-1)\alpha}{s}\left((1-\zeta_j)\frac{(s-1)(1-\alpha)}{s}\right)^{\frac{s}{1+(s-1)\alpha}}$ are constants.
In the afternoon,

$$\begin{aligned} \max_{\{x_{j1},\ell_{j1}^T\}} \pi_{j1} := {} & p_j y_{j1} - \ell_j^P w^P - \ell_{j1}^T w_1^T - P x_{j1}, \\ \text{s.t. } & y_{j1} = \tilde{\ell}_{j1}^{\alpha} x_{j1}^{1-\alpha}, \\ & \tilde{\ell}_{j1} = \left(\omega\left(\psi\ell_j^P\right)^{\frac{\sigma-1}{\sigma}} + (1-\omega)\left(\ell_{j1}^T\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}, \\ & \frac{p_j}{P} = \left(\frac{y_{j1}}{Y_1}\right)^{-\frac{1}{s}}, \\ & \ell_{j1}^T, x_{j1} \geqslant 0. \end{aligned}$$

Substitute out p_j:

$$\max_{\{x_{j1},\ell_{j1}^T\}} \pi_{j1} := P Y_1^{\frac{1}{s}} \tilde{\ell}_{j1}^{\frac{(s-1)\alpha}{s}} x_{j1}^{\frac{(s-1)(1-\alpha)}{s}} - \ell_j^P w^P - \ell_{j1}^T w_1^T - P x_{j1}$$

The unconstrained FOCs are

$$\begin{aligned} \frac{\partial\pi_{j1}}{\partial x_{j1}} &= \frac{(s-1)(1-\alpha)}{s} P Y_1^{\frac{1}{s}} \tilde{\ell}_{j1}^{\frac{(s-1)\alpha}{s}} x_{j1}^{-\frac{1+(s-1)\alpha}{s}} - P = 0, \\ x_{j1} &= \left(\frac{(s-1)(1-\alpha)}{s}\right)^{\frac{s}{1+(s-1)\alpha}} Y_1^{\frac{1}{1+(s-1)\alpha}} \tilde{\ell}_{j1}^{\frac{(s-1)\alpha}{1+(s-1)\alpha}}, \end{aligned} \tag{2.23}$$

and

$$\frac{\partial \pi_{j1}}{\partial \ell_{j1}^{T}} = \frac{(s-1)\alpha}{s} P Y_1^{\frac{1}{s}} \tilde{\ell}_{j1}^{-\frac{\alpha+s(1-\alpha)}{s}} x_{j1}^{\frac{(s-1)(1-\alpha)}{s}} \frac{\partial \tilde{\ell}_{j1}}{\partial \ell_{j1}^{T}} - w_1^{T} = 0,$$

where

$$\frac{\partial \tilde{\ell}_{j1}}{\partial \ell_{j1}^{T}} = (1-\omega)\left(\omega\left(\psi \ell_j^{P}\right)^{\frac{\sigma-1}{\sigma}} + (1-\omega)\left(\ell_{j1}^{T}\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\sigma-1}} \left(\ell_{j1}^{T}\right)^{-\frac{1}{\sigma}} = (1-\omega)\tilde{\ell}_{j1}^{\frac{1}{\sigma}}\left(\ell_{j1}^{T}\right)^{-\frac{1}{\sigma}}.$$

Substituting out x_{j1} from (2.23):

$$x_{j1}^{\frac{(s-1)(1-\alpha)}{s}} = \left(\frac{(s-1)(1-\alpha)}{s}\right)^{\frac{(s-1)(1-\alpha)}{1+(s-1)\alpha}} Y_1^{\frac{(s-1)(1-\alpha)}{s\left(1+(s-1)\alpha\right)}} \tilde{\ell}_{j1}^{\frac{(s-1)^2\alpha(1-\alpha)}{s\left(1+(s-1)\alpha\right)}}.$$

Then the ℓ_{j1}^{T} FOC yields ℓ_{j1}^{T} as a function of ℓ_j^{P}, taking wages, the price index, and aggregate output as given:

$$w_1^{T} = \tilde{C}_1^{T} Y_1^{\frac{(s-1)(1-\alpha)}{s\left(1+(s-1)\alpha\right)}+\frac{1}{s}} \tilde{\ell}_{j1}^{\frac{(s-1)^2\alpha(1-\alpha)}{s\left(1+(s-1)\alpha\right)}-\frac{\alpha+s(1-\alpha)}{s}+\frac{1}{\sigma}} \left(\ell_{j1}^{T}\right)^{-\frac{1}{\sigma}}.$$

where $\tilde{C}_1^{T} \equiv (1-\omega)\frac{(s-1)\alpha}{s}\left(\frac{(s-1)(1-\alpha)}{s}\right)^{\frac{(s-1)(1-\alpha)}{1+(s-1)\alpha}}$ is a constant. Simplify the $\tilde{\ell}_{j1}$ exponent by combining the first and second terms:

$$\frac{(s-1)^2\alpha(1-\alpha) - \left(\alpha + s(1-\alpha)\right)\left(1+(s-1)\alpha\right)}{s\left(1+(s-1)\alpha\right)},$$

The second term is

$$\begin{aligned}\left(s-(s-1)\alpha\right)\left(1+(s-1)\alpha\right) &= s(1+(s-1)\alpha) - (s-1)\alpha(1+(s-1)\alpha),\\ &= s + s(s-1)\alpha - (s-1)\alpha - (s-1)^2\alpha^2.\end{aligned}$$

Expand the numerator:

$$\begin{aligned}&s^2\alpha(1-\alpha) - 2s\alpha(1-\alpha) + \alpha(1-\alpha) - \left(s + s(s-1)\alpha - (s-1)\alpha - (s-1)^2\alpha^2\right)\\ &= s^2\alpha - s^2\alpha^2 - 2s\alpha + 2s\alpha^2 + \alpha - \alpha^2 - \left(s + s^2\alpha - s\alpha + s\alpha - \alpha + s^2\alpha^2 - 2s\alpha^2 + \alpha^2\right).\\ &= s^2\alpha - s^2\alpha^2 - 2s\alpha + 2s\alpha^2 + \alpha - \alpha^2 - s - s^2\alpha + 2s\alpha - \alpha + s^2\alpha^2 - 2s\alpha^2 + \alpha^2,\\ &= -s,\end{aligned}$$

so the $\tilde{\ell}_{j1}$ exponent is $-\frac{1}{1+(s-1)\alpha} + \frac{1}{\sigma}$. Similarly, the Y_1 exponent simplifies to $\frac{1}{1+(s-1)\alpha}$. Then ℓ_{j1}^{T} is

implicitly a function of ℓ_j^P, the wage w_1^T, and aggregate output Y_1 through the equation

$$w_1^T = \tilde{C}_1^T Y_1^{\frac{1}{1+(s-1)\alpha}} \tilde{\ell}_{j1}^{\frac{1}{\sigma}-\frac{1}{1+(s-1)\alpha}} \left(\ell_{j1}^T\right)^{-\frac{1}{\sigma}}. \tag{2.24}$$

Taking as given the $\{\ell_j^P\}$ choices from the morning, solve for w_1^T by aggregating across firms using labor market clearing and the household's labor supply equation:

$$\int_{j=0}^{1} \ell_{j1}^T dj = \ell_1^T = 1 - 2\ell^P - \ell_0^T.$$

Substitute w_1^T back in to obtain ℓ_{j1}^T. Then use the x_{j1} FOC to solve for x_{j1} as a function of ℓ_j^P and ℓ_{j1}^T. These FOCs do not depend on the η term, nor on the shock ϵ_j. These only affect the factoring decision in the morning.

If there were no penalty for default, all firms factor the lower bound from (2.21) as long as $\mu^F > 0$, and otherwise are indifferent between factoring any amount.[14] Firms still adjust by reducing period 0 production relative to period 1 production (in turn reducing permanent labor demand as R_j^F increases). Without loss of generality, assume that firms factor the bare minimum $B_j^F = \ell_j^P w^P + \ell_{j0}^T w_0^T$, so the factoring spread μ^F is akin to a tax of R_j^F on period 0 labor. All terms in the objective function are scaled by β from the perspective of the morning, so we can drop them because multiplying by β is a uniform transformation. The morning problem is equivalent to

$$\begin{aligned}\max_{\{y_{jt}, p_j, x_{jt}, \ell_j^P, \ell_{jt}^T\}} \mathbb{E}_0 \pi_j &= (1-\zeta_j)\left(p_j y_{j0} - R_j^F B_j^F\right) - P x_{j0} + \left(p_j y_{j1} - \ell_j^P w^P - \ell_{j1}^T w_1^T - P x_{j1}\right), \\ \text{s.t. } B_j^F &= \ell_j^P w^P + \ell_{j0}^T w_0^T,\end{aligned}$$

We proceed with a similar derivation to (2.24) for ℓ_{j0}^T. First define

$$\tilde{\ell}_{j0} := \left(\omega\left(\ell_j^P\right)^{\frac{\sigma-1}{\sigma}} + (1-\omega)\left(\ell_{j0}^T\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}.$$

Take the FOC with respect to material inputs:

$$\begin{aligned}\frac{\partial \mathbb{E}_0 \pi_j}{\partial x_{j0}} &= (1-\zeta_j)\frac{(s-1)(1-\alpha)}{s} P Y_0^{\frac{1}{s}} \tilde{\ell}_{j0}^{\alpha\frac{s-1}{s}} x_{j0}^{-\frac{1+(s-1)\alpha}{s}} - P = 0, \\ x_{j0} &= \left((1-\zeta_j)\frac{(s-1)(1-\alpha)}{s}\right)^{\frac{s}{1+(s-1)\alpha}} Y_0^{\frac{1}{1+(s-1)\alpha}} \tilde{\ell}_{j0}^{\frac{(s-1)\alpha}{1+(s-1)\alpha}},\end{aligned} \tag{2.25}$$

and

$$\frac{\partial \mathbb{E}_0 \pi_j}{\partial \ell_{j0}^T} = \frac{(s-1)\alpha}{s} P Y_0^{\frac{1}{s}} \tilde{\ell}_{j0}^{-\frac{\alpha+s(1-\alpha)}{s}} x_{j0}^{\frac{(s-1)(1-\alpha)}{s}} \frac{\partial \tilde{\ell}_{j0}}{\partial \ell_{j0}^T} - R_j^F w_0^T = 0.$$

[14] The reasoning is that factoring reduces expected profit. If there is no insurance value to factoring, since firms are risk-neutral, then firms factor as little as needed to satisfy constraints.

Using the substitution $\frac{\partial \tilde{\ell}_{j0}}{\partial \ell_{j0}^T} = (1-\omega)\tilde{\ell}_{j0}^{\frac{1}{\sigma}} \left(\ell_{j0}^T\right)^{-\frac{1}{\sigma}}$,

$$R_j^F w_0^T = \tilde{C}_0^T Y_0^{\frac{1}{1+(s-1)\alpha}} \tilde{\ell}_{j0}^{\frac{1}{\sigma} - \frac{1}{1+(s-1)\alpha}} \left(\ell_{j0}^T\right)^{-\frac{1}{\sigma}}, \tag{2.26}$$

where $\tilde{C}_0^T \equiv (1-\omega)\frac{(s-1)\alpha}{s}\left((1-\zeta_j)\frac{(s-1)(1-\alpha)}{s}\right)^{\frac{(s-1)(1-\alpha)}{1+(s-1)\alpha}}$ is a constant.
Now take the derivative of the objective function $\mathbb{E}_0\pi_j$ with respect to ℓ_j^P:

$$\begin{aligned}\frac{\partial \mathbb{E}_0\pi_j}{\partial \ell_j^P} = (1-\zeta_j)\Bigg[&\frac{(s-1)\alpha}{s} P Y_0^{\frac{1}{s}} \tilde{\ell}_{j0}^{-\frac{\alpha+s(1-\alpha)}{s}} x_{j0}^{\frac{s-1}{s}(1-\alpha)} \frac{\partial \tilde{\ell}_{j0}}{\partial \ell_j^P} - R_j^F w^P\Bigg] \\ &+ \frac{(s-1)\alpha}{s} P Y_1^{\frac{1}{s}} \tilde{\ell}_{j1}^{-\frac{\alpha+s(1-\alpha)}{s}} x_{j1}^{\frac{s-1}{s}(1-\alpha)} \frac{\partial \tilde{\ell}_{j1}}{\partial \ell_j^P} - w^P = 0.\end{aligned}$$

Substitute out

$$\begin{aligned}\frac{\partial \tilde{\ell}_{j1}}{\partial \ell_j^P} &= \omega \psi^{\frac{\sigma-1}{\sigma}} \tilde{\ell}_{j1}^{\frac{1}{\sigma}} \left(\ell_j^P\right)^{-\frac{1}{\sigma}}, \\ \frac{\partial \tilde{\ell}_{j0}}{\partial \ell_j^P} &= \omega \tilde{\ell}_{j1}^{\frac{1}{\sigma}} \left(\ell_j^P\right)^{-\frac{1}{\sigma}},\end{aligned}$$

and x_{jt} from (2.25) and (2.23) to obtain

$$\begin{aligned}0 = \tilde{C}_0^P Y_1^{\frac{1}{1+(s-1)\alpha}} \tilde{\ell}_{j1}^{\frac{1}{\sigma} - \frac{1}{1+(s-1)\alpha}} \left(\ell_j^P\right)^{-\frac{1}{\sigma}} - w^P \\ + (1-\zeta_j)\left[\tilde{C}_1^P Y_0^{\frac{1}{1+(s-1)\alpha}} \tilde{\ell}_{j0}^{\frac{1}{\sigma} - \frac{1}{1+(s-1)\alpha}} \left(\ell_j^P\right)^{-\frac{1}{\sigma}} - R_j^F w^P\right],\end{aligned} \tag{2.27}$$

where $\tilde{C}_0^P \equiv \omega\psi^{\frac{\sigma-1}{\sigma}}\frac{(s-1)\alpha}{s}\left(\frac{(s-1)(1-\alpha)}{s}\right)^{\frac{(s-1)(1-\alpha)}{1+(s-1)\alpha}}$ and $\tilde{C}_1^P \equiv \omega\frac{(s-1)\alpha}{s}\left((1-\zeta_j)\frac{(s-1)(1-\alpha)}{s}\right)^{\frac{s}{1+(s-1)\alpha}}$ are constants.

In combination with (2.26) and (2.24), this equation pins down ℓ_j^P when taking as given the factoring price R_j^F and aggregate equilibrium outcomes $\{Y_0, Y_1, w^P, w_0^T, w_1^T\}$. Since $\frac{\partial \mathbb{E}_0\pi_j}{\partial \ell_j^P}$ is decreasing in ℓ_j^P, $\frac{\partial \mathbb{E}_0\pi_j}{\partial \ell_{j0}^T}$ is decreasing in ℓ_{j0}^T, and $\frac{\partial \pi_{j1}}{\partial \ell_{j1}^T}$ is decreasing in ℓ_{j1}^T, the following algorithm suffices to obtain labor demand:

- Outer loop: bisection method over ℓ_j^P with (2.22)
- Inner loop: given the guess of ℓ_j^P, use bisection with (2.26) to obtain ℓ_{j0}^T, and bisection with (2.24) to obtain ℓ_{j1}^T.

Solving for the Equilibrium

From outer-most to inner-most loop,

1. Iterate over w^P for (2.14); if permanent labor demand is greater than supply, then increase w^P, otherwise decrease w^P.

2. Iterate over w_0^T and w_1^T for (2.15); the ℓ_{j0}^T FOC does not directly depend on w_1^T, and the ℓ_{j1}^T FOC does not directly depend on w_0^T, so given a guess of w^P, temporary labor demand can be equated to temporary labor supply in the morning and afternoon in parallel.

3. Iterate over guesses of Y_0 and Y_1 so that (2.13) holds. Because household consumption c_t only depends on the wages and not directly on Y_t, and x_{jt} is increasing in Y_t with first-order elasticity $\frac{1}{1+(s-1)\alpha} < 1$, a quick bisection suffices for Y_t.

4. Solve for the allocations $\{\ell_j^P, \ell_{j0}^T, \ell_{j1}^T\}_{\zeta_j \sim G_\zeta}$ following the algorithm from the previous section.

2.8.2 Calibration

The following is a list of each parameter and a justification for its calibration.

- $\alpha = 0.43$: Cobb-Douglas share on labor vs intermediate inputs. Payments to labor are 2.997 $\cdot 10^{12}$ USD, vs intermediate input purchases by firms are 4.040 $\cdot 10^{12}$ USD. Excluding firms in the trade & wholesale sector, for which intermediate input purchases are almost as high as revenue.

- $\psi = 1.31$ is the gain to experience for permanent workers vs temporary workers (in the afternoon vs the morning). The average ratio of existing employee hourly wage to new hire hourly wage is 1.77 for permanent and 1.34 for temporary employees. The mean ratio of the permanent to temporary ratio is 1.31. We purposely do not control for tenure because this primarily reflects that permanent employees spend longer at firms.

- $\mu^F = 1.13$ is the factoring spread. The mean federal funds rate (SELIC) was 7.83%. The default rate to FIDCs was 10.30% (conservatively calculated as the amount unpaid at due). The weighted average interest rate from FIDCs is 33.29%. So the spread is 1.3329/1.1813

- $\sigma = 1.80$: Elasticity of substitution between temporary and permanent employees. In a static model, this is the answer to "given a change in the ratio of temporary to permanent hourly wage, how much does a firm's ratio of temporary to permanent employees change?" From month to month, the permanent wage and number of employees barely changes by design, so σ is the coefficient of $\log \ell_{j1}^T$ on $\log w_1^T$, net of firm and month fixed effects.

- $\omega = 0.89$. CES share parameter on permanent employees. Aggregate equation (2.7) over all firms, and let L^P and L^T denote the total hours supplied of permanent and temporary employees, respectively. Then,

$$\frac{L^P}{L^T} = \left(\frac{\omega}{1-\omega}\right)^\sigma \implies \omega = \frac{1}{1 + \left(\frac{L^T}{L^P}\right)^{\frac{1}{\sigma}}} = \frac{1}{1 + 0.0816^{0.56}}.$$

- $\nu = 0.009$ is the relative labor preference term. At the worker by month level, de-meaning by worker and month fixed effects, the mean wage for temporary workers is 0.29 BRL per hour higher than for permanent workers, or 1.8% higher. For workers to be indifferent on the margin between permanent and temporary labor, then $\frac{\partial \mathcal{L}^{HH}}{\partial \ell^P} = \frac{\partial \mathcal{L}^{HH}}{\partial \ell^T}$. From the household FOCs,

$$\frac{\partial \mathcal{L}^{HH}}{\partial \ell^P} = \frac{\partial \mathcal{L}^{HH}}{\partial \ell^T} \iff \nu + \frac{w^P}{\bar{w}} = -\nu + \frac{w^T}{\bar{w}} \iff \nu = \frac{w^T - w^P}{2\bar{w}},$$

  where $\lambda_2 = \frac{1}{\bar{w}}$ is the marginal utility of consumption; we assume one total unit of labor supply, so consumption equals the weighted mean wage $\bar{w}$.

- $\xi = 5.48$ is the exponential disutility of labor supply, and $\frac{1}{\xi - 1}$ is the Frisch elasticity. We follow the Central Bank of Brazil calibration of its SAMBA DSGE model, see Table 4 of Fasolo et al. (2024), corresponding to a Frisch elasticity of 0.22.

- $s = 11$ is the elasticity of substitution across goods. We follow the Central Bank of Brazil calibration of its SAMBA DSGE model. See Table 4 of Fasolo et al. (2024).

- $\beta = 0.979$ is the discount rate between the morning and afternoon. The mean days outstanding of factored receivables is 121 days, while the average overnight interest rate was 7.62%.

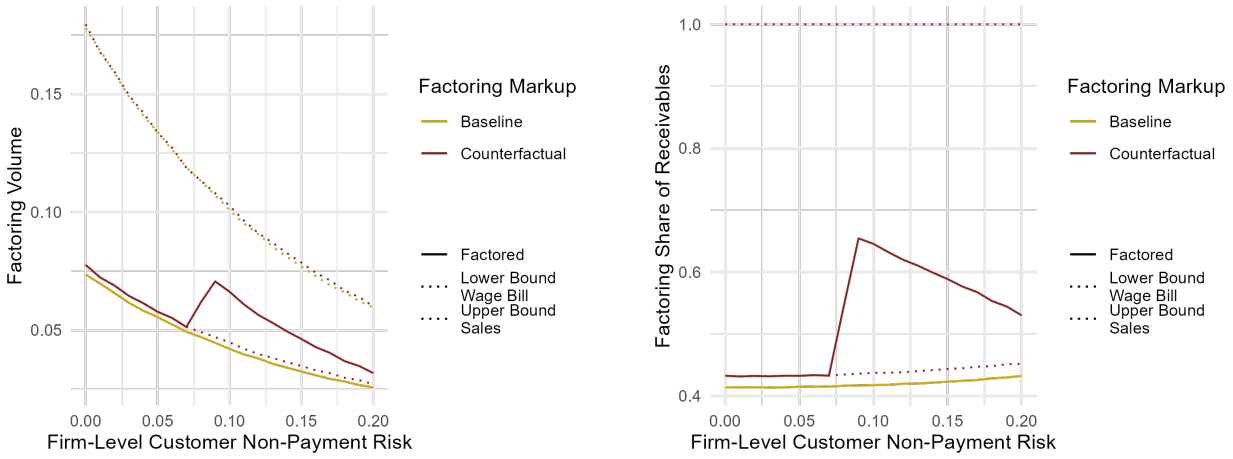
- $\eta = 0.25$ is the cost of default, following Glover (2016).

2.8.3 Additional Model Results

Figure B7 and Figure B8 compare the distributions of outcomes across the baseline equilibrium, with $\mu^F = 1.13$, and the counterfactual equilibrium, with perfect competition $\mu^F = 1$ between factors.

Figure B7 shows that the model can replicate the empirical factoring summary statistic in Figure 2.4b, that firms with moderately low credit score factor the largest share of receivables. On the left, Figure B7a shows the absolute amount factored, as well as the lower bound in equation (2.10) and the upper bound in equation (2.9) across the distribution of factoring risk ζ_j on the horizontal axis. On the right, Figure B7b shows the same outcomes normalized by the upper bound of morning receivables $p_j y_{j0}$.

Figure B7: Factoring Demand over the Factoring Risk Distribution



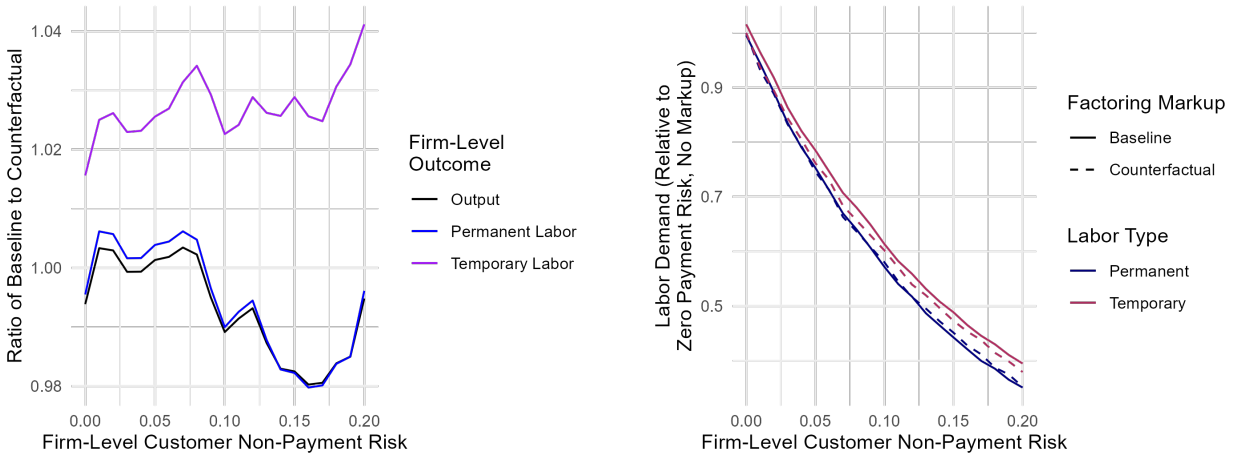
(a) Absolute Factoring

(b) Relative Factoring

Notes: These figures show the model-implied factoring demand in absolute terms (left) across the distribution of firm risk ζ_j, as well as in relative terms compared to the morning receivables $p_j y_{j0}$ (right). In gold are the baseline values, with $\mu^F = 1.13$, and in red are the counterfactual outcomes, with $\mu^F = 1$.

Figure B8 compares outcomes between the baseline and counterfactual equilibria, across the distribution of factoring risk ζ_j on the horizontal axis. In Figure B8a, a value of 1 indicates a given firm has the same outcome between the two equilibria, while a value above 1 indicates the outcome has higher value in the baseline equilibrium, where factoring spreads are higher. Figure B8a shows that all firms have greater demand for temporary labor in the baseline versus the counterfactual equilibrium, but the difference across equilibria is greater for the riskier firms. By comparison, output and permanent labor demand are higher in the counterfactual equilibrium for riskier firms but not the less risky firms. Figure B8b shows the normalized value of labor demand between the baseline and counterfactual equilibria, showing that as the factoring spread increases, moving from the dotted to the dashed line, permanent labor demand decreases and temporary labor demand increases. Also, the absolute decrease in labor demand as factoring risk ζ_j increases is greater for permanent labor than for temporary labor under both equilibria.

Figure B8: Distributional Comparison of Outcomes across Equilibria



(a) Ratio of Outcomes in across Equilibria

(b) Labor Demand

Notes: These figures show the model-implied outcomes across the distribution of factoring risk ζ_j. On the left are the ratios of firm-level outcomes between the baseline and counterfactual equilibria. A value of 1 indicates a given firm has the same outcome between the two equilibria, while a value above 1 indicates the outcome has higher value in the baseline equilibrium, where factoring spreads are higher. On the right are the values of permanent and temporary labor demand, where the counterfactual labor demand for both types of labor for the firms with $\zeta_j = 0$ is normalized to 1.

2.8.4 Invoice Network Regressions

We previously showed that firm-level characteristics (node as opposed to edge) do not explain most of the variation in trade credit contracts. To help us classify firms' exposure to other firms for the spillover regression, we use the edge-level boleto data, which we call the "invoice network." Let $y_{j,k,t}$ be a payment or trade credit variable from seller j to buyer k in year t:

$$y_{j,k,t} = \alpha_j + \alpha_k + \alpha_t + \gamma X_{j,k} + \epsilon_{j,k,t}$$

The highlighted coefficients are γ, either interactions of buyer and seller characteristics, or characteristics (like relationship length) that vary at the edge level. This is restricted to the set of around 500k firms whose receivables are ever purchased by funds, which is further restricted by the 250k firms that engage in firm-to-firm transactions (all other firms are aggregated into a single composite firm. We aggregate to the year level for computational efficiency. We classify firms by whether the payment inflow (across all periods) is above median, "large," or otherwise "small."

The takeaways are that relative to the base category of small sellers with small buyers, trade credit is offered more often from large sellers: 5.8 percentage points more than small sellers to small buyers. The final column only serves as context for the magnitudes: there are 25% fewer payments from small sellers to big buyers, 36% more payments from big sellers to big buyers, and 27% more payments from big sellers to small buyers.

2.8.5 Firm-to-Firm Spillovers

We now turn our attention to whether a given firm's factoring supply shock transmitted to other firms through the supply chain. There are two dimensions of supply chain spillovers: through buyers (who receive trade credit and purchase more) and sellers (who extend trade credit and have more sales). Using the earliest full year of firm-to-firm boleto transactions x_{j,k,t_0} in 2019,[15] with the notation j sells to k, we define firm k's exposure $e^S_{k,t}$ to its suppliers fund flow shocks $e_{j,t}$, and likewise firm j's exposure $e^B_{j,t}$ to its buyers' fund flow shocks $e_{k,t}$:

$$e^S_{k,t} = \sum_j \underbrace{\frac{x_{j,k,t_0}}{\sum_{k'} x_{j,k',t_0}}}_{\substack{\text{k's share of} \\ \text{j's sales}}} e_{j,t}, \qquad e^B_{j,t} = \sum_k \underbrace{\frac{x_{j,k,t_0}}{\sum_{j'} x_{j,k',t_0}}}_{\substack{\text{j's share of} \\ \text{k's intermediate inputs}}} e_{k,t}. \tag{2.28}$$

The regressions are

$$y_{j,t} = \alpha_j + \alpha_t + e_{j,t} + e^B_{k,t} + e^S_{j,t} + \epsilon_{j,t},$$

Table 2.15 shows that there are no spillovers to revenue, expenditure, wages, debt issuance, nor interest rate.

[15] We only use the 2019 sales because subsequent years' transactions are endogenous to firms' decisions in responses to shocks; while FIDCs existed before 2019, our summary statistics show that FIDC purchases were a small share of factoring before 2019.

Table 2.15: Firm Network Spillover Regressions of Main Outcomes

	(1)	(2)	(3)	(4)	(5)
	Log Revenue	Log Expenditure	Log Wage Bill	Log Debt Issuance (All Debt)	IR (All Debt)
$e_{j,t}$	0.0075***	0.0047***	0.0008·	0.0212***	−0.2477***
	(0.0021)	(0.0007)	(0.0005)	(0.0010)	(0.0348)
$e^{S}_{j,t}$	−0.0000	0.0000	0.0000	−0.0000	−0.0010
	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0022)
$e^{B}_{j,t}$	−0.0020	0.0000	−0.0006·	0.0000	−0.0534
	(0.0018)	(0.0007)	(0.0004)	(0.0026)	(0.0459)
Num. Obs.	2,168,496	2,216,215	1,552,077	2,220,249	2,220,249
Num. Firms	124,396	125,949	102,016	126,028	126,028
Num. Months	62	62	48	62	62

Notes: ***p < 0.001; **p < 0.01; *p < 0.05; ·p < 0.1. All regressions use data from the BCB. All regressions use firm and month fixed effects, with standard errors clusted at the firm level. The explanatory variables are the fund flow shock for firm j, the indirect impulse from suppliers, and the indirect impulse from buyers, where indirect impulse equals the shock multiplied by sales and expenditure shares, respectively, calculated from the 2019 firm-to-firm payments network. The response variables are the log revenue, proxied by payment inflows, log intermediate input expenditure, proxied by payment outflows to firms, log wage bill, log debt issuance (amount borrowed) from all financing types, and the interest rate on all debt issuance. Note that 72.5% of all debt issuance has maturity under 1 year, and 49.9% of all debt issuance is factoring.

However, Table 2.16 shows that there are small impacts on trade credit: when firm j's buyers experience high demand for receivables,

Table 2.16: Firm Network Spillover Regressions of Trade Credit Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	Maturity Offer (Days)	Percentage Receive (%)	Log Value Receive	Maturity Receive (Days)	Percentage Offer (%)	Log Value Offer
$e_{j,t}$	0.0231	−0.0001	0.0047***	−0.0271·	0.0003	0.0049***
	(0.0160)	(0.0000)	(0.0009)	(0.0159)	(0.0002)	(0.0011)
$e^{S}_{j,t}$	−0.0003	−0.0000	−0.0001*	−0.0004	−0.0000*	0.0000
	(0.0002)	(0.0000)	(0.0000)	(0.0004)	(0.0000)	(0.0000)
$e^{B}_{j,t}$	−0.0070	−0.0002***	0.0003	−0.0237	−0.0005**	0.0006
	(0.0199)	(0.0000)	(0.0010)	(0.0162)	(0.0002)	(0.0016)
Num. Obs.	2,220,249	2,220,249	2,194,852	2,220,249	2,220,249	1,986,658
Num. Firms	126,028	126,028	125,458	126,028	126,028	114,664
Num. Months	62	62	62	62	62	62

Notes: All regressions use data from the BCB. All regressions use firm and month fixed effects, with standard errors clusted at the firm level. The explanatory variables are the fund flow shock for firm j, the indirect impulse from suppliers, and the indirect impulse from buyers, where indirect impulse equals the shock multiplied by sales and expenditure shares, respectively, calculated from the 2019 firm-to-firm payments network. The response variables are the firm by month level mean maturity of trade credit, offered and received, as well as the share of receivables with at least 15 days maturity, the effective lower bound for factoring.

*** p < 0.001; ** p < 0.01; * p < 0.05; · p < 0.1

Table 2.17: Firm Network Spillover Regressions of Number of Employee Outcomes

	(1)	(2)	(3)	(4)
	Log Number of Employees (Total)	Log Number of Employees (New Hire)	Log Number of Employees (Permanent)	Log Number of Employees (Temporary)
$e_{j,t}$	0.0006	0.0028^{**}	0.0018^{***}	-0.0027^{**}
	(0.0004)	(0.0009)	(0.0004)	(0.0011)
$e^{S}_{j,t}$	0.0000	0.0001	0.0000	−0.0000
	(0.0000)	(0.0001)	(0.0000)	(0.0000)
$e^{B}_{j,t}$	-0.0006^{*}	0.0043^{***}	0.0016^{***}	-0.0055^{***}
	(0.0003)	(0.0011)	(0.0004)	(0.0011)
Num. Obs.	1,558,883	737,313	1,554,656	425,453
Num. Firms	102,300	73,405	102,147	41,635
Num. Months	48	48	48	48

Notes: All regressions use data from the BCB. All regressions use firm and month fixed effects, with standard errors clusted at the firm level. The explanatory variables are the fund flow shock for firm j, the indirect impulse from suppliers, and the indirect impulse from buyers, where indirect impulse equals the shock multiplied by sales and expenditure shares, respectively, calculated from the 2019 firm-to-firm payments network. The response variables come from restricted access month-level RAIS data. An employee is defined as new if the employee began working at the firm this month and not in the previous year.

Table 2.18: Firm Network Spillover Regressions of Hours Employed Outcomes

	(1)	(2)	(3)	(4)
	Log Wage (Hourly)	Log Employment (Hours Worked)	Log Employment (Hours Worked by New Hires)	Log Employment (Hours Worked by Existing Emp.)
$e_{j,t}$	−0.0004	0.0006	0.0027**	0.0007·
	(0.0004)	(0.0004)	(0.0009)	(0.0004)
$e^{S}_{j,t}$	−0.0000	0.0000	0.0001	0.0000
	(0.0000)	(0.0000)	(0.0001)	(0.0000)
$e^{B}_{j,t}$	0.0007*	−0.0007*	0.0040***	−0.0004
	(0.0003)	(0.0003)	(0.0011)	(0.0003)
Num. Obs.	1,551,880	1,557,393	735,744	1,547,853
Num. Firms	102,014	102,241	73,344	101,843
Num. Months	48	48	48	48

Notes: All regressions use data from the BCB. All regressions use firm and month fixed effects, with standard errors clusted at the firm level. The explanatory variables are the fund flow shock for firm j, the indirect impulse from suppliers, and the indirect impulse from buyers, where indirect impulse equals the shock multiplied by sales and expenditure shares, respectively, calculated from the 2019 firm-to-firm payments network. The response variables come from restricted access month-level RAIS data. An employee is defined as new if the employee began working at the firm this month and not in the previous year.

Table 2.19: Firm Network Spillover Regressions of Debt Issuance Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	Log Debt Issuance (Debt Under 1 Year)	Log Debt Issuance (Debt Over 1 Year)	Log Debt Issuance Factoring All Issuance	Log Debt Issuance Credit Line (Unsecured)	Log Debt Issuance Credit Line (Secured)	Log Debt Issuance (Loans Over 1 Year)
$e_{j,t}$	0.0217***	−0.0005	0.0229***	−0.0015	0.0012	−0.0014
	(0.0010)	(0.0027)	(0.0011)	(0.0015)	(0.0018)	(0.0023)
$e^{S}_{j,t}$	0.0000	−0.0008	−0.0001	−0.0006*	−0.0000	−0.0009
	(0.0001)	(0.0008)	(0.0001)	(0.0003)	(0.0005)	(0.0006)
$e^{B}_{j,t}$	−0.0025	0.0052	0.0005	−0.0156***	−0.0104*	0.0037
	(0.0024)	(0.0050)	(0.0025)	(0.0039)	(0.0046)	(0.0047)
Num. Obs.	2,220,249	316,676	2,220,249	568,378	308,534	268,596
Num. Firms	126,028	47,166	126,028	47,500	26,213	44,143
Num. Months	62	62	62	62	62	62

Notes: All regressions use data from the BCB. All regressions use firm and month fixed effects, with standard errors clusted at the firm level. The explanatory variables are the fund flow shock for firm j, the indirect impulse from suppliers, and the indirect impulse from buyers, where indirect impulse equals the shock multiplied by sales and expenditure shares, respectively, calculated from the 2019 firm-to-firm payments network. The response variables are log debt issuance by category of debt. Column 1 is the subset with maturity of up to 365 days. Column 2 is factoring. Column 3 and 4 are unsecured and secured credit lines, where issuance is defined as any drawdown of the credit line, not a change in the credit limit. Column 5 is loans with maturity of over 365 days. 89% of loans have maturity of over 1 year. Across all firms in Brazil, loans are the highest issuance form of long-term debt, and bonds are second highest.

Table 2.20: Firm Network Spillover Regressions of Interest Rate Outcomes

	(1)	(2)	(3)	(4)	(5)
	IR (Debt Under 1 Year)	IR (Debt Over 1 Year)	IR Credit Line (Unsecured)	IR Credit Line (Secured)	IR (Loans Over 1 Year)
$e_{j,t}$	-0.2641^{***}	0.2561^{**}	-0.7602^{***}	0.0413	0.0084
	(0.0351)	(0.0789)	(0.1188)	(0.0425)	(0.0176)
$e^{S}_{j,t}$	-0.0010	-0.0018	0.0007	0.0242	-0.0007
	(0.0022)	(0.0060)	(0.0217)	(0.0167)	(0.0026)
$e^{B}_{j,t}$	-0.0702	0.1226	-1.2077^{***}	0.1598^{*}	$0.0439^{\cdot}$
	(0.0465)	(0.0854)	(0.3058)	(0.0721)	(0.0256)
Num. Obs.	2,220,249	316,676	568,378	308,534	268,596
Num. Firms	126,028	47,166	47,500	26,213	44,143
Num. Months	62	62	62	62	62

Notes: All regressions use data from the BCB. All regressions use firm and month fixed effects, with standard errors clusted at the firm level. The explanatory variables are the fund flow shock for firm j, the indirect impulse from suppliers, and the indirect impulse from buyers, where indirect impulse equals the shock multiplied by sales and expenditure shares, respectively, calculated from the 2019 firm-to-firm payments network. The response variables are the interest rates by category of debt. Column 1 is the subset with maturity of up to 365 days. Columns 2 and 3 are unsecured and secured credit lines, where issuance is defined as any drawdown of the credit line, not a change in the credit limit. Column 4 is loans with maturity of over 365 days.

Table 2.21: Firm Network Spillover Regressions of Default Rate Outcomes

	(1)	(2)	(3)	(4)
	Default Rate Rec. Factoring (to Banks, %)	Default Rate Other (to Banks, %)	Default Rate Rec. Factoring (to FIDCs, %)	Default Rate Other (to FIDCs, %)
$e_{j,t}$	-0.0172^{**}	0.0064	-0.2246^{**}	-0.0059
	(0.0063)	(0.0170)	(0.0806)	(0.0085)
$e^{S}_{j,t}$	0.0001^{*}	-0.0003	0.0003	-0.0001
	(0.0000)	(0.0004)	(0.0016)	(0.0006)
$e^{B}_{j,t}$	0.0128^{***}	0.0124	-0.0048	-0.0184^{*}
	(0.0032)	(0.0130)	(0.0427)	(0.0087)
Num. Obs.	1,698,059	1,698,059	762,576	762,576
Num. Firms	85,474	85,474	70,938	70,938
Num. Months	62	62	61	61

Notes: $^{***}p<0.001$; $^{**}p<0.01$; $^{*}p<0.05$. All regressions use data from the BCB. All regressions use firm and month fixed effects, with standard errors clusted at the firm level. The explanatory variables are the fund flow shock for firm j, the indirect impulse from suppliers, and the indirect impulse from buyers, where indirect impulse equals the shock multiplied by sales and expenditure shares, respectively, calculated from the 2019 firm-to-firm payments network. The response variables are the default rates on recourse factoring and other debt issued by banks and FIDCs. The default rate is defined to be the percentage of debt that was not paid on its due date; note that this is lower than the percentage of debt that the creditor eventually collects. The issuance-weighted default rate for recourse factoring to banks is 0.40% and the issuance-weighted default rate for recourse factoring to FIDCs is 10.3%.

Table 11: If my buyers receive a lower interest rate because of their fund flow shock, then I actually receive a higher interest rate from FIDCs even though the buyer is less likely to default, but not from banks. The FIDC does not observe the firm network, and any given FIDC is unlikely to be large enough to finance a firm and a large share of its buyers, so the FIDC does not know that the buyer shock occurred. Columns 1-3 are a puzzle.

Table 12 shows that total revenue and expenditure, and Table 13 shows that total trade credit offered and received, do not respond to the buyer shock. But we know from the main reg tables that the buyers buy more and receive more trade credit (with slightly shorter maturity but otherwise trade credit terms unchanged), so there is a composition change in which firms sell more to positively shocked buyers and less to other buyers. Unclear to me how this would affect the FIDC's pricing, since the firm's total receivables to sell to the FIDC are unchanged.

Column 4 of Table 11, banks' interest rate, makes sense because - Firms tend to borrow from boleto registration banks, so the banks partially observe the change in payment flows - If my suppliers or buyers get positively shocked, I receive a lower interest rate because some of the

positive impact spills over to me. More so for the buyers, which is what I expect because recourse factoring risk still depends partly on the buyer, and a positively shocked buyer is less risky (and we know from Table 8 that default rate is causally increasing in interest rate)

The main challenge with spillovers for IV regression is that there is no reasonable pair of x and y variables for which the z-variables satisfy the exclusion restriction. Intuitively, buyer exposure $e^{B}_{k,t}$ is primarily through the receipt of trade credit from shocked suppliers and the linked intermediate input expenditure, while seller exposure $e^{S}_{k,t}$ is primarily through extending trade credit and sales to shocked buyers. Neither of these operate through the factoring interest rate channel.

Chapter 3

Quantile Mixture Models: Estimation and Inference

This chapter is jointly authored with Luis Alvarez.[1]

3.1 Introduction

Consider the model:

$$Q_\mu(u) \coloneqq \int Q_\theta(u)\mu(d\theta)\,, u \in [0,1]\,, \tag{3.1}$$

where $\{Q_\theta : \theta \in \Theta\}$ is a known family of *quantile functions* indexed by a a measurable space (Θ, Σ), and μ is an unknown (signed) measure on (Θ, Σ) such that the resulting Q_μ is a quantile function. Equation (3.1) defines a *nonparametric quantile mixture model*. Such models have a wide range of statistical applications. In causal inference, they can be used to assess the distributional effects of aggregate shocks, by constructing a counterfactual quantile function for an exposed unit based on the quantile function of non-treated units, as in the *distributional synthetic controls* (Gunsilius 2023). In financial applications, a mixture of suitable quantile basis functions can be used to extrapolate the tails of an asset return distribution (Gourieroux and Jasiak 2008; Karvanen 2006). Finally, and as we further argue below, nonparametric quantile mixture models can be seen as a convenient tool in the estimation of *nonparametric density mixtures*, a class of models which has received increasing attention in Econometrics. Finally, and as we further argue below, nonparametric quantile mixture models can be seen as a convenient tool in the estimation of *nonparametric density mixtures*, a class of models which has received increasing attention in Econometrics and Statistics (Armstrong, Kolesár, and Plagborg-Møller 2022; Gu and Koenker 2023; Ignatiadis and Wager 2022; Kline, Rose, and Walters 2022).

The main tool in estimating nonparametric *density* mixture models is the nonparametric maximum likelihood estimator (NPMLE) of Kiefer and Wolfowitz 1956. In spite of its attractive statistical properties (Polyanskiy and Wu 2020), this approach suffers from inferential and

[1]We would like to thank Alberto Abadie, Isaiah Andrews, Bruno Ferman, Ricardo Masini, Eduardo Mendes, Anna Mikusheva, Stephen Morris, Whitney Newey, Christopher Palmer, Vitor Possebom, Robert Townsend, Jaume Vives, Henry Zhang, and seminar participants at MIT and FGV-EESP for their useful comments and suggestions. Any errors are our sole responsibility. Alvarez gratefully acknowledges financial support from Fapesp grant 2022/16691-1.

computational hurdles. The NPMLE still lacks a formal frequentist inferential theory (Ignatiadis and Wager 2022). Computationally, the nonconvexity of the optimization program imposes challenges, which have estimulated several attempts at computing approximate solutions (L. Feng and Dicker 2018; Jagabathula, Subramanian, and Venkataraman 2020; Koenker and Mizera 2014; Train 2008). Motivated by these concerns, this paper aims to introduce nonparametric quantile mixture models as an attractive counterpart to density mixtures.

We begin by defining nonparametric quantile mixtures. We show that, similarly to nonparametric density mixtures, nonparametric quantile mixtures enjoy interesting approximation properties, being able to approximate sufficiently well-behaved quantile functions. We then develop a framework for conducting estimation and inference on nonparametric quantile mixture models. Our proposed estimator is a sieve-like version of the generalized method of L-moments estimator (GMLM) of Alvarez, Chiann, and Morettin 2023. Introduced by Hosking 1990, L-moments are robust alternatives to standard moments that characterize distributions with finite first moment. In our setting, the proposed estimator amounts to finding mixture weights that minimize a weighted distance between sample and theoretical L-moments. When these weights are constrained to belong to a convex set, this amounts to solving a quadratic program with convex constraints, which can be performed efficiently in standard statistical software. Another interesting feature of the GMLM is that, in parametric and some semiparametric settings, this approach to estimation has been shown to perform well in finite samples, whilst still retaining some notion of asymptotic efficiency (Alvarez and Biderman 2022; Alvarez, Chiann, and Morettin 2023).

Building upon our proposed estimator, we establish an inferential procedure for mixture weights and functionals thereof based on a novel bootstrap for quadratic minimizers, which may be of independent interest. Our sieve-like approach allows for the number of basis functions to be a function of the sample size. For valid inference, we rely an undersmoothing condition, whereby the number of basis functions used in the mixture is sufficiently large so as to control approximation bias. Alternatively, if one is willing to place restrictions on the true quantile process, we may replace the undersmoothing condition with bias-aware inference, whereby bounds on the approximation bias are computed to conduct conservative inference (Armstrong and Kolesár 2021; Ignatiadis and Wager 2022; Noack and Rothe 2021). We also note that our approach to inference explicitly allows for regularization, which may be preferable to *ad-hoc* selection methods (Masini 2022).

Inference in our setting is challenging, as it is generally not possible to find explicit limit distributions for estimators when the number of parameters diverges (Chernozhukov, Chetverikov, and Kato 2017a) and existing bootstrap methods are not applicable when we might have solutions at the boundary of the parameter set Θ (see Chernozhukov, Chetverikov, Kato, and Koike 2023 for a recent survey). Our analysis addresses these challenges by relying upon a novel strategy that combines the convexity arguments originally applied to (low-dimensional) M-estimators with convex objective functions (Kato 2009; Pollard 1991), with an approach to inference based on the concept of *strong approximation*. Strong approximations have been increasingly used to conduct inference in nonstandard or high-dimensional settings (Chernozhukov, Chetverikov, and Kato 2014; Chernozhukov, Chetverikov, Kato, and Koike 2023; Fang et al. 2023).

Key to our strategy is a novel result that bounds the strong approximation of quadratic

minimizers in terms of the constituent elements of the program, along with a restricted eigenvalue condition. Building upon a lemma in Fan, Masini, and Medeiros 2022, we show that, when combined with anticoncentration inequalities, our results enable us to derive high-level sufficient conditions that ensure asymptotic validity of our proposed inferential approach. Whilst we apply these tools to inference in quantile mixture models, we highlight that they may have potential applications on other problems of interest such as LASSO and Ridge regression.

We then show how the analysis can be specialized under additional assumptions on the smoothness of the quantile functions, thus providing sufficient rates for inference in specific applications. For example, in an unconstrained estimation setting, and under some assumptions, our results show that the number of mixture weights can grow much faster than the sample size. When considering the case of ridge regularization, we show that the number of mixtures used (p) and the sample size (n) must satisfy, up to logarithmic factors, $p/n \to 0$ a rate similar to the one obtained by Belloni et al. 2015 in an (unconstrained) sieve-regression setting. In order to demonstrate the latter result, we develop an extension of the anticoncentration inequalities for Gaussian Random variables available in the literature to the projection of Gaussian random variables onto Euclidean balls, another result of independent interest.

We then consider two applications of our quantile mixtures. First, we show how they can be used to recover estimates (and conduct inference) in nonparametric **density** mixture models. Key to this approach is the relation between the derivative of a quantile function and the density of a random variable, which we use to "invert" our quantile mixtures onto density mixtures.

As a second application, we show that, as a direct byproduct of our theory, we are able to provide a valid inferential method in the distributional synthetic control setup of Gunsilius 2023. The crucial condition to the validity of our approach in this environment is a weight "dilution" condition, which requires the oracle weights attached to each control to spread sufficiently quickly. A similar condition exists in the literature on synthetic controls (Ferman 2021). To the best of our knowledge, formal inference methods were not yet available in the distributional synthetic control setting.[2]

Finally, to illustrate the usefulness of our method, we provide an empirical application on the distributive effects of a large-scale environmental disaster, the Brumadinho Barrage rupture in Minas Gerais, Brazil, on the local wage distribution. The impact of this the disarter is an open question as its effects remain uncertain, with potential negative effects mitigated by reparations, investments, and economic recovery efforts. The main empirical challenge lies in the fact that the rupture affected both the city of Brumadinho and neighboring municipalities simultaneously. This simultaneous shock hinders the availability of a clear control group, making direct comparisons difficult and motivating the use of the distributional synthetic control method, especially since there is a large pool of potential control units (other non-affected municipalities). Our results uncover a range of effects across percentiles of the wage distribution, which we argue are consistent with *displacement effects*, whereby median-earning jobs are replaced by low-paying contracts.

[2]Gunsilius 2023 suggests using placebos to assess uncertainty in his setup, though he does not provide a formal justification to it.

Overview The remainder of the paper is organized as follows. In Section 3.2 we define nonparametric quantile mixtures and discuss its approximation properties, as well as its connection with the distributional synthetic controls of Gunsilius 2023. Section 3.3 introduces our estimation procedure. The asymptotic inferential theory for mixture weights and linear functionals thereof is presented in Section 3.4. Section 3.5 presents the algorithm to implement our bootstrap procedure. In Section 3.6, we discuss applications to the empirical Bayes and distributional synthetic control problems. Section 3.7 presents the results of our empirical application. Section 3.8 concludes. The Appendices contain the proofs of the main results, as well as our lemma on the strong approximation of a quadratic minimizer and the anti-concentration inequality of the projection of Gaussian random variables onto Euclidean balls. We also discuss the choice of tuning parameters in our bootstrap procedure, present additional results on the approximation properties of quantile mixtures and discuss further details on the empirical application.

3.2 Quantile mixture models

In this section, we define a quantile mixture and discuss its approximation properties. We begin by recalling the definition of a quantile function.

Definition 2. *A function* $Q : [0,1] \mapsto \mathbb{R} \cup \{-\infty, \infty\}$ *is a* quantile function *if it is nondecreasing and continuous on the left.*

As it is well known, given a distribution function $F : \mathbb{R} \mapsto [0,1]$, the *generalized inverse*

$$Q_F(u) \coloneqq \inf\{x \in \mathbb{R} : F(x) \geqslant u\}, \quad u \in [0,1].$$

defines a quantile function.

We are now ready to define a quantile mixture.

Definition 3. *Given a family of quantile functions* $\{Q_\theta : \theta \in \Theta\}$ *indexed by a a measurable space* (Θ, Σ)*, and a signed measure* μ *on* (Θ, Σ)*, the map:*

$$Q_\mu(u) \coloneqq \int Q_\theta(u)\mu(d\theta), u \in [0,1],$$

defines a quantile mixture *if* $Q_\mu(u)$ *exists as an extended real number for every* $u \in [0,1]$ *and the resulting* Q_μ *is a quantile function.*

A quantile mixture combines a family of quantile functions through a signed measure μ. Notice that our definition imposes that the resulting mixture is itself a quantile function, which is a desirable feature in our main applications. We also remark that, if μ is a *measure*, then any well-defined Q_μ is necessarily a quantile mixture.

Definition 4. *A quantile mixture model is a pair* $(\mathcal{G}, \mathcal{M})$*, where* $\mathcal{G}$ *is a family of quantile functions indexed by a measurable space* (Θ, Σ)*, and* $\mathcal{M}$ *is a subset of*

$$\{\mu \text{ is a signed measure on } (\Theta, \Sigma) : Q_\mu \text{ is a quantile function}\}.$$

It is well known that some density mixture models enjoy great approximation properties, being able to approximate quite general classes of densities with arbitrary error under specific norms (see H. D. Nguyen and McLachlan 2019 and T. T. Nguyen et al. 2020 for results on mixtures of densities from location-scale families). We conclude this section by providing examples of quantile mixture models that similarly exhibit interesting approximation properties.

Example 1 (Polynomial quantiles). *For $n \in \mathbb{N}$, define the vector of polynomials:*

$$J_n(u) := \begin{pmatrix} 1 \\ u \\ u^2 \\ \vdots \\ u^n \end{pmatrix}.$$

It follows from monotone approximation theory (Shvedov 1981) that, for any $p \in [1, \infty]$ and quantile function $Q \in L^p[0,1]$, and for every $n \in \mathbb{N}$, there exists $\theta_n^* \in \mathbb{R}^n$ such that ${\theta_n^*}' J_n$ is a quantile function and:

$$\|Q(\cdot) - {\theta_n^*}' J_n(\cdot)\|_{L_p[0,1]} \leqslant C\omega_{2,p}(Q, n^{-1}),$$

where $C > 0$ is an absolute constant and $\omega_{2,p}(Q, \delta) = \sup_{0<h\leqslant\delta} \|Q(\cdot + 2\delta) - Q(\cdot + \delta) - (Q(\cdot + \delta) - Q(\cdot))\|_{L_p[0,1]}$ is the second modulus of continuity. If Q is absolutely continuous, with density $q \in L^p[0,1]$, then we can further show that DeVore and Lorentz 1993:

$$\|Q(\cdot) - {\theta_n^*}' J_n(\cdot)\|_{L_p[0,1]} \leqslant Cn^{-1}\omega_p(q, n^{-1}),$$

where $\omega_p(q, \delta) = \sup_{0<h\leqslant\delta} \|q(\cdot + \delta) - q(\cdot)\|_{L_p[0,1]}$ is the (first) modulus of continuity.

Example 2 (Pareto mixtures). *Consider the class of generalized Pareto distributions with shape parameter $k \in \mathbb{R}$. In this case, the associated quantile functions are given by (Hosking and Wallis 1987):*

$$Q_k(u) = \begin{cases} \frac{(1-(1-u)^k)}{k}, & \text{if } k \neq 0 \\ -\log(1-u), & \text{if } k = 0 \end{cases}.$$

Since $\log(1-x)$ is analytic on $[0,1]$, we are able to show that the class of quantile mixtures on the Pareto family is able to reproduce any polynomial on $[0,1]$. It then follows from Example 1 that, for a given $p \in [1, \infty]$ and every quantile function $Q \in L^p[0,1]$, and for every $\epsilon > 0$, there exists a quantile mixture Q_{μ^*} on the family of Pareto distributions such that:

$$\|Q(\cdot) - Q_{\mu^*}(\cdot)\|_{L^p[0,1]} \leqslant C\omega_{2,p}(Q, \epsilon).$$

Example 3 (Extreme Value mixtures). *Consider the class of generalized extreme value distributions with shape parameter $k \in \mathbb{R}$. In this case, the associated quantile functions are given by (Hosking, Wallis, and Wood 1985):*

$$Q_k(u) = \begin{cases} \frac{(1-(-\log u)^k)}{k}, & \text{if } k \neq 0 \\ -\log(-\log u), & \text{if } k = 0 \end{cases}.$$

Notice that mixtures on this class are able to reproduce any polynomial of $-\log(u)$*. Consequently, for a given* $p \in [1, \infty]$*, every* $\underline{u} \in (0,1)$ *and quantile* $Q \in L^p[0,1]$*, we have that, for every* $\epsilon > 0$*, there exists a quantile mixture* Q_{μ^*} *such that:*

$$\|Q(\cdot) - Q_{\mu^*}(\cdot)\|_{L^p[\underline{u},1]} \leqslant C\omega_{2,p}(Q, \epsilon).$$

By assuming that the lower tail of Q *asymptotically behaves as a member of the extreme value family, we can then extend this approximation to the entire interval* $[0,1]$*.*

Example 4 (Approximation through nonnegative measures)**.** *The previous examples rely on possibly signed measures to construct quantile mixtures that arbitrarily approximate a target quantile function. One question is whether similar approximations can be constructed by relying solely on nonnegative measures. In Appendix 3.12, we construct an example of a quantile mixture model that achieves general approximation properties under nonnegative weighting under additional assumptions on the quantile function.*

Example 5 (Distributional synthetic controls)**.** *Consider a population of* $J+1$ *units. Let* $Q_{j,t}$ *denote the quantile function of the distribution of an outcome of interest in unit* j *at period* t*. For example,* $Q_{j,t}$ *may denote the quantile function of the wage distribution in state* j *at year* t*, with* $(J+1)$ *being the number of states in the country of analysis. Suppose that a policy is implemented at state* $j = 0$ *beginning at period* t^**. Gunsilius 2023 provides conditions under which, if there exists a set of weights* $(w_j)_{j=1}^J \in \Delta^{J-1}$ *such that:*

$$Q_{0,t^*-1} = \sum_{j=1}^{J} w_j Q_{j,t^*-1},$$

then the quantity,

$$\tilde{Q}_{0t} \coloneqq \sum_{j=1}^{J} w_j Q_{jt}, \quad t \geqslant t^*,$$

yields a valid counterfactual for the quantile function at $j = 0$ *in the absence of treatment. Such* distributional synthetic controls *may be used to assess the effect of policies on the entire distribution of an outcome of interest.*

3.3 Proposed estimator

Observation: In what follows, all random variables are defined in the same probability space $(S, \mathcal{S}, \mathbb{P})$.

In this section, we introduce an estimation procedure for quantile mixture models. Specifically, given a sample estimator $\hat{Q}_n$ of a target quantile function Q, we propose to estimate mixture weights by solving:

$$\hat{\theta}_n \in \operatorname{arginf}_{\theta\in\Theta_p}\left(\int_{\underline{p}_n}^{\overline{p}_n}(\hat{Q}_n(u)-\theta'\boldsymbol{J}_{p,n}(u))\boldsymbol{P}_L(u)du\right)' W_{L,n}\left(\int_{\underline{p}_n}^{\overline{p}_n}(\hat{Q}_n(u)-\theta'\boldsymbol{J}_{p,n}(u))\boldsymbol{P}_L(u)du\right), \tag{3.2}$$

where $\boldsymbol{J}_{p,n}(u) = (J_{1,n}(u), J_{2,n}(u), \dots J_{p,n}(u))'$ is a vector of p quantile functions; Θ_p is a convex subset of $\mathbb{R}^p$; $\boldsymbol{P}_L(u) = (P_1(u), P_2(u), \dots, P_L(u))'$ is a vector of L quantile weighting functions, with $\{P_l\}_{l\in\mathbb{N}}$ forming an orthonormal basis on $L^2[0,1]$; $W_{L,n}$ is an $L \times L$ symmetric positive semidefinite weighting matrix; and $0 \leqslant \underline{p}_n < \overline{p}_n \leqslant 1$ are trimming constants. Estimator (3.2) is a sieve-like version of the generalized method of L-moments (GMLM) estimator in Alvarez, Chiann, and Morettin 2023. Introduced by Hosking 1990, the r-th L-moment of a distribution function F is defined as $\lambda_r := \int_0^1 Q_F(u)P^*_{r-1}(u)du$, with P^*_l being the l-th shifted Legendre polynomial on $[0,1]$. L-moments provide robust alternatives to standard moments; they also characterize any distribution function with finite first moment. In a parametric setting, Alvarez, Chiann, and Morettin 2023 show that, under some conditions, an optimally-weighted GMLM estimator is asymptotically efficient as the sample size and the number of L-moments used in estimation (L) diverge. Moreover, by properly choosing L as a function of the sample size, they show that it is possible to improve (in a mean-squared error sense) over maximum likelihood estimation in finite samples.[3]

In light of its attractive statistical properties, we propose to use a GMLM approach in the estimation of quantile mixture models. We also note that, in our setting, such approach is highly advantageous from the computational viewpoint. Indeed, (3.2) is a quadratic program with convex constraints, which can be solved efficiently. Possible choices of Θ_p include:

1. **unconstrained mixtures:** $\Theta_p = \mathbb{R}^p$;
2. **ridge regularization:** $\Theta_p = B_{(\mathbb{R}^p, \|\cdot\|_2)}(0, M)$;
3. **lasso regularization:** $\Theta_p = B_{(\mathbb{R}^p, \|\cdot\|_1)}(0, M)$;
4. **nonnegative weights:** $\Theta_p = \mathbb{R}^p_+$; and
5. **simplex weights:** $\Theta_p = \Delta^{p-1}$.

In our setup, the quantile "series" functions $\boldsymbol{J}_{p,n}$ can be either nonstochastic, as in usual series estimation, or stochastic, as in distributional-synthetic-control-type applications. The weighting matrix $W_{L,n}$ is possibly estimated – we discuss possible choices of weights later on. Finally, we allow for the possibility of trimming, by specifying constants $0 \leqslant \underline{p}_n < \overline{p}_n \leqslant 1$, as this may be useful in applications with heavy-tailed data and in some extrapolation exercises.

In the next sections, we provide a valid inferential theory on the estimands $\theta_{0,n}$ of $\hat{\theta}_n$, and functionals thereof, as n and (possibly) p and L diverge. Our theory is purposefully generic, in order to accommodate the different types of applications we have in mind. Conceptually, we rely on strong approximations of the quantile function $\hat{Q}_n$ to a Gaussian process in order to approximate

[3]This approach to estimation has also been shown to produce efficient estimators in a class of semiparametric models of treatment effects (Alvarez and Biderman 2022).

the distribution function of $\hat{\theta}_n$ with the distribution of a minimiser of (3.2) where $\hat{Q}_n$ is replaced by a Gaussian random variable. Strong approximations of the empirical quantiles of a scalar sample with common marginal distribution have been derived in the iid setting by Csorgo and Revesz 1978, and extended to the stationary mixing setting by Fotopoulos and Ahn 1994 and Yoshihara 1995. We speculate that these results may be combined with statistical learning techniques, such as sample splitting and debiasing (Chernozhukov, Chetverikov, Demirer, et al. 2018), to construct valid strong approximations to more complex quantile estimators, such as the empirical quantiles obtained from the prediction errors of a first-step algorithm, which may be useful in risk management applications (see Section 3.8 for further discussion). While our theory is ample enough to accommodate any such approximation, we do not pursue the construction of this type of coupling in our paper, as we do not require it in our main applications.

3.4 Inferential approximation

3.4.1 Inference on mixture weights

We begin by presenting general inferential results on the distributional approximation of our GMLM estimator. We then specialize to specific bases and choice sets. We implicitly index p and L by n, thus allowing these quantities to diverge with n, and consider limits as $n \to \infty$. In what follows, we work under the high-level assumption:

Assumption 4 (Existence of tight Gaussian Approximation)**.** *There exists a sequence of zero-mean Gaussian processes, $B_{0,n}$, $n \in \mathbb{N}$, defined on $(S, \mathcal{S}, \mathbb{P})$ and indexed by $[0,1]$, such that, as $n \to \infty$:*

$$\|\sqrt{n}(\hat{Q}_n - Q) - B_{0,n}\|_{L^2[\underline{p}_n, \overline{p}_n]} = O_{\mathbb{P}}(r_n)\,. \tag{3.3}$$

Moreover, this sequence of Gaussian processes is tight in $L^2_{[\underline{p}_n, \overline{p}_n]}$, in the sense that:

$$\|B_{0,n}\|_{L^2_{[\underline{p}, \overline{p}]}} = O_{\mathbb{P}}(1)\,.$$

As briefly mentioned in the previous section, Csorgo and Revesz 1978 derived strong approximation results in the case where $\hat{Q}_n$ are empirical quantiles from a random sample with quantile function Q. In this case, under the assumptions of their Theorem 6, one may take $0 = \underline{p}_n < \overline{p}_n = 1$ and have $r_n = n^{-1/2}\log(n)^{\alpha}$ for some $\alpha > 0$.[4] In this case, the Gaussian processes $B_{0,n}$ are zero-mean with common covariance kernel $\Gamma_0(i,j) = Q'(i)Q'(j)(i \wedge j - ij)$, which shows the sequence is tight in $L^2_{[\underline{p}_n, \overline{p}_n]}$.

We also need to control the possible estimation error of the weighting matrix $W_{n,L}$.

Assumption 5 (Estimation error of weighting matrices)**.** *There exists a sequence of nonstochastic positive semidefinite (psd) matrices $\Omega_{n,L}$, $n \in \mathbb{N}$, such that, as $n \to \infty$, $\|\Omega_{n,L}\|_2 = O(1)$ and $\|W_{n,L} - \Omega_{n,L}\| = O_{\mathbb{P}}(s_n)$.*

[4] The authors also have weaker results for trimming constants $\underline{p}_n$ ande $\overline{p}_n$ that converge, respectively, to 0 and 1 at a rate. These may be used in the sample-size-dependent trimming case.

We split our analysis into two cases: (i) nonstochastic basis; and (ii) stochastic basis.

Nonstochastic basis

In this section, the basis functions $\boldsymbol{J}_{n,p}$ are taken to be nonstochastic. Our target estimand is given by:

$$\theta_{0,n} \in \operatorname{arginf}_{\theta\in\Theta_n} \|Q(\cdot) - \theta'\boldsymbol{J}_{n,p}(\cdot)\|_{L^2[\underline{p}_n,\overline{p}_n]}, \tag{3.4}$$

and we note that we can write:

$$\sqrt{n}(\hat{\theta}_n - \theta_n) \in \operatorname{arginf}_{x\in\mathcal{X}_n} \left\| \int_{\underline{p}_n}^{\overline{p}_n} (\sqrt{n}(\hat{Q}_n(u) - Q(u)) - x'\boldsymbol{J}_{p,n}(u) + D_n(u))\boldsymbol{P}_L(u)du \right\|^2_{2,W_{L,n}}, \tag{3.5}$$

where, for a symmetric psd matrix A, $\|x\|_{A,2} = \sqrt{x'Ax}$, $\mathcal{X}_n = \sqrt{n}(\Theta_n - \theta_{0,n})$, and $D_n(u) = \sqrt{n}(Q(u) - \theta'_{0,n}\boldsymbol{J}_{n,p}(u))$ is the approximation bias. In light of representation (3.5) and Assumptions 4 and 5, we are led to consider the following approximation:

$$\sqrt{n}(\theta^*_n - \theta_{0,n}) \in \operatorname{arginf}_{x\in\mathcal{X}_n} \left\| \int_{\underline{p}_n}^{\overline{p}_n} (B_{0,n}(u) - x'\boldsymbol{J}_{p,n}(u) + D_n(u))\boldsymbol{P}_L(u)du \right\|^2_{2,\Omega_{L,n}}. \tag{3.6}$$

As we show later on, representation (3.6) may be used as a basis for inference, under some assumptions. The bias term $D_n(u)$, even though unknown, may be bounded by using approximation results such as Example 1 or identification assumptions such as in Example 5.[5] Alternatively, one may consider an "undersmoothed" approximation:

$$\sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n}) \in \operatorname{arginf}_{x\in\mathcal{X}_n} \left\| \int_{\underline{p}_n}^{\overline{p}_n} (B_{0,n}(u) - x'\boldsymbol{J}_{p,n}(u))\boldsymbol{P}_L(u)du \right\|^2_{2,\Omega_{L,n}}. \tag{3.7}$$

Proposition 2. *Suppose that* $d_n := \|D_n\|_{L^2[\underline{p}_n,\overline{p}_n]}$ *is bounded. For* $z \in \mathbb{R}^p$*, define the restricted eigenvalue around* z *as :*

$$\lambda_{z,n} := \inf_{s\in\mathcal{X}_n} \frac{(s-z)' \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du \right)' \Omega_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du \right)(s-z)}{\|s-z\|_2^2}. \tag{3.8}$$

Finally, let:

$$\rho_n := \lambda_{\max}\left(\int_{\underline{p}_n}^{\overline{p}_n} (\boldsymbol{J}_{n,p}(u)\boldsymbol{J}_{n,p}(u)')du \right).$$

We then have that:

[5] See Remark 6 below.

1. *Smoothened case: for any sequences* $(\delta_n)_{n\in\mathbb{N}}$, $(M_n)_{n\in\mathbb{N}}$ *and* $(c_n)_{n\in\mathbb{N}}$ *such that:*

$$\begin{aligned} M_n \frac{\lambda_{0,n}}{\sqrt{\rho_n}} &\to \infty\,, \\ \mathbb{P}[\lambda_{\sqrt{n}(\theta_n^*-\theta_{0,n}),n} \leqslant c_n] &\to 0\,, \\ \frac{c_n\delta_n^2}{(\delta_n+M_n)\sqrt{\rho_n}(r_n\vee s_n)} &\to \infty\,, \\ \frac{c_n\delta_n^2}{(\delta_n+M_n)^2\rho_n s_n} &\to \infty\,, \end{aligned} \tag{3.9}$$

we have that

$$\mathbb{P}[\|\sqrt{n}(\hat{\theta}_n-\theta_{0,n})-\sqrt{n}(\theta_n^*-\theta_{0,n})\| \geqslant \delta_n] \to 0\,.$$

2. *Undersmoothened case:for any sequences* $(\delta_n)_{n\in\mathbb{N}}$, $(M_n)_{n\in\mathbb{N}}$ *and* $(c_n)_{n\in\mathbb{N}}$ *such that:*

$$\begin{aligned} M_n \frac{\lambda_{0,n}}{\sqrt{\rho_n}} &\to \infty\,, \\ \mathbb{P}[\lambda_{\sqrt{n}(\theta_n^*-\theta_{0,n}),n} \leqslant c_n] &\to 0\,, \\ \frac{c_n\delta_n^2}{(\delta_n+M_n)\sqrt{\rho_n}(r_n\vee d_n\vee s_n)} &\to \infty\,, \\ \frac{c_n\delta_n^2}{(\delta_n+M_n)^2\rho_n s_n} &\to \infty\,, \end{aligned} \tag{3.10}$$

we have that

$$\mathbb{P}[\|\sqrt{n}(\hat{\theta}_n-\theta_{0,n})-\sqrt{n}(\tilde{\theta}_n^*-\theta_{0,n})\| \geqslant \delta_n] \to 0\,.$$

Proof. See Appendix 3.9.2. □

Proposition 2 provides rates of the approximation of (3.5) in terms of (3.6) (or (3.7)). It expresses these rates in terms of control of the norm of the approximating solution, as well as control of the restricted eigenvalues of the approximating program.[6] The proof of Proposition 2 is deferred to Appendix 3.9, where it follows from a lemma on the strong approximation of a minimizer of a quadratic program, which may be of independent interest.

In the next subsections, we will combine Proposition 2 with Lemma 3 below to provide valid inferential tools on mixture weights. This lemma bounds the Kolmogorov distance (in a subset of $\mathcal{B}(\mathbb{R}^p)$) between two random variables taking values in $\mathbb{R}^p$, in terms of two components: *strong approximation* and *anticoncentration*.

Lemma 3. *Let* X *and* Y *be two random variables on* $(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), \mathbb{P})$. *Then, for any* $\mathcal{C} \subseteq \mathcal{B}(\mathbb{R}^p)$, *we have that:*

$$\sup_{A\in\mathcal{C}} |\mathbb{P}[X\in A]-\mathbb{P}[Y\in A]| \leqslant \inf_{s\in[2,\infty]} \inf_{\delta\geqslant 0} \{\mathbb{P}[\|X-Y\|_2 \geqslant \delta] + \Psi_s(\mathcal{C};\delta)\}\,,$$

[6] Restricted eigenvalue conditions are common in the literature on high dimensional linear models (Bickel, Ritov, and Tsybakov 2009).

where $\Psi_s(\mathcal{C};\delta) := \sup_{A\in\mathcal{C}} \mathbb{P}[Y \in A_s^{\delta} \setminus A_s^{-\delta}]$, *where* $A_s^{\delta} = \{x \in \mathbb{R}^p : \inf_{a\in A}\|x-a\|_s \leqslant \delta\}$ *and* $A_s^{-\delta} = \mathbb{R}^p \setminus (\mathbb{R}^p \setminus A)_s^{\delta}$.

Proof. The result follows from Lemma S.14 in Fan, Masini, and Medeiros 2022, by observing that, for $x \in \mathbb{R}^p$, $\|x\|_s \leqslant \|x\|_2$ for any $s \in [2,\infty]$, and optimizing. □

By combining the above lemma with anticoncentration results on Gaussian processes and Proposition 2, we will be able to find sequences δ_n such that the Kolmogorov distance between the statistic and its approximation converges to zero. This is precisely what we need to achieve a valid distributional approximation.

Unconstrained case We begin by considering the unconstrained case $\Theta_p = \mathbb{R}^p$. In this case, the estimator admits a closed form solution given by:

$$\hat{\theta}_n = \left(\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)\right)^{-1} \times \left(\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L}\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\hat{Q}_n(u)du\right), \tag{3.11}$$

provided that the inverse exists. Similarly, for the undersmoothed approximation:

$$\tilde{\theta}_n^{*} = \left(\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' \Omega_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)\right)^{-1} \times \left(\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' \Omega_{n,L}\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\frac{B_{n,0}(u)}{\sqrt{n}}du\right), \tag{3.12}$$

In this case, the restricted eigenvalue collapses to the smallest eigenvalue, and we are able to show that:

Proposition 3. *Suppose that for constants* $C > 0$, $\gamma \in [0,1)$, *the smallest eigenvalue of* $\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' \Omega_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)$, $\lambda_{n,\min}$ *satisfies,* $\lambda_{n,\min} > C(\rho_n s_n)^{\gamma}$. *Then, if* $\rho_n s_n \to 0$, $\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)$ *is invertible with probability approaching one and:*

$$\|\sqrt{n}(\hat{\theta}_n - \theta_{0,n} - \boldsymbol{D}_n) - \sqrt{n}(\tilde{\theta}_n^{*} - \theta_{0,n})\| = O_{\mathbb{P}}((\rho_n s_n)^{1-2\gamma} \vee ((\rho_n s_n)^{-\gamma}\sqrt{\rho_n} r_n)),$$

where the bias term is given by:

$$\boldsymbol{D}_n = \left(\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u) \boldsymbol{J}_{n,p}(u)' du \right)' \Omega_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u) \boldsymbol{J}_{n,p}(u)' du \right) \right)^{-1} \times$$
$$\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u) \boldsymbol{J}_{n,p}(u)' du \right)' \Omega_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u) (Q(u) - \theta'_{0,n} \boldsymbol{J}_{n,p}(u)) du \right).$$

Proof. See Appendix 3.9.3. □

The previous proposition can be combined with anticoncentration results available in the literature to provide valid approximation in specific classes of subsets of $\mathcal{B}(\mathbb{R}^n)$. To illustrate, we consider the class of hyperrectangles, $\mathcal{C}_p = \{[\boldsymbol{a}, \boldsymbol{b}] : \boldsymbol{a}, \boldsymbol{b} \in \overline{\mathbb{R}}^p, \boldsymbol{a} \leqslant \boldsymbol{b}\}$. In this case, combining the previous proposition with Nazarov's inequality, see Theorem 1 of Chernozhukov, Chetverikov, and Kato 2017b, we obtain the following conclusion.

Corollary 2. *Suppose that the conditions in Proposition 3 hold. Suppose that the variance of* $\sqrt{n}(\tilde{\theta}^*_{j,n} - \theta_{j,0,n})$, $j = 1, \ldots, p$ *is bounded below uniformly in* j *and* n. *If, for some* $\nu > \frac{1}{2}$:

$$\log(p)^\nu \left((\rho_n s_n)^{1-2\gamma} \vee ((\rho_n s_n)^{-\gamma} \sqrt{\rho_n r_n}) \right) \to 0$$

then

$$\sup_{A \in \mathcal{C}_p} |\mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_{0,n} - \boldsymbol{D}_n) \in A] - \mathbb{P}[\sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n}) \in A]| \to 0 .$$

Proof. Nazarov's inequality applied to the random vector $\boldsymbol{v} = (\sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n})', -\sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n})')'$ implies that $\Psi_\infty(\mathcal{C}_p; \delta) \leqslant \frac{\delta}{\underline{\sigma}}(\sqrt{2\log(2p)} + 2)$, where $\underline{\sigma}^2$ is the lower bound on the variance of the approximation. The conclusion then follows from Proposition 3 and Lemma 3. □

Similar results can be obtained for other subclasses of sets, by relying on different anti-concentration results available in the literature. For example, if one wishes to provide valid approximations on the class of Euclidean balls of arbitrary center or convex subsets of $\mathbb{R}^p$; one could rely, respectively, on the anticoncentration results of Theorem 2.7 of Götze et al. 2019 or Lemma A.2 of Chernozhukov, Chetverikov, and Kato 2017a.

Remark 3 (Restrictions on L)**.** *Notice that our rates never depend on the number of L-moments* L. *This is similar to the results in Alvarez, Chiann, and Morettin 2023 in the parametric setting, where* L *may increase arbitrarily with* n. *It should be noted, however, that the assumption of the smallest eigenvalue of the matrix in Proposition 3 being positive requires at least* $L \geqslant p$.

Remark 4 (Specialization to polynomial quantiles)**.** *It is instructive to consider polynomial basis functions, i.e.* $\boldsymbol{J}_{n,p}(u) = (1, u, \ldots, u^{p-1})$. *In this case, results on the Hilbert matrix (Taussky 1949) reveal that* $\rho_n = \pi(1 + O(1/\log(n)))$. *Moreover, if we consider an identity weighting matrix, i.e.* $W_{n,L} = \mathbb{I}_L$, *then we may take* $s_n = 0$. *In addition, if we assume, similarly to the least squares series regression setup (Belloni et al. 2015; Newey 1997), that the smallest eigenvalue of the population matrix in the statement of Proposition 3 is bounded away from zero uniformly, then* $\gamma = 0$. *Finally, if* $\hat{Q}_n(u)$ *are the empirical quantiles*

from a random sample of size n from a distribution satisfying the assumptions in Theorem 6 of Csorgo and Revesz 1978, then $r_n = n^{-1/2}\log(n)^{\alpha}$ for some $\alpha > 0$ and we obtain that the approximation on the class of semi-intervals in Corollary 2 holds as soon as:

$$\frac{\log(p)^{\nu}\log(n)^{\alpha}}{n^{1/2}} \to 0\,.$$

Remark 5 (Undersmoothing)**.** *The inferential approximation (3.12) does not account for the bias term $\boldsymbol{D}_n$. Observe that, in the setting of Corollary 2, this bias may be ignored in inference under the undersmoothing condition.*

$$\log(p)^{\nu}\sqrt{n}\|\boldsymbol{D}_n\| \to 0\,.$$

To understand this condition, consider, again, the polynomial basis setup of the previous remark. In this case, Example 1 and Bessel's inequality, see Theorem 3.2.8 of Kreyszig 1978 reveal that:

$$\|\boldsymbol{D}_n\| \leqslant \tilde{C}\omega_{2,2}(Q, p^{-1})\,.$$

If we assume that Q belongs to a generalized Lipschitz space (see Section 2.9 of DeVore and Lorentz 1993 for a definition), then $\omega_{2,2}(Q,t) \leqslant Dt^{b}$ for $D, b > 0$. In this case, the undersmoothing condition subsumes to:

$$\log(p)^{\nu}\frac{\sqrt{n}}{p^{b}} \to 0\,.$$

Remark 6 (Bias-aware inference)**.** *If one is not willing to impose the undersmoothing conditions in the previous remark, it is still possible to conduct conservative inference on $\theta_{0,n}$ by bounding the bias. Specifically, Approximation Theory provides the value for the absolute constant in Example 1. Then, by placing a plausible upper bound on the generalized Lipschitz norm of Q, we may use Example 1 along with representation (3.12) to adjust the critical values for coverage even in the worst-case setting. Such a procedure is known as "bias-aware" inference and has received increasing attention from the Econometrics and Statistics literature (Armstrong and Kolesár 2021; Ignatiadis and Wager 2022; Noack and Rothe 2021). Though we recognize the possibility of conducting inference this way, we do not directly pursue it in this paper.*

Remark 7 (Optimal weighting-matrix)**.** *From the closed form solution of the approximation, it is clear that the optimal choice of weights is given by:*

$$\Omega_{n,L} = \mathbb{V}\left[\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)B_{n,0}(u)du\right)\right]^{-}\,,$$

where A^{-} denotes the generalized inverse of a matrix. As in two-step GMM, this matrix may be estimated in a first-step by first running the GMLM with identity weights and then using it to estimate $\Omega_{n,L}$. Under mild assumptions, this should not affect the overall rate of estimation.

Ridge regularization We now consider the case of ridge regularization, i.e. $\Theta_p = B_{\mathbb{R}^p, R_n}(0)$. In this case, by relying on Proposition 2, we are able to show that, under the assumptions of Proposition

3, the L-moment estimator with $\Theta_p = B_{\mathbb{R}^p, R_n}(0)$ achieves the same rates as the unconstrained estimator.

To provide valid inference in this setting, one should then combine these rates with anticoncentration results on the approximation (3.7). Notice that, under ridge regularization, (3.7) is the projection of a Gaussian random vector in the $\|\cdot\|_{\Omega_n}$-norm onto an Eucliden ball. As far as we are aware, anticoncentration results on projection of Gaussian vectors onto Euclidean balls are not available in the literature. In Appendix 3.10, we provide an approach to extend results on Gaussian anticoncentration available in the literature to the ball projection setting. We believe this approach may be of independent interest.

Combining these results, we have the following proposition.

Proposition 4. *Consider the setting with ridge constraint. Suppose that for constants $C > 0$, $\gamma \in [0,1)$, the smallest eigenvalue of $\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' \Omega_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)$, $\lambda_{n,\min}$, satisfies $\lambda_{n,\min} > C(\rho_n s_n)^\gamma$.*
Then, if $\rho_n s_n \to 0$, $\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)$ is invertible with probability approaching one and:

$$\|\sqrt{n}(\hat{\theta}_n - \theta_{0,n}) - \sqrt{n}(\theta_n^* - \theta_{0,n})\| = O_{\mathbb{P}}((\rho_n s_n)^{1-2\gamma} \vee ((\rho_n s_n)^{-\gamma}\sqrt{\rho_n} r_n)),$$
$$\|\sqrt{n}(\hat{\theta}_n - \theta_{0,n}) - \sqrt{n}(\tilde{\theta}_n^* - \theta_{0,n})\| = O_{\mathbb{P}}((\rho_n s_n)^{1-2\gamma} \vee ((\rho_n s_n)^{-\gamma}\sqrt{\rho_n}(r_n \vee d_n))).$$

Moreover, suppose that the variance of the Gaussian approximation in the unconstrained setting is bounded below uniformly. Suppose that $R_n = Gp^{-l}$ for some $l \in [0, 1/4)$. If, for some $\nu > 1/2$

$$\log(p)^\nu (\rho_n s_n)^{-\gamma} p^{1/2+l} \left((\rho_n s_n)^{1-2\gamma} \vee ((\rho_n s_n)^{-\gamma}\sqrt{\rho_n} r_n)\right) \to 0,$$
$$n(\rho_n s_n)^\gamma \to \infty,$$

then:

$$\sup_{A \in \mathcal{C}_p} |\mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_{0,n}) \in A] - \mathbb{P}[\sqrt{n}(\theta_n^* - \theta_{0,n}) \in A]| \to 0.$$

In addition, if the undersmoothing condition $\log(p)^\nu (\rho_n s_n)^{-2\gamma} p^{1/2+l} d_n \to 0$ holds, then the approximation is also valid for the undersmoothened approximation.

Proof. See Appendix 3.9.4. □

Remark 8 (Polynomial basis, cont.)**.** *Consider the setting of polynomial basis with identity weights, population matrix with eigenvalues bounded away uniformly from zero, and empirical quantiles from a random sample satisfying the Csorgo and Revesz 1978 assumptions. Suppose the radius R_n is kept fixed at $\bar{R}$. In this case, the rate requirement in the statement of the proposition subsumes to:*

$$\log(p)^\nu \log(n)^\alpha \sqrt{\frac{p}{n}} \to 0,$$

which, up to logs, is the same rate obtained by Belloni et al. 2015 in an (unconstrained) sieve-regression setting. For Q in a generalized Lipschitz class, the undersmoothing condition is given by:

$$\log(p)^{\nu}\sqrt{n}p^{1/2-b} \to 0\,,$$

thus requiring a degree of smoothness $b > 1$ for proper bias control.

Remark 9 (Relaxing regularization)**.** *In the statement of Proposition 4, we have considered either a constant or decreasing radius R_n. It is also possible to consider increasing radii, though in this case one should conduct a careful analysis of the terms in Lemma 5 in the Appendix to understand the "region" where the anticoncentration "inhabits". Since constant radius are a typical choice in constrained least squares estimation (Chernozhukov, Wüthrich, and Zhu 2021); and decreasing radii may be required in some applications, we focus on these cases.*

Stochastic basis

In this section, we consider the case of stochastic basis, i.e. $\boldsymbol{J}_{p,n}$ is a random variable. In this case, our target estimand is given by:

$$\theta_{0,n} \in \operatorname{arginf}_{\theta\in\Theta_n}\|Q(\cdot) - \theta'\boldsymbol{J}^*_{n,p}(\cdot)\|_{L^2[\underline{p}_n,\overline{p}_n]}\,, \tag{3.13}$$

where $\boldsymbol{J}^*_{n,p}$ is the estimand of $\boldsymbol{J}_{n,p}$. We propose to conduct inference conditionally on the controls $\boldsymbol{J}_{n,p}$. For that, we will require $\boldsymbol{J}_{n,p}$ to be independent of $\hat{Q}_n$, and we also need to control the estimation error in $\boldsymbol{J}_{n,p}$. These conditions are subsumed in the following assumption.

Assumption 6 (Estimation error of basis functions)**.** *$\boldsymbol{J}_{n,p}$ is a random vector, with $\hat{Q}_n$ independent of $\boldsymbol{J}_{n,p}$. Moreover, there exists a sequence of nonstochastic quantile functions $\boldsymbol{J}^*_{p,n}$, $n \in \mathbb{N}$, such that, as $n \to \infty$,*

1. $\left\|\theta'_{0,n}\sqrt{n}\left(\boldsymbol{J}_{p,n}(\cdot) - \boldsymbol{J}^*_{p,n}(\cdot)\right)\right\|_{L^2[\underline{p}_n,\overline{p}_n]} = O_{\mathbb{P}}(\xi_n).$
2. $\lambda_{\max}\left(\int_{\underline{p}_n}^{\overline{p}_n}\left(\boldsymbol{J}_{n,p}(u) - \boldsymbol{J}^*_{p,n}(u)\right)\left(\boldsymbol{J}_{n,p}(u) - \boldsymbol{J}^*_{p,n}(u)\right)' du\right) = O_{\mathbb{P}}(\varepsilon_n)$

Since we propose an inferential procedure conditionally on the data $\boldsymbol{J}_{n,p}$, estimation error enters the distributional approximation similarly to the bias term $D_n(u)$ in the nonstochastic setting. As it is generally difficult to bound the bias when $\boldsymbol{J}_{n,p}$ does not belong to a known family, we thus focus on an inferential approach that ignores the bias. Specifically, we consider the approximation:

$$\sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n}) \in \operatorname{arginf}_{x\in\sqrt{n}(\Theta_n-\theta_{0,n})}\left\|\int_{\underline{p}_n}^{\overline{p}_n}(B_{0,n}(u) - x'\boldsymbol{J}_{p,n}(u))\boldsymbol{P}_L(u)du\right\|^2_{2,\Omega_{L,n}}\,, \tag{3.14}$$

and provide conditions that ensure asymptotic validity of this approximation over a class of sets $\mathcal{C}_p$, conditionally on the data, i.e.:

$$\sup_{C \in \mathcal{C}_p} |\mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_{0,n}) \in C | \boldsymbol{J}_{n,p}] - \mathbb{P}[\sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n}) \in C | \boldsymbol{J}_{n,p}]| \xrightarrow{p} 0 .$$

To verify the above, we first present an analog of Proposition 2 in the stochastic setting.

Proposition 5. *Suppose that Assumptions 4, 5 and 6 are satisfied. Let* $d_n := \|\sqrt{n}(Q(\cdot) - \theta'_{0,n}\boldsymbol{J}^*_{n,p}(\cdot)\|_{L^2[\underline{p}_n, \overline{p}_n]}$. *For* $z \in \mathbb{R}^p$, *define the restricted eigenvalue around* z *as :*

$$\lambda_{z,n} := \inf_{s \in \mathcal{X}_n} \frac{(s-z)' \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u) \boldsymbol{J}^*_{n,p}(u)' du \right)' \Omega_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u) \boldsymbol{J}^*_{n,p}(u)' du \right) (s-z)}{\|s-z\|_2^2} . \tag{3.15}$$

Moreover, let:

$$\rho_n := \lambda_{\max} \left(\int_{\underline{p}_n}^{\overline{p}_n} (\boldsymbol{J}^*_{n,p}(u) \boldsymbol{J}^*_{n,p}(u)') du \right) .$$

We then have that, for any sequences $(\delta_n)_{n \in \mathbb{N}}$, $(M_n)_{n \in \mathbb{N}}$ *and* $(c_n)_{n \in \mathbb{N}}$ *such that:*

$$\begin{aligned} M_n \frac{\lambda_{0,n}}{\sqrt{\rho_n \vee \varepsilon_n}} &\to \infty , \\ \mathbb{P}[\lambda_{\sqrt{n}(\theta^*_n - \theta_{0,n}),n} \leqslant c_n] &\to 0 , \\ \frac{c_n \delta_n^2}{(\delta_n + M_n)\sqrt{(\rho_n \vee \varepsilon_n)}(r_n \vee d_n \vee s_n \vee \xi_n)} &\to \infty , \\ \frac{c_n \delta_n^2}{(\delta_n + M_n)^2 (\rho_n \vee \varepsilon_n) s_n} &\to \infty , \end{aligned} \tag{3.16}$$

we have:

$$\mathbb{P}[\|\sqrt{n}(\hat{\theta}_n - \theta_{0,n}) - \sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n})\| \geqslant \delta_n] \to 0 .$$

In addition, the convergence holds conditionally on $\boldsymbol{J}_{n,p}$, *i.e.*

$$\mathbb{P}[\|\sqrt{n}(\hat{\theta}_n - \theta_{0,n}) - \sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n})\| \geqslant \delta_n | \boldsymbol{J}_{n,p}] \xrightarrow{p} 0 .$$

Proof. See Appendix 3.9.5. □

The previous proposition provides rates for the proposed approximation. These rates depend, crucially, on the estimation error of the basis functions and its interaction with the oracle $\theta_{0,n}$, as reflected by the constants ξ_n and ϵ_n. In Section 3.6.2, we provide conditions that bound these rates when $\boldsymbol{J}_{n,p}$ are empirical quantiles from random samples of p populations. In this setup, proper control of the estimation error is achieved under a condition that ensures the weights $\theta_{0,n}$ are diluted across basis functions. As we remark in Section 3.6.2, a similar condition appears in the Synthetic Control literature (Ferman 2021).

We conclude this section by providing a version of Proposition 4 to the case of stochastic basis.

Corollary 3. *Consider the setting with ridge constraint. Suppose that for constants* $C > 0$, $\gamma \in [0,1)$, *the smallest eigenvalue of* $\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}^*_{n,p}(u)'du\right)' \Omega_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}^*_{n,p}(u)'du\right)$, $\lambda_{n,\min}$, *satisfies* $\lambda_{n,\min} > C((\rho_n \vee \varepsilon_n)s_n)^\gamma$.
Then, if $(\rho_n \vee \varepsilon_n)s_n \to 0$, $\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}^*_{n,p}(u)'du\right)' W_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p^*}(u)'du\right)$ *is invertible with probability approaching one and:*

$$\|\sqrt{n}(\hat{\theta}_n - \theta_{0,n}) - \sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n})\| = O_{\mathbb{P}}((\rho_n \vee \varepsilon_n)s_n)^{1-2\gamma} \vee ((\rho_n \vee \varepsilon_n)s_n)^{-\gamma}\sqrt{(\rho_n \vee \varepsilon_n)}(r_n \vee d_n \vee \xi_n))).$$

Moreover, suppose that the variance of the Gaussian approximation in the unconstrained setting is bounded below uniformly. Suppose that $R_n = Gp^{-l}$ *for some* $l \in [0, 1/4)$. *If, for some* $\nu > 1/2$

$$\log(p)^\nu((\rho_n \vee \varepsilon_n)s_n)^{-\gamma}p^{1/2+l}\left((\rho_n \vee \varepsilon_n)s_n)^{1-2\gamma} \vee ((\rho_n \vee \varepsilon_n)s_n)^{-\gamma}\sqrt{(\rho_n \vee \varepsilon_n)}(r_n \vee d_n \vee \xi_n)\right) \to 0,$$

$$n((\rho_n \vee \varepsilon_n)s_n)^\gamma \to \infty,$$

then:

$$\sup_{A \in \mathcal{C}_p} |\mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_{0,n}) \in A] - \mathbb{P}[\sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n}) \in A | \boldsymbol{J}_{n,p}]| \xrightarrow{p} 0.$$

Proof. The proof is analogous to that of Proposition 4, but applied conditionally on $\boldsymbol{J}_{n,p}$. □

3.4.2 Inference on linear functionals of mixture weights

In several settings, direct interest is not on the mixture weights, but on linear functionals thereof. Specifically, one may be interested in conducting inference on the functional

$$T(x) \coloneqq \omega(x)'\theta_{0,n}, \quad x \in \mathcal{X}, \tag{3.17}$$

where $(\mathcal{X}, \mathcal{L}, \Pi)$ is a measure space. In this case, confidence sets for T may be constructed by approximating the distribution of a suitable L^p-norm. Specifically, for $p \in [1, \infty]$, we consider the scalar random variable:

$$\mathcal{T}_p \coloneqq \sqrt{n}\|\omega(\cdot)'(\hat{\theta}_n - \theta_{0,n})\|_{L^p(\Pi)}. \tag{3.18}$$

If the quantiles of $\mathcal{T}_p$ were known, a valid $(1-\alpha)$ uniform confidence band for T could be constructed as:

$$C_{T,1-\alpha} = \left[\omega(\cdot)'\hat{\theta}_n - \frac{1}{\sqrt{n}}Q_{\mathcal{T}_p}(1-\alpha), \omega(\cdot)'\hat{\theta}_n + \frac{1}{\sqrt{n}}Q_{\mathcal{T}_p}(1-\alpha)\right].$$

In practice, the distribution of $\mathcal{T}_p$ is unknown. One is thus tempted to use the distributional approximations discussed in the previous subsection to compute the quantiles. For example, if one considers an undersmoothened approximation, one could consider approximating the distribution of $\mathcal{T}_p$ with:

$$\tilde{\mathcal{T}}_p^* = \sqrt{n}\|\omega(\cdot)'(\tilde{\theta}_n^* - \theta_{0,n})\|_{L^p(\Pi)}.$$

Can we ensure this approximation provides valid inference? Application of Lemma 3, along with the Cauchy-Schwarz inequality, reveals that:

$$\sup_{c\in\mathbb{R}} |\mathbb{P}[\mathcal{T}_p \leqslant c] - \mathbb{P}[\tilde{\mathcal{T}}_p^* \leqslant c]| \leqslant \inf_{\delta \geqslant 0} \left\{ \mathbb{P}\left[\|\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n^*)\|_2 \geqslant \frac{\delta}{\left(\int_{\mathcal{X}} \|\omega(x)\|_2^p \Pi(dx)\right)^{1/p}} \right] + \sup_{d\in\mathbb{R}} \mathbb{P}[d \leqslant \tilde{\mathcal{T}}_p^* \leqslant d + \delta] \right\}.$$

The above inequality shows that to approximate the quantiles of $\mathcal{T}_p$, one has to balance the strong approximation between $\sqrt{n}(\hat{\theta}_n - \theta_{0,n})$ and $\sqrt{n}(\tilde{\theta}_n^* - \theta_{0,n})$ with the anticoncentration of $\mathcal{T}_p^*$. We have provided rates for the former in the previous subsection. As for the latter point, bounds on the anticoncentration of $\mathcal{T}_p^*$ may be obtained, in the unconstrained setting, by applying Theorem 2.1 of Chernozhukov, Chetverikov, and Kato 2014.[7] These bounds may be extended to the ridge setting by a similar approach as the one presented in Appendix 3.10.

3.5 A practical algorithm for conducting inference

Based on the discussion in Section 3.4, we propose the following algorithm for conducting inference. Suppose that we wish to approximate the quantity $\mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_{0,n}) \in B]$, where $B \in \mathcal{B}(\mathbb{R}^p)$. We can then proceed as follows:

Algorithm 1 Approximating the distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$

1: Fix $S \in \mathbb{N}$ and $\gamma_n > 0$.
2: Estimate $\hat{\theta}_n$ by solving (3.2).
3: Estimate the variance matrix $\mathbb{V}_n := 2\left(\frac{\gamma_n}{\sqrt{n}}\right)^2 \mathbb{V}\left[\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u) B_{n,0}(u) du\right)\right]$ using an estimator $\hat{V}_n$.
4: Estimate
$$\hat{b} := \gamma_n \int_{\underline{p}_n}^{\overline{p}_n} \left(\hat{Q}(u) - \hat{\theta}' \boldsymbol{J}_{n,p}(u)\right) \boldsymbol{P}_L(u) du$$
5: **for** s=1 to S **do**
6: Draw $Z_s \sim \mathcal{N}\left(\hat{b}, \hat{V}_n\right)$.
7: Find and store $\mu_s^* \in \operatorname{argmin}_{x\in\gamma_n(\Theta_n - \hat{\theta}_{0,n})} \|Z_s - \int_{\underline{p}_n}^{\overline{p}_n} x' \boldsymbol{J}_{p,n}(u) \boldsymbol{P}_L(u) du\|_{W_{L,n}}^2$
8: **end for**
9: For $B \in \mathcal{B}(\mathbb{R}^p)$, estimate $\mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_0) \in B]$ by $\frac{1}{S}\sum_{s=1}^S \mathbb{1}\left\{\mu_s^* \in \frac{\gamma_n}{\sqrt{n}} B\right\}$.

[7] In the case $p = 2$, one may alternatively use the results in Götze et al. 2019.

Algorithm 1 provides a simulation method for approximating the quantity $\mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_0) \in B]$. For simplicity, the algorithm abstracts from bias-aware considerations, thus relying on an undersmoothened approximation. The algorithm requires two hyperparameters: (i) the number of simulations S; and (ii) a tuning parameter γ_n for setting the choice set, whose choice we discuss further below.

In Section 3.4, we have provided conditions such that $\mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_0) \in B]$ is asymptotically approximated by $\mathbb{P}[\sqrt{n}(\tilde{\theta}^*_n - \theta_0) \in B]$, uniformly over a class $B \in \mathcal{C}$. Note, however, that the quantity $\mathbb{P}[\sqrt{n}(\tilde{\theta}^*_n - \theta_0) \in B]$ depends on three unknowns, namely the variance matrix $\mathbb{V}_n = \mathbb{V}\left[\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)B_{n,0}(u)du\right)\right]$, the weighting matrix $\Omega_{L,n}$ and the centering point $\theta_{0,n}$ of the choice set $\mathcal{X}_n$. Algorithm 1 suggests a "plug-in" approach, whereby these quantities are replaced by estimators $\hat{V}_n$, $W_{L,n}$ and $\hat{\theta}_n$. It also replaces $\sqrt{n}$ in the definition of the choice set $\mathcal{X}_n$ with a tuning parameter $\gamma_n > 0$. Such modification is needed in constrained settings to control how estimation error of the centering point of $\mathcal{X}_n$ affects the quality of the approximation of the true choice set, especially with regard to the "bindingness" of the constraints, see for related bootstrap strategies in different constrained settings Chernozhukov, Newey, and Santos 2023; Hong and Li 2020; Li 2021.

The choice of γ_n depends crucially on the local geometry of $\mathcal{X}_n$ around $\theta_{0,n}$. For example, in the unconstrained setting ($\Theta_n = \mathbb{R}^p$), $\mathcal{X}_n = \mathbb{R}^p$ irrespectively of the value of $\theta_{0,n}$ or γ_n. As a consequence, in this case, validity of Algorithm 1 hinges solely on proper estimation of $\mathbb{V}_n$ and $\Omega_{L,n}$. More generally, the value of $\theta_{0,n}$ affects the geometry of $\mathcal{X}_n$. In this case, Appendix 3.11 outlines a general approach which, building upon Lemma 4 in Appendix 3.9, provides conditions which can be used to find a sequence γ_n in specific settings. We also refer the reader to Cattaneo, Y. Feng, and Titiunik 2021 for a discussion on the role of properly accounting for the local geometry of $\mathcal{X}_n$ when bootstrapping in a constrained setting.

3.6 Theoretical Applications

3.6.1 Empirical Bayes

We consider the Empirical Bayes problem, which we briefly outlined in the introduction. Suppose the researcher has access to a sample of identically distributed real random variables, which admit a marginal Lebesgue density f satisfying:

$$f(y) = \int_{\Xi} \phi(y;\xi)G(d\xi), \quad y \in \mathbb{R}, \tag{3.19}$$

where $\Phi := \{\phi(\cdot;\xi) : \xi \in \Xi\}$ is a known parametric family of densities on $\mathbb{R}$, $(\Xi, \mathcal{J})$ is a measurable space, and G is an unknown probability measure on $(\Xi, \mathcal{J})$. Our interest lies in estimating G. For example, one may have access to a sample of noisy standardized measurements of school quality ξ_i for n schools, $Y_i|\xi_i \sim N(\xi_i, 1)$, $i = 1, \ldots, n$; and would like to estimate the population distribution of school quality G, so as to select the schools above 95th percentile of school quality. Alternatively, the researcher may have access to noisy measurements on treatment effects of a policy across different

sites, and would like to estimate G to understand the population distribution of these effects. Finally, the researcher may need to estimate G to perform shrinkage on some preliminary noisy measure (Armstrong, Kolesár, and Plagborg-Møller 2022).

In estimating model (3.19), Efron 2014 distinguishes between two possible approaches. In f-modelling, one estimates a model for f, and then inverts the map $G \mapsto f_G$ given by (3.19) to find an estimate of G. When the model Φ corresponds to a location family, this procedure subsumes to the well known *density deconvolution problem*. In contrast, in G-modelling, one uses the structure given by (3.19) to directly estimate G, e.g. by applying the nonparametric MLE of Kiefer and Wolfowitz 1956.

Efron 2016 argues that fully nonparametric f- or G-modelling approaches are often undesirable, as rates of convergence can be poor. In the f-modelling setting, this corresponds to known problems of nonparametric density deconvolutions (Fan 1991; Meister 2009). In the G-modelling setup, the nonparametric MLE, in spite of enjoying "optimality" properties (Polyanskiy and Wu 2020), can also perform poorly. Perhaps equally or more importantly, it is not yet clear how to conduct frequentist uncertainty quantification based on the nonparametric MLE (Ignatiadis and Wager 2022).

To circumvent the problems with a fully nonparametric approach, Efron 2016 proposes a sieve G-modelling approach. Motivated by his discussion, and given the general approximation properties of quantile mixture models discussed in Section 3.2, as well as the attractive statistical properties of L-moments, we propose a sieve f-modelling approach based on our estimation method. For that, we rely on the observation that, for a quantile function Q_F obtained upon inversion of a strictly increasing and differentiable distribution function F, we have that:

$$Q_F'(u) = \frac{1}{F'(Q_F(u))}.$$

In light of this observation, we propose the following steps to estimate G.

1. Choose nonstochastic differentiable basis functions $\boldsymbol{J}_{n,p}$ and estimate a quantile mixture model by (3.2).
2. Estimate $f(Q_F(u))$ by $\widehat{f(Q_F(u))} = \frac{1}{\hat{\theta}_n' \partial_u \boldsymbol{J}_{n,p}(u)}$.
3. Invert (3.19) to obtain $\hat{G}$.

To conduct frequentist uncertainty quantification on G or functionals thereof, one can rely on Algorithm 1. Specifically, for a given confidence level $(1-\alpha)$, if the resulting confidence set for (a functional of) G can be written as $\{\theta \in \Theta_n : \sqrt{n}(\hat{\theta}_n - \theta) \in D\}$, where D is a convex set such that $\mathbb{P}[\{\sqrt{n}(\tilde{\theta}_n^* - \theta_{0,n}) \in D\}] \geqslant 1-\alpha$, then the validity of the approach in conducting inference can be ascertained by proceeding similarly to Section 3.4.2. In this case, one must rely, in the unconstrained setting, on an anticoncentration inequality for convex sets, see Lemma A.2 of Chernozhukov, Chetverikov, and Kato 2017a. In the ridge setting, this inequality may be extended using the arguments in Section 3.10. The assumption that D is convex is not restrictive: note that the inversion step consists on the application of a linear operator on $\hat{f}$. Combined with the fact that $x \mapsto 1/x$ is convex (thus exhibiting convex lower contour sets), we can construct convex confidence sets for G.

Remark 10. *Bias-aware inference has been recently advocated in the Empirical Bayes setup by Ignatiadis and Wager 2022. We note that, by placing bounds on the approximation error of the sieve-quantile model on the true* Q*, it may be possible to conduct bias-aware inference on* G*. It should be noted, however, that strong restrictions may be needed for these bounds to be informative. Indeed, since the differentiation operator is unbounded in general function spaces, the construction of informative bounds may require strong restrictions on the candidate function space for* Q*. As we remarked earlier on, we do not pursue bias-aware inference in this paper.*

3.6.2 Distributional synthetic controls

In this section, we show how our methodology can be applied to the distributional synthetic control setting of Gunsilius 2023, which we briefly overviewed in Example 5.[8] We note that Gunsilius 2023 does not introduce formal inference methods in his distributional setup: in the article, there is a brief suggestion of using placebo to assess the uncertainty in estimates, though no formal justification is given to it.

We start by recasting the problem in the notation of Section 3.4. We consider a setting where we only have one pre-intervention period, which we denote by t^*-1. The stochastic basis functions $\mathbf{J}_{t^*-1,n,p}$ represent the empirical quantile functions of the p control units at time t^*-1. These empirical quantiles are estimators of the true population quantile functions $\mathbf{J}^*_{t^*-1,n,p}$, which are unknown. The empirical quantile function of the treated unit in the pre-treatment period is $\hat{Q}_{t^*-1}$, and Q_{t^*-1} is its population counterpart. We begin with the identification assumption in our distributional setting.

Assumption 7 (Identification Assumption of Distributional Synthetic Control)**.** *There exists* $\theta_{0,n} \in B_{\mathbb{R}^p,1}(0)$ *such that* $Q_{t^*-1} = \theta'_{0,n}\boldsymbol{J}^*_{t^*-1,n,p}$.

Assumption 7 requires that, if the population counterparts were known, it would be possible to perfectly replicate the quantile function of the treated unit in the pre-treatment period as a mixture of the controls. We assume that the weights belong to the unit Euclidean ball, which motivates the use of ridge regularization in the analysis.

The next assumption constrains the estimation error of the empirical quantile functions.

Assumption 8 (Sampling)**.** $\hat{Q}_n$ *is the empirical quantile function from a random sample of size* n *from a continuous distribution* F_0*. Similarly, for each* $i = 1, \dots, p$*.* $\mathbf{J}_{t^*-1,n,i}$ *is the empirical quantile function from a random sample of size* n_i *from a continuous distribution function* F_i*. Finally, for each* $k \in \{0, 1, \dots, p\}$*, we assume that:*

1. F_k *is twice differentiable on* (a_k, b_k)*, where* $a_k = \sup\{x : F_k(x) = 0\}$*,* $b_k = \inf\{x : F_k(x) = 1\}$*, and* $F'_k \neq 0$ *on* (a_k, b_k)*.*
2. $\sup_{a_k<x<b_k} F_k(x)(1-F_k(x))\left|\frac{F''(x)}{F'^2(x)}\right| \leqslant \alpha_k$*, for some* $\alpha_k > 0$*.*

[8] Its worth noting that, when there is only one pre-intervention period, the estimator proposed by Gunsilius 2023 is equivalent to (3.2) when $L = \infty$, $W_{n,L} = \mathbb{I}_L$, $\underline{p}_n = 0$, $\overline{p}_n = 1$ and $\Theta_n = \Delta^{p-1}$. More generally, by using the optimal weighting scheme, we expect improvements over Gunsilius 2023 estimator.

Assumption 8 requires that the population distributions satisfy the assumptions of Theorem 6 of Csorgo and Revesz 1978, which provides rates for the strong approximation of empirical quantile functions.

Assumption 9 ("Dilution condition")**.** *We require that:*

1. $\min_{i\in 1,\dots,p} n_i \to \infty$.
2. *There exists* $\underline{f} > 0$ *such that,* $\inf_{i\in\mathbb{N}} \inf_{x\in\mathbb{R}} F_i'(x) \geqslant \underline{f}$.
3. *For some* $\mu > 1$ *we have*
$$\sum_{i=1}^{p} \theta_{0,i,n}^2 \frac{n}{n_i} = O_{\mathbb{P}}\left(\frac{1}{p^{\mu}}\right)$$

Assumption 9 provides conditions that allow us to properly control the estimation error of the basis functions. Specifically, under the Csorgo and Revesz conditions in Assumption 8 and Assumption 9, we are able to show that:

$$\left\| \theta_{0,n}' \sqrt{n} \left(\boldsymbol{J}_{p,n}(\cdot) - \boldsymbol{J}_{p,n}^*(\cdot) \right) \right\|^2_{L^2[\underline{p}_n, \overline{p}_n]} = O_{\mathbb{P}}\left(\frac{1}{p^{\mu-1}}\right),$$

ensuring that estimation error of the basis functions may be safely ignored when the number of controls is large. Crucially, Assumption 9 requires the oracle weights to be sufficiently diluted across controls, so the sampling error of any control unit becomes asymptotically negligible. A similar condition appears, either explicitly or implicitly, in the literature on inference in Synthetic Controls, where randomness of the outcomes in the donor pool is abstracted from when conducting inference (Ferman 2021).

Under the previous assumptions, we can derive explicit convergence rates by directly appealing to Corollary 3.

Corollary 4. *Suppose that for constants* $C > 0$, $\gamma \in [0,1)$, *the smallest eigenvalue of* $\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}^*(u)'du\right)' \Omega_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}^*(u)'du\right)$, $\lambda_{n,\min}$ *satisfies,* $\lambda_{n,\min} > C(\rho_n s_n)^{\gamma}$. *Suppose that the assumptions in Proposition 5 and Assumptions 7, 8 and 9 hold. Moreover, suppose the variance of the Gaussian Approximation in the unconstrained setting is bounded below uniformly,* $\varepsilon_n \leqslant \rho_n$, *and* $R_n = 1$ *for all* n. *If*

$$\log(p)^{\nu}(\rho_n s_n)^{-\gamma} p^{1/2} \left((\rho_n s_n)^{1-2\gamma} \vee (\rho_n s_n)^{-\gamma} \sqrt{\rho_n} \frac{\log(n)^{\alpha}}{n^{1/2}} \right) \to 0,$$

$$n(\rho_n s_n)^{\gamma} \to \infty,$$

$$\log(p)^{\nu} \frac{\sqrt{n}}{p^{b}} \to 0,$$

then:

$$\sup_{A\in\mathcal{C}_p} \left| \mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_{0,n}) \in A | \boldsymbol{J}_{n,p}] - \mathbb{P}[\sqrt{n}(\tilde{\theta}_n^* - \theta_{0,n}) \in A | \boldsymbol{J}_{n,p}] \right| \xrightarrow{p} 0.$$

3.7 Empirical Application

In this section, we apply our methodology to assess the effects of an environmental catastrophe on the local wage distribution of the affected municipality.

3.7.1 Brumadinho Dam Disaster

Context

On January 25, 2019, a devastating tailings dam collapse occurred near the city of Brumadinho, located in the state of Minas Gerais, Brazil. This incident resulted in the loss of 270 lives and caused extensive contamination of the neighboring Paraopeba River. The victims primarily consisted of employees of Vale, the largest mining conglomerate in Latin America, which operated both the dam and the mines providing materials for its construction.

The release of mining waste resulted in significant environmental and socioeconomic damage. Brumadinho suffered the most severe impact as the closest city to the collapsed dam. However, according to official authorities, an additional 25 cities were also affected. Damages resulted in the loss of hectares of forest and vegetation cover, as well as the contamination of the Paraopeba River with elevated levels of copper, negatively impacting its ecosystem and local agriculture. The city of Brumadinho experienced a decline in revenue due to the halting of mining operations, adversely affecting the local economy and government finances.

Vale faced legal and financial repercussions, including penalties and the obligation to provide compensation to affected individuals and communities.[9] According to the judicial agreement reached in 2021, Vale is required to allocate approximately 7.58 billion USD towards reparations, of which 4.4 billion USD has already been disbursed, with a substantial portion allocated to economic restitution for the affected cities.[10] Amidst all the repercussions, the labor market consequences of such an event still remain an important economic and policy question.

Empirical question: distributive consequence on wage earners

The distributive impact of the events in Brumadinho on the wages of formal sector workers is an open empirical question. The economic disruptions resulting from the dam collapse would negatively affect job opportunities but it is unclear which workers are more exposed to the shock. However, the allocation of reparations and investments in the affected areas as well as income support programs would go in the other direction, helping mitigate the impacts. The overall economic recovery and

[9] Upon the initial news of the disaster, the stock price of Vale on the B3 (São Paulo Stock Exchange) fell over 10%. It then fell another 24% on the next trading day, January 28, corresponding to $19 billion in lost market capitalization (Source: Reuters). At the same time, bond prices fell, reflecting an increased risk of non-payment by Vale, as one of the consequences included a court order that froze 11.8 billion reais (USD 3.1 billion at the time) in Vale's accounts, comprising roughly half of Vale's cash on hand.

[10] For comprehensive and detailed information regarding affected cities and the reparations, please refer to the official website of the State Government of Minas Gerais. For the expenditure report on reparations by Vale, please refer to the official website of the company. Both sources are in Portuguese.

reconstruction efforts may generate new jobs and attract investments, benefiting workers across different skill levels.

The main empirical challenge that motivates the choice of the distributional synthetic control method is the nature of the shock's impact, affecting at once both the city of Brumadinho and neighboring municipalities. The lack of a clear control group hinders a direct comparison and motivates the need to construct the synthetic control unit. The approach discussed in Section 3.6.2 is particularly suitable as a large number of potential control units (other municipalities) favors regularization and correct inference is needed to assess the effects on the entire distribution and validate the identification Assumption 7 by evaluating the fit on pre-shock period data not used in the analysis.

3.7.2 Data

Our primary outcome of interest is the distribution of average monthly wages for private sector workers in each municipality. We use the publicly accessible Brazilian employer-employee matched data (RAIS) to construct the empirical quantile function for each municipality $\times$ year pair. RAIS database, also known as Relação Anual de Informações Sociais (Annual Social Information Report) is a comprehensive dataset maintained by the Brazilian the Ministry of Labor and Employment and contains information on formal sector employment in the country, including both public and private sector data.[11] We use data from 2017-2021, details are on Appendix 3.13.

3.7.3 Empirical Strategy

We set the treatment year to 2019, $t^* = 2019$, and use as potential control units other municipalities within the same state as Brumadinho (Minas Gerais) that were not affected by the barrage rupture ($p = 827$). We solve for the synthetic control estimator as in (3.2):

$$\hat{\theta}_n \in \operatorname{arginf}_{\theta \in B_{\mathbb{R}^p,1}(0)} \left\| \left(\int_0^1 (\hat{Q}_{2018}(u) - \theta' \boldsymbol{J}_{2018,p}(u)) \boldsymbol{P}_L(u) du \right) \right\|_{2,\mathbb{I}_L}^2, \tag{3.20}$$

where $\hat{Q}_{2018}(\cdot)$ is Brumadinho's empirical average monthly wage quantile function in the year 2018, $\boldsymbol{J}_{2018,p}(\cdot)$ are the empirical wage quantile functions of the control units. We set $L = 1000$ and use the identity weighting matrix.

To construct confidence sets for $\boldsymbol{J}^*_{t,p}(\cdot)'\theta_{0,n}$, the counterfactual distribution at year t, we rely on Algorithm 1. Specifically, we set $\gamma_n = \sqrt{n}$, where $n = 9,632$ is the number of observations in Brumadinho in 2018; and consider $S = 500$ simulations. We then use these simulations to construct pointwise confidence intervals for $\boldsymbol{J}^*_{t,p}(u)'\theta_{0,n}$ at different years and quantile levels u.

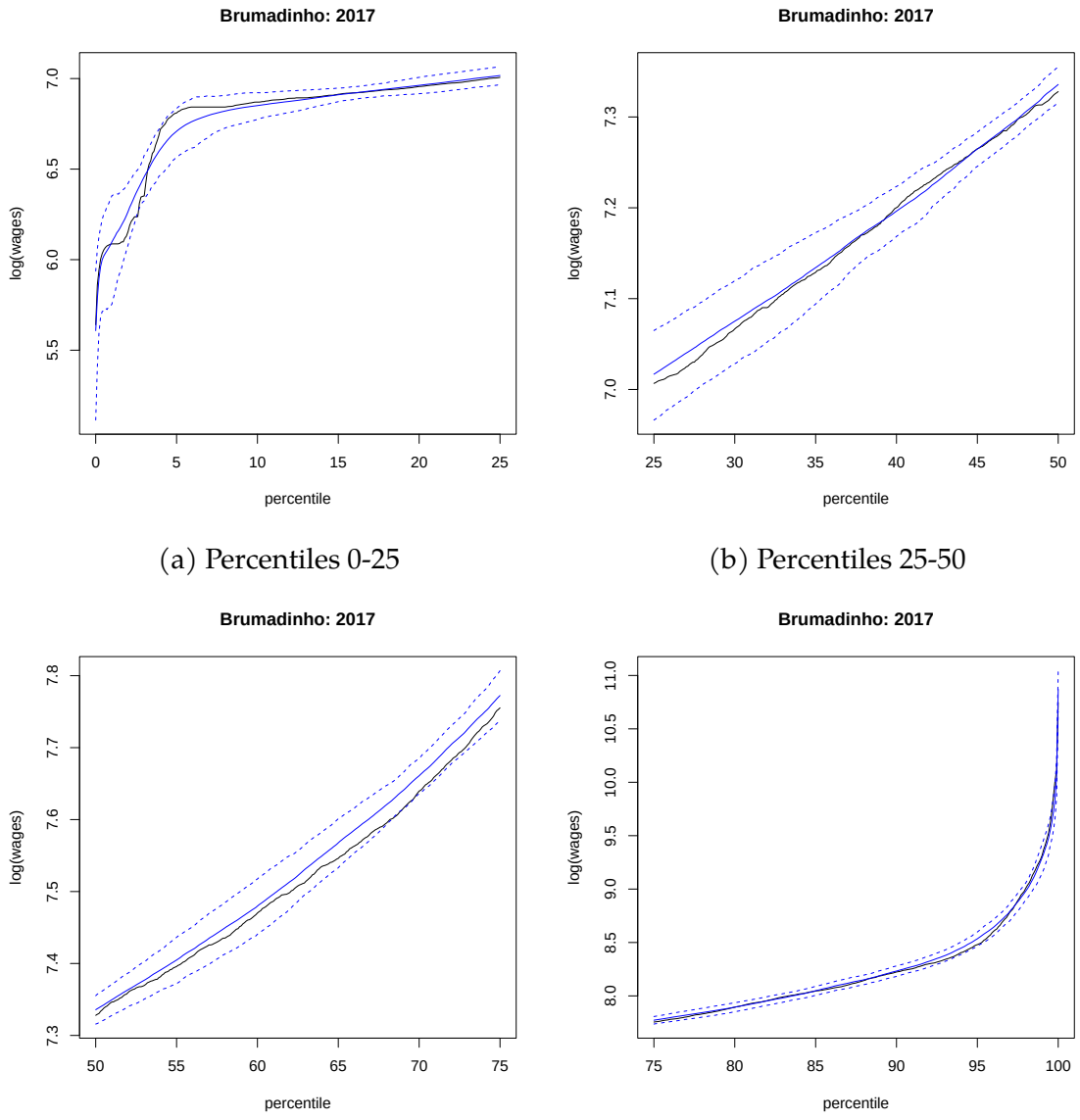
[11] The dataset is collected annually from employers through mandatory reporting requirements. Employers are required to provide detailed data about their employees, including their employment status, occupation, earnings, and other relevant indicators and demographic characteristics. Within the publicly accessible version of the dataset, we can observe for a given municipality the entire wage distribution of formal sector workers (but not individual IDs).

3.7.4 Results

We conduct two types of analysis. First, as a check on the plausibility of Assumption 7, we evaluate how well the weights in 2018 replicate the wage distribution of Brumadinho in 2017. Next, we evaluate the distributional effects at years $t \in \{2019, 2020, 2021\}$. The observed empirical quantile functions are reported in black solid lines, whereas the estimated counterfactuals are reported in blue, along with 95% pointwise confidence intervals for the counterfactual in blue dashed lines.

Pretraining Fit

Figure 3.1 reports the observed quantile function of Brumadinho in 2017 (black line), along with the counterfactual that uses the weights estimated in 2018 (blue solid line) and the associated 95% pointwise confidence intervals (blue dashed line). Overall, pretreatment fit is good, though 2018 weights have some difficulty in reproducing the distribution at the lowermost percentiles.



(a) Percentiles 0-25

(b) Percentiles 25-50

(c) Percentiles 50-75

(d) Percentiles 75-100

Figure 3.1: Pretraining fit (2017).

Treatment Effects

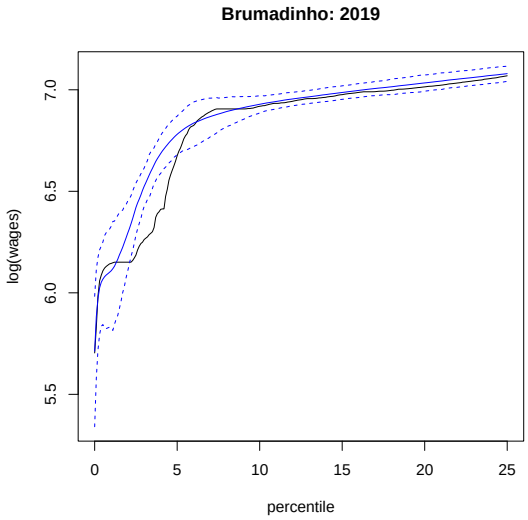
Figure 3.2 reports distributional treatment effects in 2019. We note that the observed quantiles remain below the counterfactual at the smallest percentiles, with this difference being statistically significant around percentiles 3-5. In the range 6-30, observed and counterfactual quantiles remain close to each other, after which the observed quantiles *cross* the counterfactual, remaining above it in the range 35-65, with the difference being (slightly) significant in the range 50-60. After the 65th quantile, observed and counterfactual quantiles return to being close to each other.

Distributional effects in 2020 (Figure 3.3) and 2021 (Figure 3.4) exhibit quite different patterns.

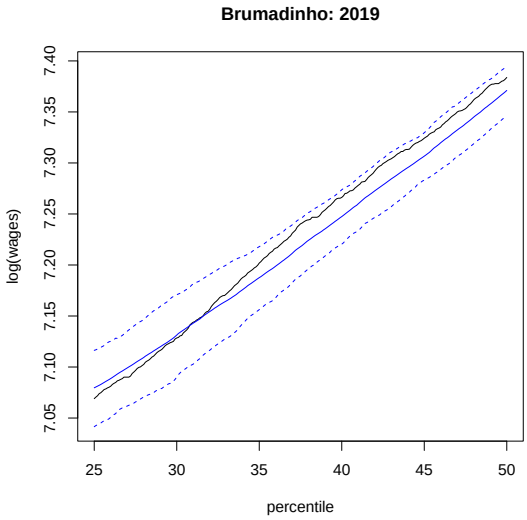
In 2020, there is still a slight, though barely significant, decrease in the observed quantiles, vis-às-vis the counterfactual, around the 5th percentile. In contrast, after the median, the observed quantile is consistently (and significantly), above the counterfactual. In 2021, effects at the lower tail mostly disappear, with the observed counterfactual being above the counterfactual around the median and, especially, after the 95th percentile.

Overall, our results uncover a range of distributional effects on the wage distribution. The crossing of distributions, which is observed in 2019 and, to a lesser degree, in 2020, is consistent with a *displacement effect*, whereby intermediate-paying wages are replaced by low-payment contracts. Such a phenomenon leads simultaneously to a decrease in the lowest quantiles, and an increase from the median on. The counterfactual and wage distributions in 2021 exhibit a pattern which more closely resembles first-order stochastic dominance, with quantiles being above the counterfactual around the median and especially in the upper tail. Such pattern is compatible with two possible explanations. First, an increase in the median may be explained by a loss of intermediate-paying jobs, with *no subsequent replacement* by low-earning contracts. The increase in the upper tail could also be driven by a similar phenomenon, though an alternative explanation is also plausible. In Brumadinho, most private high-earning jobs are offered by Vale. Inasmuch as Vale workers are able to extract higher concessions from the company after the rupture, this could lead to an increase in the upper tail.[12]

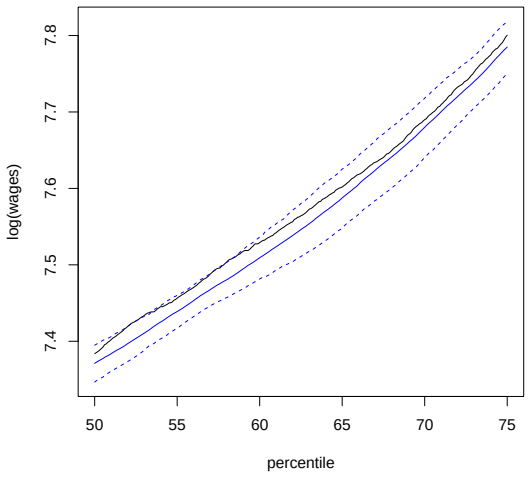
[12] In a similar vein, the increase in the median could be alternatively driven by an increase in median-paying jobs three years into the barrage rupture, e.g. due to sectoral changes in the local economy driven by the disaster.



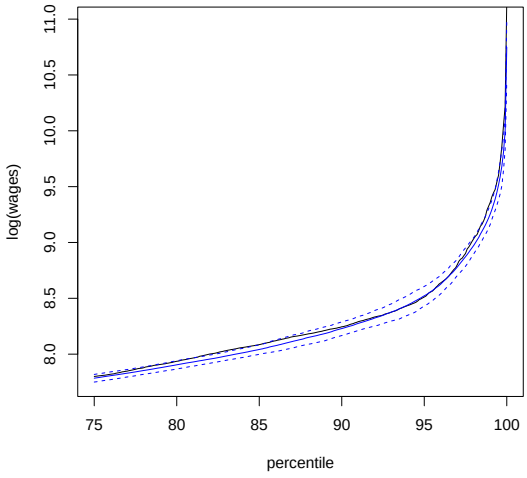
(a) Percentiles 0-25



(b) Percentiles 25-50

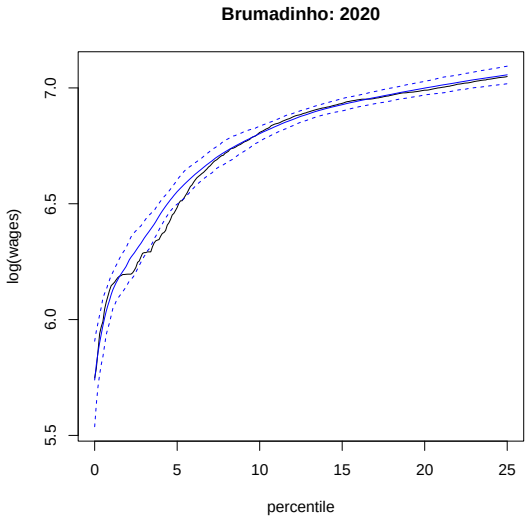


(c) Percentiles 50-75

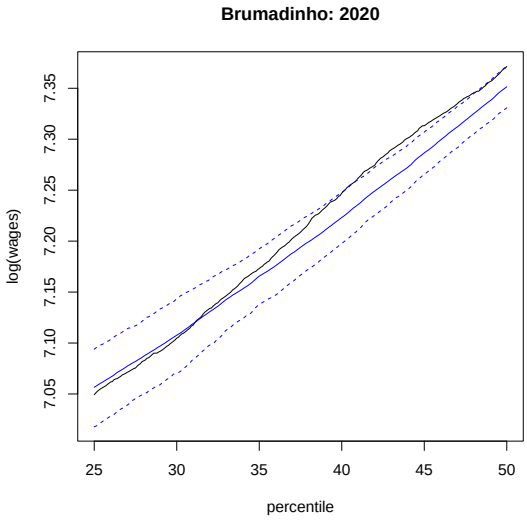


(d) Percentiles 75-100

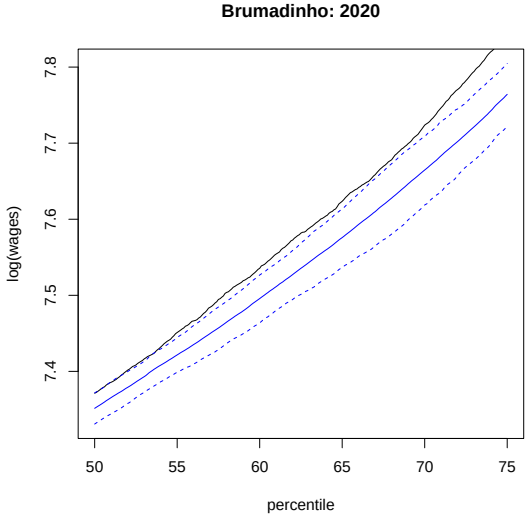
Figure 3.2: Treatment effects (2019).



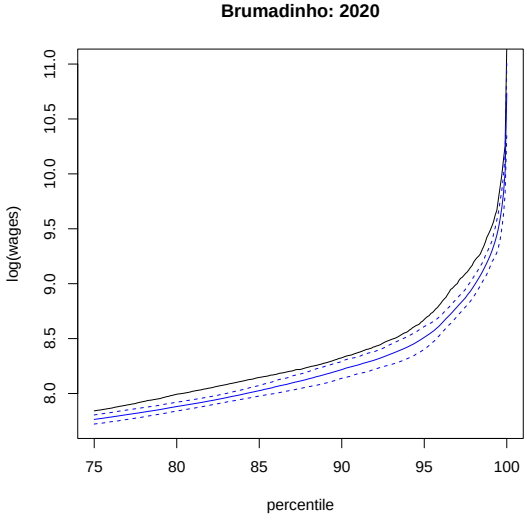
(a) Percentiles 0-25



(b) Percentiles 25-50

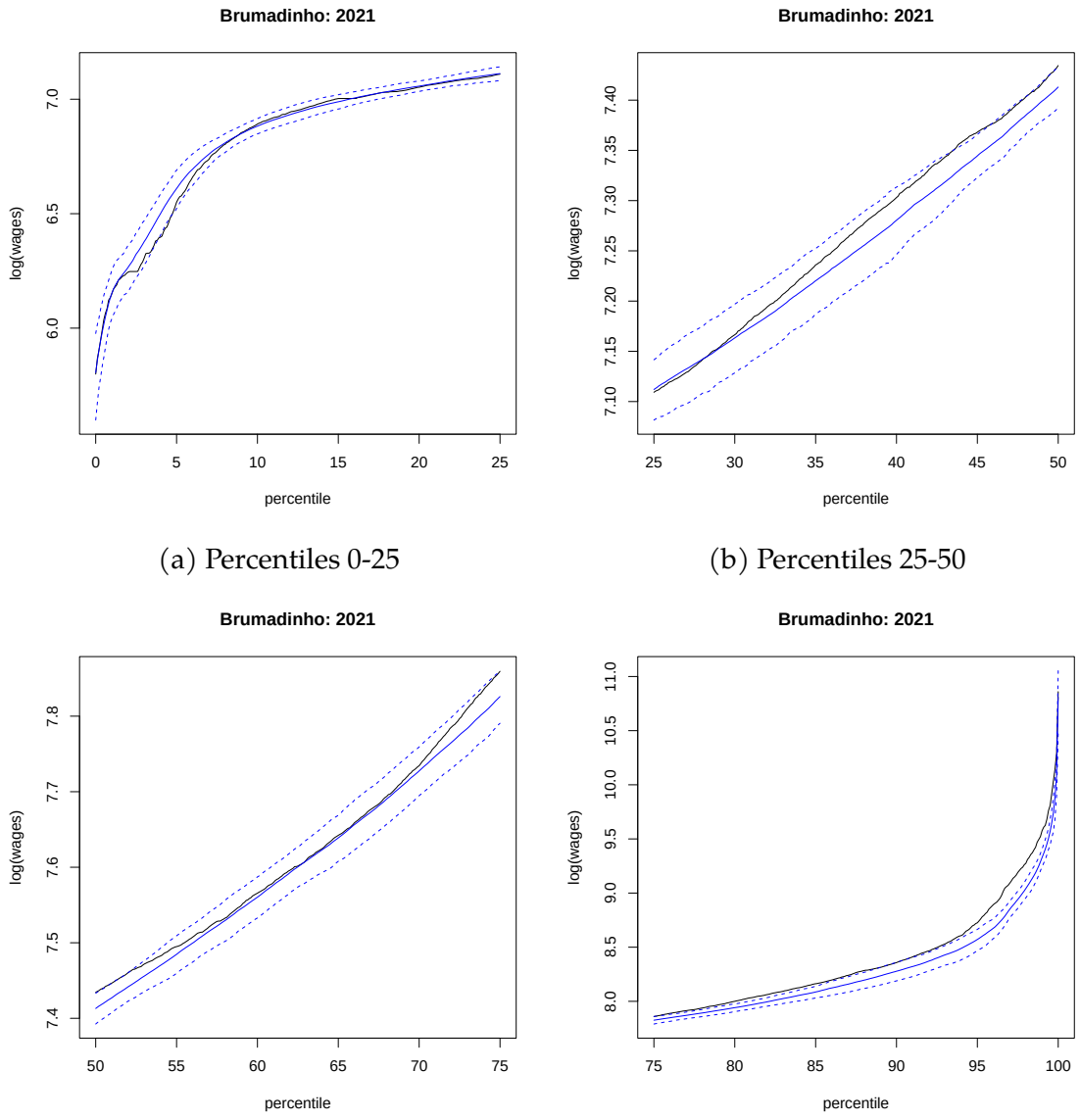


(c) Percentiles 50-75



(d) Percentiles 75-100

Figure 3.3: Treatment effects (2020).



(a) Percentiles 0-25

(b) Percentiles 25-50

(c) Percentiles 50-75

(d) Percentiles 75-100

Figure 3.4: Treatment effects (2021).

3.8 Conclusion

In this paper, we have introduced nonparametric quantile mixture models as an attractive counterpart to nonparametric density mixture models. We have shown that, similarly to density mixture models, nonparametric quantile mixtures exhibit interesting approximation properties. They are also closely connected to the distributional synthetic controls recently proposed by Gunsilius 2023.

We have introduced estimation and inference tools for nonparametric mixtures by relying on

the concept of *L-moments*. Introduced by Hosking 1990, L-moments are robust alternative alternative to standard moments that characterise distributions with finite first moments. The estimation of models by matching a weighted distance between sample and theoretical L-moments has been shown to produce statistically efficient estimators in parametric (Alvarez, Chiann, and Morettin 2023) and semiparametric (Alvarez and Biderman 2022) settings. In this paper, we have extended this estimation approach to a sieve setting, which we show leads to a computationally convenient estimator with tractable statistical properties.

We develop a full inferential theory for our proposed estimator – including a general approach to constructing confidence sets on mixture weights –, by relying on the concept of strong approximation. In so doing, we make two contributions to statistical theory, which we believe may be of independent interest. First, we introduce a lemma that bounds the strong approximation of a quadratic optimizer by an approximating program in terms of its constituent elements. Inasmuch as strong approximations are an ever more prevalent tool in devising inferential procedures in nonstandard or high-dimensional settings (Cattaneo, Farrell, and Y. Feng 2020; Chernozhukov, Chetverikov, and Kato 2014; Chernozhukov, Newey, and Santos 2023; Fang et al. 2023), this result may be useful in other contexts. Second, we develop a strategy that enables extending anticoncentration inequalities for Gaussian random variables available in the literature to the projection of Gaussian random variables onto Euclidean balls. Such procedure is useful for providing inferential guarantees in constrained estimation settings.

As theoretical applications of our proposed methodology, we show how quantile mixtures may be used to recover estimates (and confidence sets) of a nonparametric *density* model. We also show that, as a direct byproduct of our theory, one is able to provide a valid inferential procedure in the distributional synthetic control setup of Gunsilius 2023, where formal inference procedures were previously unavailable. As an empiricl application, we apply our inferential procedure in order to assess the distributional impacts of the Brumadinho barrage rupture. Our approach enables us to uncover a rage of effects across the wage distribution.

We believe there are several additional applications to our proposed methodology. For example, in a risk manamagement setting, one may be tempted to model the quantile function of the prediction error of an algorithm as a mixture of extreme value distributions in order to better understand the likelihood of extreme returns. Even though our theory is general enough to accomodate this setting, our results require a strong approximation to the adopted quantile estimator, which is generally unavailable when these are empirical quantiles from the residuals of a first-step. As we discuss in Section 3.3, it appears possible to extend the strong approximation of Csorgo and Revesz 1978 to this setting by relying on sample-splitting and cross-fitting. We intend to pursue such extensions in future research.

References

Alvarez, Luis and Ciro Biderman (2022). *Semiparametric analysis of randomized experiments using L-moments*.

Alvarez, Luis, Chang Chiann, and Pedro Morettin (2023). *Inference in parametric models with many L-moments*. arXiv: 2210.04146 [stat.ME].

Armstrong, Timothy B. and Michal Kolesár (2021). "Finite-Sample Optimal Estimation and Inference on Average Treatment Effects Under Unconfoundedness". In: *Econometrica* 89.3, pp. 1141–1177. DOI: https://doi.org/10.3982/ECTA16907. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA16907. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA16907.

Armstrong, Timothy B., Michal Kolesár, and Mikkel Plagborg-Møller (2022). "Robust Empirical Bayes Confidence Intervals". In: *Econometrica* 90.6, pp. 2567–2602. DOI: https://doi.org/10.3982/ECTA18597. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA18597. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA18597.

Belloni, Alexandre, Victor Chernozhukov, Denis Chetverikov, and Kengo Kato (2015). "Some new asymptotic theory for least squares series: Pointwise and uniform results". In: *Journal of Econometrics* 186.2. High Dimensional Problems in Econometrics, pp. 345–366. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2015.02.014. URL: https://www.sciencedirect.com/science/article/pii/S030440761500038X.

Bickel, Peter J., Ya'acov Ritov, and Alexandre B. Tsybakov (2009). "Simultaneous analysis of Lasso and Dantzig selector". In: *The Annals of Statistics* 37.4, pp. 1705–1732. DOI: 10.1214/08-AOS620. URL: https://doi.org/10.1214/08-AOS620.

Cattaneo, Matias D., Max H. Farrell, and Yingjie Feng (2020). "Large sample properties of partitioning-based series estimators". In: *The Annals of Statistics* 48.3, pp. 1718–1741. DOI: 10.1214/19-AOS1865. URL: https://doi.org/10.1214/19-AOS1865.

Cattaneo, Matias D., Yingjie Feng, and Rocio Titiunik (2021). "Prediction Intervals for Synthetic Control Methods". In: *Journal of the American Statistical Association* 116.536. PMID: 35756161, pp. 1865–1880. DOI: 10.1080/01621459.2021.1979561. eprint: https://doi.org/10.1080/01621459.2021.1979561. URL: https://doi.org/10.1080/01621459.2021.1979561.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins (Jan. 2018). "Double/debiased machine learning for treatment and structural parameters". In: *The Econometrics Journal* 21.1, pp. C1–C68. ISSN: 1368-4221. DOI: 10.1111/ectj.12097. eprint: https://academic.oup.com/ectj/article-pdf/21/1/C1/27684918/ectj00c1.pdf. URL: https://doi.org/10.1111/ectj.12097.

Chernozhukov, Victor, Denis Chetverikov, and Kengo Kato (2014). "ANTI-CONCENTRATION AND HONEST, ADAPTIVE CONFIDENCE BANDS". In: *The Annals of Statistics* 42.5, pp. 1787–1818. ISSN: 00905364. URL: http://www.jstor.org/stable/43556344 (visited on 06/29/2023).

— (2017a). "CENTRAL LIMIT THEOREMS AND BOOTSTRAP IN HIGH DIMENSIONS". In: *The Annals of Probability* 45.4, pp. 2309–2352. ISSN: 00911798. URL: http://www.jstor.org/stable/26362255 (visited on 06/21/2023).

— (2017b). *Detailed proof of Nazarov's inequality*. arXiv: 1711.10696 [math.ST].

Chernozhukov, Victor, Denis Chetverikov, Kengo Kato, and Yuta Koike (2023). "High-dimensional data bootstrap". In: *Annual Review of Statistics and Its Application* 10, pp. 427–449.

Chernozhukov, Victor, Whitney K. Newey, and Andres Santos (2023). "Constrained Conditional Moment Restriction Models". In: *Econometrica* 91.2, pp. 709–736. DOI: https://doi.org/10.3982/ECTA13830. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA13830. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA13830.

Chernozhukov, Victor, Kaspar Wüthrich, and Yinchu Zhu (2021). "An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls". In: *Journal of the American Statistical Association* 116.536, pp. 1849–1864. DOI: 10.1080/01621459.2021.1920957. eprint: https://doi.org/10.1080/01621459.2021.1920957. URL: https://doi.org/10.1080/01621459.2021.1920957.

Csorgo, Miklos and Pal Revesz (1978). "Strong Approximations of the Quantile Process". In: *The Annals of Statistics* 6.4, pp. 882–894. DOI: 10.1214/aos/1176344261. URL: https://doi.org/10.1214/aos/1176344261.

DeVore, Ronald A and George G Lorentz (1993). *Constructive approximation*. Vol. 303. Springer Science & Business Media.

Efron, Bradley (2014). "Two Modeling Strategies for Empirical Bayes Estimation". In: *Statistical Science* 29.2, pp. 285–301. ISSN: 08834237, 21688745. URL: http://www.jstor.org/stable/43288481 (visited on 06/29/2023).

— (Feb. 2016). "Empirical Bayes deconvolution estimates". In: *Biometrika* 103.1, pp. 1–20. ISSN: 0006-3444. DOI: 10.1093/biomet/asv068. eprint: https://academic.oup.com/biomet/article-pdf/103/1/1/24331584/asv068.pdf. URL: https://doi.org/10.1093/biomet/asv068.

Fan, Jianqing (1991). "On the optimal rates of convergence for nonparametric deconvolution problems". In: *The Annals of Statistics*, pp. 1257–1272.

Fan, Jianqing, Ricardo Masini, and Marcelo C. Medeiros (2022). *Bridging factor and sparse models*. arXiv: 2102.11341 [econ.EM].

Fang, Zheng, Andres Santos, Azeem M. Shaikh, and Alexander Torgovitsky (2023). "Inference for Large-Scale Linear Systems With Known Coefficients". In: *Econometrica* 91.1, pp. 299–327. DOI: https://doi.org/10.3982/ECTA18979. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA18979. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA18979.

Feng, Long and Lee H. Dicker (2018). "Approximate nonparametric maximum likelihood for mixture models: A convex optimization approach to fitting arbitrary multivariate mixing distributions". In: *Computational Statistics & Data Analysis* 122, pp. 80–91. ISSN: 0167-9473. DOI: https://doi.org/10.1016/j.csda.2018.01.006. URL: https://www.sciencedirect.com/science/article/pii/S0167947318300070.

Ferman, Bruno (2021). "On the Properties of the Synthetic Control Estimator with Many Periods and Many Controls". In: *Journal of the American Statistical Association* 116.536, pp. 1764–1772. DOI: 10.1080/01621459.2021.1965613. eprint: https://doi.org/10.1080/01621459.2021.1965613. URL: https://doi.org/10.1080/01621459.2021.1965613.

Fotopoulos, SB and SK Ahn (1994). "Strong approximation of the quantile processes and its applications under strong mixing properties". In: *Journal of Multivariate Analysis* 51.1, pp. 17–45.

Götze, Friedrich, Alexey Naumov, Vladimir Spokoiny, and Vladimir Ulyanov (2019). "Large ball probabilities, Gaussian comparison and anti-concentration". In: *Bernoulli* 25.4A, pp. 2538–2563. DOI: 10.3150/18-BEJ1062. URL: https://doi.org/10.3150/18-BEJ1062.

Gourieroux, C. and J. Jasiak (2008). "Dynamic quantile models". In: *Journal of Econometrics* 147.1. Econometric modelling in finance and risk management: An overview, pp. 198–205. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2008.09.028. URL: https://www.sciencedirect.com/science/article/pii/S0304407608001358.

Gu, Jiaying and Roger Koenker (2023). "Invidious Comparisons: Ranking and Selection as Compound Decisions". In: *Econometrica* 91.1, pp. 1–41. DOI: https://doi.org/10.3982/ECTA19304. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA19304. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA19304.

Gunsilius, F. F. (2023). "Distributional Synthetic Controls". In: *Econometrica* 91.3, pp. 1105–1117. DOI: https://doi.org/10.3982/ECTA18260. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA18260. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA18260.

Hong, Han and Jessie Li (2020). "The numerical bootstrap". In: *The Annals of Statistics* 48.1, pp. 397–412. DOI: 10.1214/19-AOS1812. URL: https://doi.org/10.1214/19-AOS1812.

Hosking, J. R. M. (1990). "L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 52.1, pp. 105–124. ISSN: 00359246. URL: http://www.jstor.org/stable/2345653 (visited on 06/13/2023).

Hosking, J. R. M. and J. R. Wallis (1987). "Parameter and Quantile Estimation for the Generalized Pareto Distribution". In: *Technometrics* 29.3, pp. 339–349. ISSN: 00401706. URL: http://www.jstor.org/stable/1269343 (visited on 06/11/2023).

Hosking, J. R. M., J. R. Wallis, and E. F. Wood (1985). "Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments". In: *Technometrics* 27.3, pp. 251–261. ISSN: 00401706. URL: http://www.jstor.org/stable/1269706 (visited on 06/12/2023).

Ignatiadis, Nikolaos and Stefan Wager (2022). "Confidence Intervals for Nonparametric Empirical Bayes Analysis". In: *Journal of the American Statistical Association* 117.539,

pp. 1149–1166. DOI: 10.1080/01621459.2021.2008403. eprint: https://doi.org/10.1080/01621459.2021.2008403. URL: https://doi.org/10.1080/01621459.2021.2008403.

Jagabathula, Srikanth, Lakshminarayanan Subramanian, and Ashwin Venkataraman (2020). "A conditional gradient approach for nonparametric estimation of mixing distributions". In: *Management Science* 66.8, pp. 3635–3656.

Karvanen, Juha (2006). "Estimation of quantile mixtures via L-moments and trimmed L-moments". In: *Computational Statistics & Data Analysis* 51.2, pp. 947–959. ISSN: 0167-9473. DOI: https://doi.org/10.1016/j.csda.2005.09.014. URL: https://www.sciencedirect.com/science/article/pii/S0167947305002513.

Kato, Kengo (2009). "Asymptotics for argmin processes: Convexity arguments". In: *Journal of Multivariate Analysis* 100.8, pp. 1816–1829. ISSN: 0047-259X. DOI: https://doi.org/10.1016/j.jmva.2009.02.008. URL: https://www.sciencedirect.com/science/article/pii/S0047259X09000426.

Kiefer, J. and J. Wolfowitz (1956). "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters". In: *The Annals of Mathematical Statistics* 27.4, pp. 887–906. DOI: 10.1214/aoms/1177728066. URL: https://doi.org/10.1214/aoms/1177728066.

Kline, Patrick, Evan K Rose, and Christopher R Walters (2022). "Systemic discrimination among large US employers". In: *The Quarterly Journal of Economics* 137.4, pp. 1963–2036.

Koenker, Roger and Ivan Mizera (2014). "Convex Optimization, Shape Constraints, Compound Decisions, and Empirical Bayes Rules". In: *Journal of the American Statistical Association* 109.506, pp. 674–685. DOI: 10.1080/01621459.2013.869224. eprint: https://doi.org/10.1080/01621459.2013.869224. URL: https://doi.org/10.1080/01621459.2013.869224.

Kreyszig, Erwin (1978). "Introductory functional analysis with applications. Jonh Wiley & Sons". In: *Inc., New York*.

Li, Jessie (May 2021). "The Proximal Bootstrap for Finite-Dimensional Regularized Estimators". In: *AEA Papers and Proceedings* 111, pp. 616–20. DOI: 10.1257/pandp.20211036. URL: https://www.aeaweb.org/articles?id=10.1257/pandp.20211036.

Masini, Ricardo (2022). "Distributional Counterfactual Analysis in High-Dimensional Setup". In: *arXiv preprint arXiv:2202.11671*.

Meister, Alexander (2009). *Deconvolution Problems in Nonparametric Statistics*. Springer Berlin Heidelberg. DOI: 10.1007/978-3-540-87557-4. URL: https://doi.org/10.1007/978-3-540-87557-4.

Newey, Whitney K. (1997). "Convergence rates and asymptotic normality for series estimators". In: *Journal of Econometrics* 79.1, pp. 147–168. ISSN: 0304-4076. DOI: https://doi.org/10.1016/S0304-4076(97)00011-0. URL: https://www.sciencedirect.com/science/article/pii/S0304407697000110.

Nguyen, Hien D. and Geoffrey McLachlan (2019). "On approximations via convolution-defined mixture models". In: *Communications in Statistics - Theory and Methods* 48.16,

pp. 3945–3955. DOI: 10.1080/03610926.2018.1487069. eprint: https://doi.org/10.1080/03610926.2018.1487069. URL: https://doi.org/10.1080/03610926.2018.1487069.

Nguyen, T. Tin, Hien D. Nguyen, Faicel Chamroukhi, and Geoffrey J. McLachlan (2020). "Approximation by finite mixtures of continuous density functions that vanish at infinity". In: *Cogent Mathematics & Statistics* 7.1. Ed. by Lishan Liu, p. 1750861. DOI: 10.1080/25742558.2020.1750861. eprint: https://doi.org/10.1080/25742558.2020.1750861. URL: https://doi.org/10.1080/25742558.2020.1750861.

Noack, Claudia and Christoph Rothe (2021). *Bias-Aware Inference in Fuzzy Regression Discontinuity Designs*. arXiv: 1906.04631 [econ.EM].

Pollard, David (1991). "Asymptotics for least absolute deviation regression estimators". In: *Econometric Theory* 7.2, pp. 186–199.

Polyanskiy, Yury and Yihong Wu (2020). *Self-regularizing Property of Nonparametric Maximum Likelihood Estimator in Mixture Models*. arXiv: 2008.08244 [math.ST].

Shvedov, Alexei Sergeevich (1981). "Orders of coapproximation of functions by algebraic polynomials". In: *Mathematical notes of the Academy of Sciences of the USSR* 29, pp. 63–70.

Taussky, Olga (Jan. 1949). "A remark concerning the characteristic roots of the finite segments of the Hilbert Matrix". In: *The Quarterly Journal of Mathematics* os-20.1, pp. 80–83. ISSN: 0033-5606. DOI: 10.1093/qmath/os-20.1.80. eprint: https://academic.oup.com/qjmath/article-pdf/os-20/1/80/4667491/os-20-1-80.pdf. URL: https://doi.org/10.1093/qmath/os-20.1.80.

Train, Kenneth E (2008). "EM algorithms for nonparametric estimation of mixing distributions". In: *Journal of Choice Modelling* 1.1, pp. 40–69.

Yoshihara, Kenichi (1995). "The Bahadur representation of sample quantiles for sequences of strongly mixing random variables". In: *Statistics & Probability Letters* 24.4, pp. 299–304. ISSN: 0167-7152. DOI: https://doi.org/10.1016/0167-7152(94)00187-D. URL: http://www.sciencedirect.com/science/article/pii/016771529400187D.

Appendix C

3.9 Proof of main results in Section 3.4

3.9.1 A lemma on the approximation of quadratic minimizers

Before proving the results in Section 3.4, we state an auxiliary lemma, which concerns the strong approximation of quadratic minimizers. This lemma may be of independent interest:

Lemma 4 (Strong approximation of quadratic minimizers)**.** *Consider the program:*

$$\psi \in \min_{\xi \in \mathcal{X}} -\xi' S + \frac{1}{2}\xi' W \xi,$$

where $\mathcal{X}$ is a convex subset of $\mathbb{R}^d$, S is a $d \times 1$ random vector, and W is a $d \times d$ symmetric random matrix. Consider the alternative program:

$$\psi^* \in \min_{\xi \in \mathcal{X}} -\xi' Z + \frac{1}{2}\xi' \Omega \xi,$$

where Z is a $d \times 1$ random vector and Ω is a $d \times d$ symmetric random matrix. Define the restricted eigenvalue:

$$\lambda_{\psi^*}(\Omega) = \inf_{\xi \in \mathcal{X}} \frac{(\xi - \psi^*)'\Omega(\xi - \psi^*)}{\|\xi - \psi^*\|_2^2}.$$

We then have that, for any $\delta > 0$:

$$\begin{aligned}\mathbb{P}[\|\psi - \psi^*\|_2 > \delta] \leqslant \inf_{c>0, M>0, \kappa \in (0,1)} \Big\{ & \mathbb{P}\left[\|S - Z\|_2 \geqslant \kappa \frac{c}{2} \frac{\delta^2}{(\delta + M)}\right] + \\ & \mathbb{P}\left[\|W - \Omega\|_2 \geqslant (1-\kappa)\frac{c}{2}\frac{\delta^2}{(\delta + M)^2}\right] + \mathbb{P}\left[\|\psi^*\| \geqslant M\right] + \mathbb{P}[\lambda_{\psi^*}(\Omega) \leqslant c]\Big\}\end{aligned} \tag{3.21}$$

Proof. Let $g^*(\xi) := -\xi' Z + \frac{1}{2}\xi'\Omega\xi$. We begin by noticing that, by convexity, for any $\xi \in \mathcal{X}$, we must have that:

$$(-Z + \Omega\psi^*)'(\xi - \psi^*) \geqslant 0.$$

The latter implies that:

$$g^*(\xi) - g^*(\psi^*) \geqslant -\psi^{*'}\Omega(\xi - \psi^*) + \frac{1}{2}\xi'\Omega\xi - \frac{1}{2}\psi^{*'}\Omega\psi^* = \frac{1}{2}(\xi - \psi^{*'})'\Omega(\xi - \psi^*).$$

It then follows, by the definition of the restricted eigenvalue, that:

$$g^*(\xi) - g^*(\psi^*) \geqslant \frac{\lambda_{\theta_n^*}(\Omega_n)}{2} \|\hat{\theta}_n - \theta_n^*\|_2^2.$$

Next, proceeding similarly to the proof of Theorem 2 of Kato 2009, we obtain the following bound, for each $\delta > 0$:

$$\mathbb{P}[\|\psi - \psi^*\|_2 > \delta] \leqslant \mathbb{P}[\Delta(\delta) > \lambda_{\psi^*}(\Omega)\delta^2/2], \tag{3.22}$$

where $\Delta(\delta) = \sup_{\xi \in \mathcal{X}: \|\xi - \psi^*\|_2 \leqslant \delta} |-\xi'(S - Z) + (1/2)\xi'(W - \Omega)\xi|$. Note that, for any $c > 0$ and $M > 0$:

$$\mathbb{P}[\|\psi - \psi^*\|_2 > \delta] \leqslant \mathbb{P}[(M+\delta)\|S - Z\|_2 + (M+\delta)^2\|\Omega - W\|_2/2 > c\delta^2/2] + \mathbb{P}[\|\psi^*\| \geqslant M] + \mathbb{P}[\lambda_{\psi^*}(\Omega) \leqslant c],$$

and, then, by the union bound, for any $\kappa \in (0, 1)$

$$\mathbb{P}[\|\psi - \psi^*\|_2 > \delta] \leqslant \mathbb{P}\left[\|S - Z\|_2 \geqslant \kappa \frac{c}{2} \frac{\delta^2}{(\delta + M)}\right] + \mathbb{P}\left[\|W - \Omega\|_2 \geqslant (1 - \kappa)\frac{c}{2}\frac{\delta^2}{(\delta + M)^2}\right] + \mathbb{P}\left[\|\psi^*\| \geqslant M\right] + \mathbb{P}[\lambda_{\psi^*}(\Omega) \leqslant c].$$

Optimizing leads to the desired result. □

Lemma 4 decomposes the approximation of a quadratic minimizer by another quadratic minimizer onto four terms: the first two terms concern the approximation of the linear and quadratic terms of the target program; the third term restricts the norm of the approximating solution; and the final term limits the "restricted eigenvalue" of the approximating program. We note that the third term, which concerns the norm of the approximating solution, may be further bounded by applying Lemma 4 recursively. Specifically, one may bound the norm of the approximation ψ^* in terms of a nonstochastic approximation (where random variables are replaced by their expected values). In the next subsection, we explore this idea to prove a general result on the rates of our L-moment estimator.

3.9.2 Proof of Proposition 2

We apply Lemma 4 to our problem. We consider the "undersmoothed" approximation $\sqrt{n}(\tilde{\theta}_n^* - \theta_{0,n})$. The approximation $\sqrt{n}(\theta_n^* - \theta_{0,n})$ is similar and therefore omitted. To apply the lemma, we define:

$$
\begin{aligned}
S &= \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} (\sqrt{n}(\hat{Q}_n(u) - Q(u)) + D_n(u))\boldsymbol{P}_L(u)du\right) \\
W &= \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right) \\
Z &= \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' \Omega_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} B_{0,n}(u)\boldsymbol{P}_L(u)du\right) \\
\Omega &= \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' \Omega_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)
\end{aligned}
\tag{3.23}
$$

We will provide rates for each term. We first note that, for any $\phi \in \mathbb{R}^p$, Bessel's inequality yields:

$$
\left\| \left(\int_{\underline{p}_n}^{\overline{p}} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)\phi \right\|_2^2 \leqslant \int_{\underline{p}_n}^{\overline{p}_n} (\boldsymbol{J}_{n,p}(u)'\phi)^2 du = \phi' \left(\int_{\underline{p}_n}^{\overline{p}_n} (\boldsymbol{J}_{n,p}(u)\boldsymbol{J}_{n,p}(u)')du\right)\phi\,.
$$

Consequently, by the definition of the spectal norm, we obtain that:

$$
\left\| \int_{\underline{p}_n}^{\overline{p}} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du \right\|_2 \leqslant \sqrt{\rho_n}\,.
$$

We also note that Bessel's inequality yields

$$
\left\| \int_{\underline{p}_n}^{\overline{p}_n} (\sqrt{n}(\hat{Q}_n(u) - Q(u)) - B_{0,n}(u))\boldsymbol{P}_l(u)du \right\|_2 \leqslant \|\sqrt{n}(\hat{Q}_n - Q) - B_{0,n}\|_{L^2[\underline{p}_n,\overline{p}_n]} = O_{\mathbb{P}}(r_n)\,,
$$

and

$$
\left\| \int_{\underline{p}_n}^{\overline{p}_n} D_n(u)\boldsymbol{P}_l(u)du \right\|_2 \leqslant \|D_n\|_{L^2[\underline{p}_n,\overline{p}_n]} = d_n\,.
$$

Combining these facts, we obtain that:

$$
\begin{aligned}
\|Z - S\|_2 &= O_{\mathbb{P}}(\sqrt{\rho}_n(r_n \vee d_n \vee s_n))\,, \\
\|W - \Omega\|_2 &= O_{\mathbb{P}}(\rho_n s_n)\,.
\end{aligned}
$$

Combining these facts with the fact that $A_n = O_{\mathbb{P}}(1) \implies \lim_{s\to\infty} \mathbb{P}[|A_n| > s] = 0$, and applying Lemma 4 twice: first, to approximate $\sqrt{n}(\hat{\theta}_n - \theta_{0,n})$ with $\sqrt{n}(\tilde{\theta}_n^* - \theta_{0,n})$; then, to approximate $\sqrt{n}(\tilde{\theta}_n^* - \theta_{0,n})$ with 0 (the minimizer of $\min_{x\in\mathcal{X}_n} x'\Omega x$); we obtain the desired result.

3.9.3 Proof of Proposition 3

One may prove this proposition by relying on Proposition 2. Instead, we offer a direct proof. By Weyl's inequality, see page 241 of Wainwright 2019:

$$\left|\lambda_{\min}\left(\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)\right)\right.$$
$$\left.-\lambda_{\min}\left(\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' \Omega_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)\right)\right| =$$
$$O(\rho_n s_n)$$

Consequently, under our assumptions:

$$\lim_{n\to\infty} \mathbb{P}\left[\lambda_{\min}\left(\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)\right) > 0\right] = 1\,,$$

and we may work with representation (3.11).

Next, we observe that, for symmetric invertible matrices A and B:

$$\|A^{-1} - B^{-1}\| = \|A^{-1}(A-B)B^{-1}\| \leqslant \|A^{-1}\|\|A-B\|\|B^{-1}\| \leqslant \|A^{-1}\|^2\|A-B\| + \|A^{-1}\|\|A-B\|^2\,,$$

where the last inequality follows from $\|B^{-1}\| \leqslant \|A^{-1}\| + |\|A^{-1}\| - \|B^{-1}\||$ and Weyl's inequality.

Applying these results, we obtain the desired conclusion.

3.9.4 Proof of Proposition 4

The first part of the proposition follows by applying Proposition 2. The second part of the proof follows from the projection characterization of (3.7) and Example 6 in Appendix 3.10, by taking $\epsilon_p = (\log(p)^\nu(\rho_n s_n)^{-\gamma}p^{1/2+\iota})^{-1}$ and noticing that $\underline{\sigma}\sqrt{n} \leqslant \operatorname{tr}(\Sigma) \leqslant \sqrt{\|\Sigma\|_2}\sqrt{n}$, where Σ is the variance matrix of the approximation in the unconstrained setting, which can be shown to satisfy, $\|\Sigma\| = O((\rho_n s_n)^{-2\gamma})$; and $\underline{\sigma}^2$ is the lower bound on the variance.

3.9.5 Proof of Proposition 5

The proof is analogous to the proof of Proposition 2, in that we apply Lemma 4 to our problem. We start by defining $\hat{D}_n(u) := \theta'_{n,0}\left(\boldsymbol{J}_{n,p}(u) - \boldsymbol{J}^*_{n,p}(u)\right)$ and $D_n(u) := Q(u) - \theta'_{n,0}\boldsymbol{J}^*_{n,p}(u)$. Next, define:

$$
\begin{aligned}
S &= \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} (\sqrt{n}(\hat{Q}_n(u) - Q(u)) + D_n(u) + \hat{D}_n(u))\boldsymbol{P}_L(u)du\right) \\
W &= \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right) \\
Z &= \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' \Omega_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} B_{0,n}(u)\boldsymbol{P}_L(u)du\right) \\
\Omega &= \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' \Omega_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)
\end{aligned}
\tag{3.24}
$$

Following the same steps as in Appendix 3.9.2, we arrive at:

$$
\begin{aligned}
\|Z - S\|_2 &= O_{\mathbb{P}}(\sqrt{(\rho_n \vee \varepsilon_n)}(r_n \vee d_n \vee s_n \vee \xi_n)), \\
\|W - \Omega\|_2 &= O_{\mathbb{P}}((\rho_n \vee \varepsilon_n)\, s_n),
\end{aligned}
$$

and the first conclusion follows similarly as in Appendix 3.9.2. The conclusion that convergence also holds conditionally follows from Markov inequality.

3.10 Anti-concentration inequalities on the projection onto Euclidean balls

This appendix shows how to extend Gaussian anti-concentration inequalities available in the literature to the projection of Gaussian vectors onto an Euclidean ball.

We consider $Z \sim N(0, \Sigma)$ a $d \times 1$ normal random variable. Let $X := P_{B(0;M)}(Z)$ be the projection of Z onto the Euclidean ball of radius M and centered at the origin. We consider bounding the anticoncentration function, $\mathcal{L}_X^{\infty}(\epsilon; \mathcal{C})$, where:

$$
\mathcal{L}_X^{\infty}(\epsilon; \mathcal{C}) := \sup_{C \in \mathcal{C}} \mathbb{P}[X \in C_{\infty}^{\epsilon} \setminus C_{\infty}^{-\epsilon}].
$$

We begin by noting that:

$$
X = Z\mathbf{1}_{\|Z\|_2 \leqslant M} + M\frac{Z}{\|Z\|_2}\mathbf{1}_{\|Z\|_2 > M}.
$$

Let $Z^* = M\frac{Z}{\|Z\|_2}$. We note that the law of total probability yields :

$$
\mathcal{L}_X^{\infty}(\epsilon; \mathcal{C}) \leqslant \mathcal{L}_Z^{\infty}(\epsilon; \mathcal{C}) \wedge \mathbb{P}[\|Z\|_2 \leqslant M] + \underbrace{\sup_{C \in \mathcal{C}} \mathbb{P}\left[Z^* \in C_{\infty}^{\epsilon} \setminus C_{\infty}^{-\epsilon}, \|Z\|_2 > M\right]}_{= L^*(\epsilon)}
$$

The literature provides upper bounds on the first term for specific classes of sets. We seek to

bound the second term. Let $\psi^2 = \mathbb{E}[\|Z\|_2^2]$. Applying Lemma 3 with $s = \infty$ yields:

$$L^*(\epsilon) \leqslant \inf_{\delta \geqslant 0} \left\{ \mathbb{P}\left[M\|Z_j\|_\infty \left| \frac{\psi - \|Z\|_2}{\psi \|Z\|_2} \right| \geqslant \delta \right] + \sup_{C \in \mathcal{C}} \mathbb{P}\left[\frac{M}{\psi} Z \in \left(C_\infty^\epsilon \setminus C_\infty^{-\epsilon}\right)^\delta \setminus \left(C_\infty^\epsilon \setminus C_\infty^{-\epsilon}\right)^{-\delta} \right] \right\} \wedge \mathbb{P}[\|Z\|_2 > M] \,,$$

and we also have that, for any s:

$$\mathbb{P}\left[M\|Z_j\|_\infty \left| \frac{\psi - \|Z\|_2}{\psi \|Z\|_2} \right| \geqslant \delta \right] \leqslant \mathbb{P}\left[\|Z_j\|_\infty \geqslant s \right] + \mathbb{P}\left[\left| \frac{\psi - \|Z\|_2}{\psi \|Z\|_2} \right| \geqslant \frac{\delta}{Ms} \right] .$$

We will bound both terms on the right-hand-side. Let $\bar{\sigma}^2 = \max_i \mathbb{V}[Z_i]$. First, by Lemma 2.2.2 of Vaart and Wellner 1996:

$$\mathbb{P}\left[\|Z\|_\infty \geqslant s \right] \leqslant 2 \exp\left(- \frac{s^2}{K \log(d+1) \bar{\sigma}^2} \right) .$$

Next, we note that, by concavity of $x \mapsto \sqrt{x}$:[13]

$$\sqrt{|\psi^2 - \|Z\|^2|} \geqslant |\psi - \|Z\|_2|$$

Consequently, we have that:

$$\mathbb{P}\left[\left| \frac{\psi - \|Z\|_2}{\psi \|Z\|_2} \right| \geqslant \frac{\delta}{Ms} \right] \leqslant \mathbb{P}\left[|\psi^2 - \|Z\|^2| > \psi^2 \|Z\|_2^2 \frac{\delta^2}{M^2 s^2} \right] \leqslant \mathbb{P}\left[\left| \psi^2 - \|Z\|^2 \right| > \frac{\psi^2}{2} \left(\frac{\psi^2 \delta^2}{Ms^2} \wedge 1 \right) \right]$$

But then, the Hanson-Wright inequality (Theorem 1.1 of Rudelson and Vershynin 2013), yields:

$$\mathbb{P}[|\psi^2 - \|Z\|^2| > t] \leqslant 2 \exp\left(-c \min\left\{ \frac{t^2}{\|\Sigma\|_F^2}, \frac{t}{\|\Sigma\|_2} \right\} \right) \implies$$

$$\mathbb{P}\left[\left| \psi^2 - \|Z\|^2 \right| > \frac{\psi^2}{2} \left(\frac{\psi^2 \delta^2}{Ms^2} \wedge 1 \right) \right] \leqslant 2 \exp\left(-c \min\left\{ \frac{\psi^4}{4\|\Sigma\|_F^2} \left(\frac{\psi^2 \delta^2}{Ms^2} \wedge 1 \right)^2, \frac{\psi^2}{2\|\Sigma\|_2} \left(\frac{\psi^2 \delta^2}{Ms^2} \wedge 1 \right) \right\} \right) ,$$

But, by Von-Neumann trace inequality,

$$\|\Sigma\|_F^2 = \operatorname{tr}(\Sigma^2) \leqslant \|\Sigma\|_2 \operatorname{tr}(\Sigma) = \|\Sigma\|_2 \psi^2 .$$

Equalling the upper bounds corresponding to the tail inequalities yields:

[13] Recall that $\sqrt{a} + \sqrt{b} \geqslant \sqrt{a+b}$, for any $a, b \geqslant 0$.

$$\mathbb{P}\left[M\|Z_j\|_\infty \left|\frac{\psi - \|Z\|_2}{\psi\|Z\|_2}\right| \geqslant \delta\right] \leqslant 4\exp\Big(-\min\Big\{\frac{c^{1/3}\psi^2\delta^{4/3}}{K^{2/3}\log(d+1)^{2/3}\bar{\sigma}^{4/3}(4M\|\Sigma\|_2)^{1/3}}, \frac{\sqrt{c}\psi^2\delta}{K^{1/2}\log(d+1)^{1/2}\bar{\sigma}(2M\|\Sigma\|_2)^{1/2}}, \frac{c\psi^2}{4\|\Sigma\|_2}\Big\}\Big), \tag{3.25}$$

Combining the above leads to the following lemma.

Lemma 5. *For any class of sets* $\mathcal{C} \subseteq \mathcal{B}\left(\mathbb{R}^d\right)$,

$$\mathcal{L}_X^\infty(\epsilon;\mathcal{C}) \leqslant \mathcal{L}_Z^\infty(\epsilon;\mathcal{C}) \wedge \mathbb{P}[\|Z\|_2 \leqslant M] + \inf_{\delta\geqslant 0}\Big\{\overline{u}(\delta) + \sup_{C\in\mathcal{C}}\mathbb{P}\left[\frac{M}{\psi}Z \in \left(C_\infty^\epsilon \setminus C_\infty^{-\epsilon}\right)^\delta \setminus \left(C_\infty^\epsilon \setminus C_\infty^{-\epsilon}\right)^{-\delta}\right]\Big\} \wedge \mathbb{P}[\|Z\|_2 > M], \tag{3.26}$$

where

$$\overline{u}(\delta) = 4\exp\Big(-\min\Big\{\frac{c^{1/3}\psi^2\delta^{4/3}}{K^{2/3}\log(d+1)^{2/3}\bar{\sigma}^{4/3}(4M\|\Sigma\|_2)^{1/3}}, \frac{\sqrt{c}\psi^2\delta}{K^{1/2}\log(d+1)^{1/2}\bar{\sigma}(2M\|\Sigma\|_2)^{1/2}}, \frac{c\psi^2}{4\|\Sigma\|_2}\Big\}\Big).$$

The previous lemma offers an upper bound to anticoncentration in the sup-norm in terms of estimable quantities. The second term inside the infimum is itself an anticoncentration inequality of a rescaled Gaussian vector. In what follows, we illustrate our lemma by providing anticoncentration rates in the class of hyperrectangles.

Example 6 (Hyperrectangles)**.** *Consider the class of hyperrectangles* $\mathcal{C}_d$*. In this case, Nazarov's inequality reveals that:*

$$\sup_{C\in\mathcal{C}_d}\mathbb{P}\left[\frac{M}{\psi}Z \in \left(C_\infty^\epsilon \setminus C_\infty^{-\epsilon}\right)^\delta \setminus \left(C_\infty^\epsilon \setminus C_\infty^{-\epsilon}\right)^{-\delta}\right] \leqslant \frac{\psi(\epsilon+\delta)}{\underline{\sigma}M}\left(\sqrt{2\log(2d)} + 2\right).$$

Take $\epsilon = \delta$*. We can then use Lemma 5 to find a sequence* ϵ_d *such that* $\mathcal{L}_X^\infty(\epsilon_d;\mathcal{C}) \to 0$*.*

Remark 11. *It is possible to adapt the proof of Lemma 5 to bound* $\mathcal{L}_X^2(\epsilon;\mathcal{C})$*. In this case, one replaces the maximal inequality that bounds* $\mathbb{P}[\|Z\|_\infty \geqslant \delta]$ *with Proposition 1 of Hsu, Kakade, and Zhang 2012, which provides an upper bound to* $\mathbb{P}[\|Z\|_2 \geqslant \delta]$*.*

3.11 On the choice of tuning sequence γ_n

Define the following terms:

$$S_n = \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n}\left(\frac{B_{0,n}(u)-B^*_{0,n}(u)}{\sqrt{n}} + (\hat{Q}(u)-\hat{\theta}'\boldsymbol{J}_{n,p}(u))\right)\boldsymbol{P}_L(u)du\right)$$
$$W_n = \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right), \tag{3.27}$$

where $B^*_{0,n}(u)$ is an independent copy of $B_{0,n}(u)$.The bootstrap estimator is defined as:

$$\theta^B_n \in \operatorname{arginf}_{\theta\in\Theta_n} -\theta' S_n + \frac{1}{2}\left(\theta-\hat{\theta}\right)' W_n\left(\theta-\hat{\theta}\right), \tag{3.28}$$

We consider the problem of approximating the distribution of $\gamma_n(\theta^*_n-\theta_{0,n})$ with $\gamma_n(\theta^B_n-\hat{\theta}_n)$. We start with the strong approximation, note that, by the triangle inequality

$$\|\gamma_n(\theta^B_n-\hat{\theta}_n)-\gamma_n(\theta^*_n-\theta_{0,n})\| \leqslant \left\|\frac{\gamma_n\sqrt{n}(\hat{\theta}_n-\theta_{0,n})}{\sqrt{n}}\right\| + \|\gamma_n(\theta^B_n-\theta_{0,n})-\gamma_n(\theta^*_n-\theta_{0,n})\|.$$

Following the proof of Proposition 2, we have

$$\left\|\frac{\gamma_n\sqrt{n}(\hat{\theta}_n-\theta_{0,n})}{\sqrt{n}}\right\| = O_{\mathbb{P}}\left(\frac{\gamma_n}{\sqrt{n}}\left(\frac{\sqrt{\rho_n}}{\lambda_{0,n}}\vee\xi_n\right)\right),$$

where $\|\sqrt{n}(\hat{\theta}_n-\theta_{0,n})-\sqrt{n}(\theta^*_n-\theta_{0,n})\| = O_{\mathbb{P}}(\xi_n)$. Now note that the optimization problem (3.28) is equivalent to the problem:

$$\gamma_n(\theta^B_n-\theta_{0,n}) \in \operatorname{arginf}_{x\in\gamma_n(\Theta_n-\theta_{0,n})} -x'\tilde{S}_n + \frac{1}{2}x'W_nx, \tag{3.29}$$

with

$$\tilde{S}_n := \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \frac{\gamma_n}{\sqrt{n}}\left(B_{0,n}(u)+D_n(u)+\sqrt{n}(\hat{Q}(u)-Q(u))-B^*_{0,n}(u)\right)\boldsymbol{P}_L(u)du\right)$$

Then, we can appeal to Lemma 4. We conclude:

$$\|\gamma_n(\theta^B_n-\theta_{0,n})-\gamma_n(\theta^*_n-\theta_{0,n})\| = O_{\mathbb{P}}\left(\frac{\gamma_n}{\sqrt{n}}\sqrt{\rho_n}\,(r_n)\right).$$

Thus we need the additional condition that we can find a sequence γ_n, such that

$$\frac{\gamma_n}{\sqrt{n}}\left(\frac{\sqrt{\rho_n}}{\lambda_{0,n}}\vee\xi_n\vee\sqrt{\rho}_n r_n\vee\sqrt{\rho}_n s_n\right)\vee\rho_n s_n \to 0.$$

Considering the conditions for inference, this imposes the additional requirement that:

$$\frac{\gamma_n}{\sqrt{n}} \frac{\sqrt{\rho_n}}{\lambda_{0,n}} \to 0.$$

3.12 Approximation through nonnegative weights

In this appendix, we show that any nonnegative random variable whose quantile function satisfies Assumption 10 can be arbitrarily well approximated (in the L^2-norm) by a mixture of "truncated" Extreme Value quantiles with an underlying nonnegative measure.

3.12.1 Setup

We consider the problem of approximating a quantile function Q by "truncations" of quantile functions given by

$$Q_\theta(u) = \frac{1}{\theta}(-\log u)^{-1}, \quad \theta \in (0, \infty)$$

The quantile functions presented above represent the quantiles of random variables that conform to a generalized extreme value distribution, with associated scale parameter given by $\sigma = \frac{1}{\theta}$, shape parameter given by $\xi = 1$ and scale given by $\mu = \frac{1}{\theta}$.

Assumption 10. *We make additional assumptions about Q*

1. $Q(0) \geqslant 0$.
2. $Q \in L^2[0,1]$.
3. *Q is differentiable and strictly increasing in* $(0,1)$.

3.12.2 Approximation Result

Lemma 6. *Let Q be a quantile function satisfying Assumption 10 and let,*

$$\alpha_{\theta,k}(u) := \frac{1}{\theta}(-\log u)^{-1} \mathbf{1}\left[\theta \geqslant \frac{u}{k}\right] \mathbf{1}\left[\theta \leqslant \frac{u^{(1-\frac{C}{k})}}{k}\right],$$

for some $C > 0$. *There exists a sequence of nonnegative measure* $\{\mu_k^+\}_{k\in\mathbb{N}}$ *on* $(\mathbb{R}_+, \mathcal{B}[\mathbb{R}_+])$, *such that,*

$$\left\| Q - \int_{\mathbb{R}_{++}} \alpha_{\theta,k} \mu_k^+(d\theta) \right\|_{L^2[0,1]} \to 0,$$

as $k \to \infty$. *Moreover, these mixtures are themselves quantiles.*

Proof. Fix $C > 0$. Define the following functions:

$$\alpha(m) := \frac{1}{m}$$

$$h(m, x) := m^{-\log x}$$

$$g(m, k) := (em)^k$$

$$h^{-1}(z, x) := z^{-\frac{1}{\log x}}.$$

Here $h(e^{-1}, x) = e^{\log x} = x$, then we define the identity element $\eta := e^{-1}$. Define the sequence $\{m_k\}_{k\in\mathbb{N}}$, $m_k = \exp\left(\frac{C}{k} - 1\right)$ and define a sequence of functions indexed by $k \in \mathbb{N}$,

$$\alpha_k(m) = \frac{dg(m, k)}{dm}\alpha(m) = \frac{k}{m}.$$

We claim that $\alpha_k(m)$ is an approximate identity. First,

$$\int_\eta^{m_k} \frac{k}{m} = k\left(\log\left(\exp\left(\frac{C}{k} - 1\right)\right) + 1\right) = C$$

. Moreover, for every $\delta > 0$,

$$\lim_{k\to\infty} \int_{\eta+\delta}^{\infty} k\mathbf{1}\left(e^{-1} + \delta < \exp\left(\frac{C}{k} - 1\right)\right) dm = 0$$

then we can apply Lemma 7 to conclude that $\alpha_k(m)$ is a C-approximate identity. Defining

$$\left[\alpha_k * \frac{Q}{C}\right](u) := \int_\eta^{m_k} \alpha_k(m)\frac{Q(h(m, u))}{C} dm,$$

we can apply Lemma 7 to conclude

$$\left\|\alpha_k * \frac{Q}{C} - Q\right\|_{L^2[0,1]} \to 0, \quad \text{as } k \to \infty$$

Next, we show that the function $\left[\alpha_k * \frac{Q}{C}\right]$ is a quantile function. We can rewrite:

$$\left[\alpha_k * \frac{Q}{C}\right](u) = \int_\eta^{m_k} \alpha_k(m)\frac{Q(h(m, u))}{C} dm = \int_u^{u^{\left(1-\frac{C}{k}\right)}} \frac{dh^{-1}(z, u)}{dz}\alpha_k(h^{-1}(z, u))\frac{Q(z)}{C} dz =$$

$$= \int_u^{u^{\left(1-\frac{C}{k}\right)}} \frac{1}{-z\log u} z^{-\frac{1}{\log u}} \frac{k}{z^{-\frac{1}{\log u}}} \frac{Q(z)}{C} dz = \int_u^{u^{\left(1-\frac{C}{k}\right)}} \frac{k}{z}(-\log u)^{-1}\frac{Q(z)}{C} dz,$$

where the second equality follows from the change of variables formula of the Riemann Integral by

letting $z := h(m, x)$. Since Q is nonnegative, $\left[\alpha_k * \frac{Q}{C}\right]$ is non negative as well. Taking the derivative with respect to u by applying the Leibnitz Rule and simplifying we get:

$$\frac{d}{du}\left[\alpha_k * \frac{Q}{C}\right](u) = \frac{k}{C}\frac{1}{u(\log u)^2}\left(\int_u^{u^{\left(1-\frac{C}{k}\right)}} \frac{Q(z)}{z}dz + (-\log u)\left(Q\left(u^{\left(1-\frac{C}{k}\right)}\right) - Q(u)\right)\right),$$

further integrating by parts the term $\int_u^{u^{\left(1-\frac{C}{k}\right)}} \frac{Q(z)}{z}dz$, we conclude

$$\frac{d}{du}\left[\alpha_k * \frac{Q}{C}\right](u) \geqslant 0 \iff \int_u^{u^{\left(1-\frac{C}{k}\right)}} (-\log z)Q'(z)dz \geqslant 0,$$

which is true by Assumption 10. Finally, we can get the desired result, by defining $\theta := \frac{z}{k}$ and applying the change of variables formula again to conclude:

$$\left[\alpha_k * \frac{Q}{C}\right](u) = \int_0^\infty \frac{1}{\theta}(-\log u)^{-1}\mathbf{1}\left[\theta \geqslant \frac{u}{k}\right]\mathbf{1}\left[\theta \leqslant \frac{u^{\left(1-\frac{C}{k}\right)}}{k}\right]\frac{kQ(\theta k)}{C}d\theta$$

□

Remark 12. *The result goes through if we replace* $[0,1]$ *by any closed interval* $\mathcal{X} \subseteq [0,1]$ *and/or* $Q(0) \geqslant 0$ *by* $Q(1) \leqslant 1$ *in Assumption 10. In this case, the approximation is with respect to the norm of* $L^2(\mathcal{X})$.

3.12.3 Auxiliary Definitions and Lemmas

We follow the idea from H. D. Nguyen and McLachlan 2019 and T. T. Nguyen et al. 2020, to construct convolution-based approximations. Because we are working with mixture of quantiles instead of mixture of densities we have to adapt the definitions and proofs.

Approximate Identity Definition

Let $\{\mathcal{M}_k\}_{k\in\mathbb{N}}$ be a sequence of sets, where for all k, $\mathcal{M}_k \subset \mathbb{R}_+$ and let $h_k : \mathcal{M}_k \times \mathbb{R}_+ \to \mathbb{R}_+$, be a family of functions indexed by k such that, there exists $\eta \in \mathbb{R}_+$ with $h_k(x, \eta) = x$, for all $x \in \mathbb{R}_+$. We define the modified convolution operator $*$, such that

$$[f * g]_k(x) = \int_{\mathcal{M}_k} f(m)g\left(h_k(m, x)\right)dm.$$

When there is no ambiguity We will drop the subscript k of the left-hand side.

Definition 5. *Let* $k \in \mathbb{N}$, *and* k^* *be a limit point of* $\mathbb{N}$. *A family of functions* $\{\alpha_k\}_{k\in\mathbb{N}}$ *defined on* $\mathbb{R}_+$ *is called a* C*-approximate identity if*

1. $\int_{\mathcal{M}_K} |\alpha_k(m)dm| \leqslant \tilde{C}$, *for all* $k \in \mathbb{R}_+$
2. $\int_{\mathcal{M}_k} \alpha_k(m)dm = C$, *for all* $k \in \mathbb{R}_+$, *where* $C \neq 0$ *is a constant.*

3. $\int_{\mathcal{M}_k / B_\eta(\delta)} |\alpha_k(m)| dm \underset{k \to k^*}{\longrightarrow} 0$ *for every* $\delta > 0$.

4. $\eta \in \mathcal{M}_k$, *for all* $k \in \mathbb{R}_+$

where $B_\eta(\delta)$ *denotes the ball centered at* η *with radius* δ.
Next, we provide a useful lemma for constructing approximate identities.

Lemma 7. *Let* $g : \mathbb{R}_+ \times \mathbb{N} \to \mathbb{R}_+$, *and* $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$, *if:*

1. $\frac{dg(m,k)}{dm}$ *exists and is positive almost everywhere,*

2. *There exists a sequence* $\{m_k\}_{k \in \mathbb{N}}$, *such that, for every* k, $\mathcal{M}_k = [\eta, m_k]$, $m_{k+1} \leqslant m_k$, $\lim_{k \to \infty} m_k = \eta$,

$$\int_\eta^{m_k} \frac{dg(m,k)}{dm} \alpha(g(m,k)) dm = C$$

for some $C \neq 0$ *and, for every* $\delta > 0$,

$$\lim_{k \to \infty} \int_{\eta + \delta}^{\infty} \frac{dg(m,k)}{dm} \alpha(g(m,k)) \mathbf{1}\left(n + \delta < m_k\right) dm = 0$$

then

$$\alpha_k(x) := \frac{dg(m,k)}{dm} \alpha(g(m,k))$$

is a C*—approximate identity with* $k^* = \infty$.

Proof. The proof is direct by verifying that the definition holds.

□

Generic Approximation

Lemma 8. *Let* α_k *be a* C*-approximate identity for some* $k^* \in [0, \infty]$. *If* $\mathcal{X} \subseteq \mathbb{R}_+$, $f \in \mathcal{L}_2(\mathcal{X})$, *and, for all* $k \in \mathbb{N}$, $f(h_k(m,x))$ *is continuous in* m *and uniformly continuous in a neighborhood of* η, *then* $\left\| \alpha_k * \frac{f}{C} - f \right\|_2 \to 0$ *as* $k \to k^*$.

Proof.

$$\begin{aligned}
&\left|\left[\alpha_k * \frac{f}{C}\right](x) - f(x)\right| = \left|\int_{\mathcal{M}_k} \alpha_k(m)\frac{f(h_k(m,x))}{C}dm - f(x)\right| \\
&= \frac{1}{|C|}\left|\int_{\mathcal{M}_k} \alpha_k(m)[f(h_k(m,x)) - f(x)]dm\right| \quad \text{(by 2.)} \\
&\leqslant \frac{1}{|C|}\int_{\mathcal{M}_k} \left|\alpha_k(m)[f(h_k(m,x)) - f(x)]\right| dm \\
&= \frac{1}{|C|}\int_{\mathcal{M}_k} |\alpha_k(m)|^{\frac{1}{2}}|[f(h_k(m,x)) - f(x)]||\alpha_k(m)|^{\frac{1}{2}} dm \\
&\leqslant \frac{1}{|C|}\left(\int_{\mathcal{M}_k} |\alpha_k(m)|[f(h_k(m,x)) - f(x)]^2 dm\right)^{\frac{1}{2}}\left(\int_{\mathcal{M}_k} |\alpha_k(m)|dm\right)^{\frac{1}{2}} \quad \text{(by C-S)} \\
&\leqslant \frac{1}{\sqrt{|C|}}\left(\int_{\mathcal{M}_k} |\alpha_k(m)|[f(h_k(m,x)) - f(x)]^2 dm\right)^{\frac{1}{2}} \quad \text{(by 1.).}
\end{aligned}$$

We can bound the L_2 norm:

$$\begin{aligned}
&\int_{\mathcal{X}}\left|\left[\alpha_k * \frac{f}{K}\right](x) - f(x)\right|^2 dx \leqslant \frac{1}{\sqrt{|C|}}\int_{\mathcal{X}}\int_{\mathcal{M}_k} |\alpha_k(m)|[f(h_k(m,x)) - f(x)]^2 dm dx \\
&= \frac{1}{\sqrt{|C|}}\int_{\mathcal{M}_k} |\alpha_k(m)|\left(\int_{\mathcal{X}} [f(h_k(m,x)) - f(x)]^2 dx\right) dm \quad \text{(Fubinis).}
\end{aligned}$$

Note that we we have that as $m \to \eta$,

$$\lim_{m\to\eta}\left(\int_{\mathcal{X}} [f(h_k(m,x)) - f(x)]^2 dx\right) = 0 \quad \text{(By continuity of } f(h_k(m,x)) \text{ with respect to } m\text{).}$$

Define $g_k(m) \equiv \left(\int_{\mathcal{X}} [f(h_k(m,x)) - f(x)]^2 dx\right)$, then $g(\eta) = 0$. Moreover,

$$\begin{aligned}
&\int_{\mathcal{M}_k} |\alpha_k(m)|g_k(m)dm \leqslant \int_{\mathcal{M}_k} |\alpha_k(m)||g_k(m)|dm \\
&= \int_{B_\eta(\delta)} |\alpha_k(m)||g_k(m)|dm + \int_{\mathcal{M}_k/B_\eta(\delta)} |\alpha_k(m)||g_k(m)|dm \\
&\leqslant \sup_{B_\eta(\delta)} |g_k(m)|\int_{B_\eta(\delta)} |\alpha_k(m)|dm + \sup_{\mathcal{M}_k/B_{\eta(\delta)}} |g_k(m)|\int_{\mathcal{M}_k/B_\eta(\delta)} |\alpha_k(m)|dm
\end{aligned}$$

Since we can bound $|g_k(m)| \leqslant M$, taking the limit we get

$$\lim_{k\to k^*}\int_{\mathcal{M}_k} |\alpha_k(m)|g_k(m)dm \leqslant C\lim_{k\to k^*}\sup_{B_\eta(\delta)} |g_k(m)| \quad \text{(By 1. and 2.)}$$

Since $g(\eta) = 0$, $f(h_k(m,x))$ is continuous in m we can pick delta such that the right-hand side is

close enough to 0. Then we conclude that

$$\lim_{k \to k^*} \int_{\mathcal{X}} \left| \left[\alpha_k * \frac{f}{C} \right](x) - f(x) \right|^2 dx = 0$$

□

3.13 Empirical Application

3.13.1 Data Source

Our data source is the publicly available version of the employer-employee matched data RAIS from 2017 to 2021. A standardized version is available to download here.

3.13.2 Data Cleaning

The construction of the empirical quantile function of the average monthly wages for private sector workers in each municipality $\times$ year proceeds as follows:

1. We exclude entries with 0 average monthly wage (column valor_remuneracao_media equal 0).
2. We exclude entries where the employee is hired as a public servant (column tipo_vinculo equal 2, 30, 31 or 35).
3. We exclude entries where the employer sector is either public administration, defense, social security, or international organizations. (first two digits of column cnae_2 equal 84 or 99).