

Essays in Economics

by

Hiroshi Toma

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in the University of Michigan
2025

Doctoral Committee:

Professor Javier Cravino, Co-Chair
Professor Andrei Levchenko, Co-Chair
Professor Kathryn Dominguez
Professor John Leahy
Professor Stephen Terry

Hiroshi Toma

htoma@umich.edu

ORCID iD: 0000-0002-2499-3006

© Hiroshi Toma 2025

DEDICATION

This is dedicated to everyone who stood by me on this long journey.

TABLE OF CONTENTS

DEDICATION	ii
LIST OF FIGURES	v
LIST OF TABLES	vii
LIST OF APPENDICES	ix
ABSTRACT	x

CHAPTER

1 The Financial Propagation Mechanism of Commodity Booms	1
1.1 Introduction	1
1.2 Setting: Global Mining Price Boom and the Peruvian Economy	4
1.3 Data	5
1.3.1 Exposure of Banks and Non-Mining Firms to the Mining Sector	7
1.4 Empirical Evidence	8
1.4.1 Facts	9
1.4.2 Additional Bank-Level Results	12
1.4.3 Additional loan-level results	13
1.4.4 Additional Firm-Level Results	14
1.5 Model	15
1.5.1 Household	16
1.5.2 Non-Tradable Intermediate Good Firms	17
1.5.3 Commodity Exporter	18
1.5.4 Non-Commodity Exportable Good Firm	18
1.5.5 Loan Aggregator	19
1.5.6 Banks	19
1.5.7 Market Clearing Conditions, Real Exchange Rate, Balance of Payments, and Real GDP	21
1.5.8 Equilibrium	22
1.5.9 Calibration	22
1.6 Quantifying the Aggregate Importance of the Financial Propagation Mechanism	24
1.6.1 Main Simulation	25
1.6.2 Simulation with Only a Commodity Price Shock	27

1.7	Conclusions	27
2	Aggregate Implications of Heterogeneous Inflation Expectations: The Role of Individual Experience	29
2.1	Introduction	29
2.2	Literature review	32
2.3	Empirical facts	35
2.4	A simple model with heterogeneous expectations	39
2.4.1	Standard Kalman filter	41
2.4.2	History-based Kalman filter	45
2.5	Aggregate implications of heterogeneous expectations	52
2.5.1	Households	52
2.5.2	Firms	55
2.5.3	Monetary policy	55
2.5.4	Calibration	56
2.5.5	Simulations	56
2.5.6	Implied heterogeneous responses	58
2.6	Optimal Taylor rules	60
2.7	Analyzing an episode of high inflation	62
2.8	Conclusions	63
3	Cultural Proximity and Inter-firm Trade	65
3.1	Introduction	65
3.2	Background, data and construction of variables	69
3.2.1	Caste and Religion in India	69
3.2.2	Data	70
3.2.3	Construction of variables	72
3.3	Stylized facts	73
3.3.1	Mechanisms	76
3.3.2	Discussion of stylized facts	88
3.3.3	Robustness	90
3.4	Quantitative Importance of Cultural Proximity	91
3.4.1	Model	92
3.4.2	Estimation and calibration	94
3.4.3	Counterfactual analysis	96
3.5	Conclusions	99
	APPENDICES	101
	BIBLIOGRAPHY	187

LIST OF FIGURES

FIGURE

1.1	Commodity price boom	5
1.2	Firm deposits and firm loans	6
1.3	Total mining weights in banks	8
2.1	Average 12-months-ahead inflation expectations	35
2.2	Lifetime average inflation rate among respondents	37
2.3	Inflation point forecast and average lifetime inflation	38
2.4	History-based Kalman-filter-based inflation forecasts by cohort	44
2.5	Observed inflation forecasts and history-based Kalman filter forecasts	52
2.6	Impulse response functions	57
2.7	Impulse response functions, optimal Taylor rule	61
2.8	Impulse response functions, forecast	63
3.1	Probability-weighted sales decomposition of largest cultural groups	68
3.2	Probability-weighted sales and purchases across cultural groups	71
3.3	Effect of cultural proximity on trade, intensive margin	74
3.4	Effect of cultural proximity on prices	76
A.1	Mining exports by mineral	101
A.2	Mining prices	102
A.3	Mining output	102
A.4	Financial propagation mechanism under alt. parameters: θ	119
A.5	Financial propagation mechanism under alt. parameters: ϕ	120
A.6	Financial propagation mechanism under alt. parameters: ε	121
A.7	Financial propagation mechanism under alt. parameters: Estimated coefficients	122
A.8	Financial propagation mechanism under alt. parameters: α	123
B.1	History-based Kalman-filter-based inflation forecasts by cohort, full sample	124
B.2	Parameter estimation, stability by cohorts	125
B.3	Impulse response functions, history-based expectations by cohort, taste shock	126
B.4	Impulse response functions, history-based expectations by cohort, cost shock	127
B.5	Impulse response functions, consumption and labor supply by cohort, history-based KF with heterogeneity, taste shock	128
B.6	Impulse response functions, consumption and labor supply by cohort, history-based KF with heterogeneity, cost shock	129
B.7	Impulse response functions, inflation rate history-based expectations by cohort, forecast	130

B.8	Impulse response functions, shocks, forecast	131
B.9	Constant Kalman gain across cohorts	138
B.10	Observed inflation forecasts and history-based Kalman filter forecasts, AR(1)	140
B.11	Observed inflation forecasts and history-based Kalman filter forecasts, alternative signals	142
B.12	Inflation rate, Europe	144
B.13	Lifetime average inflation rate among respondents, Europe	145
B.14	Inflation point forecast and average lifetime inflation rate, Europe	146
B.15	History-based Kalman-filter-based inflation forecasts by cohort, Europe	149
B.16	Observed inflation forecasts and history-based Kalman filter forecasts, Europe . . .	150
B.17	Impulse response functions, comparison with alternative history-based Kalman fil- ter operator	155
C.1	Distribution of total sales per firm	156
C.2	Histogram of estimated concentration parameters for Dirichlet distribution	157

LIST OF TABLES

TABLE

1.1	Bank-level results	10
1.2	Loan-level results	11
1.3	Firm-level results	12
1.4	Parameters	23
1.5	Main simulation	26
1.6	Commodity price shock-only simulations	28
2.1	Effects of current and experienced inflation rates on inflation expectations	40
2.2	Parameter estimation	49
2.3	Goodness of fit	51
3.1	Effect of cultural proximity on trade, intensive and extensive margins	75
3.2	Effect of cultural proximity on prices	77
3.3	Effect of cultural proximity on trade by types of good, intensive margin	78
3.4	Effect of cultural proximity on trade interacted by court quality, intensive margin	81
3.5	Effect of cultural proximity on trade by number of varieties, intensive margin	82
3.6	Cancellations	83
3.7	Effect of cultural proximity on trade by vertical hierarchies, intensive and extensive margins	84
3.8	Effect of cultural proximity after controlling for establishment age of sellers, intensive margin	86
3.9	Effect of cultural proximity on trade interacted with market concentration, intensive margin	87
3.10	Effect of cultural proximity on trade, augmented with name distance, intensive margin	89
3.11	Effect of cultural proximity on aggregate outcomes (counterfactual scenarios)	98
3.12	Change in sales by productivity quartiles	99
A.1	Mining production ranking of Peru, 2011	103
A.2	Bank statistics	104
A.3	Loan statistics	104
A.4	Firm statistics	104
A.5	Ratios and other weights	113
A.6	Main simulation, complete results	117
A.7	Commodity price shock-only simulations, complete results	118
B.1	Model calibration	132
B.2	Optimal Taylor rule parameters, cost shock	132

B.3	Optimal Taylor rule parameters, taste shock	132
B.4	AR(1) regression for monthly inflation	135
B.5	Parameter estimation, AR(1)	140
B.6	Main parameter estimation, alternative signals	141
B.7	Effects of current and experienced inflation rates on inflation expectations, Europe . .	147
B.8	History-based Kalman filter parameters, Europe	148
B.9	Parameter estimation, Europe	148
C.1	Data summary	158
C.2	Estimation for matching cost	159
C.3	Correction for selection bias, first stage	161
C.4	Correction for selection bias, second stage	162
C.5	Effect of cultural proximity on trade, intensive and extensive margins, Kullback-Leibler	163
C.6	Effect of cultural proximity on prices, Kullback-Leibler	164
C.7	Effect of cultural proximity and language on trade, intensive margin	166
C.8	Effect of cultural proximity on trade by good specialization, intensive margin	168
C.9	Effect of cultural proximity after controlling for industries, intensive margin	169
C.10	Targeted and untargeted moments	186

LIST OF APPENDICES

A Appendix to Chapter 1 101

B Appendix to Chapter 2 124

C Appendix to Chapter 3 156

ABSTRACT

This dissertation contains three chapters related to the study of diverse topics in macroeconomics and trade. Chapter 1 focuses on how commodity shocks propagate through the banking sector in small open economies; Chapter 2 studies how inflation expectations depend on past inflation experiences and how this affects the aggregate economy; and Chapter 3 analyzes the effect of cultural proximity on trade at the firm-to-firm level.

In Chapter 1, titled “The Financial Transmission Channel of Commodity Booms”, we examine the financial propagation mechanism in small open economies (SOEs) that links commodity price booms to the non-commodity sector via banks. We propose a mechanism in which a commodity price boom leads to an increase in deposits from commodity exporters into domestic banks, enabling banks to expand their loan supply to non-commodity firms. Then, these non-commodity firms increase their output. Using detailed bank-firm-loan microdata from Peru—an SOE that experienced a commodity price boom in the 2000s—we provide empirical evidence supporting this mechanism.

In Chapter 2, titled “Aggregate Implications of Heterogeneous Inflation Expectations: The Role of Individual Experience”, we show that inflation expectations are heterogeneous and depend on past individual experiences. We propose a methodology to represent the inflation expectations-formation process, where heterogeneity comes from an anchoring-to-the-past mechanism. In a New-Keynesian framework, we find heterogeneous expectations increase shock persistence.

In Chapter 3, titled “Cultural Proximity and Inter-Firm Trade”, we examine how cultural proximity shapes production networks. We combine a firm-to-firm trade dataset for an Indian state with information on cultural proximity. We find a larger cultural proximity reduces prices and fosters trade.

CHAPTER 1

The Financial Propagation Mechanism of Commodity Booms

Abstract

We examine the financial propagation mechanism in small open economies (SOEs) that links commodity price booms to the non-commodity sector via banks. We propose a mechanism in which a commodity price boom leads to an increase in deposits from commodity exporters into domestic banks, enabling banks to expand their loan supply to non-commodity firms. Then, these non-commodity firms increase their output. Using detailed bank-firm-loan microdata from Peru—an SOE that experienced a commodity price boom in the 2000s—we provide empirical evidence supporting this mechanism. We then incorporate this mechanism into an SOE model with banks to quantify its aggregate importance. After calibrating the model to the Peruvian data, our simulations suggest that the mechanism explains one tenth of the observed 65% Peruvian GDP growth in the 2003–2011 commodity price boom episode.¹

1.1 Introduction

A well-documented fact about small open economies (SOEs) is that commodity prices, or more generally terms of trade, are important sources of business cycle fluctuations [Mendoza, 1995, Kose, 2002, García-Cicco and Kawamura, 2015, Schmitt-Grohé and Uribe, 2018, Fernández et al., 2023]. The literature has identified several mechanisms by which commodity shocks propagate to the broader economy: wealth effects [Salter, 1959, Swan, 1963, Corden and Neary, 1982], sovereign interest rate premiums [Shousha, 2016, Drechsel and Tenreyro, 2018], and labor reallocation [Benguria et al., 2023], among others. At the same time, many emerging market SOEs exhibit significant financial constraints, and the domestic banking system serves as a significant amplifier of shocks in these countries [Paravisini, 2008, Schnabl, 2012, Agarwal et al., 2020, Bustos et al., 2020, Morelli et al., 2022].

¹This chapter is co-authored with Walter Cuba.

In this paper, we explore how a commodity price boom interacts with the bank loan supply in SOEs. More specifically, we propose a novel propagation mechanism that goes from commodity producers to non-commodity producers via the domestic banking system. We show that after a commodity boom, commodity exporters increase their deposit holdings in domestic banks, prompting banks to expand their loan supply to non-commodity firms. Then, non-commodity firms increase their output. Our paper complements previous studies of commodity booms in SOEs by highlighting the role of domestic banks in the local transmission of shocks from the commodity sector to the non-commodity sector.

We document three facts using detailed matched bank-firm-loan microdata from Peru, an SOE that experienced a mining commodity price boom in the 2000s. First, banks are heterogeneously exposed to the mining commodity sector through their different mining client portfolios. When mining commodity prices rise, banks with greater exposure to the mining sector receive more firm deposits. After an annual 100 percent increase in commodity prices, the average exposed bank registers an increase of 61 percent in firm deposits compared to a non-exposed bank.

Second, exploiting the heterogeneity in bank-borrower relationships, we find that more mining-exposed banks supply more loans to non-mining firms after mining commodity prices increase. After a 100 percent commodity price hike, the average exposed bank increases its loan supply by 16 percent when compared to a non-exposed bank.

Third, we document that non-commodity firms more exposed to mining commodity prices through their banks increase their output after a commodity price surge. When commodity prices double, the average exposed non-commodity firm increases its total sales by 2 percent compared to a non-exposed non-commodity firm.

To quantify the aggregate importance of this mechanism, we embed it into a static SOE model with banks and an endowment commodity sector. Consistent with our empirical findings, we assume commodity exporters hold deposits in domestic banks. Furthermore, as the data suggest, banks in our model are heterogeneous in their exposure to the commodity sector, such that each bank receives different amounts of deposits from the commodity sector after a commodity price shock.

The model contains two key financial frictions. First, non-commodity firms face a working capital constraint, which implies that they must pay for part of their expenses before production takes place. This friction creates the need for these firms to take loans from banks. Second, banks face a balance sheet constraint, so they must obtain foreign wholesale funding to cover the loans they supply when deposits and equity are not enough. Importantly, they must pay a premium on their foreign wholesale funding that is decreasing in the amount of deposits they hold. This friction links the amount of deposits held to the interest rate banks charge for their loans.

The model's tractability allows us to derive structural equations that correspond with the re-

gressions estimated in the empirical analysis. In this way, we give a structural interpretation to the estimated coefficients and recover the values for all the bank-related parameters in the model. We calibrate the remaining parameters following the SOE literature and to match the change in Peruvian GDP in the 2003–2011 commodity price boom. After simulating the non-linear model using hat algebra, we successfully match the targeted moment—GDP growth—and achieve a good fit to untargted moments.

We then conduct a counterfactual exercise in which we turn off the financial propagation mechanism to quantify its aggregate importance. More specifically, we turn off the pass-through from commodity prices to firm deposits. Comparing GDP growth in the baseline and counterfactual scenarios reveals that the financial propagation mechanism accounts for one tenth of the observed 65 percent GDP growth in Peru during the 2003–2011 commodity price boom period.

Literature review. Our paper relates to (i) the wider literature that studies the aggregate effects of terms of trade and commodity price fluctuations in SOEs [Salter, 1959, Swan, 1963, Corden and Neary, 1982, Mendoza, 1995, Kose, 2002, García-Cicco and Kawamura, 2015, Schmitt-Grohé and Uribe, 2018, Fernández et al., 2023] and (ii) the literature that studies how financial frictions interact with commodity shocks [Shousha, 2016, Drechsel and Tenreyro, 2018, Agarwal et al., 2020]. We contribute by providing evidence of a new financial propagation mechanism of commodity shocks in SOEs that works through the banking system.

We also connect with the recent literature that studies the transmission channels of a commodity price shock using microdata [Allcott and Keniston, 2018, Benguria et al., 2023, Silva et al., 2024]. Here, our contribution lies in tracing and quantifying the importance of the financial propagation mechanism using detailed bank-firm-loan matched microdata.

Our empirical strategy is close to that of Peek and Rosengren [2005], Paravisini [2008], Schnabl [2012], Iyer et al. [2014], Gilje et al. [2016], Agarwal et al. [2020], Bustos et al. [2020], Federico et al. [2023], who use detailed bank and loan microdata to study how banks propagate shocks across firms, sectors, and geographies. We take a step further and embed our proposed financial mechanism into a general equilibrium model disciplined by microdata. Thus, we not only provide empirical evidence that supports the financial propagation mechanism but also quantify its aggregate importance.

On the theoretical side, we build on the standard textbook SOE model of Uribe and Schmitt-Grohé [2017]. Our contribution is to extend it with financial frictions and heterogeneous banks. Firms need to take out loans because they are subject to a working capital constraint, following Neumeyer and Perri [2005], Jermann and Quadrini [2012]. Banks are subject to a balance sheet constraint [Jayaratne and Morgan, 2000, Kashyap and Stein, 2000, Gilje et al., 2016, Wang et al., 2022, Whited et al., 2022], as they have to obtain foreign wholesale funding to cover their loans [Shousha, 2016].

Section 1.2 provides details on the global mining price boom and the Peruvian economy. Section 1.3 introduces the data and exposure measures. Section 1.4 presents the empirical evidence. In Section 1.5 we introduce the theoretical framework. Section 1.6 quantifies the aggregate importance of the mechanism. Finally, Section 1.7 concludes.

1.2 Setting: Global Mining Price Boom and the Peruvian Economy

From the early 2000s to the mid-2010s, the world experienced a mining commodity price boom, mainly driven by a demand expansion from China [Fernández et al., 2023]. This boom represented a substantial shock for those SOEs where the mining commodity sector plays an important role. In this paper, we focus on Peru, one SOE that benefited greatly from the boom. During the boom period, the mining sector represented 70 percent of exported goods and 10 percent of real GDP.

Peru is an ideal setting for examining the financial propagation mechanism of commodity booms in SOE. First, Peru exports a diverse basket of mining commodities: copper, gold, iron, lead, silver, tin, zinc, and aluminum, among others.² Thus, we observe variation across different minerals. Second, Peru holds a modest global market share for these commodities, so it is reasonable to assume it is a price taker in these markets³ Third, despite its aggregate importance, the mining sector is weakly linked to the rest of the economy. For instance, in 2007, it only employed 1 percent of the total labor force, while it only demanded 2 percent of the domestically supplied goods and services. This facilitates the identification of the financial propagation mechanism, as increased labor income or intermediate demand are unlikely to be confounding factors.

In Figure 1.1 we report a weighted relative mining price index across all minerals that Peru exports. Mining commodity prices increased sharply during the study period. For instance, at its peak in 2011, the prices more than doubled relative to its value at the start of the boom. Figure 1.1 also shows that mining exports increased over this period, closely following the price index.⁴

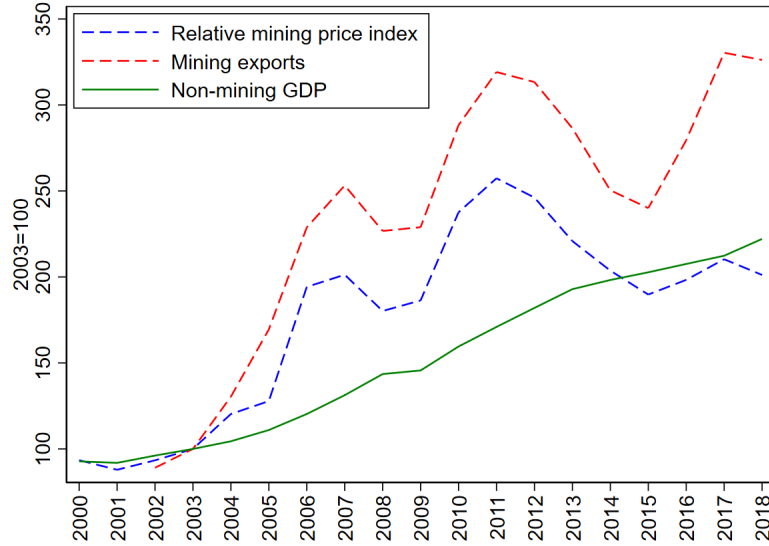
Related to this boom, Figure 1.1 additionally shows the evolution of non-mining GDP. This variable starts growing around the beginning of the boom and keeps increasing throughout the boom. While non-mining GDP is not directly linked to the mining sector, we will argue later that the two are connected through the banking system.

²Appendix Figure A.1 shows the share of each mining commodity in total mining exports.

³We present the world output shares for Peru in 2011, the peak year of the mining commodity price boom, in Appendix Table A.1.

⁴This boom was mostly price driven rather than quantity driven. We show individual mining price and quantity indices in Appendix Figures A.2 and A.3, respectively. Quantities remained relatively flat for most minerals, the only exceptions being copper and iron, which benefited late in the boom from large mining projects entering production.

Figure 1.1: Commodity price boom



Note: For the weighted relative mining price index we first normalize the dollar prices for the commodity prices and then divide them by the Peruvian imports price index. The weights we use are the average shares of each mining commodity across the sample in total mining commodity exports. For mining exports we first normalize the dollar value and then divide it by the Peruvian imports price index. The non-mining GDP is expressed in real domestic currency units and normalized.

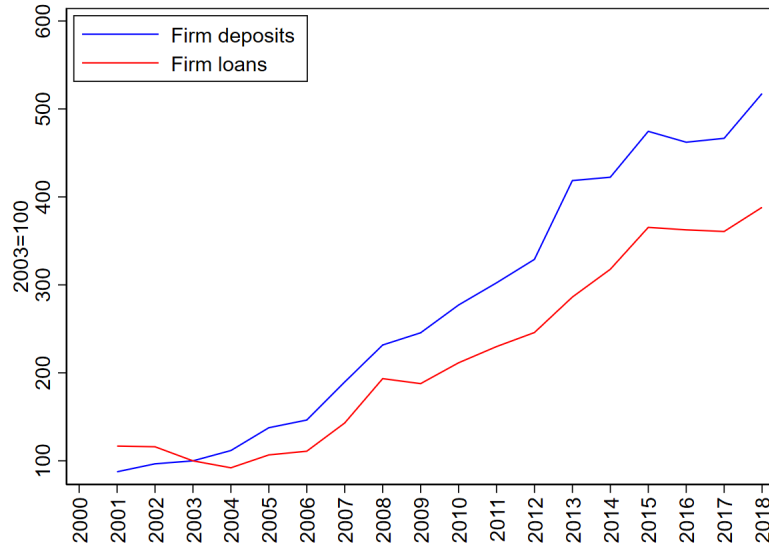
Figure 1.2 presents the evolution of relevant variables in the banking sector, namely firm deposits and firm loans. The former are the deposits coming from private sector firms to domestic banks, while the latter are the loans going from the domestic banks to the private sector firms. Both series register an increase in the early years of the boom and keep growing during the boom episode. In the next sections, we will show that part of this growth is due to banks being connected to the booming mining sector.

1.3 Data

For our empirical analysis, we combine several data sources that result in a matched bank-firm-loan dataset for the Peruvian economy from 2006 to 2018 with annual frequency. First, we use firm-level customs data extracted from ADEX Data Trade. From this dataset, we identify the mining commodity exporters and their exports.

Second, we retrieve end-of-year bank balance sheet data from the Peruvian banking regulator (SBS). We focus on firm deposits—the total amount of deposits from private sector firms held in the banks under consideration. We have information for 9 domestic banks, whose firm deposits represented 8 percent of Peruvian GDP on average over the analyzed period. We show the basic

Figure 1.2: Firm deposits and firm loans



Note: The series are expressed in real domestic currency units and normalized.

statistics in Appendix Table A.2.⁵

Third, we use the Peruvian firm survey (EEA) to recover firm-level balance sheet information. We have information on approximately 4,800 non-exporting non-mining firms that belong to the following sectors: construction, education, hotels, manufactures, restaurants, retail, services, transport and communications, and travel agencies.⁶ These firms are mostly medium to large when measured by their total sales. On average, between 2006 and 2018, their aggregate sales represented 19 percent of the GDP. We present statistics on the firms in Appendix Table A.4.

Fourth, we exploit the Peruvian credit registry. This rich dataset provides end-of-year annual loan-level information on the lending institution, the borrowing non-mining firm, and the outstanding loan amount. We focus on a subsample that considers all the loans received by the non-mining firms that appear in the firm survey. The total loans in this subsample represented 102 percent of GDP on average within the analyzed period. We give more detail about the loan-level statistics in Appendix Table A.3.

⁵The total Peruvian financial system is composed of banks and other non-banking institutions (*cajas municipales*, *cajas rurales*, and *financieras*). The 9 banks in our sample represented 87 percent of total assets, 87 percent of total firm loans, and 88 percent of total firm deposits on average between 2006 and 2018. We treat merged banks as single entities and combine their accounts. We exclude banks that exclusively serve foreign-owned firms or specialize in credit card and consumer loans, as they fall outside the scope of our analysis. Lastly, we do not consider the *cajas municipales*, *cajas rurales*, and *financieras*, as they were not permitted to hold firm deposits during the period of analysis because of regulatory restrictions.

⁶Although information on non-mining commodity-producing firms is available in the firm survey, we exclude these firms from the sample because their prices may be correlated with those of the mining sector. The excluded sectors include agriculture, electricity, fishing, and oil and gas.

Last, we retrieve international commodity prices from the IMF commodity price database. We have information on aluminum, copper, gold, iron, lead, silver, tin, and zinc prices. Metals that are not found in the commodity price database are assigned the all-metals average price index. Per the usual SOE assumption, we consider these prices to be exogenous to the Peruvian economy.

All commodity prices are divided by the Peruvian import price index and normalized to 2007=100. All other variables are converted to Peruvian soles (PEN) using the average nominal exchange rate for the period and deflated by the Peruvian GDP deflator.

1.3.1 Exposure of Banks and Non-Mining Firms to the Mining Sector

The key variable in the empirical analysis is the exposure of banks and firms to mining commodity prices. First, we identify the mining commodity exporters in the customs data, each indexed by m . We classify a firm as a mining commodity exporter if its average mining exports are greater than or equal to 50 percent of its total exports.

Next, we define ω_{bm}^{CO} as the weight of mining firm m in the total loan portfolio of bank b , with non-mining clients having 0 weight. Because our focus is on the response of deposits to a commodity boom, we would ideally use data on deposits to measure the strength of the relationship between a bank and a mining firm. Unfortunately, deposit data broken down by bank and firm do not exist. Instead, under the assumption that the bank-firm relationship influences both deposits and loans (i.e., mining firms hold deposits in the same banks from which they borrow), we use loans from a bank to a mining firm as a proxy for the strength of their deposit-based relationship.

Moreover, the weights we use are fixed to their pre-sample values to address the concern that either the banks could look for mining clients after the boom or the mining companies could look for loans after the boom.⁷ We report the sum of all mining weights across banks in Figure 1.3, in which we see that the banks under consideration are heterogeneously exposed to the mining sector.

We now combine the heterogeneous exposure of each bank b to the mining sector with the time variation of mining commodity prices. We define the weighted mining commodity price exposure index E_{bt} as

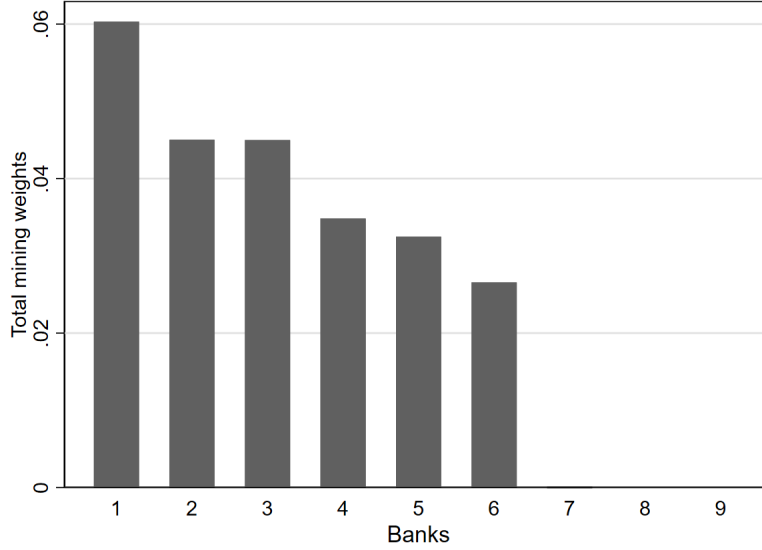
$$E_{bt} = \sum_m \omega_{bm}^{CO} \ln(p_{mt}^{CO}), \quad (1.1)$$

where ω_{bm}^{CO} is the weight of mining firm m in the total loan portfolio of bank b and p_{mt}^{CO} is the relevant mineral price for mining firm m (e.g., copper price for a copper exporter, gold price for a gold exporter).⁸ In line with the SOE assumption, we treat commodity prices as exogenous to

⁷To construct the constant weights, we take the 2004 end-of-year loan portfolio weights for each mining exporting company operating in that period.

⁸For commodity exporters that export more than one mining commodity, we take the price of the most important mineral in their export basket.

Figure 1.3: Total mining weights in banks



Note: Figure shows the total weight of mining commodity firms in the total loan portfolio of the 9 banks under consideration. We use fixed 2004 weights.

Peruvian banks and firms. Since weights ω_{bm}^{CO} are fixed, time variations in E_{bt} arise solely from changes in commodity prices, making the index exogenous to the Peruvian economy.

Likewise, we construct a measure of how non-mining firms are exposed to mining commodity prices through their associated banks. For non-mining firm i , we define the weighted mining commodity price exposure index e_{it} as

$$e_{it} = \sum_b s_{ib} E_{bt}, \quad (1.2)$$

where s_{ib} is the fixed pre-sample weight of bank b in the total loan portfolio of non-mining firm i . Banks with no exposure have a weight of 0.⁹

1.4 Empirical Evidence

The goal of this section is to empirically demonstrate the existence of the financial propagation mechanism that connects the commodity price boom to the rest of the economy. We use detailed microdata from Peru to establish three different empirical facts after a mining commodity price increase: (i) banks that were related to mining firms received more firm deposits; (ii) banks that were related to mining firms increased their loan supply to non-mining firms; and (iii) non-commodity

⁹The weights are fixed to their 2004 values.

firms related to exposed banks increased their output. Together, these facts provide a detailed step-by-step account of the financial propagation mechanism of the commodity price boom to the non-commodity real economy of Peru.

1.4.1 Facts

Fact 1: More exposed banks receive more firm deposits. We first establish that banks exposed to mining exporters experienced an increase in their deposits. We rely on the exogeneity assumption for mining commodity prices in an SOE and exploit the heterogeneous exposure to such prices across banks to estimate

$$\ln(D_{bt}) = \alpha_b + \alpha_t + \beta E_{bt} + u_{bt}, \quad (1.3)$$

where D_{bt} are the firm deposits, E_{bt} is the bank-level weighted mining commodity price exposure index from Equation 1.1, α_b is a bank fixed effect, and α_t is a period fixed effect.¹⁰

For any given bank b in our dataset it holds that the sum of total weights of mining firms $\sum_m \omega_{bm}^{CO} \neq 1$, as no bank has a loan portfolio that consists only of mining firms. Thus, our regression is subject to an incomplete weights problem, in which we do not account for the effect that the non-mining clients have on banks. This may bias the results [Borusyak et al., 2022, 2024]. To control for this under our panel setting, we include an interaction of the total bank-level weights $\sum_m \omega_{bm}^{CO}$ with period fixed effects in our regressions.

For this regression, we use data at the bank-level. Column 1 of Table 1.1 shows there is a positive elasticity between the firm deposits of a bank and its exposure to mining commodity prices. Given an average ω_{bm}^{CO} of 0.04 across exposed banks (see Figure 1.3), a 100 percent increase in mining commodity prices would result in an increase of 61 percent in firm deposits within exposed banks ($E_{bt} > 0$) compared to non-exposed banks ($E_{bt} = 0$).¹¹

While we only have information on total firm deposits across all sectors and not mining-specific firm deposits, we argue that the positive effect we find occurs mostly because the mining sector increases its firm deposit holdings after mining commodity prices increase. The identification strategy we use relies on two elements that should isolate the effects coming from the mining sector. First, the mining exposure weights ω_{bm}^{CO} guarantee we only focus on banking relationships with mining commodity firms. Second, the commodity prices p^{CO} are relevant for mining commodity firms and not necessarily firms from other sectors.¹²

¹⁰Firm deposits refer to deposits that originate from firms. Therefore, this measure excludes other types of deposits, such as those that come from households.

¹¹The calculation we perform is $\beta \times \overline{\omega_{bm}^{CO}} \times [\ln(2)] \approx 61\%$, where $\overline{\omega_{bm}^{CO}} = 0.04$ is the average mining commodity weight across exposed banks.

¹²According to professionals in the Peruvian mining industry whom we interviewed, mining companies must hold deposits in domestic banks for several operational reasons: paying suppliers, workers, and taxes, as well as financing reinvestment and other precautionary motives.

Table 1.1: Bank-level results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. var.	$\ln(D_{bt})$	$\ln(D_{bt}^{dom})$	$\ln(D_{bt}^{usd})$	$\ln(D_{bt})$	$\ln(D_{bt}^{hh})$	$\ln(L_{bt})$	Pct. L_{bt}^{def}
E_{bt}	22.171*** (4.424)	22.489*** (4.983)	20.908*** (6.619)	21.663*** (4.517)	-1.797 (4.288)	5.323* (2.876)	1.013 (0.760)
E_{bt-1}				0.655 (6.394)			
FE	Bank, period	Bank, period	Bank, period	Bank, period	Bank, period	Bank, period	Bank, period
Obs.	114	114	114	114	114	113	109
Within R2	0.249	0.222	0.174	0.249	0.167	0.106	0.304

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered by bank in parentheses. Column 1 shows the result of estimating Equation 1.3. Column 2 changes the dependent variable to deposits in domestic currency. Column 3 changes the dependent variable to deposits in US dollars. Column 4 adds a lag of the independent variable to the estimation. Column 5 changes the dependent variable to household deposits. Column 6 changes the dependent variable to total firm loans. Column 7 changes the dependent variable to the percentage of defaulted firm loans over the total firm loans.

Fact 2: More exposed banks supply more loans to non-mining firms. We now analyze whether exposed banks supply more loans to their non-mining firm clients. For this, we follow the Khwaja and Mian [2008] identification strategy, which relies on non-mining firms being matched to multiple banks. More specifically, we estimate

$$\ln(L_{ibt}) = \alpha_{ib} + \alpha_{it} + \zeta E_{bt} + u_{ibt}, \quad (1.4)$$

where L_{ibt} are the outstanding loans from bank b to non-mining firm i , E_{bt} is the bank-level weighted mining commodity price exposure index from Equation 1.1, α_{ib} is a firm-bank fixed effect that controls for the special relationship a firm-bank pair could have, and α_{it} is a firm-period fixed effect that controls for any firm-period-specific loan demand shocks. We also include an interaction of the total bank-level weights with period fixed effects to control for the incomplete weights problem.

In this part, we use data at the bank-firm level. The combination of fixed effects results in a within-firm specification, in which the sample consists of only firms that have loans with more than one bank. Then, by assuming mining commodity prices are exogenous to the Peruvian economy and exploiting the within-firm and across-bank variation, the coefficient ζ identifies the loan supply shock coming from the banks. Moreover, we trim the dependent variable at 1 and 99 percent each period.

Column 1 of Table 1.2 presents the result. We find a positive elasticity between the loans a bank gives to its non-mining clients and the exposure of that bank to mining commodity prices. For a

Table 1.2: Loan-level results

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. var.	$\ln(L_{ibt})$	$\ln(L_{ibt})$	$\ln(L_{ibt})$	$\ln(L_{ibt})$	$\mathbb{I}_{ibt}^{entry}$	$\mathbb{I}_{ibt}^{exit}$
E_{bt}	5.697** (2.479)	10.950*** (2.822)	9.633*** (3.104)	5.531** (2.470)	-0.169 (0.314)	0.098 (0.562)
E_{bt-1}			-5.035* (3.020)			
$E_{bt} \times \mathbb{I}_{ibt}^{usd}$				1.306*** (0.156)		
FE	Firm $\times$ bank, firm $\times$ period	Firm $\times$ bank, sector $\times$ region $\times$ period	Firm $\times$ bank, firm $\times$ period	Firm $\times$ bank, firm $\times$ period	Firm $\times$ bank, firm $\times$ period	Firm $\times$ bank, firm $\times$ period
Obs.	73,121	102,667	73,121	73,121	141,682	57,593
Within R2	0.002	0.001	0.002	0.005	0.004	0.003

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by firm-bank in parentheses. Column 1 shows the result of estimating Equation 1.4. Column 2 changes the set of fixed effects. Column 3 adds a lag of the independent variable to the estimation. Column 4 adds a loan dollarization dummy variable to the estimation. Column 5 examines the entry-side extensive margin. Column 6 examines the exit-side extensive margin.

non-mining firm, a 100 percent increase in mining commodity prices would correspond with an increase of 16 percent in loans supplied by exposed banks ($E_{bt} > 0$) compared to non-exposed banks ($E_{bt} = 0$).¹³

Taken together, Facts 1 and 2 suggest that banks face financial liquidity frictions that are relieved by the deposit inflow after the commodity price shock. If there were no financial frictions (i.e., if banks had costless access to liquidity), then we would not see an increase in loan supply after the shock, as banks would have already exploited all profitable lending opportunities [Gilje et al., 2016, Gilje, 2019, Bustos et al., 2020].

Fact 3: More exposed non-mining firms produce more. We now ask what occurs to the production by non-mining firms related to exposed banks. First, we measure production using total sales in deflated monetary units. Second, we maintain the exogeneity assumption for mining commodity prices and leverage the heterogeneity in loan portfolios across non-mining firms to estimate

$$\ln(p_{it}y_{it}) = \alpha_i + \alpha_t + \kappa e_{it} + u_{it}, \quad (1.5)$$

where $p_{it}y_{it}$ represents the total sales of non-mining firm i , e_{it} is the firm-level weighted mining commodity price exposure index from Equation 1.2, α_i is a firm fixed effect, and α_t is a period

¹³The calculation we perform is $\zeta \times \overline{\omega_{bm}^{CO}} \times [\ln(2)] \approx 16\%$.

Table 1.3: Firm-level results

	(1)	(2)	(3)
Dep. var.	$\ln(p_{it}y_{it})$	$\ln(p_{it}y_{it})$	$\ln(w_t^N h_{it})$
e_{it}	0.940*** (0.317)	0.711*** (0.248)	1.037*** (0.301)
e_{it-1}		0.636** (0.254)	
FE	Firm, period	Firm, period	Firm, period
Obs.	29,570	29,570	29,421
Within R2	0.002	0.002	0.001

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by firm in parentheses. Column 1 shows the result of estimating Equation 1.5. Column 2 adds a lag of the independent variable to the estimation. Column 3 changes the dependent variable to the wage bill.

fixed effect. As in the previous cases, this regression is also subject to the missing weights problem, for which we add an interaction of the total weights with period fixed effects. For this regression, we use firm-level data. We also trim the dependent variable at 1 and 99 percent in each period.

Column 1 of Table 1.3 shows the result. There is a positive elasticity between the total sales of a non-mining firm and its exposure to the mining commodity prices through banks. A 100 percent increase in mineral prices would imply a 2 percent increase in the total sales of exposed firms ($e_{it} > 0$) with respect to non-exposed firms ($e_{it} = 0$).¹⁴

Fact 3 also suggests non-mining firms face financial frictions. As with the banks, if firms were not exposed to such frictions (i.e., if they had costless access to financing), they would have already taken all the borrowing opportunities and we would not see the increase in output after the shock.

1.4.2 Additional Bank-Level Results

Firm deposits by currency. In Peru, firms can hold deposits in either domestic currency or dollars. The average of firm deposits in domestic currency over total firm deposits is 48 percent for the 2006-2018 period. To assess whether deposit responses differ by currency, we estimate Equation 1.3, but differentiate deposits by domestic currency (D_{bt}^{dom}) or dollars (D_{bt}^{usd}). For consistency, dollar-denominated deposits are converted into domestic currency. Columns 2 and 3 of Table 1.1 show that different currencies do not imply significantly different responses with respect to mining commodity prices, as the estimated coefficients are similar between them and are also close to the

¹⁴The calculation we perform is $\kappa \times \overline{\omega_{bm}^{CO}} \times \overline{s_{ib}} \times [\ln(2)] \approx 2\%$, where $\overline{s_{ib}} = 0.97$ is the average weight exposed banks have in the firm loan portfolios.

main result of Column 1.

Lagged effects. Our main specification does not account for potential lagged effects of past shocks on current deposits. To address this, we estimate Equation 1.3 incorporating a lag of the weighted mining commodity price exposure index. Column 4 of Table 1.1 shows that the lagged value of the independent variable is not statistically significant, whereas the current value remains significant. This allows us to rule out the possibility of significant lagged effects of mining price exposure on firm deposits.

Household deposits. While our main focus is on firm deposits, we also analyze the response of household deposits to the changes in mining commodity prices. Specifically, we estimate Equation 1.3, but substitute the dependent variable to the deposits made by households D_{bt}^{hh} . Column 5 of Table 1.1 shows there is not a statistically significant effect of mining commodity prices on household deposits.

Total firm loans. We estimate Equation 1.3 using the total of firm loans L_{bt} given by each bank to private sector firms as the dependent variable. Column 6 of Table 1.1 shows that total firm loans also increase after a commodity price shock. This suggests the loan-level result we found for Fact 2 also holds at the aggregate bank-level.

Defaults. An expansion of the loan supply could make banks lend to riskier firms, who may be more likely to incur in defaults. However, Column 7 of Table 1.1 shows that the percentage of defaulted firm loans relative to total firm loans at the bank level does not exhibit a statistically significant response to a commodity price shock. Thus, within our sample, we find no evidence that the increased firm loan supply led to higher default rates.

1.4.3 Additional loan-level results

Alternative fixed effects. Our within-firm identification we use [Khwaja and Mian, 2008] relies on our sample consisting only of firms that have loans from more than one bank. This excludes firms with only one banking relationship. Following Degryse et al. [2019], we can use a different combination of fixed effects to prevent this issue. In particular, we estimate Equation 1.4 but replace α_{it} with α_{srt} , where s denotes two-digit ISIC sector and r denotes geographic region. The assumption is that all firms within the same region and sector face the same loan demand shock in any given period. We present the results in Column 2 of Table 1.2. The result of the main regression holds, albeit with two differences. First, the number of observations increases because now we include firms that have loans with only one bank. Second, the magnitude of the coefficient is larger. Then, we may interpret the main regression as a lower bound for the effect of mining commodity prices on loan supply.

Lagged effects. To assess potential lagged effects of past mining price exposure on current

loans, we estimate Equation 1.4 adding a lag of the independent variable. Column 3 of Table 1.2 shows that the lagged independent variable is not statistically significant, while the current variable remains significant. We interpret this as evidence that bank exposure to mining commodity prices affects loan supply contemporaneously, with no significant lagged effects.

Firm loans by currency. In Peru, loans may be given in domestic currency or in dollars. We define indicator variable $\mathbb{I}_{ibt}^{usd}$, which equals 1 if 50 percent or more of the loans received by firm i from bank b are in dollars. In our sample, 44 percent of the loan-level observations are dollarized. Column 4 of Table 1.2 incorporates the interaction $E_{bt} \times \mathbb{I}_{ibt}^{usd}$ into the analysis. We find that, when commodity prices increase, dollar-denominated firm loans show a slightly stronger response compared to domestic currency-denominated firm loans.

Extensive margin. The focus of our analysis so far has been on the intensive margin. Here, we modify our within-firm specification to explore whether there is also an effect on the extensive margin by introducing the indicator variables $\mathbb{I}_{ibt}^{entry}$ and $\mathbb{I}_{ibt}^{exit}$. The former is equal to 1 if a new relationship is formed between firm i and bank b in period t , while the latter is equal to 1 if an existing relationship between firm i and bank b ends in period t .

In the entry sample, for a given bank-firm pair, we cover all the periods from 2006 until a relationship is formed. We estimate Equation 1.4 using $\mathbb{I}_{ibt}^{entry}$ as the dependent variable and present the results in Column 5 of Table 1.2. We do not find a statistically significant effect of mining commodity prices on the extensive margin. Within our medium to large firm sample, loan supply responds to mining commodity prices on the intensive margin but not on the extensive margin.

For the exit sample, we cover all existing bank-firm pairs from 2006 until they dissolve. We estimate Equation 1.4 with $\mathbb{I}_{ibt}^{exit}$ as the dependent variable and show the results in Column 6 of Table 1.2. We find no statistically significant response of exits to mining commodity prices.

1.4.4 Additional Firm-Level Results

Lagged effects. We examine whether lagged values of the independent variable affect the current production of non-commodity firms. We estimate Equation 1.5 with an additional lag of the weighted mining commodity price exposure index. We present the results in Column 2 of Table 1.3, in which we see that the current and lagged values of the independent variables are statistically significant. Moreover, we see that the sum of both coefficients is close to that of the baseline specification in Column 1, suggesting that the baseline specification captures both the contemporaneous and lagged effects of mining price shocks.

Wage bill. The main input in our theoretical model is labor. We explore the relationship between non-mining firm wage bill $w_t^N h_{it}$ and exposure to mining commodity prices. Since the number of workers or hours worked is not precisely recorded in our dataset, we use the wage bill

as a proxy for labor. Column 3 of Table 1.3 shows the results of estimating Equation 1.5 with the wage bill as the dependent variable. We find a positive elasticity between the wage bill of a non-mining firm and its exposure to mining commodity prices through banks. This suggests that mining-exposed firms not only produce more when mining commodity prices go up but also increase their hiring.

1.5 Model

In this section, we introduce a static SOE model that captures the financial propagation mechanism of commodity booms and quantifies its aggregate importance. The model consists of a representative household, a discrete number of non-tradable intermediate goods producers, a non-commodity exports producer, a loan aggregator, a discrete number of banks, and a representative commodity exporter. The household owns all the firms and banks. Importantly, the non-tradable intermediate goods producers are the model-equivalent of the non-commodity firms in the data.

The economy has a fixed endowment of the commodity good which it wholly exports at the given international price. At the same time, the economy exports and imports non-commodity goods, while banks can obtain wholesale funds from foreign lenders. We take the non-commodity importable good as the numeraire. Moreover, the non-commodity exportable good can be consumed domestically as part of the tradable consumption good. Any excess supply of the non-commodity exportable good is exported.¹⁵ Following the SOE assumption, the price for the commodity good, the exportable non-commodity good, and importable non-commodity good are taken as given.

In line with our empirical findings, we assume that commodity exporters hold deposits in domestic banks and that they will increase such deposits after an increase in commodity prices. Moreover, as indicated by the data, banks exhibit varying degrees of exposure to the commodities sector. Each bank will receive different amounts of deposits from the commodity sector according to its exposure after a shock to commodity prices.

Motivated by our empirical results, we introduce two key financial frictions in the model. First, non-tradable intermediate firms produce with a linear technology that uses only labor. However, they face a working capital constraint. This means these firms have to pay for a fraction of their wage bill before production takes place [Neumeyer and Perri, 2005, Jermann and Quadrini, 2012]. Then, non-tradable intermediate firms need to take out loans from banks to finance such expenses.

Second, banks face a balance sheet constraint [Wang et al., 2022, Whited et al., 2022]. Banks typically fund their loans with deposits and equity. However, when deposits and equity are not enough, banks must turn to foreign wholesale money markets [Shousha, 2016]. Because foreign

¹⁵Likewise, any excess demand of the non-commodity exportable good is covered by imports.

lenders are concerned that domestic banks might default, domestic banks must pay a premium. This premium is decreasing in the amount of deposits held. Importantly, this friction connects the interest rates that banks charge on loans to the amount of deposits held.

The frictions we introduce in the model have several effects. After a commodity price boom that increases deposit holdings, banks will require less foreign wholesale funding. First, this will reduce the premiums banks pay, so they will charge lower interest rates to the domestic firms. Second, because the stock of foreign wholesale funding falls, there are fewer financial outflows to the foreign sector and more resources in the domestic economy, which implies imports will be larger. Importantly, imports are used as an input to produce the domestic tradable consumption good, which itself is complementary with the domestic non-tradable consumption good. Then, an increase of imports also drives up the domestic production of the tradable and non-tradable goods.

Timewise, the model operates as follows. At the beginning of the single period, the commodity price is revealed. The commodity producers export their endowment at the given price and deposit part of their income flow in domestic banks. We assume this deposit bears no interest. Then, all agents decide how much to consume, how much labor to supply, and how much to produce. Households fund banks with equity. Given the previous actions, banks take foreign wholesale funding and lend to non-tradable intermediate firms. At the end of the single period, firms produce and pay back loans. Banks pay back wholesale funding and return deposits. Then, the household receives the profits from the firms and the banks. Finally, the household consumes.

1.5.1 Household

The representative household consumes the final consumption good c at price p , supplies labor h^N to the non-tradable intermediate firms at wage w^N , supplies labor h^X to the non-commodity exporting firm at wage w^X , and owns the firms and the banks, for which it receives total profits π . Moreover, the household provides equity K_b to fund the banks. Because our focus is not on equity, we will define it exogenously in our simulations.

Following the SOE literature, we assume the household has GHH preferences [Greenwood et al., 1988], so labor supply is independent of consumption. The household solves the problem

$$\max_{c, h^N, h^X} \frac{1}{1-v} \left(c - \frac{(h^N)^{1+\psi}}{1+\psi} - \frac{(h^X)^{1+\psi}}{1+\psi} \right)^{1-v},$$

subject to

$$pc + \sum_b K_b = w^N h^N + w^X h^X + \pi,$$

where v measures the relative risk aversion and ψ is the inverse of the labor supply elasticity.

The final consumption good is a bundle of the non-tradable consumption good c^N and the tradable consumption good c^T , such that

$$c = \left[(\Lambda^{NT})^{\frac{1}{\gamma}} (c^N)^{\frac{\gamma-1}{\gamma}} + (1 - \Lambda^{NT})^{\frac{1}{\gamma}} (c^T)^{\frac{\gamma-1}{\gamma}} \right]^{\frac{\gamma}{\gamma-1}},$$

where γ is the elasticity of substitution and Λ^{NT} is the weight of the non-tradable good. We calibrate the elasticity of substitution so that the tradable and the non-tradable consumption goods are complements.

The non-tradable consumption good c^N is a bundle composed of a discrete number of differentiated non-tradable intermediate consumption goods c_i so that

$$c^N = \left[\sum_i \Lambda_i^{\frac{1}{\sigma}} c_i^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}},$$

where σ is the elasticity of substitution and Λ_i denotes the weights, with $\Lambda_i \geq 0$ and $\sum_i \Lambda_i = 1$. We calibrate this bundle such that non-tradable intermediate consumption goods are substitutes. Also, the price of the non-tradable consumption good is p^N .

Lastly, the tradable consumption good c^T is a bundle composed of non-commodity exportable consumption good c^X , and an importable consumption good c^M , so that

$$c^T = \left[(\Lambda^{XM})^{\frac{1}{\chi}} (c^X)^{\frac{\chi-1}{\chi}} + (1 - \Lambda^{XM})^{\frac{1}{\chi}} (c^M)^{\frac{\chi-1}{\chi}} \right]^{\frac{\chi}{\chi-1}},$$

where χ is the elasticity of substitution and Λ^{XM} is the weight of the non-commodity exportable good. We calibrate the elasticity of substitution to be equal to 1, so that this bundle has a Cobb-Douglas functional form. The price of the tradable consumption good is p^T .

1.5.2 Non-Tradable Intermediate Good Firms

There is a discrete number of non-tradable intermediate good firms indexed by $i \in [1, \dots, I]$ that engage in monopolistic competition and are owned by the household. These firms represent the non-commodity firms from our dataset. Each firm employs labor h_i to produce y_i and charge price p_i . We further assume these firms all face the same productivity A , and the market for each non-tradable intermediate good is in equilibrium, such that $c_i = y_i$.

These firms are also subject to a working capital constraint, which means they must pay a proportion θ of the wage bill $w^N h_i$ before production takes place [Neumeyer and Perri, 2005, Jermann and Quadrini, 2012]. Thus, these firms need to take out loans from banks to finance their production expenses. The firms produce and pay off their loans plus interest at the end of the

period.

For a given non-tradable intermediate firm i , the optimization problem is

$$\max_{p_i, y_i, h_i} \pi_i = p_i y_i - (1 + \theta r_i^L) w^N h_i,$$

subject to

$$y_i = \Lambda_i \left(\frac{p_i}{p^N} \right)^{-\sigma} c^N,$$

$$y_i = A h_i,$$

where r_i^L is the loan interest rate each firm faces, and p^N is the price of the non-tradable consumption bundle.

Additionally, for each non-tradable intermediate firm we define loan demand as

$$L_i = \theta w^N h_i.$$

1.5.3 Commodity Exporter

We assume there is a representative commodity exporter, fully owned by the household.¹⁶ It receives an endowment $y^{CO} = 1$ and exports it at exogenous price p^{CO} . This firm does not hire labor to operate or use intermediate inputs.¹⁷

As suggested by Fact 1, we assume that the commodity exporter holds deposits in the domestic banks. Additionally, the amount of deposits held is sensitive to the commodity prices. We discuss this feature more extensively in the bank section. As a simplification, we further assume these deposits bear no interest rate. At the end of the period, the commodity producer gets its deposits back from banks and transfers the export revenue to households.

1.5.4 Non-Commodity Exportable Good Firm

In the model, there is a perfectly competitive non-commodity exportable goods producer, which uses labor h^X to produce y^X . The problem it faces is

$$\max_{y^X, h^X} \pi^X = p^X y^X - w^X h^X,$$

¹⁶In Appendix A.2.8 we explore the possibility of the commodity exporters paying part of their revenues as dividends to foreign owners.

¹⁷This is based on the observation that, in the Peruvian setting, the mining sector employs few workers and demands few domestically supplied goods and services.

subject to

$$y^X = Ah^X,$$

where this firm displays the same level of productivity A as the non-tradable intermediate firms. As a producer of a tradable good, this firm charges the price p^X , which is exogenously set in the foreign markets.

1.5.5 Loan Aggregator

Non-tradable intermediate firms obtain their loans L_i from firm-specific and perfectly competitive loan aggregators, which are owned by the household. Each loan aggregator specializes in serving a single non-tradable intermediate firm $i \in [1, \dots, I]$ by sourcing bank-level loans L_{ib} from banks b at interest rate r_b^L to convert them into loan bundle L_i using CES technology [Gerali et al., 2010, Andrés et al., 2013, Ulate, 2021] and charge interest rate r_i^L . The CES assumption reflects two facts from firm and bank relationships. First, firms may obtain loans from more than one bank. Second, there is imperfect substitutability across different banks as they may be specialized in certain sectors or may have special relationships with specific firms [Chodorow-Reich, 2014, Cong et al., 2019, Paravisini et al., 2023].

The problem of the loan aggregator is

$$\max_{L_{ib}} (1 + r_i^L) L_i - \sum_b (1 + r_b^L) L_{ib},$$

subject to loan bundle

$$L_i = \left[\sum_b s_{ib}^{\frac{1}{\varepsilon}} L_{ib}^{\frac{\varepsilon-1}{\varepsilon}} \right]^{\frac{\varepsilon}{\varepsilon-1}},$$

where s_{ib} are fixed weights and ε is the elasticity of substitution across banks.¹⁸ The calibration we follow for the simulations sets the bank-level loans that compose the bundle as substitutes. Moreover, for a given firm i , it must hold that $s_{ib} \geq 0$ and $\sum_b s_{ib} = 1$. We allow the weights s_{ib} to be zero, so firms do not necessarily have to source loans from all banks.

1.5.6 Banks

There is a discrete number of banks indexed by $b \in [1, \dots, B]$, with each offering a differentiated variety of loans. Banks operate in a monopolistically competitive fashion and are owned by households. Banks choose loans L_b , interest rate r_b , and foreign wholesale funding N_b subject to their

¹⁸Column 4 of Table 1.2 shows there is no effect of the mining commodity price shock on the extensive margin of loans. Thus, we do not consider this margin within the model.

balance sheet and the loan demand coming from loan aggregators.

Banks are subject to a balance sheet friction [Gilje et al., 2016, Wang et al., 2022, Whited et al., 2022]. Their balance sheet has loans on the asset side, which have to be funded on the liabilities side through firm deposits D_b from commodity producers, equity K_b , and foreign wholesale funding N_b .¹⁹ This implies that when deposits and equity are insufficient to fully cover loans, domestic banks have to go to foreign money lenders to obtain foreign wholesale funding [Shousha, 2016].

The cost of the foreign wholesale funding has two parts.²⁰ First, the cost depends on the foreign interest rate r . Second, because domestic banks could default, foreign lenders will charge them a premium $\phi \ln \left(\frac{\mathcal{B}_b}{D_b} \right)$, where ϕ is a cost function parameter and $\mathcal{B}_b$ is a constant that serves as a benchmark. The functional form $\phi \ln \left(\frac{\mathcal{B}_b}{D_b} \right)$ captures the idea that when deposits are abundant, the need for foreign wholesale funding is lower and domestic banks are charged lower premiums in foreign markets.²¹

The problem banks solve is

$$\max_{r_b^L, L_b, N_b} \pi_b = (1 + r_b^L) L_b - D_b - \left[1 + r + \phi \ln \left(\frac{\mathcal{B}_b}{D_b} \right) \right] N_b,$$

subject to balance sheet

$$L_b = D_b + K_b + N_b,$$

and total loan demand

$$L_b = \sum_i \left[\left(\frac{1 + r_b^L}{1 + r_i^L} \right)^{-\varepsilon} s_{ib} L_i \right].$$

The optimal interest rate r_b^L charged is

$$1 + r_b^L = \frac{\varepsilon}{\varepsilon - 1} \left[1 + r + \phi \ln \left(\frac{\mathcal{B}_b}{D_b} \right) \right],$$

where the introduction of the balance sheet friction connects the interest rate banks charge with the amount of firm deposits they hold. The larger the amount of deposits held, the lower the premium paid by the bank for the foreign wholesale funding and the lower the interest rate charged to firms.

Consistent with the empirical evidence from Equation 1.3, we posit the following reduced-form

¹⁹As stated above, because our focus is not on equity, in our simulations we set equity K_b exogenously to match its observed change in the data.

²⁰This cost functional form may be motivated by different microfoundations. In Appendix A.2.1 we outline one such possible microfoundation.

²¹We may think of constant benchmark $\mathcal{B}_b$ as capturing the reputation of the domestic bank in foreign markets or the relationships the domestic bank may have with foreign lenders. For technical reasons related to recovering the model parameters from our empirical results, we require $\frac{\mathcal{B}_b}{D_b} > 1$ and $\phi \ln \left(\frac{\mathcal{B}_b}{D_b} \right) \approx 0$ (see Appendix A.2.5).

relationship between deposits D_b and commodity prices p^{CO} :

$$\ln(D_b) = \alpha_b + \beta \omega_b^{CO} \ln(p^{CO}),$$

where α_b is a constant and ω_b^{CO} is the total exposure to the commodity exporting sector of bank b . As suggested by the data, this equation captures the idea that commodity exporters hold deposits in domestic banks and will increase their holdings when their income increases.

1.5.7 Market Clearing Conditions, Real Exchange Rate, Balance of Payments, and Real GDP

In this section we first list the equilibrium conditions for the markets of the model. The equilibrium for the non-tradable labor is

$$h^N = \sum_i h_i,$$

while total labor h is defined as

$$h = h^X + h^N.$$

For the non-tradable consumption good we have

$$c^N = y^N,$$

and for non-tradable intermediate goods we have that

$$c_i = y_i,$$

for all i .

We further define total non-commodity output Y^{NCO} as

$$Y^{NCO} = y^X + y^N.$$

On the loan side, the equilibrium condition is

$$L_b = \sum_i L_{ib},$$

for all b .

We define the real exchange rate rer as

$$rer = \frac{1}{p},$$

where a decrease means a real appreciation of the domestic goods relative to the foreign ones.

The balance of payments equation is

$$p^{CO}y^{CO} + p^X(y^X - c^X) - c^M = \sum_b \left[r + \phi \ln \left(\frac{\mathcal{B}_b}{D_b} \right) \right] N_b,$$

where $p^{CO}y^{CO} + p^X(y^X - c^X) - c^M$ is the trade balance and $\sum_b \left[r + \phi \ln \left(\frac{\mathcal{B}_b}{D_b} \right) \right] N_b$ is the financial account.

Last, we define real GDP as

$$\overline{p}GDP = \overline{p}c + \overline{p^{CO}}y^{CO} + \overline{p^X}(y^X - c^X) - \overline{p^M}c^M,$$

where overlines denote constant prices.

1.5.8 Equilibrium

We solve the model using the exact hat algebra method of Dekle et al. [2008]. For a given variable x in levels, we define $\widehat{x} = \frac{x'}{x}$ as the change between final state x' and initial state x . Likewise, for a given interest rate i we define $1 + \widehat{i} = \frac{1+i'}{1+i}$.

Given a commodity price shock $\widehat{p^{CO}}$, the market foreign interest rate $\widehat{r}$, TFP $\widehat{A}$, equity $\left\{ \widehat{K}_b \right\}_b$, parameters, and ratios, the equilibrium of the model is a set of prices $\widehat{p}, \widehat{p^N}, \widehat{p^T}, \left\{ \widehat{p}_i \right\}_i, \widehat{w^N}, \widehat{w^X}, \widehat{r}, \widehat{er}$, a set of interest rates $\left\{ \widehat{r}_i^L \right\}_i, \left\{ \widehat{r}_b^L \right\}_b$, and a set of allocations $\widehat{c}, \widehat{c^N}, \widehat{c^T}, \widehat{c^X}, \widehat{c^M}, \widehat{y^N}, \widehat{y^X}, \widehat{y^{NCO}}, \widehat{h}, \widehat{h^N}, \widehat{h^X}, \left\{ \widehat{y}_i, \widehat{h}_i, \widehat{L}_i \right\}_i, \left\{ \widehat{L}_b, \widehat{D}_b, \widehat{N}_b \right\}_b, \left\{ \widehat{L}_{ib} \right\}_{i,b}$ such that the final consumption good, non-tradable good, loan, and labor markets clear and the bank balance sheets hold. We present the equilibrium conditions in Appendix A.2.2 and the conditions expressed in terms of hat algebra in Appendix A.2.3.

1.5.9 Calibration

We divide the parameters into the non-bank block, the bank block, weights, ratios, and shocks. We present the non-bank block parameters, the bank block parameters, weights and shocks in Table 1.4, while we relegate the ratios to Appendix A.2.4. Based on the matched bank-firm-loan dataset from Peru, we calibrate the model for 5 representative non-tradable firms, 9 banks and a representative commodity exporter.²² For the non-bank block, we calibrate the intratemporal parameters using standard values in the SOE literature [Uribe and Schmitt-Grohé, 2017], while we

²²We sort non-commodity firms from the firm survey according to their sales and average them by thirds. Each representative non-tradable firm in the model is calibrated to match the average firm by third from the data.

Table 1.4: Parameters

Non-bank block		
ψ	Inverse of labor supply elasticity	0.455
γ	Elasticity of substitution in final good basket	0.5
χ	Elasticity of substitution in tradable good basket	1
σ	Elasticity of substitution in NT good basket	5
Λ^{NT}	Weight of non-tradables in final consumption basket	0.58
Λ^{XM}	Weight of exportables in tradable consumption basket	0.1
Λ_i	Weight of firms in NT consumption basket	By firm
Bank block		
β	Commodity price to deposit elasticity	22.171
ε	Elasticity of substitution across banks in loan bundle	3.879
ϕ	Bank cost function parameter	0.066
θ	Working capital financed by loans	0.160
Bank weights in firm loan portfolio		
s_{ib}	Individual loans / Total loans, by firm	By firm and bank
Bank exposure to commodity prices		
ω_b^{CO}	Mining loans / Total loans, by bank	By bank
Shocks		
$\widehat{p^{CO}}$	Commodity price shock	2.57
$\widehat{A}$	TFP	1.13
$\widehat{r}$	Foreign interest rate	-0.02
$\widehat{K_b}$	Bank equity	2.22

Note: All ratios calculated using deflated data and expressed in domestic currency. Imports and exports consider traded goods only. Because foreign wholesale funding N_b is not properly measured in the data, we calculate $N_b = L_b - D_b - K_b$ as a proxy. We provide details on the data underlying the shocks in Appendix A.2.6.

set the elasticity of substitution in the non-tradable basket so firms have a markup of 25 percent. For the ratios, we calibrate them to match the averages in Peruvian data in the period 2006-2018.

Next, we provide detail on how we obtain the values for the bank block parameters based on the data and the regressions of Section 1.4, as well as on how we calibrate the shocks.

Working capital financed by banks θ . We proceed in two parts. First, from the firm survey, we recover total expenses Z , which include intermediate consumption, investment, wage bill, and financial expenses. Let $\theta r^L Z$ be the financial expenses and r^L the interest rate paid.

Second, we need to obtain a value for r^L . The only interest rate data available for Peru in our period of analysis is the average real interest rate for loans across all banks. We take the average

between 2006 and 2018 to set $r^L = 0.11$.²³ With this, we recover a firm-level value for θ . Finally, we calculate the median across all firms and years to find $\theta = 0.16$.

Commodity price to deposit elasticity β . We refer to the estimation of Equation 1.3 to recover this elasticity, where we found $\beta = 22.171$. This implies that a hike in the mining commodity price increases the amount of deposits held by the mining-exposed banks.

Elasticity of substitution across banks in loan bundle ε and bank cost function parameter ϕ . To recover these two parameters, we start from the loan-level and firm-level equilibrium conditions of the model and derive expressions that correspond with Equations 1.4 and 1.5. We show the detailed steps in Appendix A.2.5.

Thus, we are able to express the estimated coefficients ζ and κ in terms of the structural parameters of the model so that $\zeta = \varepsilon\phi\beta$ and $\kappa = (\sigma - 1)\theta\phi\beta$. Given that we know the numerical values of estimated coefficients β , ζ , and κ , as well as calibrated parameters σ and θ , we find $\varepsilon = 3.879$ and $\phi = 0.066$.

Shocks. In the next section, we simulate a commodity price shock $\widehat{p^{CO}} = 2.57$, which is equal to the observed change in the weighted relative mining price index across all minerals that Peru exports between 2003 and 2011. We choose 2003 as the starting point because it marks the beginning of the boom, while we pick 2011 as the end point of the simulation because it is when mining commodity prices reached their peak, as seen in Figure 1.1.

We target the observed change in Peruvian GDP between 2003 and 2011 by choosing the change in TFP. We set $\widehat{A} = 1.13$ for all exportable and non-tradable intermediate firms.²⁴ Within the model, we also have additional shocks that help us match the data. We set these shocks to match their observed changes in the data between 2003 and 2011. We set the market foreign interest rate $\widehat{r} = -0.02$, and the equity $\widehat{K_b} = 2.22$ for all b .²⁵ Lastly, our focus being on commodity price shocks, we assume that the price of the exportable non-commodity good does not vary, so $\widehat{p^X} = 1$.

1.6 Quantifying the Aggregate Importance of the Financial Propagation Mechanism

In this section, we conduct simulations using the hat algebra version of the model to quantify the aggregate importance of the financial propagation mechanism during commodity booms. Specifically, we first simulate the model with the financial propagation mechanism activated, using the calibration and shocks defined in the previous section. Next, we simulate the model under the

²³This is a weighted average of domestic currency-denominated and dollar-denominated loans.

²⁴As a comparison, the change in Peruvian TFP at constant national prices in between 2003 and 2011 according to the Penn World Table is 1.12.

²⁵We provide details on the data underlying the shocks in Appendix A.2.6.

same calibration and shocks but with the financial propagation mechanism turned off. We define the aggregate importance of the financial propagation mechanism as the difference in GDP growth rates between these two scenarios.

1.6.1 Main Simulation

Table 1.5 reports the results of the main simulation. Columns 1 and 2 show the moments coming from the data and the model, respectively. First, our baseline simulation hits the targeted moment. The baseline simulation predicts a GDP growth $\widehat{GDP}$ of 65 percent, as we see in the data. Second, related to the untargeted moments, we focus on the change in consumption $\widehat{c}$, in non-commodity output $\widehat{y^{NCO}}$, in imports $\widehat{c^M}$, real exchange rate $\widehat{rer}$, average loans $\widehat{L_b}$, average deposits $\widehat{D_b}$, average wholesale funding $\widehat{N_b}$, and average loan interest rate $\widehat{r_b^L}$. Our baseline simulation gets fairly close to the untargeted moments.²⁶

Within our model, a commodity price boom that raises firm deposits reduces the need for banks to rely on foreign wholesale funding. First, this decreases the premiums banks pay for foreign wholesale funding, enabling them to charge lower interest rates to non-tradable intermediate firms. This lowers the production costs of non-tradable intermediate firms and constitutes an interest rate channel effect that drives GDP upward.

Second, because the amount of foreign wholesale funding decreases, there are fewer financial outflows to the foreign sector and more resources remain in the domestic economy. This leads to an increase in imports, which also raises consumption of the tradable good. Because tradable goods are complements with non-tradable goods, the demand for non-tradable goods rises as well. This is a wealth channel effect that also contributes to GDP growth.

To quantify the aggregate importance of the financial propagation mechanism of commodity booms, we perform a counterfactual simulation in which we simulate the model with the same calibration and shocks as before but turn off the financial propagation mechanism. In terms of the model, this means setting $\beta = 0$. Therefore, firm deposits do not vary when there is a mining commodity price shock. We present the results of the counterfactual simulation in Column 3, while Column 4 shows the difference between the baseline and the counterfactual scenarios.

Under the counterfactual scenario, GDP still grows, but at a lower rate than in the baseline scenario. The difference between the two cases, which we interpret as the aggregate importance of the financial propagation mechanism of commodity booms, is 6 percentage points or approximately one tenth of the observed GDP growth in the baseline case. We now turn to explain this difference. In the counterfactual scenario, there is still a positive wealth shock to the economy that generates an increase in loan demand. However, with the financial propagation mechanism turned off, deposits

²⁶We provide details on the data-based moments in Appendix A.2.6.

Table 1.5: Main simulation

Variable	(1) Data, 2003-2011	(2) Baseline: Fin. mech.	(3) CF: No fin. mech. ($\beta = 0$)	(4) Baseline minus CF
Targeted moment				
$\widehat{GDP}$	1.65	1.65	1.59	0.06
Untargeted moments				
$\widehat{c}$	1.83	2.00	1.89	0.11
$\widehat{y^{NCO}}$	1.71	1.66	1.61	0.05
$\widehat{c^M}$	2.24	2.15	2.00	0.14
$\widehat{rer}$	0.77	0.87	0.89	-0.02
Ave. $\widehat{L_b}$	2.34	2.46	2.23	0.23
Ave. $\widehat{D_b}$	3.02	2.49	1.00	1.49
Ave. $\widehat{N_b}$	8.54	4.24	6.65	-2.41
Ave. $\widehat{r_b^L}$	-0.05	-0.08	-0.02	-0.06

Note: Column 1 shows data moments. Column 2 shows moments from baseline simulation with financial propagation mechanism. Column 3 shows moments from counterfactual simulation with no financial propagation mechanism. Column 4 shows the difference between baseline and counterfactual. All moments are expressed in percentage variations between initial and final states. We report the complete set of results in Appendix A.2.7. For the bank-related variables, the model-based moments are averages computed using loan-based weights of each bank in total loans. These weights are calibrated according to observed data. For these same variables, the empirical untargeted moments are based on the change of the aggregate variable.

do not increase, so banks now require more foreign wholesale funding. First, this drives up their premiums and interest rates charged. Thus, non-tradable intermediate firms do not take out loans as much as in the baseline. Second, a higher amount of debt increases the financial outflows to the foreign sector and reduces the resources available in the domestic economy, which implies fewer imports and lower demand for non-tradable goods. These two channels reduce the GDP growth rate.²⁷

We interpret this result as a lower bound of the aggregate importance of the financial propagation mechanism of commodity booms. First, the model is calibrated using data from medium to large firms coming from the firm survey. This excludes small firms, who presumably are more financially constrained than their larger counterparts and would react more strongly to an increased loan supply. Second, our simulation includes other shocks, which decreases the relative importance of the financial propagation mechanism.

²⁷In Appendix A.2.8 we explore the sensitivity of the model results under alternative parameterizations.

1.6.2 Simulation with Only a Commodity Price Shock

In this section, we isolate the effects of a commodity price shock by simulating the model with only this shock, excluding any other disturbances. Thus, the model economy is subject to a pure wealth increase coming from the commodity price shock, more in line with the standard SOE macroeconomic models. To do so, we conduct the simulation by introducing a positive commodity price shock $\widehat{p}^{CO} = 2.57$ to the model, under the same baseline calibration outlined earlier, but turning off the other shocks (i.e., $\widehat{A} = 1$, $\widehat{r} = 0$, and $\widehat{K}_b = 1$ for all b).

Table 1.6 presents the results after the commodity price shock. Under the baseline case, with the financial propagation mechanism turned on, GDP grows by 32 percent. When we turn off the financial propagation mechanism, GDP grows by 25 percent. Thus, the financial propagation mechanism accounts for 7 percentage points or approximately one fifth of what we observe in the baseline case. Alternatively, we can say that the financial propagation mechanism amplifies the GDP growth rate observed in the counterfactual case by approximately 28 percent.

The channels behind this result remain the same as before: The increase in deposits reduces the need of banks to obtain foreign wholesale funding, which results in lower financial outflows and lower interest rates. However, in the absence of the other complementary shocks from the main simulation that helped us match the data, the financial propagation mechanism becomes more important in driving the baseline GDP growth.

1.7 Conclusions

This paper proposes a novel financial propagation mechanism in SOEs in which commodity price booms affect the non-commodity sector through domestic banks. Following a boom, exporters increase their deposits in domestic banks, enabling banks to expand loans to non-commodity firms. In turn, non-commodity firms increase their output.

Using bank-firm-loan microdata from Peru, an SOE that went through a mining commodity price boom in the 2000s, we find empirical support for this mechanism. We document that banks are heterogeneously exposed to the mining sector, as each holds a different mining client composition. After an exogenous rise in mining commodity prices, banks with higher exposure to the mining sector see increased firm deposits and expand their loan supply to non-mining firms. Furthermore, non-mining firms exposed to mining prices through their banks register a boost in output.

We incorporate this mechanism into a static SOE model with banks and an endowment commodity sector to quantify the aggregate importance of this mechanism. We calibrate the model to Peruvian data and replicate key economic moments. A counterfactual analysis suggests the finan-

Table 1.6: Commodity price shock-only simulations

	(1)	(2)	(3)
Variable	Baseline: Fin. mech.	CF: No fin. mech. ($\beta = 0$)	Baseline minus CF
$\widehat{GDP}$	1.32	1.25	0.07
$\widehat{c}$	1.65	1.51	0.14
$\widehat{y^{NCO}}$	1.23	1.17	0.06
$\widehat{c^M}$	1.87	1.68	0.19
$\widehat{rer}$	0.78	0.81	-0.03
Ave. $\widehat{L_b}$	2.35	2.02	0.33
Ave. $\widehat{D_b}$	2.49	1.00	1.49
Ave. $\widehat{N_b}$	6.13	7.79	-1.66
Ave. $\widehat{r_b^L}$	-0.06	0.00	-0.06

Note: Column 1 shows moments from baseline simulation with financial propagation mechanism. Column 2 shows moments from counterfactual simulation with no financial propagation mechanism. Column 3 shows the difference between baseline and counterfactual. All moments are expressed in percentage variations between initial and final states. We report the complete set of results in Appendix A.2.7. Averages for the bank-related model variables are computed using loan-based weights of each bank in total loans. These weights are calibrated according to observed data.

cial propagation mechanism accounts for one tenth of the observed 65 percent GDP growth in Peru during the 2003–2011 boom.

CHAPTER 2

Aggregate Implications of Heterogeneous Inflation Expectations: The Role of Individual Experience

Abstract

We show that inflation expectations are heterogeneous and depend on past individual experiences. We propose a history-based expectations-augmented Kalman filter to represent consumers' heterogeneous inflation expectations, where heterogeneity comes from an anchoring-to-the-past mechanism. Using survey data, we show that the model replicates US consumers' inflation expectations and its heterogeneity across cohorts. We introduce this mechanism into a New Keynesian model and find that heterogeneous expectations anchor aggregate responses to the agents' inflation history, producing sluggish expectations dynamics. Central banks should be active to prevent inflationary episodes that agents will remember far into the future.¹

2.1 Introduction

Inflation expectations matter for decisions at both the firm and the household levels [Coibion et al., 2023, 2020, Hajdini et al., 2022b]. Given their importance, there is an increased interest in measuring them and exploring what determines their formation process. The recent literature shows that individuals form inflation expectations, for instance, based on their recent buying experience [D'Acunto et al., 2021] and historical experiences regarding aggregate inflation [Malmendier and Nagel, 2016]. While most studies on this topic provide empirical evidence regarding how differences in inflation expectations arise at the individual level and their effects on different micro-level decisions, there is less understanding of the aggregate implications associated with the inflation expectations' heterogeneity observed in the data. There is a noticeable gap in the literature between the empirical micro-level findings and macroeconomic models.

¹This chapter is co-authored with Mathieu Pedemonte and Esteban Verdugo.

This paper aims to fill this gap. First, we show that individuals’ inflation expectations depend on their history of inflation, confirming the main empirical finding of Malmendier and Nagel [2016] but using new US and international evidence.² Using detailed micro-level data, we find that (i) inflation expectations are heterogeneous across cohorts, (ii) inflation experiences are clustered by age, (iii) individual inflation history is positively correlated with inflation expectations, and (iv) there are no differences between cohorts in updating to current information, after controlling for their own history.

These facts suggest that consumers use their history of inflation to form expectations by positively weighing their past inflationary experiences.³ We propose a model of inflation expectation-formation process where consumers weigh past experiences. We depart from the full information rational expectations model by presenting a framework where individuals use their past inflation histories and compare them to current information rationally gathered from signals.

Under the proposed framework, inflation expectations have two components: a forecast made with current shared information between agents and an idiosyncratic referential term that depends on individuals’ past experiences. We structurally estimate the parameter governing the expectations-formation process, using the history of forecasts consumers would make in the model according to signals associated with their shopping experiences. As all respondents share current information, we control for the common forecast using time-fixed effects. The resulting estimation for the coefficient is less than zero, implying that consumers positively weigh their inflation history to the detriment of current unexpected news.⁴

We model the shared component following a standard signal-extraction procedure. Based on works exploring reference prices through consumers’ shopping experiences [D’Acunto et al., 2021] and recent evidence showing that consumers use some food prices to form inflation expectations [D’Acunto and Weber, 2022], we use the lagged inflation rate of the food component of the CPI as a signal of the non-observed aggregate inflation variable.⁵ Then, we construct a cohort-specific inflation expectations measure using (i) the already computed shared component, (ii) an

²While Malmendier and Nagel [2016] use the University of Michigan’s Survey of Consumers (MSC), in sections 2.3 and 2.4 of this paper, we provide external validity to their result by using a different data source with panel characteristics: the Survey of Consumer Expectations (SCE) from the Federal Reserve Bank of New York. In addition, in Appendix B.7, we show that the finding is consistent at the international level even after controlling for common cohort characteristics.

³This could be interpreted as in the availability bias [Tversky and Kahneman, 1973], where consumers would use the information available (their experience) to form expectations, over-weighing their history instead of using more complex models that are necessary to form expectations about future economic variables.

⁴This result does not imply that agents cannot overreact to current news. Our empirical exercise shows that the proposed expectation modeling framework explains the heterogeneity observed across cohorts. However, the presence of a common component in the modeling allows for a common overreaction to some current news.

⁵Our findings are robust to using more specific signals such as CPI’s food at home or dairy components, which, according to D’Acunto and Weber [2022], are the prices that US consumers primarily consider when forming inflation expectations.

idiosyncratic element related to the inflation history of the cohort, and (iii) the coefficient value estimated in the empirical section of the paper.

With our cohort-specific inflation expectations measure, we obtain model-based forecasts that closely match the inflation expectations observed in the data across cohorts and time. Regressing our model-based cohort-specific expectations on the individual expectations observed in survey data delivers a statistically significant coefficient of 1.069. Although the model-based inflation forecasts do not use micro-level information on inflation expectations, they predict consumers' survey data remarkably well. By using these two simple consumer references — shopping experience and history of inflation — we can effectively model survey households' expectations. This paper shows that consumer expectations surveys contain relevant and meaningful information that can be modeled.

As in Bianchi et al. [2021] and L'Huillier et al. [2021], we incorporate our proposed expectation formation process into a New Keynesian model.⁶ Our proposed framework differs from and complements these works. Bianchi et al. [2021] and L'Huillier et al. [2021] use a rational expectations operator, such that agents overreact to any news in a representativeness heuristic way [Kahneman and Tversky, 1972] and where full information characterizes the most likely scenario.⁷ We model a history-based heuristic instead. We explicitly incorporate agents' history into the model and use it as their referential information. Due to this approach, we can exploit cohort-specific time-evolving history and show that consumers use this *availability* heuristic in their expectation formation process. Hence, agents' inflation experiences are over-sampled [Bordalo et al., 2023] as the past is over-extrapolated [Angeletos et al., 2021].

We then explore the macro implications of this micro-level heterogeneity in expectations, where we allow households to form expectations according to the proposed framework. In our model, while old generations have their expectations shaped mainly by their past, new generations are highly influenced by recent developments. We find that heterogeneous expectations anchor the aggregate inflation and output gap response to agents' history. At the same time, they also increase the duration of the effects of the shocks.⁸ After an inflationary shock, the model produces hump-shaped expectations. This reaction is consistent with over-extrapolation as in Angeletos et al. [2021]. Consumers react slowly to the inflationary shock in the first few periods, as they are tied

⁶They include the diagnostic expectation model proposed by Bordalo et al. [2020] into a Dynamic Stochastic General Equilibrium (DSGE) model. While Bordalo et al. [2020]'s model is an overreaction model, we use and extend their framework to allow for a negative distortion parameter, in their case, the diagnosticity parameter. In the following sections, we explain our microfoundation. Notice that Bianchi et al. [2021] and L'Huillier et al. [2021] solutions also apply to our type of model.

⁷See Benjamin [2019] for a discussion on the implications of the assumptions of the diagnostic expectations model.

⁸While we could incorporate other forms of heterogeneity and biases related to experience-based mechanisms on the firm or the government side of the economy, in this paper, we introduce non-rational heterogeneous expectations only at the household level. We opt for this option mainly because of data availability motives.

to their reference from the steady state. This first reaction is consistent with randomized controlled trial evidence showing that consumers weigh their priors and do not react one-to-one to current or future inflation news [Weber et al., 2023]. After witnessing higher inflation (and new cohorts entering the economy in a high-inflation environment), consumers get tied to this new inflationary reference and over-extrapolate the shock, overreacting in their forecast. From the aggregate results, we infer individual heterogeneous consumption and labor reactions.

We perform an optimal Taylor rule exercise where the central bank seeks to minimize the expected volatility of the economy by optimally choosing the parameters of a Taylor rule. When we allow for heterogeneous expectations in the model, agents have long memories and remember current shocks far into the future. After a negative supply shock or a positive demand shock, the optimal response of the central bank is to be more active compared to the full information rational expectations case. In this manner, the monetary authority prevents inflation from rising and prevents agents from incorporating a high inflation episode into their memories.

This paper has important implications for explaining past inflation dynamics and learning about the future consequences of recent economic developments. Since 2021, a new cohort of consumers worldwide has been experiencing relatively high inflation for the first time. According to our findings, this high-inflation episode could have consequences in the medium run since consumers incorporate this episode into their history of inflation, adjusting future expectations.⁹ Our framework shows that accommodating high inflation produces higher and more persistent inflation expectations, which in turn generate a higher and more persistent inflation rate in the future. Our findings help us understand why inflation has persisted in the past, why consumers' inflation expectations are persistent today, what to expect from episodes of unusually high inflation, and how central banks should react to such episodes.

The rest of the paper is organized as follows. Section 2.2 discusses recent works on the topic. Section 2.3 provides empirical results regarding consumers' heterogeneity in inflation expectations. We empirically model inflation expectations depending on the history of inflation experienced by cohorts in Section 2.4. Section 2.5 discusses the aggregate implications arising from heterogeneous inflation expectations. Section 2.6 shows results obtained for an optimal Taylor rule exercise. We analyze the high-inflation episode of 2021 through the lens of our theoretical model in Section 2.7. Finally, Section 2.8 concludes.

2.2 Literature review

Recent macroeconomic models show the relevance of heterogeneity in explaining aggregate fluctuations. However, the focus has been almost exclusively on households' financial constraints

⁹In a recent work, Salle et al. [2024] find that people use their lifetime inflation experiences to form expectations.

[Kaplan et al., 2018]. Moreover, there are few studies on the aggregate role of expectations heterogeneity. Although surveys show significant heterogeneity across firms [Coibion et al., 2018] and households inflation expectations [Hajdini et al., 2022c], few works study its macroeconomic implications. A notable exception is Afrouzi [2020], which shows that heterogeneity in firm-level inflation expectations, coming from different levels of attention due to endogenous information acquisition on competitors' beliefs, amplifies monetary non-neutrality. Our paper focuses on the heterogeneity of expectations from the household side. In our framework, households' heterogeneous inflation expectations anchor the response of aggregate variables to agents' memories, increasing the persistence of the effects of shocks. Therefore, an energetic reaction from monetary authorities prevents current high inflation and prevents agents from incorporating high-inflation episodes into their memories, thus preventing higher future inflation expectations.

Recent empirical literature shows that households' inflation expectations depart from full information rational expectations and are heterogeneous. Relevant to our paper, Malmendier and Nagel [2016] document that households present learning from past inflation mechanisms when forming inflation expectations. People who have experienced higher inflation rates in the past have higher future inflation expectations. Salle et al. [2024] also find that people use their lifetime inflation experiences to form expectations. Those experiences affect how they interpret recent data and their trust on the monetary authority. Therefore, heterogeneity of inflation expectations naturally arises due to different experiences with past inflation rates. Malmendier et al. [2021] discuss similar results, while other works focus on exploring other heterogeneity sources across firms and consumers [Hajdini et al., 2022c].

Households' inflation expectations depend on more than just past experiences. Evidence shows they also respond to other variables such as professional forecasts [Carroll, 2003], prices exposure [D'Acunto et al., 2021], and socioeconomic characteristics [D'Acunto et al., 2022], among others.¹⁰ This departure from full information rational expectations and the presence of heterogeneity is relevant since households' inflation expectations affect a broad set of households' decisions [Roth and Wohlfart, 2020, Hajdini et al., 2022b]. For instance, Malmendier and Nagel [2016] shows that inflation expectations influence individuals' financial decisions, while Coibion et al. [2023] confirms that inflation expectations partly determine households' spending on durable goods.

The model proposed in this paper features monetary policy, a learning from the past mechanism, and overlapping generations, connecting the paper to several strands of the theoretical literature. First, our model closely follows the literature enclosing behavioral New Keynesian models such

¹⁰Although our focus is on the household side of the economy, evidence shows that professional forecasters depart from rational expectations too [Coibion and Gorodnichenko, 2015, Bordalo et al., 2020, Gáti, 2020]. On the other hand, due to data availability, evidence on firms' expectations is notably scarce [Candia et al., 2022].

as the ones proposed by Branch and McGough [2010], Gabaix [2020], Gáti [2020], and Pfäuti and Seyrich [2023]; as well as behavioral overlapping generations models as in Adam [2003]. By considering overlapping generations in a New Keynesian context, we relate, for instance, to Gali [2021]. By stating that cohorts show heterogeneity in expectations because of different experiences, we connect to papers unrelated to monetary policy but where different cohorts have different beliefs about the future, such as Collin-Dufresne et al. [2017] and Kuchler and Zafar [2019].

Our model of non-rational and heterogeneous expectations is also inspired by the diagnostic expectations literature, as introduced in Bordalo et al. [2018, 2019, 2020].¹¹ Bianchi et al. [2021] and L’Huillier et al. [2021] provide recent applications of this framework in macro monetary settings. We use that setting to incorporate our findings into a general equilibrium model in a tractable way.

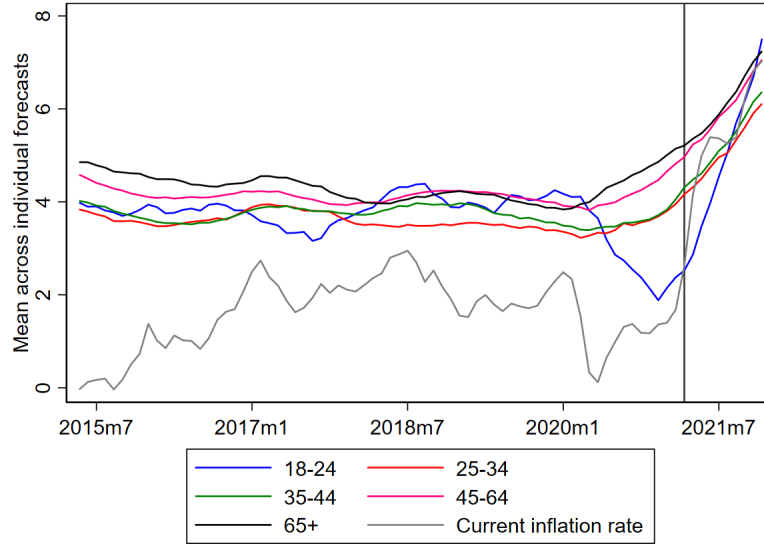
Besides diagnostic expectations, a large body of literature analyzes the implications of departing from the full information rational expectations assumption. Among the main examples, we include the set of papers related to the imperfect information approach [Mankiw and Reis, 2011], the complex systems/animal spirits/heuristic approach [De Grauwe, 2011, Jump et al., 2019], the sticky information approach [Mankiw and Reis, 2002], and the adaptive learning approach [Marcet and Sargent, 1989, Evans and Honkapohja, 2001]. However, only some of these papers have studied how heterogeneity in expectations arises and its macroeconomic consequences.

Our findings are closely related to those of Malmendier and Nagel [2016], who use adaptive learning to approximate cohorts’ heterogeneous inflation expectations. Instead, we opt for a constant forecasting revision Kalman filter model. While we also rely on constant gain, the selection of parameters in our framework is primarily data-driven. Our approach allows us to incorporate a current shared forecast component using a standard Kalman filter and a structure incorporating past inflation experiences.

Our paper shows that cohorts do not adjust their expectations differently in response to current inflation news. Our results suggest that younger cohorts implicitly put more weight on current information but adjust to the news data similarly to older cohorts. In other words, younger cohorts do not react more strongly to inflation news than older cohorts. In that sense, our modeling follows Malmendier and Nagel [2016] by incorporating a method consistent with the availability heuristic by using past experiences and allowing agents to use the new information to form expectations. Following recent evidence showing that agents form expectations based on their shopping experiences [D’Acunto et al., 2021, D’Acunto and Weber, 2022], we use a Kalman filter approach where agents get signals from current food prices to model the expectation formation process. In addition, our approach is flexible enough to be incorporated into a general equilibrium framework as

¹¹Our approach is also related to broader signal extraction/noisy information approaches such as the ones proposed in Woodford [2001] and Blanchard et al. [2013], among others.

Figure 2.1: Average 12-months-ahead inflation expectations



Note: The figure shows the 12-month moving average for the 10 percent and 90 percent trimmed mean for each cohort using the point forecast. We use population weights. Data goes from June 2013 to December 2021. Ages correspond to the interviewee's age at the time of the survey. The vertical line denotes March 2021. The current inflation rate measure is based on the monthly YoY percentage variation of the CPI.

Source: Survey of Consumer Expectations, Federal Reserve Bank of New York.

some other recent studies [Bianchi et al., 2021, L'Huillier et al., 2021].

2.3 Empirical facts

This section reviews some empirical facts related to heterogeneous inflation expectations at the household level. We show how these expectations are correlated with past experiences regarding the aggregate inflation variable. These empirical facts motivate and guide the theoretical model of the paper.

Malmendier [2021] and D'Acunto et al. [2022] document that consumers' experiences influence their inflation expectations. Thus, individual experiences are a source of expectation heterogeneity. This paper focuses on how aggregate inflation experiences influence idiosyncratic inflation expectations, as in Malmendier and Nagel [2016]. In the US economy, this heterogeneity in inflation expectations became more evident after the high-inflation episode of 2021, when inflation surged after 30 years of low and stable rates.

For this section, we use data from the Survey of Consumer Expectations (SCE) of the Federal Reserve Bank of New York. This data set is a US-wide rotating panel with information between

March 2013 and December 2021, where each respondent is surveyed for a maximum of 12 contiguous months. This data set is handy for our purposes because it provides high-frequency data on American households' inflation expectations in two different periods of the US economy. In particular, we focus our analysis on respondents' 12-months-ahead point forecast. The 12-months-ahead inflation rate is computed as the inflation rate existing between the current month and 12 months after the current month.¹²

Fact 1: Inflation expectations are heterogeneous across cohorts.

Figure 2.1 shows the mean 12-months-ahead inflation forecast by cohort. The heterogeneity across cohorts is evident. The oldest (65+) and the second oldest cohort (45-64) have higher mean inflation expectations throughout most of the sample. Those cohorts experienced a period of high inflation in the 60s, 70s, and early 80s. Regarding inflation forecast value, these cohorts are followed by intermediate cohorts (25-34 and 35-44), who experienced the stable and low inflation rates of the 90s, 00s, and 10s. Finally, the youngest cohort (18-24) shows the most volatile mean, following the current inflation rate most of the time. The mean value of this cohort notably increased after the high-inflation episode of 2021, surpassing older cohorts' expectations.

Fact 2: Inflation experiences are clustered by age.

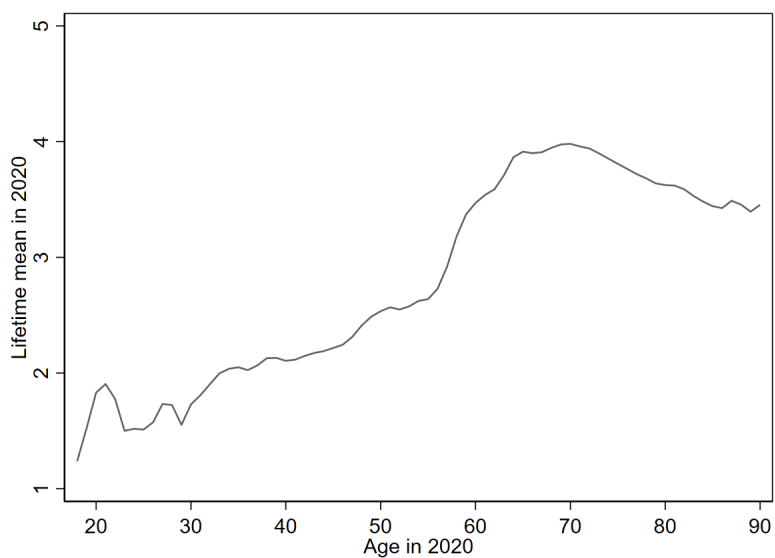
Figure 2.2 plots the average lifetime inflation rate people have experienced according to their age in the years 2020 and 2021. In the US, the average lifetime inflation rates are clustered by age.

The heterogeneity of average experienced inflation rates across cohorts results from the different inflation-related events Americans have gone through. Older cohorts have experienced events such as the Great Inflation period (1965-1982), characterized by high and persistent inflation. Thus, these cohorts have a higher lifetime average inflation rate, regardless of the year we calculate. Meanwhile, intermediate cohorts have experienced low and stable inflation rates throughout the 80s, 90s, 00s, and 10s. Therefore, they present lower values on the lifetime average inflation rate. For older and intermediate cohorts, experiencing the high-inflation episode of 2021 did not significantly affect their lifetime average inflation rate.

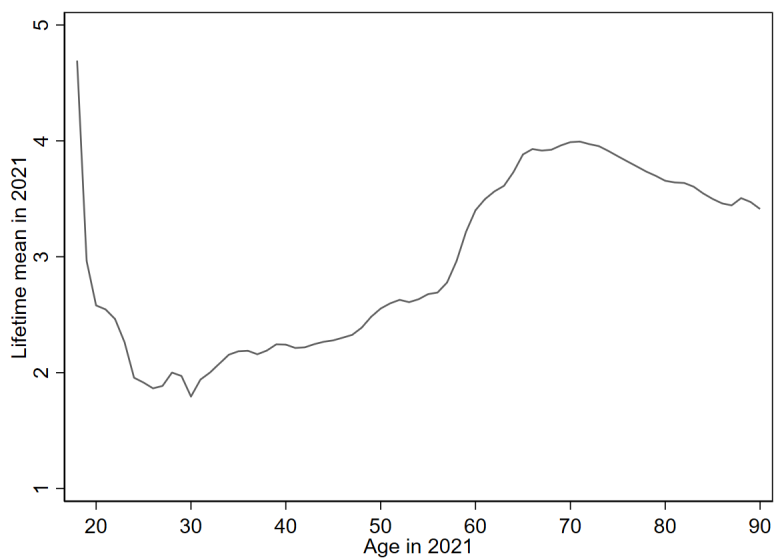
In contrast, youngest cohorts show a significant change between 2020 and 2021. Until 2020, the youngest cohorts had not experienced high inflation, showing low lifetime average inflation rates. However, after being exposed to the high-inflation episode of 2021, their lifetime average inflation rate dramatically increased.

¹²Specifically, consumers are asked to answer first the following question: "The next few questions are about inflation. Over the next 12 months, do you think that there will be inflation or deflation?" After indicating if they forecast inflation or deflation, they are asked to give a numerical answer to the following question: "What do you expect the rate of inflation/deflation to be over the next 12 months? Please give your best guess. "

Figure 2.2: Lifetime average inflation rate among respondents



(a) 2020

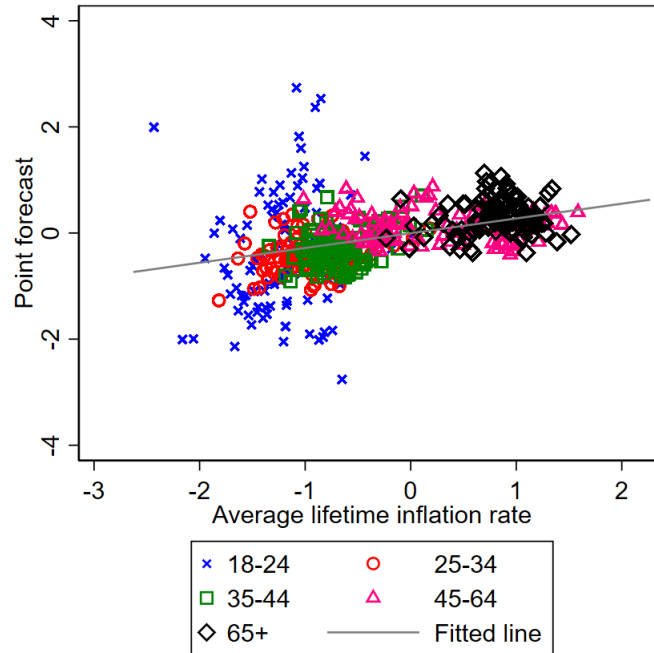


(b) 2021

Note: The figure shows the mean of the monthly YoY CPI-based inflation rate that people of the age indicated in the years 2020 and 2021 have experienced in their lifetimes, starting when they were age 18.

Source: Bureau of Labor Statistics.

Figure 2.3: Inflation point forecast and average lifetime inflation



Note: The figure shows binned scatterplot across lifetime average inflation bins. We residualized the variables by respondent gender and commuting zone. Data goes from June 2013 to December 2021. Ages correspond to the interviewee's age at the time of the survey. The average lifetime inflation rate measure is based on the monthly YoY percentage variation of the CPI.

Source: Survey of Consumer Expectations, Federal Reserve Bank of New York and Bureau of Labor Statistics.

Fact 3: A higher average lifetime inflation rate is correlated with a higher point forecast.

Tying together both previous empirical facts, Figure 2.3 shows that people who have experienced higher average inflation rates during their lifetimes, when surveyed, tend to give a higher inflation point forecast.¹³ We formally test this result in Table 2.1. Columns 3 and 4 of this table conclude that the inflation experienced significantly affects individuals' inflation expectations, even after controlling for the current environment and individual characteristics.

This fact provides empirical support for the literature on learning from past experiences [Malmendier and Nagel, 2016, Malmendier, 2021, Malmendier et al., 2021, Malmendier and Wachter, 2022] pointing to a possible source of heterogeneity in inflation expectations: past experiences with the aggregate inflation variable.

While the evidence we provide here is for the US economy, in Appendix B.7, we find similar evidence for a panel of European countries. With this European data set, we show that the pattern (i) is present beyond the US and (ii) does not arise from cohorts' systematic characteristics but because of cross-country heterogeneous inflation experiences. In the panel of countries we use, we observe different inflation histories across countries that are not necessarily similar to the

¹³We control for observable characteristics of the respondent except for the age and period variables.

US experience. Again, we find that average inflation experience positively relates to individual inflation expectations, even after including country-time fixed effects and, more importantly, cohort fixed effects. This last set of fixed effects controls for the fact that cohorts can have biases because of motives related to their age. In this sense, Hajdini et al. [2022a] find similar evidence using a survey for a panel of countries.

Fact 4: After controlling for the average lifetime inflation rate, younger cohorts do not react more strongly to inflation news than older cohorts.

We test whether younger generations react more strongly to the current economic environment after controlling for their average lifetime inflation. This exercise aims to inform the modeling approach we will take. The marginal contribution of new information for younger cohorts is higher, as they have a shorter history. Assuming constant weights implies that new information will matter more, especially in the context of persistent inflation. From a modeling perspective, we want to know whether cohorts react differently to news, conditional on their cohort-dependent marginal contribution of new information. These findings help us to discriminate between models.

We test this hypothesis through individual-level regressions presented in Table 2.1. Similar to the results of other papers using information treatment [Hajdini et al., 2022b], Column 1 of this table shows that all individuals react to current inflation events. These results also confirm the existence of a positive relationship between the inflation forecasts and average lifetime inflation rates, as we saw previously in Figure 2.3, even after considering current inflation.

To study whether there are different reactions across cohorts, we run regressions that consider interactions of current inflation with a cohort indicator variable. Column 2 of Table 2.1 shows the corresponding results. After controlling for average lifetime inflation, the interaction term has no statistically significant effect. There are no different reactions to current inflation news across cohorts. We confirm the finding by performing an F-test where the null hypothesis is that all interactions are jointly equal to zero. The test gives a p-value of 0.39, so we cannot reject the null hypothesis. In addition, this conclusion holds in a sample of European countries, as we show in Table B.7 of Appendix B.7. Together, these results suggest that the main source of heterogeneity across cohorts comes from different past experiences regarding inflation.¹⁴

2.4 A simple model with heterogeneous expectations

In this section, we propose a history-based expectations-augmented Kalman filter as the process by which agents form their inflation expectations. We begin with a simple model that provides a

¹⁴Furthermore, in Appendix B.4 we assume the structure of a standard Kalman filter and provide empirical evidence that provides support for the hypothesis of different cohorts not reacting differently to inflation signals.

Table 2.1: Effects of current and experienced inflation rates on inflation expectations

Dep. var.: Inflation expectations	(1)	(2)	(3)	(4)
Average lifetime inflation	0.332*** (0.026)	0.269*** (0.065)	0.299*** (0.025)	0.250*** (0.022)
Current inflation	0.524*** (0.016)	0.632*** (0.118)		
Cohort 25-34		-0.080 (0.310)		
Cohort 35-44		-0.131 (0.302)		
Cohort 45-64		-0.010 (0.319)		
Cohort 65+		0.062 (0.325)		
Current inflation $\times$ 25-34		-0.163 (0.128)		
Current inflation $\times$ 35-44		-0.099 (0.126)		
Current inflation $\times$ 45-64		-0.080 (0.122)		
Current inflation $\times$ 65+		-0.141 (0.126)		
Time FE	No	No	Yes	Yes
Controls	No	No	No	Yes
Observations	105,415	105,415	105,415	105,402
R-squared	0.057	0.058	0.091	0.198

Note: Table shows regressions where the dependent variable is inflation expectations according to the Survey of Consumer Expectations (SCE) of the Federal Reserve Bank of New York. Column 1 shows controls by the average lifetime inflation of respondents of a given age at each period of time and the last inflation measure. Column 2 follows (1) but adds cohort fixed effects and the interaction of those cohort fixed effects with the current inflation. Column 3 follows Column 1 but adds time fixed effects and, hence, omits the current inflation variable. Column 4 follows Column 1 but adds time fixed effects and demographic controls. The demographic controls are income, gender, Hispanic origin, race, educational level, numerical proficiency, and commuting zone. The current inflation rate measure is based on the monthly YoY percentage variation of the CPI. Standard errors clustered by age. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The dependent variable is trimmed, dropping the lower and upper 10 percent of answers in each period.

good starting point where differences in agents' personal experiences do not imply heterogeneity in expectations. Given the absence of private information, we show that the observed heterogeneity cannot arise from a standard Kalman filter. Then, by introducing a history-based expectations-augmented Kalman filter, we explain how the private inflation history distorts the expectations, producing inflation forecast heterogeneity. Moreover, we estimate the corresponding distortion parameter and close the section by comparing the heterogeneous rates of inflation expectations generated by our proposed framework and those observed in the data and presented in Figure 2.1.

2.4.1 Standard Kalman filter

2.4.1.1 Setup

The economy is populated by different cohorts indexed by i . These cohorts are heterogeneous in their dates of birth and the inflation history they have experienced. Since there is no heterogeneity within cohorts, a single representative agent summarizes the situation of each one of these groups. In a given period $t + 1$, the level of inflation π_{t+1} is defined according to the following random walk process¹⁵

$$\pi_{t+1} = \pi_t + \varepsilon_t,$$

where ε_t is a normally independent and identically distributed inflation shock. We assume that, in a given period t , agents wish to forecast the future inflation rate π_{t+1} , but they only observe a noisy signal of this variable. In other words, the agents face a standard signal extraction problem. To simplify the analysis, in a given period t , we assume that the signal s_t is defined as

$$s_t = \zeta \pi_{t+1} + v_t,$$

where the coefficient $\zeta \geq 0$ denotes the pass-through existing between the unobserved variable π_{t+1} and its corresponding signal s_t , and v_t is a signal noise. We assume that this noise is a normally independent and identically distributed variable. Moreover, we allow for a non-zero covariance between both shocks to consider the existence of some elements causing correlated movements in both the observed signal and the unobserved variable. Therefore, we have

¹⁵We opt for a random walk process instead of a first-order auto-regressive (AR(1)) specification because the data cannot reject the hypothesis that the monthly inflation rate has a unit root. We provide a more thorough discussion in Appendix B.3. Also, see Pivetta and Reis [2007] for a discussion on the very high persistence of the (quarterly) inflation rate in the US. To complete our analysis, in Appendix B.5, we show the model's results when the inflation series follows an AR(1) process. These results are very similar to those found under the random walk assumption.

$$\begin{pmatrix} \varepsilon_t \\ v_t \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon v} \\ \sigma_{\varepsilon v} & \sigma_v^2 \end{pmatrix} \right).$$

As a further simplification, we assume the model has no private information. In other words, all of the agents receive precisely the same signal. Since the agents face a standard signal extraction problem, we assume they generate inflation forecasts using this variable's conditional expected value. More precisely, given their information set in period t , the agents apply a linear Kalman filter to forecast inflation in period $t + 1$. Therefore, the inflation's predicted value is given by

$$\mathbb{E}_{i,t}^{KF} [\pi_{t+1}] = (1 - \zeta K) \mathbb{E}_{i,t-1}^{KF} [\pi_{t+1}] + K s_t, \quad (2.1)$$

where K denotes a standard Kalman gain.¹⁶ Conditional on agents' past and current signals, the Kalman filter approach allows us to characterize the forecasted distribution of the unobserved variable π_{t+1} in any period t . When the signal is perfectly revealing about the true state of the variable, we observe $\zeta = 1$, $v_t = 0$, and $\sigma_{\varepsilon v} = 0$. Therefore, we conclude $K = \zeta K = 1$ and $\mathbb{E}_{i,t}^{KF} [\pi_{t+1}] = s_t = \pi_{t+1}$. The presence of a signal noise induces $K \in [0, 1)$ even without a correlation between both error terms.

Regarding long-run values of inflation expectations, from the Kalman-based forecast equation and using the random walk structure associated with the inflation variable, we conclude that given $h \geq 1$, we must have

$$\mathbb{E}_{i,t}^{KF} [\pi_{t+h}] = \mathbb{E}_{i,t}^{KF} [\pi_{t+1}].$$

Finally, and considering the parameter $\gamma = (1 - \zeta K) \in [0, 1]$, the Kalman filter prediction can be written recursively as

$$\mathbb{E}_{i,t}^{KF} [\pi_{t+1}] = \gamma^{t+1} \mathbb{E}_{i,-1}^{KF} [\pi_0] + K \sum_{j=0}^t \gamma^{t-j} (\zeta \pi_{j+1} + v_j)$$

Therefore, using this simple version of the model, we conclude that higher values of past inflation imply a higher forecasting value of this same variable. However, agents' personal experiences are not associated with heterogeneity in expectations. According to this model, agents who lived

¹⁶As usual in the literature, this signal-to-noise ratio is defined such that it minimizes the variance of the prediction error associated with the unobserved variable, i.e., $\pi_{t+1} - \mathbb{E}_{i,t}^{KF} [\pi_{t+1}]$. The Kalman gain that solves this optimization problem depends on the covariance existing between the error associated with the observed signal and the unobserved variable and the constant $\Sigma_{t+1|t-1} = \text{Var} [\pi_{t+1} - \mathbb{E}_{i,t-1}^{KF} [\pi_{t+1}]]$. See Cheung [1993] for an example of a Kalman gain that considers these terms. Regarding the constant $\Sigma_{t+1|t-1}$, it can be shown that it must satisfy $(\Sigma_{t+1|t-1} - \sigma_\varepsilon^2) (\zeta^2 \Sigma_{t+1|t-1} + \sigma_v^2 + 2\zeta \sigma_{\varepsilon v}) - (\sigma_v^2 \Sigma_{t+1|t-1} - \sigma_{\varepsilon v}^2) = 0$. We opt for a constant Kalman gain as it is a convenient assumption for the dynamic model explored in the following sections. However, we find similar results when we use time-varying gains.

through episodes of high inflation forecast an inflation value identical to those who lived through episodes of low inflation. Given a starting point assumption where every agent observes the initial level of inflation, i.e. $\mathbb{E}_{i,-1}^{KF}[\pi_0] = \pi_0$ for every agent i , we conclude that $\mathbb{E}_{i,t}^{KF}[\pi_{t+1}] = \mathbb{E}_t^{KF}[\pi_{t+1}]$ must hold for every agent. In what follows, to simplify the analysis, we assume $\zeta = 1$.

To identify an appropriate signal for the empirical counterpart of the expectations formation model, we follow the evidence presented in D’Acunto et al. [2021], which shows that agents use their consumption experience to form expectations. More specifically, Campos et al. [2022], using the University of Michigan’s Survey of Consumers, concludes that consumers highly weigh CPI’s food components when forming inflation expectations. Dietrich [2022] finds similar evidence using different data sources. Although differences in the composition of individual expenditures could lead to differences in perceived inflation [D’Acunto et al., 2021], according to data from the US Bureau of Labor Statistics, the share of food expenditure has no relevant variation across US cohorts. This fact supports our assumption of a common signal on the underlying aggregate inflation represented by the food component of the CPI. Finally, D’Acunto and Weber [2022] show that consumers use food prices as a reference to form inflation expectations. Therefore, we use the food component of the CPI as a common inflation signal for all consumers.

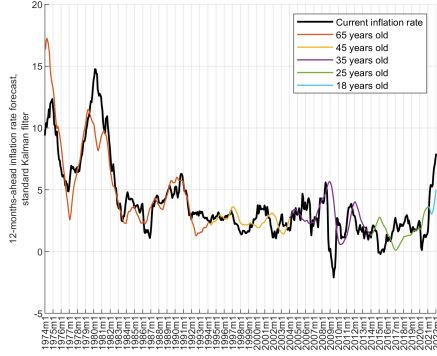
Because of timing issues, we use the lagged year-over-year percentage variation of the CPI’s food component. Consumers use past changes in food prices as a signal to forecast future aggregate inflation. For example, suppose that in December, an agent forecasts aggregate inflation for the next 12 months, and we presume that consumers make this prediction at the beginning of the month. In that case, we assume this agent considers November’s food inflation to forecast. Therefore, for the empirical counterpart of the expectations formation model, we assume $s_t = \pi_{t-1}^{food}$ where π_{t-1}^{food} denotes food inflation in period $t - 1$. D’Acunto and Weber [2022] show that this behavior is consistent with consumers’ use of food prices to form expectations, as they show that consumers weigh changes of observed reference food prices, in their case milk, to form inflation expectations. Hence, we use annual changes in the CPI’s food component for the baseline specification primarily because of data availability motives and consistency with recent evidence. In Appendix B.6, we show that our results are robust to alternative signals also consistent with the recent empirical evidence on reference prices.

Using monthly data on aggregate and food inflation, we obtain $\sigma_\varepsilon^2 = 0.15$, $\sigma_v^2 = 4.09$ and $\sigma_{\varepsilon v} = -0.03$, concluding $K = 0.1751$.¹⁷

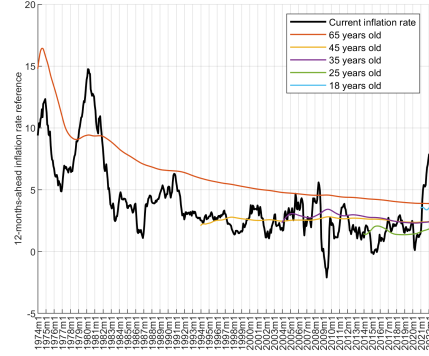
We now perform a forecasting exercise using monthly inflation data and distinguishing agents by cohorts. Given the recursive structure of the Kalman filter and to initialize the cohort-specific forecasting process, we assume that in the period in which the cohort representative agent reaches

¹⁷We detail how we obtain this calibration in Appendix B.3, while Appendix B.4 provides empirical support for the Kalman gain being constant across cohorts.

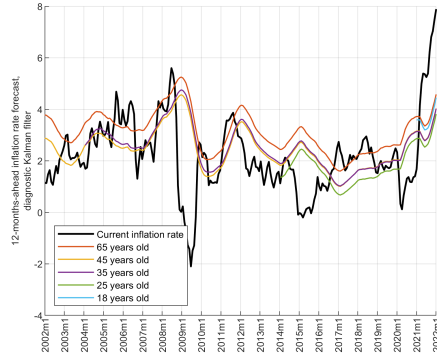
Figure 2.4: History-based Kalman-filter-based inflation forecasts by cohort



(a) Standard Kalman-filter-based inflation forecasts



(b) Inflation rate reference



(c) History-based Kalman-filter-based inflation forecasts

Note: Panel (a) shows the Kalman filter forecast for the common component for selected cohorts, differentiated by their age in 2021. Panel (b) shows the references for selected cohorts obtained according to the Kalman filter and given the history of inflation experienced by the corresponding age group. Panel (c) shows forecasts for selected cohorts according to the Kalman-filter-augmented expectations and considering the estimate for θ from Column 1 of Table 2.2. Selected cohorts are differentiated by their age in 2021. We further assume that each cohort starts forecasting when they become 18 years old. The current inflation rate measure is based on the monthly YoY percentage variation of the CPI.

adulthood and begins forecasting, she uses the previous period's Kalman filter expected value as a starting point. We denote the period when agent i starts forecasting as period k_i . Given the starting point assumption where the initial level of inflation is common knowledge, we have $\mathbb{E}_{i,k_i-1}^{KF}[\pi_{k_i}] = \mathbb{E}_{k_i-1}^{KF}[\pi_{k_i}]$ for every agent i . Panel (a) of Figure 2.4 presents the 12-month-ahead inflation forecasts of different cohorts produced according to the standard Kalman filter. This figure plots the actual inflation rate and the forecast made by some selected cohorts. The standard Kalman filter cannot generate the heterogeneous pattern in expectations observed in the data (i.e., Figure 2.1). Under this framework, inflation expectations evolve following an identical process across cohorts. Therefore, we need to move to a more sophisticated framework to replicate the facts observed in the data.

2.4.2 History-based Kalman filter

2.4.2.1 Setup

In this section, we depart from the standard Kalman filter framework. We adopt a model of non-Bayesian beliefs where consumers revise their forecast using a reference based on their experienced inflation history.¹⁸ In this framework, consumers receive signals and compare them to their history of forecasts for the next period. By making that comparison when forecasting, consumers can put more weight on the new information they receive or on their past history-specific information. Our framework allows for both scenarios.

When consumers put more weight on the likelihood of new information for forecasting, our setting would be as in Bordalo et al. [2019], related with the representativeness heuristic. When consumers put more weight on their past history-specific information for forecasting, it would be associated with the availability heuristic [Tversky and Kahneman, 1973]. That is, when forming expectations, consumers use their available information related in this case to their history, instead of sampling the complex current information that is needed to form future expectations about the economy. This version of the model can also be related to the conservatism bias [Edwards, 1968], as consumers put less weight on the signal when forming their beliefs.

This setting aims to model the source of the heterogeneity in inflation expectations, not to make a general statement about expectations formation processes. This source of heterogeneity is relevant, consistent across surveys of expectations, and it complements other findings on inflation expectation formation. The flexibility of this framework allows us to connected to our empirical setting and to learn about how people use their history of inflation when they form expectations.

We denote the true conditional distribution of the unknown inflation variable in a given period t as $f(\pi_{t+1} | \mathcal{I}_t)$, where the term $\mathcal{I}_t$ denotes the common set of information available in that period. Given this definition, we assume that the inflation's belief distribution for agent i is given by

$$f_{i,t}^\theta(\pi_{t+1}) = f(\pi_{t+1} | \mathcal{I}_t) D_{i,t}^\theta(\pi_{t+1}) Z_{i,t}, \quad (2.2)$$

where

$$D_{i,t}^\theta(\pi_{t+1}) = \left[\frac{f(\pi_{t+1} | \mathcal{I}_t)}{f(\pi_{t+1} | \mathcal{I}_{i,t}^{ref})^{\frac{1}{(t-k_i)}}} \right]^\theta.$$

¹⁸We can think of our framework as a generalization of the diagnostic expectations framework. The generalization comes due to the fact that we do not restrict the parameter θ to be strictly positive. Thus, in terminology of the diagnostic expectations, we may find overreaction ($\theta > 0$), underreaction ($\theta < 0$), or neither ($\theta = 0$). A similar generalization of diagnostic expectations theory is discussed in Liao [2023].

In this setup, the parameter $\theta \in \mathbb{R}$ governs the level of distortion that the likelihood ratio $[D_{i,t}^\theta(\pi_{t+1})]^\frac{1}{\theta}$ introduces into agents' beliefs. The normalizing parameter $Z_{i,t}^{-1} = \int f_{i,t}^\theta(\pi_{t+1}) d\pi_{t+1}$ ensures that the distribution $f_{i,t}^\theta(\pi_{t+1})$ integrates to one in every period t and for every agent i . Each agent compares the current common information set $\mathcal{I}_t$ against her referential information set $\mathcal{I}_{i,t}^{ref}$. Later, we assume that this idiosyncratic referential information set relates to each agent's past inflation experiences. As mentioned above, the parameter θ captures the level of distortion associated with the model. In the standard Kalman filter framework, we have $\theta = 0$, which implies $D_{i,t}^\theta(\pi_{t+1}) = 1$. When this is the case, there is no distortion in beliefs, and $f_{i,t}^\theta(\pi_{t+1}) = f(\pi_{t+1} | \mathcal{I}_t)$. When $\theta \neq 0$, beliefs are distorted, as we discuss below.

Notice that the no private information assumption implies that the information associated with the true conditional distribution is equal for everyone. Therefore, under the proposed signal extraction framework, in any period t , the expected value associated with the true conditional distribution of the unknown inflation variable is common to every agent and corresponds to $\mathbb{E}_t^{KF}[\pi_{t+1}]$. Given the normality assumption on the error term of the inflation process, the true conditional distribution satisfies $f(\pi_{t+1} | \mathcal{I}_t) \sim \mathcal{N}(\mathbb{E}_t^{KF}[\pi_{t+1}], \sigma_\pi^2)$. As we explain below, this normality result implies that the distribution associated with the referential information set $\mathcal{I}_{i,t}^{ref}$ is normal too. Given both of these results, we can show that agent i 's distribution satisfies $f_{i,t}^\theta(\pi_{t+1}) \sim \mathcal{N}(\mathbb{E}_{i,t}^\theta[\pi_{t+1}], \sigma_\pi^2)$, where the mean value has the following linear structure

$$\mathbb{E}_{i,t}^\theta[\pi_{t+1}] = \mathbb{E}_t^{KF}[\pi_{t+1}] + \theta \left(\mathbb{E}_t^{KF}[\pi_{t+1}] - \mathbb{E}_{i,t}^{ref}[\pi_{t+1}] \right), \quad (2.3)$$

where $\mathbb{E}_{i,t}^\theta[\pi_{t+1}]$ is the distorted forecast associated with the belief distribution $f_{i,t}^\theta(\pi_{t+1})$, and $\mathbb{E}_{i,t}^{ref}[\pi_{t+1}]$ is the expected value obtained according to the distribution associated with the referential information set $\mathcal{I}_{i,t}^{ref}$.¹⁹ We define this linear composition of the standard Kalman filter as our history-based-augmented Kalman filter.

Up to this point, our variables are history-independent. Remember that the standard Kalman filter is Markovian since it only requires the previous period's beliefs to be computed. However, this simple filter does not reproduce our set of empirical facts. We now introduce a role for past experiences through the reference term $\mathbb{E}_{i,t}^{ref}[\pi_{t+1}]$. For the cohort i 's representative agent, we define the reference term as

$$\mathbb{E}_{i,t}^{ref}[\pi_{t+1}] = \frac{\sum_{j=1}^{t-k_i} \mathbb{E}_{i,t-j}^{KF}[\pi_{t+1}]}{(t - k_i)}, \quad (2.4)$$

where k_i denotes the period in which cohort i starts forecasting. Notice that the reference term contains all past expectations of this cohort-specific agent about future inflation. Since the reference term uses food prices as a signal, it contains the history of how consumers of a given

¹⁹See Appendix B.2 for a more detailed derivation.

cohort used the food prices they were exposed to forecast aggregate inflation. In Panel (b) of Figure 2.4, we show how the inflation rate reference defined according to Equation 2.4 evolves for some selected cohorts. Older cohorts, which have gone through episodes of higher inflation during their lifetimes, have higher reference points when compared to younger cohorts that have not experienced relevant inflationary episodes.

Now, we turn the discussion into the parameter θ . When $\theta > 0$, agents “overreact” to the current information they receive relative to their prior beliefs. For instance, if agents receive news about inflation being different from what they expected in the past, they will overreact to this news by adding more weight to current information, under-weighting their inflation history. If $\theta < 0$, agents “underreact” to the information just received, placing more weight on their references relative to current information. In this case, if agents observe inflation values different than their priors, they will tend to anchor their current expectations to their past beliefs. As discussed in Section 2.3, recent economic developments show that older consumers had higher inflation expectations when aggregate inflation was low, such as the period between 2010 and 2020. Meanwhile, when inflation was high in 2021, consumers reacted moderately, not increasing their expectations significantly. This idea goes in line with agents anchoring their expectations to their experience or, in our framework, with a parameter satisfying $\theta < 0$. Lastly, when $\theta = 0$, there are no distortions in beliefs and $\mathbb{E}_{i,t}^\theta [\pi_{t+1}] = \mathbb{E}_t^{KF} [\pi_{t+1}]$ for every cohort i .

Given the random walk structure assumed for the inflation variable, and considering $h \geq 1$, we conclude $\mathbb{E}_t^{KF} [\pi_{t+h}] = \mathbb{E}_t^{KF} [\pi_{t+1}]$. Since the reference term is a linear composition of expected values associated with the true conditional distribution, we must have $\mathbb{E}_{i,t}^{ref} [\pi_{t+h}] = \mathbb{E}_{i,t}^{ref} [\pi_{t+1}]$. Therefore, when $h \geq 1$, we observe $\mathbb{E}_{i,t}^\theta [\pi_{t+h}] = \mathbb{E}_{i,t}^\theta [\pi_{t+1}]$.

2.4.2.2 Estimation and forecasting exercise

In this section, we estimate the parameter θ directly from the data. Under our current assumptions, from Equation 2.3 we know that the cohort i ’s forecast for period $t + 12$ corresponds to

$$\mathbb{E}_{i,t}^\theta [\pi_{t+12}] = (1 + \theta) \mathbb{E}_t^{KF} [\pi_{t+12}] - \theta \mathbb{E}_{i,t}^{ref} [\pi_{t+12}]. \quad (2.5)$$

Since the first term is common across all cohorts, it can be captured by a time-fixed effect γ_t . Regarding the distorted expectation, we use data on the 12-month-ahead forecast of each agent m from cohort i presented in the SCE survey. We denote this variable as $\mathbb{E}_{m,i,t}^{SCE} [\pi_{t+12}]$. Finally, to compute $\mathbb{E}_{i,t}^{ref} [\pi_{t+12}] = (t - k_i)^{-1} \sum_{j=1}^{k_i} \mathbb{E}_{i,t-j}^{KF} [\pi_{t+12}]$ we return to the standard Kalman filter from Section 2.4.1 and recover the different terms $\mathbb{E}_{i,t-j}^{KF} [\pi_{t+12}]$, which are optimal forecasts under the given setup. Given this setting, we run the following regression

$$\mathbb{E}_{m,i,t}^{SCE} [\pi_{t+12}] = \gamma_t + \varphi \mathbb{E}_{i,t}^{ref} [\pi_{t+12}] + \varepsilon_{m,i,t}. \quad (2.6)$$

We present the results in Table 2.2. Column 1 shows our main specification, from which we obtain $\hat{\theta} = -\hat{\varphi} = -0.374$. Because the estimated parameter satisfies $\hat{\theta} < 0$, we conclude that when agents forecast, they positively weigh their reference sets (i.e., their past inflation history, their priors coming from their experiences with food prices). In the remaining columns of Table 2.2, we show that the negative coefficient is robust to the inclusion of control variables.²⁰

This result contrasts with the diagnostic expectations literature (for instance, Bordalo et al. 2020), where a positive θ is the usual result using a similar empirical structure. However, this positive θ result is based on a history-based expectations formation process with a different reference point. Instead of using agents' past inflation history, the diagnostic expectations literature usually uses the one-period lagged information set as their reference point. Moreover, their finding is also conceptually different from ours, as most of the previous empirical literature on diagnostic expectations relies on surveys of professional forecasters, who are better informed about the economy than the households surveyed in the SCE.²¹

The value of the parameter θ is also conditional on the expectation formation process. In Bordalo et al. [2019, 2020], current news are processed with a Kalman filter, so, in a sense, current forecasts are partly processed with past information. With that modeling approach, surprises with respect to a reference based on the previous period also come from a reference based on the available individual experience, as suggested by Tversky and Kahneman [1973].

However, this result changes when we consider other forms for current expectations instead of a Kalman filter, such as rational expectations (for instance, Bianchi et al. 2021 and L'Huillier et al. 2021). In that case, agents know the model's features, so their reference is based on everything contained within the model, which is considered more likely (as in Kahneman and Tversky 1972). Given the assumption of full information, agents will overreact to any shock that was not expected in the previous period; as agents understand the shock and its effects as soon as they see it, it is immediately more likely.²²

We take a different approach, as we explicitly model the reference of the agents given their cohort. Agents also optimally generate a forecast with a Kalman filter and a signal we provide. Then, they combine this optimal forecast with their past experience to produce their final forecasts.

²⁰In Figure B.2 of Appendix B.1 we show the results of estimating Equation 2.6 but leaving one cohort out of the sample in successive regressions. We find that the coefficient of interest remains stable throughout the successive estimations and, thus, does not depend on the inclusion or exclusion of any particular cohort in the sample.

²¹While it has not been central to the previous discussion in the broader diagnostic expectations literature, the result of $\theta < 0$ is not necessarily a rare sight. For instance, Bordalo et al. [2020] find $\theta < 0$ for the forecasts of a subset of variables in the Survey of Professional Forecasters. For a further discussion see Liao [2023].

²²See Benjamin [2019] for a discussion of these assumptions

Table 2.2: Parameter estimation

	(1)	(2)	(3)	(4)
$\mathbb{E}_{i,t}^{ref} [\pi_{t+12}]$	0.374*** (0.041)	0.384*** (0.034)	0.268*** (0.036)	0.210*** (0.032)
Time FE	Yes	Yes	Yes	Yes
Controls	No	Gender, commuting zone	Gender, commuting zone, HH income	Gender, commuting zone, HH income, educational degree
Observations	83,991	83,969	83,969	83,969
R-squared	0.100	0.168	0.194	0.211

Note: Table shows results of Regression 2.6. $\mathbb{E}_{i,t}^{ref} [\pi_{t+12}]$ is the reference constructed for a respondent of age i as explained in the main text. We use population weights. Column 1 has only a time fixed effect as an additional control. Columns (2), (3) and (4) add different levels of controls. Robust standard errors in parentheses. Standard errors clustered by age. Dependent variable trimmed at 10 percent and 90 percent in each period. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Our framework allows for a common component that might come from the news [Carroll, 2003], common price experiences [D’Acunto et al., 2021], among other reasons. While the time fixed effect would capture any of those factors, we model it with a common adjustment to their price experience. Then, we explicitly model the reference given their history. Our findings are consistent with consumers using an availability heuristic to form expectations. In this case people use their history of forecasts to positively weigh them given the current developments in prices. When events deviate from what an agent has experienced, they tend to revise their expectation toward their historical forecasts. Our empirical results strongly point in that direction. This finding is consistent with consumers’ survey-based evidence that shows that consumers, after receiving news about the economy, do not overweigh those news, keeping a positive weight on their prior or reference, as shown, for example, in Weber et al. [2023].

Using our estimate for the parameter, we perform a forecasting exercise using the Kalman filter. Panel (c) of Figure 2.4 shows the 12-months-ahead inflation forecasts for different cohorts. We can see that the history-based Kalman filter can generate a heterogeneous pattern across cohorts, similar to the one we saw in the data in Figure 2.1.²³ First, older cohorts show higher inflation expectations than the rest of the cohorts throughout most of the sample, based on their experiences of high inflation in the 60s, 70s, and early 80s. Second, the intermediate cohorts show lower inflation expectations than the other cohorts because they went through the stable and low inflation

²³We show the forecasts for the full sample in Figure B.1 in Appendix B.1.

rates of the 90s, 00s, and 10s. The youngest cohort shows the highest inflation expectations after being exposed to the high inflation rates of 2021.

To study the external validity of our results, in Appendix B.7, we repeat the exercises but using data from the Consumer Expectations Survey of the European Central Bank, which contains observations for six countries: Belgium, France, Germany, Italy, the Netherlands, and Spain. We find evidence that corroborates our main findings. The inflation expectations are also heterogeneous in Europe and partially explained by the history of consumers' inflation experiences. We find support for the use of a history-based Kalman filter as a way of modeling heterogeneous inflation expectations.

Moreover, given the cross-country panel structure, we control for a common cohort fixed effect in this exercise, as in Hajdini et al. [2022a]. This is a highly relevant step since common cohort characteristics (i.e. different patterns of inflation exposure over the life cycle) could affect our results. By controlling for cohort fixed effects, we rule out those common cohort characteristics and cleanly exploit differences in the inflation experienced by the different cohorts across different countries. We find similar results after adding our set of controls, implying that the heterogeneity does not stem from common cohort characteristics but from different inflation experiences in different countries. Our findings are also valid for Europe.

2.4.2.3 Goodness of fit

This subsection studies how the time and individual variation of our history-based Kalman filter expectations compare to the data. We run a regression where the dependent variable is the survey individual inflation expectation, while the right-hand side variable is our history-based Kalman filter expectation. Our formulation has two components: one comes from the standard Kalman filter that uses common food inflation data as a signal, and the second arises from cohorts references and a parameter θ estimated using survey data. As we use a time fixed effect for the parameter estimation, individual survey data does not provide the time variation observed in our inflation expectations measure. Column 1 of Table 2.3 shows that the slope between the survey forecast and the forecast produced by our history-based Kalman filter is 1.069. This result confirms that our inflation expectations formulation effectively reproduces consumers' expectations. After considering the time and cross-section variation, our estimates provide a good prediction of heterogeneous inflation expectations.

An alternative measure to explain the heterogeneity in survey data inflation expectations is the cohort's lifetime average inflation rate $\overline{\pi}_{i,t} = (t - k_i + 1)^{-1} \sum_{j=0}^{k_i} \pi_{t-j}$. Column 2 of Table 2.3 shows that the history of inflation by cohort can also predict part of the variation in the data. Also, we can consider the lagged inflation rate $\pi_{i,t-1}$ as a predictor of inflation expectations, akin to following an adaptive expectations process. Column 3 of Table 2.3 presents that the lagged

Table 2.3: Goodness of fit

	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{E}_{i,t}^{\theta} [\pi_{t+12}]$	1.069*** (0.056)			1.068*** (0.058)	0.931*** (0.048)	0.871*** (0.050)
$\overline{\pi}_{i,t}$		0.310*** (0.038)		0.003 (0.040)		0.109*** (0.036)
$\pi_{i,t-1}$			0.535*** (0.026)		0.470*** (0.024)	0.479*** (0.024)
Observations	83,991	83,991	83,991	83,991	83,991	83,991
R-squared	0.041	0.006	0.043	0.041	0.074	0.075

Note: Table shows results of a regression where the dependent variable is consumers' inflation expectations according to the Survey of Consumer Expectations (SCE) of the Federal Reserve Bank of New York. We use population weights. $\mathbb{E}_{i,t}^{\theta} [\pi_{t+12}]$ is our estimated measure of inflation expectations. $\overline{\pi}_{i,t}$ is average inflation expectations. $\pi_{i,t-1}$ is the lagged inflation rate. Standard errors clustered at the date-of-birth level in parentheses. Dependent variable trimmed at 10 percent and 90 percent in each period.

*** p<0.01, ** p<0.05, * p<0.1.

inflation also has predictive power when it comes to inflation expectations.

Furthermore, we can make the history-based Kalman filter compete with the other two measures and see which one better predicts the forecasts observed in the data. Column 4 of Table 2.3 shows that in a horserace with the lifetime average inflation rate, our model forecast is superior for explaining the observed heterogeneity. The coefficient for our measure is close to one and statistically significant. In contrast, the coefficient for the history of inflation by cohort approaches zero in magnitude and becomes statistically insignificant.

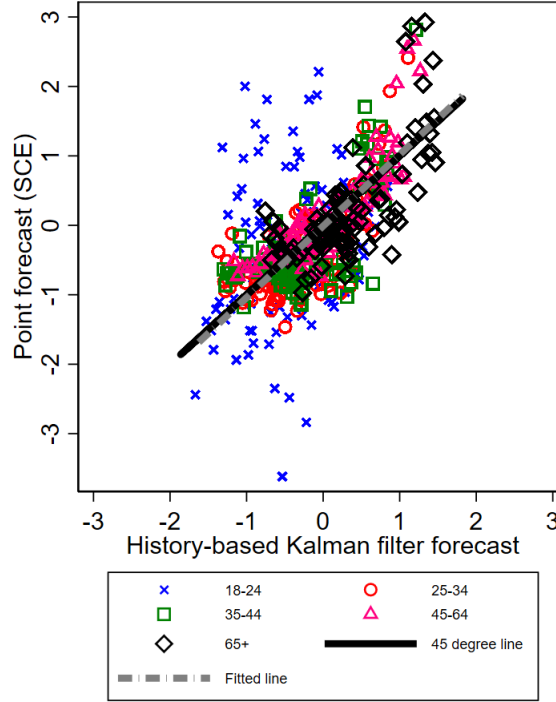
Column 5 of Table 2.3 performs a similar exercise against the lagged inflation rate. While both measures are statistically significant, the history-based Kalman filter has a larger coefficient. Given that both variables have the same scale, we find that the history-based Kalman filter adds more information relative to the lag when it comes to predicting observed inflation expectations.

Finally, Column 6 of Table 2.3 shows the comparison between the three measures. The history-based Kalman filter has a larger coefficient in comparison to the other measures, which confirms the previous two results.

Figure 2.5 visually presents the results of Column 1 in Table 2.3. The slope associated with the regression and a 45-degree line are very close. Our measure effectively models the time and cross-sectional variation of consumers' inflation expectations.

Overall, our measure shows a very good fit with the data. We can replicate heterogeneous inflation expectations at the individual level using a relatively simple model of expectations. We can effectively replicate the time and cross-sectional variation on individual survey data inflation expectations by constructing a simple model based on current inflation and heterogeneous histories.

Figure 2.5: Observed inflation forecasts and history-based Kalman filter forecasts



Note: Figure shows binned scatterplot across history-based Kalman filter forecasts (x-axis) and point forecast inflation expectations according to the Survey of Consumer Expectations (SCE) of the Federal Reserve Bank of New York (y-axis). Variables demeaned by the intercept. Data go from June 2013 to December 2021. SCE variable trimmed at 10 percent and 90 percent in each period.

2.5 Aggregate implications of heterogeneous expectations

This section presents an overlapping generations monetary model based on an expectation formation process that replicates the heterogeneity observed in inflation expectations data (i.e., Figure 2.1). We assume that agents follow the history-based Kalman filter and the assumptions introduced in Section 2.4.2. The long history of references inherent to this approach allows for different past inflation experiences to shape different current inflation expectations across cohorts.

2.5.1 Households

On the demand side, we assume that an infinite number of cohorts populates the economy. Every cohort comprises a continuum of households whose situation is summarized by a single cohort-specific representative agent. These cohorts are heterogeneous in their date of birth and past inflation experiences. We follow the perpetual youth approach presented in Blanchard [1985] and Yaari [1965] to model the different cohorts. This approach considers that households have uncertainty regarding the date on which they will die. They only know they face a mortality rate of $\lambda \in (0, 1)$ every period. At the same time, a new cohort of size λ is born every period. Therefore, in a given

period t , the size of a cohort born in period $k \leq t$ corresponds to $\lambda(1 - \lambda)^{t-k}$.

We assume households form their expectations using the history-based Kalman filter presented in Equation 2.3, and the assumptions from Section 2.4.2 hold. Agents assume that the output gap and the inflation variable are random walk processes. As in Gutiérrez-Daza [2022], agents cannot directly observe these variables; instead, they observe signals on each of these.²⁴ Under these assumptions, the representative household of cohort i solves

$$\max_{\{C_{i,j}, L_{i,j}, B_{i,j+1}\}_{j \geq t}} \left(\frac{C_{i,t}^{1-\sigma}}{1-\sigma} - \frac{L_{i,t}^{1+\eta}}{1+\eta} \right) + \mathbb{E}_{i,t}^\theta \left[\sum_{j>t} [\beta(1-\lambda)]^{j-t} \left(\frac{C_{i,j}^{1-\sigma}}{1-\sigma} - \frac{L_{i,j}^{1+\eta}}{1+\eta} \right) \right],$$

subject to

$$P_t C_{i,t} + (1 - \lambda) \frac{B_{i,t+1}}{(1 + i_t)} = W_t L_{i,t} + B_{i,t} + T_{i,t},$$

where $C_{i,t}$ is consumption, $L_{i,t}$ is the labor supply, $B_{i,t}$ are nominal savings in a risk-less bond, P_t is the aggregate price level, W_t are the nominal wages, $T_{i,t}$ are transfers, and i_t is the nominal interest rate associated with this economy's bond. Also, β is the discount factor, σ is the intertemporal elasticity of substitution, and η is the inverse of the Frisch elasticity.

The transfers $T_{i,t}$ are crucial to our model, as they incorporate two different mechanisms. First, as in Blanchard [1985] and Yaari [1965], we assume that households insure themselves to receive a flow of income every period they are alive. Then, when they die, the insurance company takes away any residual wealth. Thus, there are no accidental bequests. Second, as in Mankiw and Reis [2006], we assume that the income households receive each period from the insurance company is such that households start each period with the same wealth and that the nominal savings market clears. Therefore, we do not have to worry about wealth distribution. Lastly, and as a way of closing the model, the transfers also incorporate the benefits of firms producing intermediate goods.

Additionally, the transversality condition of the problem is

$$\lim_{T \rightarrow \infty} \frac{(1 - \lambda)^T}{\prod_{h=0}^{T-1} (1 + i_{t+h})} B_{i,t+T} = 0.$$

We introduce expectations in a general equilibrium setting as in Bianchi et al. [2021] and L'Huillier et al. [2021]. The expectations operator used by households is $\mathbb{E}_{i,t}^\theta[\cdot]$, which works under the assumptions made in Section 2.4.2. Therefore, the first-order conditions of this problem are

²⁴For simplicity, we assume that these signals correspond to the lagged value of each of the forecasted variables.

$$(P_t C_{i,t}^\sigma)^{-1} = \beta (1 + i_t) \mathbb{E}_{i,t}^\theta \left[(P_{t+1} C_{i,t+1}^\sigma)^{-1} \right], \quad (2.7)$$

$$C_{i,t}^\sigma L_{i,t}^\eta = \frac{W_t}{P_t}, \quad (2.8)$$

where the first equation corresponds to our Euler equation, and the second denotes a standard labor supply condition.

To simplify the problem, we further assume that the agents know the current state of the economy, and for any log-linearized variable x_t we have $\mathbb{E}_{i,t}^\theta [x_{t-h}] = x_{t-h}$ when $h \geq 0$. Otherwise, as L'Huillier et al. [2021] explains, agents will over or underreact when predicting this variable's present or past values. The log-linearization of Equation 2.7, following the history-based Kalman filter presented in Equation 2.3, results in

$$c_{i,t} = (1 + \theta) \mathbb{E}_t^{KF} [c_{i,t+1}] - \frac{1}{\sigma} (i_t - (1 + \theta) \mathbb{E}_t^{KF} [\pi_{t+1}]) - \theta \left(\mathbb{E}_{i,t}^{ref} [c_{i,t+1}] + \frac{1}{\sigma} \mathbb{E}_{i,t}^{ref} [\pi_{t+1}] \right), \quad (2.9)$$

where lower cases denote deviations from the steady state.²⁵ This is the IS curve associated with a given cohort i .

2.5.1.1 Aggregation

Since the steady-state value is the same across the different cohorts, the aggregate consumption gap c_t corresponds to the weighted sum of all the different cohort-level consumption gaps. Therefore, we establish

$$c_t = \lambda \sum_{k \geq 0} (1 - \lambda)^k c_{k,t}. \quad (2.10)$$

We further assume that household k , when forecasting its future individual consumption gap, believes that all the other households will behave similarly such that $\mathbb{E}_{k,t}^{ref} [c_{k,t+1}] = \mathbb{E}_{k,t}^{ref} [c_{t+1}]$. Incorporating Equation 2.9 into Equation 2.10 and using the market clearing condition in the goods market, we conclude:

²⁵ An intermediate step in the log-linearization of Equation 2.7 results in $c_{i,t} = \mathbb{E}_{i,t}^\theta [c_{i,t+1}] - \frac{1}{\sigma} (i_t - \mathbb{E}_{i,t}^\theta [\pi_{t+1}]) + \frac{1}{\sigma} \sum_{j=0}^{k_i+1} (\mathbb{E}_{i,t}^\theta [\pi_{t-j}] - \pi_{t-j})$, where the last term results from the fact that $\mathbb{E}_{i,t}^\theta [X_{t+1} Z_t] \neq \mathbb{E}_{i,t}^\theta [X_{t+1}] Z_t$ [L'Huillier et al., 2021]. The term $\mathbb{E}_{i,t}^\theta [\pi_{t-j}]$ implies that agents also have distorted beliefs about the past. We assume $\mathbb{E}_{i,t}^\theta [\pi_{t-j}] = \pi_{t-j}$ for $j \geq 0$. That is, the beliefs about the past align with the observed variables. We follow this assumption because we focus on how the reference distortions affect beliefs about the future, not the past. Importantly, this assumption also makes the solution of the model computationally simpler, as we do not have to keep track of the beliefs of each cohort for each past period. Therefore, we drop the last term and obtain Equation 2.9.

$$y_t = (1 + \theta) \mathbb{E}_t^{KF} [y_{t+1}] - \frac{1}{\sigma} (i_t - (1 + \theta) \mathbb{E}_t^{KF} [\pi_{t+1}]) - \theta \lambda \sum_{k \geq 0} (1 - \lambda)^k \left(\mathbb{E}_{k,t}^{ref} [y_{t+1}] + \frac{1}{\sigma} \mathbb{E}_{k,t}^{ref} [\pi_{t+1}] \right). \quad (2.11)$$

Equation 2.11 is the IS curve associated with our model. It equals the standard IS curve plus two different distortions associated with parameter θ . In this version of the IS curve, the past matters in the sense that current realizations are affected by the history of inflation of the cohorts.²⁶

2.5.2 Firms

On the supply side, we assume that a final good producer operates in a perfectly competitive market. This firm produces using a continuum of intermediate goods as inputs. Therefore, there is a continuum of intermediate goods producers, each one operating under monopolistic competition. These intermediate goods producers are subject to Calvo pricing frictions. We assume these firms follow rational expectations when setting their prices because they are model-consistent. Thus, we adopt the standard firms' framework associated with the New Keynesian setting to obtain the standard New Keynesian Phillips curve.²⁷

2.5.3 Monetary policy

The central bank sets the interest rate following a standard Taylor rule. Then, we have

$$\frac{(1 + i_t)}{(1 + \bar{i})} = \left[\frac{(1 + \pi_t)}{(1 + \bar{\pi})} \right]^{\chi_\pi} \left[\frac{Y_t}{\bar{Y}} \right]^{\chi_y}, \quad (2.12)$$

where the bars denote steady state values and χ_π and χ_y represent the central bank's reaction to deviations from the steady state of the inflation rate and output, respectively.

After log-linearizing, the model is summarized by

$$y_t = (1 + \theta) \mathbb{E}_t^{KF} [y_{t+1}] - \frac{1}{\sigma} (i_t - (1 + \theta) \mathbb{E}_t^{KF} [\pi_{t+1}]) - \theta \lambda \sum_{k \geq 0} (1 - \lambda)^k \left(\mathbb{E}_{k,t}^{ref} [y_{t+1}] + \frac{1}{\sigma} \mathbb{E}_{k,t}^{ref} [\pi_{t+1}] \right) + u_t^{taste}, \quad (2.13)$$

$$\pi_t = \frac{(1 - \phi)(1 - \phi\beta)}{\phi} (\sigma + \eta) (y_t + u_t^{cost}) + \beta \mathbb{E}_t [\pi_{t+1}], \quad (2.14)$$

$$i_t = \chi_\pi \pi_t + \chi_y y_t, \quad (2.15)$$

²⁶Notice that if agents do not over or underreact when forming expectations, i.e. $\theta = 0$, the IS curve reduces to $y_t = \mathbb{E}_t^{KF} [y_{t+1}] - \frac{1}{\sigma} (i_t - \mathbb{E}_t^{KF} [\pi_{t+1}])$.

²⁷We present the details associated with the supply side of the economy in Appendix B.8.

where Equation 2.13 is the dynamic IS curve augmented with a taste shock, Equation 2.14 is the Phillips curve, and Equation 2.15 is the Taylor rule. Notice the Phillips curve follows the rational expectations operator $E_t[\cdot]$, while the IS curve results from following the history-based Kalman filter operator $E_t^\theta[\cdot]$.

We consider a cost shock u_t^{cost} and a taste shock u_t^{taste} that behave as an AR(1) process, such that

$$u_t^{cost} = \rho_{cost} u_{t-1}^{cost} + \epsilon_t^{cost}, \quad (2.16)$$

$$u_t^{taste} = \rho_{taste} u_{t-1}^{taste} + \epsilon_t^{taste}, \quad (2.17)$$

where ρ_{cost} and ρ_{taste} are the persistence parameter and ϵ_t^{cost} and ϵ_t^{taste} are the unexpected innovations.

2.5.4 Calibration

The model is calibrated to a monthly frequency. The parameters from Table B.1 in Appendix B.1 show a fairly standard calibration. We calibrate the price stickiness parameter ϕ so that the expected duration of a given price quote is 12 months. We also calibrate the mortality rate λ so that the expected life span is 80 years.²⁸

Regarding the history-based Kalman filter, we need to calibrate the Kalman gain K and the parameter θ . We calibrate both according to the results from Sections 2.4.1 and 2.4.2. We must make an additional assumption around these two parameters. While we only used inflation rate data in the previous sections, we assume these parameters also hold for the output gap.

2.5.5 Simulations

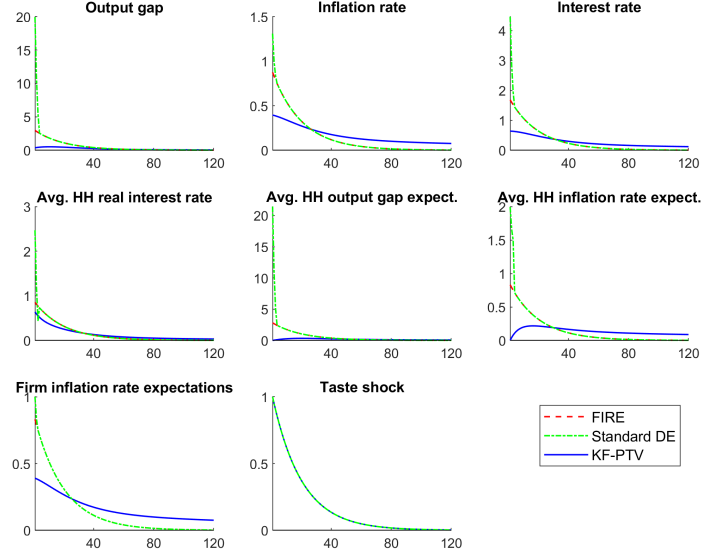
In this section we compare three different cases: (i) households form their expectations according to full information rational expectations (FIRE), (ii) households form their expectations according to a standard diagnostic expectations operator with overreaction,²⁹ and (iii) households form their expectations according to the history-based Kalman filter from Equation 2.3.³⁰

²⁸Because we assume that agents become economically active and relevant at age 18, this means agents expect to consume and work for 62 years

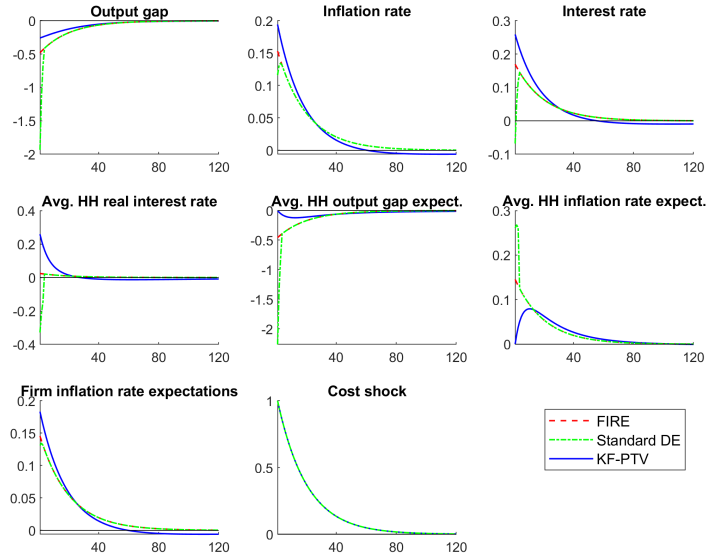
²⁹We include the standard diagnostic expectations model because it serves as an useful benchmark that features overreaction. Because our model is a generalization of the diagnostic expectations model, following L'Huillier et al. [2021], we set the parameter $\theta = 0.992$ to obtain the desired overreaction result.

³⁰For our Kalman filter operator we assume each cohort differs in its beliefs. In Appendix B.9 we analyze the consequences of dropping this assumption and having all cohorts follow the same beliefs while keeping the Kalman filter structure.

Figure 2.6: Impulse response functions



(a) Taste shock



(b) Cost shock

Note: Figure shows impulse response functions for a selected group of variables after the mentioned shocks. The red dashed line shows the results for the case of the full information rational expectations model (FIRE), the green dotted line shows the results of a standard diagnostic expectations operator and the solid blue line shows the history-based Kalman filter model. For the standard diagnostic expectations case we assume that agents use the expectations operator $\mathbb{E}_t^{\theta,s}[X_{t+h}] = \mathbb{E}_t[X_{t+h}] + \varpi(\mathbb{E}_t[X_{t+h}] - \mathbb{E}_{t-3}[X_{t+h}])$ with $\varpi > 0$. Horizontal axis denotes months after the shock.

Figure 2.6 presents the impulse response functions to a taste shock and a cost shock. In Panel (a), we see that inflation and the output gap increase after a taste shock. The FIRE case shows the usual reaction, and the standard diagnostic expectations model presents an overreaction that lasts three periods, consistent with the reference used. In the case of the history-based Kalman filter model, we first see a milder reaction in terms of inflation and output gap. Agents remember the past (in this case, the steady state), so their expectations tend to stay closer to such value. This allows the central bank to reduce the size of its hike in the interest rate, resulting in a slightly lower real interest rate. In addition, firms reduce the size of their price increase. Overall, under the history-based Kalman filter, agents anchor their expectations to the past, reducing the magnitude of the responses on impact.

While inflation is lower on impact with the history-based Kalman filter, it takes longer to return to its steady state values when compared to the other cases (the FIRE and standard diagnostic expectations cases have inflation going back to a steady state at the same pace of the persistence of the shock). This is because, with the history-based Kalman filter, agents remember the high inflation period. This effect is exacerbated by new cohorts that have only experienced inflation above the steady state.

Panel (b) of Figure 2.6 shows the responses for a cost shock. In the history-based Kalman filter case, there are two forces going in opposite directions. There is a high persistence in inflation, but we also have consumers who remember the zero output gap of the steady state, which reduces the pressure on prices. Thus, the IS curve becomes more inelastic to the shock with the history-based Kalman filter. Then, in this economy, rational firms can raise prices by more than they would under FIRE. This is followed by a central bank that must raise the interest rate more strongly than in the rational economy.

One thing to notice in the history-based Kalman filter case is that while the household inflation rate expectations have a hump shape, actual inflation does not. This is because firms are always rational. Thus, their expectations follow the shock very closely (no hump shape because of the AR(1) nature of the shock), and they set prices accordingly. The hump-shaped expectations are consistent with the evidence provided by Angeletos et al. [2021]. Consumers underreact to the inflationary shock in our setting initially, as their memories are tied to the steady state. After the inflationary shock, they incorporate the inflationary episode in their memories, over-extrapolating it, as in Afrouzi et al. [2023].

2.5.6 Implied heterogeneous responses

After obtaining the aggregate responses to the shocks, we now check the responses at the cohort-specific level implied by the aggregate variable paths of the previous section. Thus, we use the

individual level equations to see how the informational frictions of the model generate heterogeneous consumption and labor patterns depending on the past experiences of the cohorts. We must stress this result just shows a reaction starting akin to a partial equilibrium exercise taking the aggregate variable paths as given and is not the result of full heterogeneous agents model.

First, we calculate the within-model response of the output gap and inflation rate expectations for each cohort according to the framework of Section 2.4.2 and the aggregate variable paths following the shocks studied in the previous section. Figures B.3 and B.4 of Appendix B.1 present the heterogeneity in expectations across cohorts under the history-based Kalman filter. We see that, after both shocks, the expectations of the older cohorts remain anchored to the steady state, whereas the young react more strongly.

Then, we obtain cohort-specific consumption gap according to Equation 2.9, which contains common and heterogeneous components across cohorts. Similarly, we obtain the cohort-specific paths of labor supply gap from the individual labor supply problem. Figure B.5 in Appendix B.1 shows the consumption and labor supply gap responses of the different cohorts after a taste shock. We see that older cohorts consume less and work more, relative to the younger ones. For this result, there are two forces working simultaneously. In first place, the output gap expectations are lower for older cohorts, which brings their consumption down. In second place, the inflation rate expectations are lower for older cohorts, which relates to a higher real interest rate and reinforces the lower consumption. Because the older cohorts believe the future aggregate output gap will be lower and perceive a higher real interest rate, they end up supplying more labor relative to the young.

Figure B.6 of Appendix B.1 shows the responses of the consumption and labor supply gap across cohorts after a cost shock. Older cohorts consume more and supply less labor than the young. Contrary to the previous exercise, in this case this outcome is explained by two channels going in opposite directions. First, older cohorts have higher output gap expectations than younger ones, which drives their consumption upwards. Second, older cohorts have lower inflation rate expectations that translate into higher real interest rates, such that their consumption goes down. In the end, we see that the first channel dominates over the second. That is, even though older cohorts perceive a higher real interest rate, their more optimistic beliefs over the future output gap drive their consumption upwards relative to the young cohorts.

We highlight that these results are specific to this exercise, where the economy starts in the steady state. The heterogeneous responses are determined by the heterogeneous experiences of households which, given the history-based biases when forming expectations, create heterogeneous ex-ante real interest rates. In Section 2.7 we explore a data-based inflation scenario that will result in different ex-ante real interest rates across cohorts.

2.6 Optimal Taylor rules

Now we analyze the use of an optimal Taylor rule in each of the different cases from the previous section. The Taylor rule we use in this section is

$$i_t = \chi_\pi^* \pi_t + \chi_y^* y_t. \quad (2.18)$$

We assume that the central bank chooses the time-invariant parameters χ_π^* and χ_y^* such that it solves the problem

$$\min \mathbb{E}_t [\pi_t^2 + \vartheta y_t^2],$$

subject to the equations of the model in Section 2.5 and ϑ is the weight of the output gap in the objective function.³¹ That is, the central bank, given the model setup, seeks to minimize the volatility of both the inflation rate and the output gap. Notice that we assume that the central bank has rational expectations.

The optimal parameters are dependent on which shocks exist in the model (cost or taste). Therefore, we will have two sets of parameters, one for each shock.³²

We start by analyzing the response to a cost shock under the optimal Taylor rule. Panel (b) of Figure 2.7 shows that, when responding to this shock, the central bank faces the usual trade-off between the output gap and the inflation rate. After the cost shock, the inflation rate goes up and the output gap goes down after the interest rate hikes. In this particular exercise we will have that, given the relative importance of the output gap in the objective function, the central bank will favor reducing the volatility of the inflation rate.

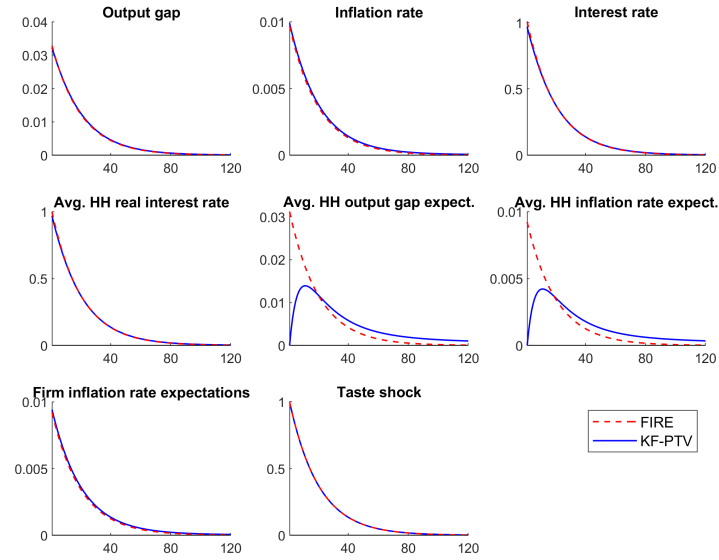
After the shock hits, the central bank becomes more active in the history-based Kalman filter than in the FIRE case. This behavior is because inflation history plays a role when there are history-based expectations. The central bank knows that people will remember the current shock far into the future, affecting future inflation expectations. By being more active under the history-based Kalman filter case, the central bank can quickly lower inflation expectations. While in the baseline results from Figure 2.6 the inflation expectations remained high for a long period, the optimal Taylor rule brings them down. It even generates deflation expectations that later spill over to the observed inflation rate.

Panel (a) of Figure 2.7 shows the impulse response functions to a taste shock and an optimal response from the central bank. As it is well known in the literature, upon a taste shock, the optimal response of the central bank is to raise the interest rate strongly. What follows is that the central

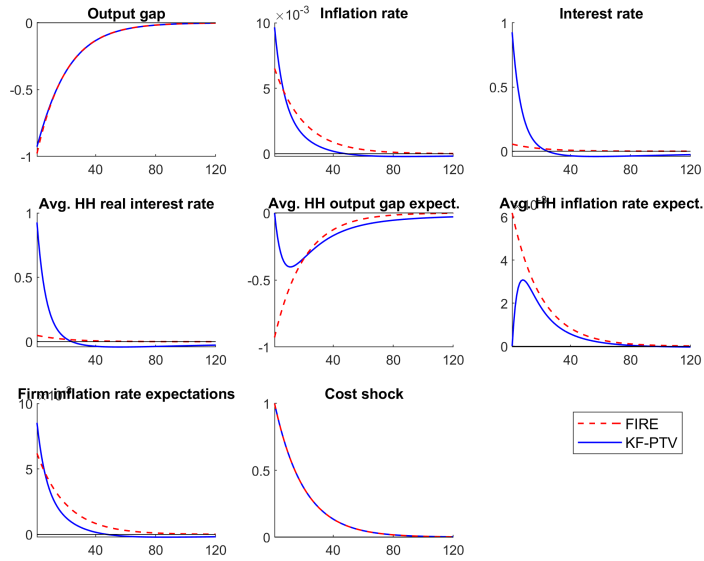
³¹Following Gali [2015] we define $\vartheta = \frac{(1-\phi)(1-\phi\beta)(\sigma+\eta)}{\phi\varepsilon} = 0.0017$.

³²Tables B.2 and B.3 in Appendix B.1 show the optimal parameters.

Figure 2.7: Impulse response functions, optimal Taylor rule



(a) Taste shock



(b) Cost shock

Note: Figure shows impulse response functions for a selected group of variables after the mentioned shocks under an optimal Taylor rule. The red dashed line shows the results for the case of the full information rational expectations model (FIRE) and the solid blue line shows the history-based Kalman filter model. Horizontal axis denotes months after the shock.

bank manages to bring down both the output gap and the inflation rate to their steady state values.

We find no significant difference in the central bank's response between the FIRE and history-based cases. The optimal Taylor rule case says that the central bank should always be active when facing a taste shock, no matter the type of expectations agents have. In this manner, the central bank can close the output gap and lower the inflation rate faster than the baseline results. Moreover, with the active stance recommended by the optimal Taylor rule, the output gap and inflation expectations are positive but very close to zero in all cases.

2.7 Analyzing an episode of high inflation

This section analyzes the model's behavior after the high-inflation episode of 2021.³³ To do so, we feed the model from Section 2.5 with actual monthly data on the output gap, inflation rates, and interest rates up to December 2021.³⁴ Afterward, we produce forecasts using the different versions of the model.³⁵

We first analyze the history-based Kalman filter case. Figure B.7 of Appendix B.1 shows the inflation rate expectations by cohort according to our model and the data. Before 2021, older cohorts had the highest inflation expectations. This result occurs because older cohorts experienced the high-inflation episodes of the 60s, 70s, and early 80s. They are followed by the intermediate cohorts and finally by the youngest cohorts, who experienced low and stable inflation rates from the 90s to the 10s. With the inflationary shock, things changed as newer cohorts experienced a significant part of their lives in a high-inflation environment.

Figure 2.8 presents how variables evolve based on our model and the data. After 2021, when we allow for a history-based Kalman filter, average inflation expectations are higher and more persistent with respect to other cases. This result is because agents remember and anchor their expectations to what they experienced in the past. Hence, the central bank must react more strongly.

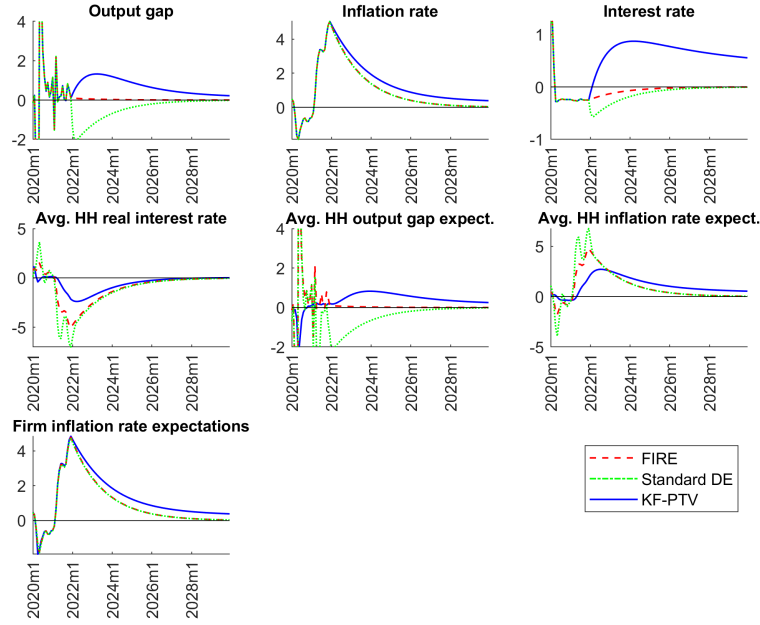
Because of their historical reference, agents that follow a history-based Kalman filter remember the high inflation episode far into the future when compared to the other cases case. As a consequence, the observed inflation rate and the interest rate also remain higher for longer.

³³In this part, we go back to the basic calibration of Table B.1.

³⁴We use monthly series from March 1967 to December 2021. We go as far as the data allows us to build the agents' references. We use the National Activity Index (CFNAI) from the Federal Reserve Bank of Chicago for the output gap. For the interest rate, we use the effective federal funds rate. For the inflation rate, we use the CPI 12-month percentage change.

³⁵We present the shocks that, according to our model, explain the observed data in Figure B.8 in Appendix B.1.

Figure 2.8: Impulse response functions, forecast



Note: Figure shows impulse response functions for a selected group of variables according to the model and the data (up to December 2021). The red dashed line shows the results for the case of the full information rational expectations model (FIRE), the green dotted line shows the results of a standard diagnostic expectations operator and the solid blue line shows the history-based Kalman filter model. For the standard diagnostic expectations case we assume that agents use the expectations operator $\mathbb{E}_t^{\theta, s} [X_{t+h}] = \mathbb{E}_t [X_{t+h}] + \varpi (\mathbb{E}_t [X_{t+h}] - \mathbb{E}_{t-3} [X_{t+h}])$ with $\varpi > 0$. Horizontal axis denotes months.

2.8 Conclusions

This paper studies the macroeconomic consequences of heterogeneous inflation expectations. We first show that inflation expectations are heterogeneous across cohorts. We introduce a Kalman filter augmented with history-based expectations to model the inflation forecast formation process. We structurally estimate the relevant parameter, concluding that individuals effectively consider their past inflation histories when forecasting.

Our expectation formation process includes two known and relevant aspects of consumers' expectations processes. On one side, we consider the history of inflation each cohort has experienced. On the other side, we consider the current inflation experience, particularly with salient prices such as grocery prices. Our proposed framework can predict the heterogeneity prevalent in consumers' inflation expectations across cohorts remarkably well, even though we only feed our model with monthly CPI inflation data and an estimated parameter. We show that consumer inflation expectations data can be modeled, predicted, and contain meaningful information.

Our modeling approach also has the advantage of being flexible, such that we incorporate it in a general equilibrium model. This heterogeneous expectation process has relevant aggregate

implications. Heterogeneous expectations anchor aggregate response to agents' history of inflation while increasing the persistence of the effect of the shocks.

This result has implications for monetary policy: when inflation starts rising, the optimal response of monetary authorities is to take an active stance, as agents have a long history and remember current shocks far into the future. An energetic response of the central bank under inflationary pressures prevents current inflation from rising and, more importantly, prevents agents from incorporating high-inflation episodes into their memories, thus preventing higher future inflation expectations.

Our results also have relevant implications for the current macroeconomic environment. The model suggests that the 2021 high-inflation episode, even though it may be transitory, could have long-lasting effects: new cohorts incorporate the high-inflation episode into their memories of inflation, adjusting future inflation expectations upwards.

CHAPTER 3

Cultural Proximity and Inter-firm Trade

Abstract

Emerging economies sometimes have low-quality institutions, generating micro-level trade frictions. In these settings, firms may rely on informal institutions, such as cultural proximity, to overcome such frictions. Using new microdata on firm-to-firm trade from India with information on prices, transactions, and caste and religious connections, we find that higher cultural proximity reduces prices and fosters trade at intensive and extensive margins. When exploring potential mechanisms, we find evidence supporting the role of cultural proximity in alleviating contracting frictions. We formalize these findings by building an inter-firm trade model with cultural proximity between firm owners. Our counterfactual exercises indicate that an economy composed of culturally closer firms features lower costs, lower prices, higher sales, and higher welfare than an economy with culturally distant firms.¹

3.1 Introduction

Non-economic forces, such as *culture*— religion, language, values, etc.— may drive economic outcomes. The role of culture on agent behavior has been well documented in entrepreneurship, loan access, labor markets, marriage, development, and cross-country international trade [Fisman et al., 2017, Goraya, 2023, Munshi and Rosenzweig, 2016, Rauch, 1996, Rauch and Trindade, 2002, Startz, 2016]. Nevertheless, the mechanisms by which cultural proximity shapes trade between firms remain less understood. Understanding how and why cultural proximity affects inter-firm trade potentially allows policy-makers to better leverage social inclusion programs and foster economic development.

We first provide empirical evidence on the role of cultural proximity in inter-firm trade. To do this, we leverage a unique dataset of transactions between firms from a large Indian state,

¹This chapter is co-authored with Brian C. Fujiy and Gaurav Khanna.

along with data on firm owners' names and their cultural proximity derived from India's caste and religious system.² We report three new stylized facts. First, culturally closer firms report higher sales between them: the higher the cultural proximity, the higher the trade on the intensive margin. Second, culturally closer firms are more likely to trade with each other. That is, the higher the cultural proximity, the higher the trade on the extensive margin as well. Third, firms that are culturally further apart report higher unit prices in their transactions. All these results are robust to an array of high-dimensional fixed effects, including seller and buyer fixed effects, origin-by-destination fixed effects (and for specifications with product and time, seller-by-product, and product-by-month fixed effects).

We then turn to explore various mechanisms, and find evidence most in line with the importance of contract enforcement. First, we show that the effect we find of cultural proximity on trade is driven by differentiated products, which often rely on either formal or informal contract enforcement [Nunn, 2007, Rauch, 1999].³ Second, cultural proximity matters the most when institutional quality, proxied by the court quality in the districts where the trade partners belong, is particularly low. We argue that, in a setting with low institutional quality and poor contract enforcement, firms that trade differentiated goods rely on informal institutions (i.e. cultural proximity) as a substitute for the imperfect formal ones. We understand these findings as evidence that cultural proximity may relate to contract enforcement and trust [Munshi, 2019, 2014].⁴

We further find that the more varieties a firm sells or buys, the more the trade intensity is affected by cultural proximity. We posit that the larger the number of different varieties a firm sells or buys, the more firms it has to negotiate with, which increases the contracting frictions it faces. Then, to minimize the contracting frictions they face, firms will rely more on trading with culturally closer firms they trust.

We do not find sufficient evidence that hierarchies (and preference-based discrimination) across social groups matter, or that linguistic distance and the specialization in certain goods matters for our cultural proximity results. To analyze whether our results are caused by vertical social hierarchies and discrimination across cultural groups, we study asymmetric effects in those transactions where one firm is placed higher than the other based on the caste-based hierarchy, allowing us to test for preference-based discrimination across the social hierarchy. In other tests, we find our results are less likely to be driven by firms sharing the same language or sharing specialization in the production of certain goods.

²Due to a confidentiality agreement, we cannot disclose the name of the Indian state.

³Differentiated goods do not trade in exchanges and are not homogeneous, but are branded and specific to certain producing firms. In a country with market imperfections as India, firms can easily renege on their commitments. Suppliers and buyers in differentiated goods markets are not easily replaceable. In such cases, trade will increase when firms trust and know each other, that is, when they are culturally close.

⁴Munshi [2019] uses survey data to show that Indians trust people from their caste.

We next conduct a counterfactual analysis to study the importance of cultural proximity for inter-firm trade. We build a quantitative trade model with cultural proximity between firms. Firms optimally decide whom to trade with subject to matching fixed costs, and how much to trade with subject to iceberg trade costs. In line with our empirical findings, we allow these costs to depend (positively or negatively) on how culturally close firms are.⁵

The model derives equations that precisely match their empirical counterparts in the previous section. We use these equations to estimate the key parameters of the model: the semi-elasticity of the trade cost to cultural proximity and the semi-elasticity of matching cost to cultural proximity. Our model allows us to estimate both of these parameters externally. In line with our stylized facts, we find a negative semi-elasticity of both the intensive and extensive margin of trade to cultural proximity. This implies the closer two firms are in cultural terms, the lower the trade and matching costs are. Therefore, the higher the cultural proximity for a pair of firms, the higher the trade is on both the intensive and extensive margins, and the lower the prices charged.

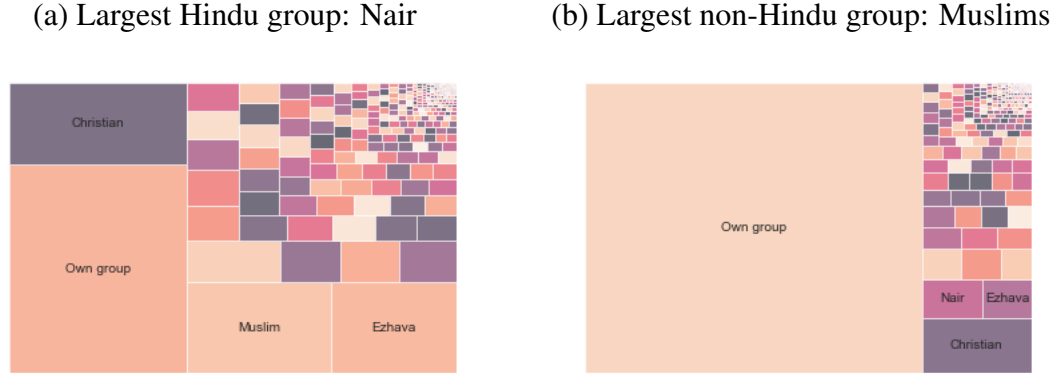
We then use the model and estimated parameters to quantify the implications for welfare and other aggregate outcomes of implementing different policies. First, we evaluate the effects of social mixing/inclusion (i.e. firms become culturally the closest possible) and social isolation policies (i.e. firms become culturally the furthest possible). Second, we study the effects of a policy that reduces contracting frictions, such that firms rely less on cultural proximity when trading (i.e. trade and matching costs become less sensitive to cultural proximity). We find that welfare increases by 1.76 percent under a diversity-friendly social inclusion policy. In contrast, welfare falls by 1.45 percent when we evaluate the effects of social isolation or exclusion. Finally, we show that policies that reduce contracting frictions raise welfare by 0.87 percent by reducing the reliance of trade on cultural links.

The analysis of cultural proximity is especially relevant for developing countries, where agents face several contracting frictions and, consequently, rely more on non-economic forces. In particular, India has a society that follows the parameters of a caste system, which also intertwines with the different religious groups.⁶ In this case, cultural proximity naturally arises as a product of the inherent hierarchical structure of the caste system and the different religions. Related to this, Figure 3.1 shows an example of how trade between cultural groups occurs, in a selected subset of

⁵On the extensive margin, cultural proximity between firms can reduce matching costs since culture encodes useful information for sellers to decide who to sell to [Balmaceda and Escobar, 2017, Ali and Miller, 2016]. On the intensive margin, sellers charge a premium to buyers arising from contracting frictions. Given the risk of renegeing on the contract or delaying payment, sellers determine the charged premium to buyers, and so affect the intensive margin of trade [Boehm and Oberfield, 2020]

⁶In this paper, we consider the caste system and religious groups as a proxy for cultural groups. There is a large historical legacy for the caste system to be considered a discrimination device, which we consider. Even though there is an active agenda of the government to implement policies that hinder caste-based discrimination, it is still used by Indians as a way to determine how similar individuals are between them.

Figure 3.1: Probability-weighted sales decomposition of largest cultural groups



Notes: The figure shows the decomposition across buyers for the largest Hindu and non-Hindu cultural groups measured by probability-weighted sales. For instance, the Nair and Muslims accounted for 5.39 and 19.68 percent of total probability-weighted sales, respectively.

our data. We can see that there are cultural groups that are bound to trade more or less with other cultural groups. We thus ask whether cultural proximity, measured as the cultural group-based distance between firms, can determine trade.

We contribute to two strands of the literature. First, we speak to the role of cultural proximity on trade [Bandyopadhyay et al., 2008, Guiso et al., 2009, Macchiavello and Morjaria, 2015, Rauch, 1996, Rauch and Casella, 2003, Rauch and Trindade, 2002, Richman, 2006, Schoar et al., 2008, Startz, 2016, Zhou, 1996, Boken et al., 2022]. Cultural differences in international trade or across administrative regions also interplay with non-cultural barriers. In contrast, we use transaction-level firm-to-firm data to explore the effect of culture on trade that does not rely on cross-border variation. This allows us to isolate the mechanisms at a granular level, and to provide a framework that quantifies the role of cultural proximity for inter-firm trade.⁷ We also connect to the importance of cultural or social proximity on other economic outcomes, like entrepreneurship [Goraya, 2023], finance [Fisman et al., 2017], the composition of the board of directors [Faia et al., 2021], and labor markets [Munshi and Rosenzweig, 2016]. We complement this work by examining how culture affects trade frictions. Understanding the source of these frictions or even leveraging them is key to adequately determining the consequences of economic policy.

Second, we contribute to the literature on social cohesion [Alan et al., 2021, Alesina and Giu-

⁷In ongoing work, Boken et al. [2022] also shows the role of cultural proximity for inter-firm trade. We mainly distinguish ourselves by leveraging data on prices, which allows us to estimate how cultural proximity influences inter-firm trade through the alleviation of contracting frictions. Additionally, we explore a rich set of mechanisms, as our data contains information to test how cultural proximity matters for inter-firm trade through discrimination (caste hierarchies), order cancellations, firm survival over age, and how cultural proximity matters for firm-level complexity. Leveraging other data, we implement a Heckman selection bias correction model to estimate trade elasticities following Helpman et al. [2008].

liano, 2015, Alesina et al., 2021, Alesina and Reich, 2015, Bazzi et al., 2019, Depetris-Chauvin et al., 2020, Gradstein and Justman, 2019, Ritzen et al., 2000] and contract enforcement policies.⁸ We contribute by implementing a policy counterfactual analysis that quantifies the aggregate effects of education and court quality policies through cultural proximity.

The rest of the paper is structured as follows. In Section 3.2 we provide a brief review of the caste system in India, describe our new datasets, and explain how we construct firm-level trade and cultural proximity variables. In Section 3.3, we report our stylized facts, study the mechanisms driving the effect of cultural proximity on trade, and robustness checks. In Section 3.4, we briefly describe the model, how we estimate the model, and perform a counterfactual analysis. Section 3.5 concludes.

3.2 Background, data and construction of variables

3.2.1 Caste and Religion in India

India has a society that is heavily influenced by the parameters of a caste system: a hierarchical system that has prevailed in the country since around 1,500 BC and that still rules its economy. According to this classification, people are classified across four possible groups called *Varnas*. From the most to the least privileged in hierarchical order, the four Varnas are *Brahmins*, *Kshatriyas*, *Vaishyas*, and *Shudras*. The Brahmins have historically enjoyed the most privileges, and are traditionally comprised of priests and teachers. The Kshatriyas are next in the hierarchy, usually associated with a lineage of warriors. The Vaishyas are third and are related to businessmen such as farmers, traders, among others. Finally, the Shudras are the most discriminated against and are the caste formed to be the labor class. Below these groups in the socio-economic hierarchy, were marginalized groups called Dalits.

At the same time, Varnas are comprised of sub-groups called *Jatis*, determined by factors such as occupation, geography, tribes, or language. In that sense, using Jatis as castes is appropriate for studying economic networks [Munshi, 2019], and from here on, we use the notion of Jatis when referring to castes.

We also consider religious groups to define other cultural groups. The caste system is inherently based on Hindu religion, the predominant religion in India. While other religions in India did not historically follow the caste system, they do relate to it today: the other non-Hindu religions work as cultural groups of their own. We leverage information on firm owners belonging to both caste and non-Hindu religious groups to construct our measure of cultural proximity.

⁸Contracting frictions can be either formal or informal. We show that informal channels, such as cultural proximity, matter in the aggregate when implementing policies.

3.2.2 Data

Firm-to-firm trade. We obtain a new firm-to-firm trade dataset for a large Indian state provided by the state’s corresponding tax authority.⁹ We use daily transaction-level data from January 2019 to December 2019, as long as at least one node of the transaction (either origin or destination) was in the state. This data exists due to the creation of the E-Way bill system in India in April 2018, where firms register the movements of goods online for tax purposes. This is a major advantage over traditional datasets collected for tax purposes in developing countries since the E-Way bill system was created to significantly increase tax compliance.¹⁰

The state has a diversified production structure, roughly 50 percent urbanization rates, and high levels of population density. To compare its size in terms of standard firm-to-firm transaction datasets, the population of this Indian state is roughly three times the population of Belgium, seven times the population of Costa Rica, and double the population of Chile. In addition, we can uniquely measure product-specific prices for each transaction, along with the usual measures of the total value traded.

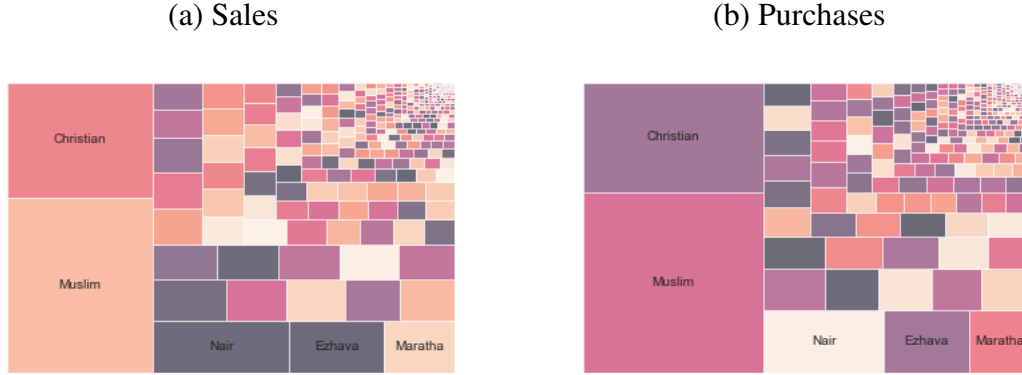
Each transaction reports a unique tax code identifier for both the seller and buyer. We use these identifiers to merge this data with other firm-level datasets. We also have information on all the items contained within the transaction, the value of the transaction, the 6-digit HS code of the traded items, the quantity of each item, and the units the quantity is measured in. Since the data report both value and quantity of traded items, we construct unit values for each transaction. Each transaction also reports the pincode (zip code) location of both selling and buying firms. By law, any person dealing with the supply of goods and services whose transaction value exceeds 50,000 Rs (600 USD) must generate E-way bills. Transactions that have values lower than 600 USD can also be registered, but it is not mandatory. There are three types of recorded transactions: (i) within-state trade, (ii) across-state trade, and (iii) international trade. For this paper, we ignore international trade.

Firm owner names. Information about the name of firm owners comes from two different sources. The first is also provided by the tax authority of the Indian state, which is a set of firm-level characteristics for firms registered within our large state. Among these variables, we are provided with the name of the owner, directors and/or representatives of the firm.

⁹While we use the term ‘firm’ in most parts of the paper, these data are actually at the more granular establishment level.

¹⁰Tax evasion rates are thought to have fallen with the E-Way bill system, given how it is implemented. A selling establishment must online register the transaction, and print out a receipt that the driver of the transportation (usually a truck) must carry with them while transporting the product. If the driver is stopped or checked at any of the numerous checkpoints, and fails to produce a receipt, the goods are confiscated. Furthermore, the earlier VAT regime was only for large firms, and the new GST system was aimed at including smaller firms too. For more details about the new E-Way bill system, see <https://docs.ewaybillgst.gov.in/>

Figure 3.2: Probability-weighted sales and purchases across cultural groups



Notes: Figure shows the decomposition of the probability-weighted sales and purchases across the 452 cultural groups in our dataset. The size of the rectangles reflects the share of sales and purchases.

To obtain firm-level characteristics of firms not registered in this state, we scrape the website *IndiaMART*,¹¹ the largest e-commerce platform for business-to-business (B2B) transactions in India. The website is comprised of firms of all sizes. By 2019, the website registered around 5-6 million sellers scattered all around India. Most importantly, this platform provides the name of the owner of the firm and the unique tax code identifier. Thus, we use the platform to obtain these variables for out-of-state firms.

Matching owner names to cultural groups. We follow Bhagavatula et al. [2018] to match owner names to their Jatis (if the owners are of Hindu religion) or to their religion (in case the owners are not Hindu). Their procedure consists of using scraped data from Indian matrimonial websites that contain information on names, castes, and religions. They train a sorting algorithm that uses names as inputs and gives a probability distribution across cultural groups per name as outputs. We match these probability distributions to each owner's name in our dataset. Notice that our notion of cultural group-belonging is probabilistic and not deterministic. This probabilistic approach is more relevant to our setup since, when firm owners trade with each other, they do not know each other's cultural group *ex ante*. Our sample finally consists of 452 cultural groups.

Merged dataset. For the analytical part, we merge the three previous datasets. We end up with a sample that contains information from 21,497 unique firms, of which there are 10,010 sellers and 16,403 buyers. We drop any registered transaction in which the seller and the buyer are the same firm. In total, the dataset comprises approximately 800 thousand transactions equivalent to 1.2 billion US dollars. Each firm is linked to a unique pincode. Finally, we assign a sector to each firm

¹¹<https://www.indiamart.com/>

based on the ISIC codes of the goods sold. We provide more aggregate and firm-level statistics in the Appendix Table C.1.

There is a large degree of heterogeneity across firms and cultural groups. Appendix Figure C.1 shows that there is a significant dispersion in sales per firm, although most of the firms within the dataset can be classified as small. Furthermore, to provide a summary of the heterogeneity of cultural groups present in the firm-to-firm trade data, we show the distribution of probability-weighted sales and purchases across cultural groups in Figure 3.2.

3.2.3 Construction of variables

Inter-firm trade variables. The firm-to-firm dataset provides information at the transaction level between any two registered firms. More specifically, we have information on (i) transaction-level unique identifiers, (ii) seller and buyer unique identifiers, (iii) the 6-digit HS description of the traded goods in each transaction, (iv) the total value of the transaction in rupees per type of good involved in each transaction and (v) the number of units sold of each good in each transaction.

For every seller/buyer pair, we construct total sales, the total number of transactions, and unit values. For the total sales, we add up all the sales between each given pair of firms in our sample. We do the same with the total number of transactions. To obtain prices, we calculate the unit values. We first calculate the total amount sold and the total units sold of each good at the 6-digit HS level between each given pair of firms in our sample. Then, we divide the total amount sold by the number of units sold for each good.

Cultural proximity. Consider the set $\mathcal{X}$ of cultural groups, where $|\mathcal{X}| = X = 452$ in our final dataset. Since not all names are deterministically matched to a cultural group, each firm in our dataset has a discrete probability distribution over the set X of cultural groups. In particular, every firm ν has a probability distribution $\rho_\nu = [\rho_\nu(1), \dots, \rho_\nu(X)]$, such that $\sum_{x=1}^X \rho_\nu(x) = 1$. In this part, we distinguish between the probability distribution over cultural groups of the seller and the probability distribution over cultural groups of the buyer. Define $\rho_\nu(x)$ as the probability of seller ν of belonging to cultural group x . Similarly, define $\rho_\omega(x)$ as the probability of buyer ω of belonging to cultural group x . Based on these two distributions, we construct the following measure of cultural proximity: the Bhattacharyya coefficient [Bhattacharyya, 1943].

The Bhattacharyya coefficient between seller ν and buyer ω measures the level of overlap be-

tween two different probability distributions.¹² We define it as

$$BC(\nu, \omega) = \sum_{x=1}^X \sqrt{\rho_\nu(x) \rho_\omega(x)}.$$

Because $0 \leq \rho_\nu(x) \leq 1$ and $0 \leq \rho_\omega(x) \leq 1$, we have that $0 \leq BC(\nu, \omega) \leq 1$. On the one hand, $BC(\nu, \omega) = 0$ means the seller has a completely different probability distribution from the buyer's. In our context, this means the seller and the buyer have almost no chance of belonging to the same cultural group or that their cultural proximity is the farthest. On the other hand, $BC(\nu, \omega) = 1$ means the seller has the same probability distribution as the buyer. This implies that the seller has the same probability of belonging to a group of certain cultural groups as the buyer or that their cultural proximity is the closest possible.¹³ In robustness checks, we use the Kullback and Leibler [1951] divergence measure (Appendix C.2.2). All our results are qualitatively similar, and statistically significant when doing so.

3.3 Stylized facts

Fact 1: Cultural proximity fosters trade. We first discuss results related to the intensive margin of the firm-to-firm trade. Figure 3.3 shows the residualized scatterplots between the Bhattacharyya coefficient and two intensive margin measures: total sales between two firms and total transactions between two firms. The scatterplots show a higher Bhattacharyya coefficient (buyer and seller are probabilistically more alike in their cultural group) is related to a higher amount of sales and transactions.

We confirm the findings using a gravity equation. For transactions from firm ν to firm ω in our sample we estimate

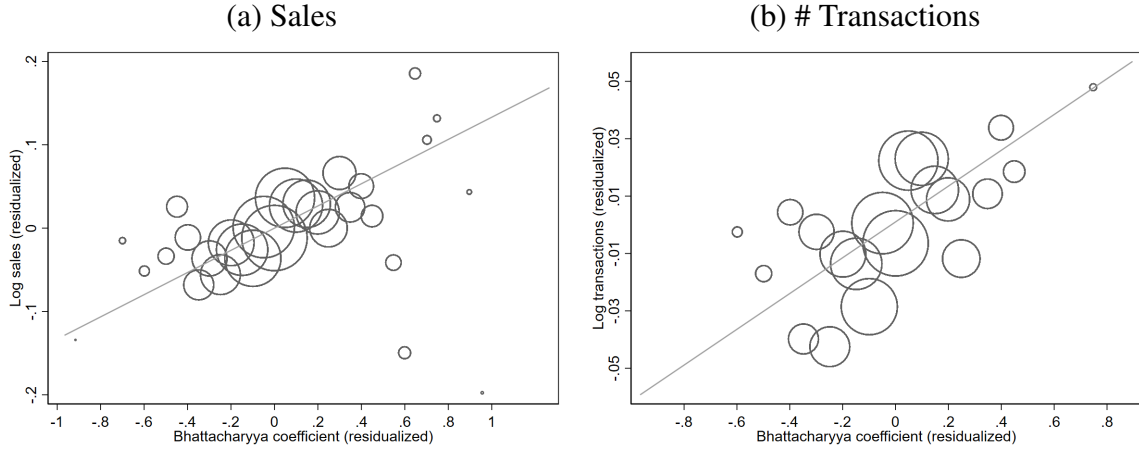
$$\ln y(\nu, \omega) = \iota_\nu + \iota_\omega + \delta BC(\nu, \omega) + \eta \Xi(\nu, \omega) + \varepsilon(\nu, \omega), \quad (3.1)$$

where $y(\nu, \omega)$ is either the total sales $n(\nu, \omega)$ or total transactions $t(\nu, \omega)$ from seller ν to buyer ω , $BC(\nu, \omega)$ is the Bhattacharyya coefficient, $\Xi(\nu, \omega)$ denotes geographic distance control variables

¹²Notice the Bhattacharyya coefficient is not the Bhattacharyya distance, which is defined as $BD(s, b) = -\log(BC(s, b))$. We prefer the Bhattacharyya coefficient because it is easier to interpret.

¹³For our purposes, it is important that the cultural proximity measure we use is symmetric. To see why, consider an example where, in our dataset, we have a transaction between a seller ν and a buyer ω , from which we obtain $BC(\nu, \omega)$. Further, assume that in our dataset, we record a second transaction in which the roles of the firms revert (i.e. the buyer becomes the seller and vice versa), so we calculate $BC(\omega, \nu)$. Regardless of the roles the firms take in this second transaction, we want their cultural proximity to remain constant, as the membership of cultural groups is fixed. This goal is achieved through the means of a symmetric proximity measure, and the Bhattacharyya coefficient complies with this symmetry requirement, as $BC(\nu, \omega) = BC(\omega, \nu)$.

Figure 3.3: Effect of cultural proximity on trade, intensive margin



Notes: Results are residualized of seller fixed effects, buyer fixed effects, and log distance. Equally distanced bins formed over the horizontal axis. The size of the bubbles represents the number of transactions in each bin. The higher the Bhattacharyya coefficient, the culturally closer the two firms are.

(such as the Euclidean distance between pincodes in which the firms are located $dist(\nu, \omega)$ or a district-level origin-destination fixed effect ι_{od}), ι_ν and ι_ω are seller and buyer fixed effects. Columns 1-4 of Table 3.1 present the results of the intensive margin estimation, which confirm the preliminary findings from Figure 3.3. Columns 1 and 2 show that, on average, there will be a higher amount of sales and transactions between a pair of firms when these firms are more alike in cultural terms. Columns 3 and 4 show that these results remain strong after including origin-destination fixed effects, which account for geographic distance but also control for other features that might arise between a pair of locations, such as different terrains, different languages, location-specific cultural ties, historical ties, etc.

Fact 2: Cultural proximity increases the likelihood of trading. Next, we estimate the extensive margin relationship. Given the size of our full dataset, the number of potential extensive margin links is computationally large. For tractability, we modify our sample. In the first place, we construct a sample with all possible combinations of in-state buyers and in-state sellers with cultural group information. Then, we proceed to drop all potential transactions that include unfeasible sectoral combinations. That is, we drop the combinations of firms that are involved in productive sectors that never recorded a transaction in the data. Finally, we drop all unfeasible transactions based on distance. This is to say, we drop the combinations of firms where the seller is further away than the maximum recorded distance for the in-state buyer or vice versa.

With this sample, we construct a trade indicator variable $tr(\nu, \omega)$ which is equal to 1 if there is any kind of trade between firms ν and ω , and 0 otherwise. With this variable, we estimate a

Table 3.1: Effect of cultural proximity on trade, intensive and extensive margins

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable	Log Sales	Log Transactions	Log Sales	Log Transactions	Trade Indicator	Trade Indicator
<i>BC</i>	0.100*** (0.033)	0.066** (0.027)	0.129*** (0.034)	0.076*** (0.028)	0.0009*** (0.0001)	0.0010*** (0.0001)
Log dist.	-0.023 (0.015)	-0.065*** (0.011)			0.0001 (0.0000)	
Obs.	32,678	32,678	32,843	32,843	5,606,627	5,628,290
Adj. R2	0.415	0.359	0.410	0.356	0.617	0.0106
FE	Seller, buyer	Seller, buyer	Seller, buyer, origin×dest.	Seller, buyer, origin×dest.	Seller, buyer	Seller, buyer, origin×dest.

Notes: Columns 1, 2, 3 and 4 show the results of estimating Equation 3.1. Columns 5 and 6 show the results of estimating Equation 3.2. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales and transactions refer to the total sales and total number of transactions for a given pair of seller and buyer, respectively. The trade indicator is equal to 1 if there was at least 1 recorded transaction for a given pair of seller and buyer, and it is equal to 0 otherwise. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are two-way clustered at the seller and buyer level. Standard errors in parentheses. The higher the Bhattacharyya coefficient, the culturally closer the two firms are. The number of observations varies between specifications due to the dropping of observations separated by a fixed effect [Correia et al., 2019].

gravity-type specification:

$$tr(\nu, \omega) = \iota_\nu + \iota_\omega + \delta BC(\nu, \omega) + \eta \Xi(\nu, \omega) + \varepsilon(\nu, \omega, t). \quad (3.2)$$

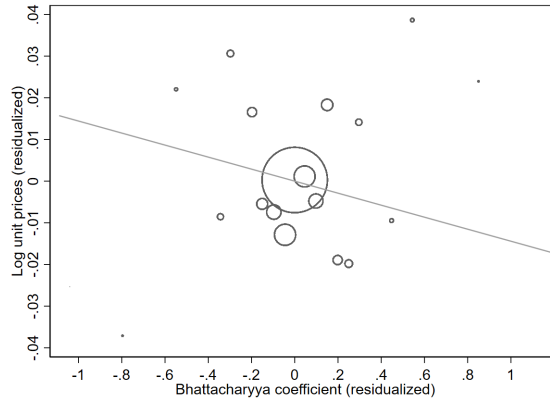
Columns 5-6 of Table 3.1 present the extensive margin results. We find that the higher the Bhattacharyya coefficient, the more likely it is that two given firms will trade.

Fact 3: Cultural proximity lowers prices. Figure 3.4 now uses buyer-seller-product-month groups and shows the residualized scatterplots between the similarity measure and the unit prices. We see the higher the Bhattacharyya coefficient between two firms involved in a transaction, the lower the price that will be charged. To confirm the results, we work with transaction-level data and estimate:

$$\ln p_g(\nu, \omega, t) = \iota_{\nu g} + \iota_{gt} + \iota_\omega + \delta BC(\nu, \omega) + \eta \Xi(\nu, \omega) + \epsilon_g(\nu, \omega), \quad (3.3)$$

where $p_g(\nu, \omega, t)$ is the unit value of good g (at the 6-digit HS classification) sold by firm ν to firm ω in month t , $\iota_{\nu g}$ is a seller-good fixed effect and ι_{gt} is a good-month fixed effect. We present the results in Table 3.2, which confirms the previous findings from Figure 3.4: the culturally closer,

Figure 3.4: Effect of cultural proximity on prices



Notes: Results residualized of seller fixed effects and HS code fixed effects. Sectors are defined according to the 6-digit HS classification. Equally distanced bins formed over the horizontal axis. The size of the bubbles represents the number of transactions in each bin. The higher the Bhattacharyya coefficient, the culturally closer the two firms are.

the lower the unit value of the transactions.

3.3.1 Mechanisms

3.3.1.1 Contracting frictions

Differentiated goods and contract enforcement. To understand the underlying forces driving these empirical patterns, we explore the role of cultural proximity in alleviating contracting frictions. The lack of contract enforcement in developing countries inhibits trade as sellers or buyers may not comply with the terms of the contract. For instance, the buyer could hold up the seller by withholding payment after the buyer receives the shipped goods. Differentiated or relationship-specific goods are subject to more severe hold-up problems, as there are fewer alternative buyers and sellers of such products, and buyers may withhold payment knowing that these goods are not useful outside of the relationship. In that sense, differentiated goods rely on better contract enforcement.

Contract enforcement can be either formal (e.g. courts) or informal (e.g. cultural proximity). Most of the literature has focused on the role of formal institutions in enforcing contracts. We hypothesize and show evidence that cultural proximity alleviates contracting frictions for differentiated goods when formal contract enforcement is lacking [Nunn, 2007].

To bring in information about the type of product, we disaggregate our data at the transaction level. Then, we classify the goods into differentiated goods and non-differentiated goods based on

Table 3.2: Effect of cultural proximity on prices

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable	Log Prices	Log Prices	Log Prices	Log Prices	Log Prices	Log Prices
<i>BC</i>	-0.069** (0.033)	-0.069** (0.033)	-0.066** (0.033)	-0.045* (0.023)	-0.040* (0.023)	-0.039* (0.022)
Log dist.	0.023 (0.016)	0.023 (0.016)	0.028* (0.017)			
Obs.	230,744	230,744	226,645	235,001	236,617	230,900
Adj. R2	0.932	0.932	0.935	0.933	0.925	0.936
FE	Seller×HS, Seller×HS, Seller×HS, Seller×HS, Seller×HS, Seller×HS, buyer buyer, buyer, buyer, buyer, buyer, buyer, month month×HS origin×dest. month, month×HS, origin×destorigin×dest.					

Notes: This table shows the results of estimating Equation 3.3. Good g is defined according to 6-digit HS classification. Prices trimmed by 4-digit HS code at 5 and 95 percent. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Prices refer to unit prices of goods sold. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are multi-way clustered at the seller, 4-digit HS and origin-destination level. Standard errors are in parentheses. The higher the Bhat-tacharyya coefficient, the culturally closer the two firms are. The number of observations varies between specifications due to the dropping of observations separated by a fixed effect [Correia et al., 2019].

the classification developed by Rauch [1999].¹⁴ We estimate the following specification:

$$\ln n_g(\nu, \omega, t) = \iota_{\nu g} + \iota_{gt} + \iota_{\omega} + \delta BC(\nu, \omega) + \xi(BC(\nu, \omega) \times \mathbb{I}_g^{diff}) + \eta \Xi(\nu, \omega) + \epsilon_g(\nu, \omega), \quad (3.4)$$

where $n_g(\nu, \omega, t)$ are the sales going from firm ν to firm ω of good g in month t and $\mathbb{I}_g^{diff}$ is an indicator for differentiated goods.¹⁵ Columns 2 and 3 of Table 3.3 present the results. Our findings suggest that the baseline results of cultural proximity increasing trade are mostly driven by differentiated goods.

To understand the channel for why trade in differentiated goods depends on cultural proximity, we turn to the analysis of contract enforcement. We posit that, when facing poor contract enforcement and poor quality of institutions, firms trading differentiated goods must rely on alternative

¹⁴According to Rauch [1999] differentiated goods are the goods not traded in organized exchanges or not reference priced in commercial listings. Differentiated goods have specific characteristics that “differentiate” (i.e. specialized goods, branded goods) them from other, more homogeneous types of goods. Because of their relative uniqueness in features, these goods are not as easily replaceable as non-differentiated goods and, as such, rely more on relationship-specific types of trade. This means sellers and buyers must face search frictions to match to a suitable trade partner and will likely not abandon the commercial matches they have already made.

¹⁵We use both the conservative and liberal classifications from Rauch [1999]. The conservative classification minimizes the number of goods classified as non-differentiated and, thus, has the largest amount of differentiated goods. The liberal classification maximizes the amount of goods classified as differentiated and has the largest number of differentiated goods.

Table 3.3: Effect of cultural proximity on trade by types of good, intensive margin

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable	Log Sales	Log Sales	Log Sales	Log Sales	Log Sales	Log Sales
BC	0.099*** (0.031)	0.018 (0.050)	0.039 (0.040)	0.069** (0.027)	-0.019 (0.048)	0.013 (0.038)
$BC \times \mathbb{I}_g^{diff,con}$		0.122** (0.058)			0.139** (0.059)	
$BC \times \mathbb{I}_g^{diff,lib}$			0.097** (0.047)			0.095** (0.047)
Obs.	174,352	174,352	174,352	177,584	177,584	177,584
Adj. R2	0.852	0.852	0.852	0.853	0.853	0.853
FE	Seller×HS, Seller×HS, Seller×HS, Seller×HS, Seller×HS, Seller×HS, buyer, buyer, buyer, buyer, buyer, buyer, month×HS month×HS month×HS month×HS, month×HS, month×HS, origin×dest.origin×dest.origin×dest.					

Notes: This table shows the results of estimating Equation 3.4. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales refer to the sales of specific good g for a given pair of seller and buyer. Good g is defined according to the 6-digit HS classification. Sales were trimmed by 4-digit HS code at 5 and 95 percent. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are two-way clustered at the seller and 4-digit HS level. Standard errors in parentheses. The higher the Bhattacharyya coefficient, the more culturally closer two firms are. The number of observations varies between specifications due to the dropping of observations separated by a fixed effect [Correia et al., 2019]. $\mathbb{I}_g^{diff,con}$ indicates the good g is a differentiated one according to the conservative classification of Rauch [1999]. $\mathbb{I}_g^{diff,lib}$ indicates the good g is a differentiated one according to the liberal classification of Rauch [1999].

mechanisms that substitute the formal ones. Here, cultural proximity arises as a substitute for the trust and enforcement a well-functioning contract would have provided [Munshi, 2019, 2014].

To test this channel in our firm-to-firm setting, we use data from Ash et al. [2021], and calculate two different measures of court quality in between 2010 and 2018. First, for each court at the district level, we calculate the average number of days in between the filing of a case and its first hearing, $days^{first}$. Second, for each of these courts, we also calculate the average number of days between the filing and decision of a case, $days^{decision}$. The intuition is that the larger these two measures, the more congested and/or the worse contract enforcement is.

However, our data is at the pairwise level, so we need to construct a measure that captures the quality of the district court that corresponds to the seller on one side and to the buyer on the other side. Thus, we take the average between the quality of the district courts of the seller and buyer, to obtain measures $\overline{days^{first}}(\nu, \omega)$ and $\overline{days^{decision}}(\nu, \omega)$. Formally, we have

$$\overline{days^{first}}(\nu, \omega) = \frac{days^{first}(\nu) + days^{first}(\omega)}{2}, \quad (3.5)$$

and

$$\overline{days^{decision}}(\nu, \omega) = \frac{days^{decision}(\nu) + days^{decision}(\omega)}{2}, \quad (3.6)$$

Lastly, to simplify our analysis, we create a discrete measure of pairwise court quality, where we define dummy variables

$$\mathbb{I}^{court, first}(\nu, \omega) = \begin{cases} 1 & \overline{days^{first}}(\nu, \omega) \text{ in upper quartile} \\ 0 & \text{otherwise} \end{cases}, \quad (3.7)$$

$$\mathbb{I}^{court, decision}(\nu, \omega) = \begin{cases} 1 & \overline{days^{decision}}(\nu, \omega) \text{ in upper quartile} \\ 0 & \text{otherwise} \end{cases}, \quad (3.8)$$

such that the pairs of worst-performing courts have a value of 1.

With these definitions, we turn to estimate the following specification, first for a sample of only differentiated products, and then for a sample of exclusively non-differentiated products:

$$\ln n_g(\nu, \omega, t) = \iota_{\nu g} + \iota_{gt} + \iota_{\omega} + \delta BC(\nu, \omega) + \xi_1 (BC(\nu, \omega) \times \mathbb{I}^{court}(\nu, \omega)) + \eta \Xi(\nu, \omega) + \epsilon_g(\nu, \omega), \quad (3.9)$$

Table 3.4 shows the results of estimating Equation 3.9. Columns 1 and 7 show that cultural proximity is important overall for differentiated goods. In contrast, we do not find such an effect in Columns 4 and 7 for non-differentiated goods. This result echoes those of Table 3.3.

However, in Columns 2, 3, 8, and 9 we find that the impact of cultural proximity on trade is particularly pronounced in places with court congestion for differentiated products. However, this effect of court quality is absent for non-differentiated products, as Columns 5, 6, 11, and 12 show.

We interpret this as a suggestion that firms rely on cultural proximity as a source of trust in places where institutions do not work well. In other words, cultural proximity acts as an informal substitute for the trust that a well-functioning contract would provide, as argued by Nunn [2007]. Differentiated goods are branded and specific to certain producing firms. In a country with market imperfections, firms can easily renege on their commitments. This is particularly exacerbated in regions with poor court quality and low contract enforcement. Unlike homogeneous goods, firms in differentiated good markets are not easily replaceable. As a result, firms buying or selling differentiated goods will only trade with firms they know and trust, and perhaps are culturally close.¹⁶ As we show, this is particularly exacerbated in areas where the court system is delayed

¹⁶This relates to Rauch [1999], who mentions that search frictions (i.e. having to look for a trustworthy supplier) are more important to the trade of differentiated goods than to the trade of non-differentiated goods.

and backed up.

Complexity. To further investigate how differentiated and complex products are particularly reliant on caste-based trade, we analyze how the cultural proximity results vary by the number of varieties of goods bought and sold by firms. We first count how many 4-digit HS codes a firm buys or sells. Then we estimate

$$\ln y(\nu, \omega) = \iota_\nu + \iota_\omega + \delta BC(\nu, \omega) + \eta \Xi(\nu, \omega) + \zeta_0 var + \zeta_1 (BC(\nu, \omega) \times var) + \varepsilon(\nu, \omega), \quad (3.10)$$

where var_ν^{sold} and var_ν^{bought} refer to the number of varieties sold and bought by the seller, while var_ω^{sold} and var_ω^{bought} refer to the number of varieties sold and bought by the buyer.

Table 3.5 presents the results, which point to the effects of cultural proximity on trade being stronger when firms buy and sell more varieties. Our interpretation of these findings is that firms that buy and sell more varieties of goods have to face more contracting frictions, caused by having to negotiate more contracts. Then, these firms, to minimize their load of contracting frictions, will rely more on trading with counterparts that they trust. Moreover, this explanation based on trust is compatible with the results related to differentiated goods and contract enforcement. In both cases, we posit that the intensity of trade is driven by trust between firms, to overcome market imperfections in India.

Cancellations. Our data is unique in that it records canceled transactions as well. Among the diverse reasons for which cultural proximity could affect trade, we can also study reneged contracts. In this section, we analyze whether it becomes more likely pairs of firms will cancel their transactions if they are far in cultural terms. We estimate the following specification:

$$\mathbb{I}_g^{cancel}(\nu, \omega, t) = \iota_{\nu g} + \iota_{gt} + \iota_\omega + \delta BC(\nu, \omega) + \eta \ln dist(\nu, \omega) + \epsilon_g(\nu, \omega, t), \quad (3.11)$$

where $\mathbb{I}_g^{cancel}(\nu, \omega, t)$ is an indicator that says if there was at least one canceled transaction going from firm ν to firm ω of good g (at the 6-digit HS classification) in month t , $\iota_{\nu g}$ is a seller-good fixed effect, ι_{gt} is a good-month fixed effect and ι_ω is a seller-level fixed effect. Here, we control for the month of the year to account for macro events that could have caused widespread cancellations.

Table 3.6 presents the results. We find that the closer firms are in cultural terms, the less likely it is that there will be a cancellation. Here we must highlight that cancellations can occur for reasons other than reneged contracts.

Table 3.4: Effect of cultural proximity on trade interacted by court quality, intensive margin

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable	Log Sales	Log Sales	Log Sales	Log Sales	Log Sales	Log Sales
Diff. good class.	Conserv.	Conserv.	Conserv.	Conserv.	Conserv.	Conserv.
Type of good	Diff.	Diff.	Diff.	Non-diff.	Non-diff.	Non-diff.
<i>BC</i>	0.093*** (0.030)	0.044 (0.036)	0.053 (0.035)	0.023 (0.055)	0.020 (0.039)	-0.005 (0.035)
$BC \times \mathbb{I}^{court,first}$		0.219** (0.100)			0.008 (0.133)	
$BC \times \mathbb{I}^{court,decision}$			0.169* (0.097)			0.066 (0.149)
Obs.	122,793	122,793	122,793	46,735	46,735	46,735
Adj. R2	0.837	0.837	0.837	0.849	0.849	0.849
FE	Seller×HS, Seller×HS, Seller×HS, Seller×HS, Seller×HS, Seller×HS, buyer, buyer, buyer, buyer, buyer, buyer, month×HS, month×HS, month×HS, month×HS, month×HS, month×HS, origin×dest.origin×dest.origin×dest.origin×dest.origin×dest.origin×dest.					
	(7)	(8)	(9)	(10)	(11)	(12)
Dep. Variable	Log Sales	Log Sales	Log Sales	Log Sales	Log Sales	Log Sales
Diff. good class.	Liberal	Liberal	Liberal	Liberal	Liberal	Liberal
Type of good	Diff.	Diff.	Diff.	Non-diff.	Non-diff.	Non-diff.
<i>BC</i>	0.104*** (0.033)	0.045 (0.040)	0.051 (0.039)	-0.029 (0.047)	-0.050 (0.036)	-0.060* (0.035)
$BC \times \mathbb{I}^{court,first}$		0.283** (0.120)			0.066 (0.107)	
$BC \times \mathbb{I}^{court,decision}$			0.227** (0.113)			0.091 (0.116)
Obs.	115,459	115,459	115,459	54,000	54,000	54,000
Adj. R2	0.830	0.830	0.830	0.853	0.853	0.853
FE	Seller×HS, Seller×HS, Seller×HS, Seller×HS, Seller×HS, Seller×HS, buyer, buyer, buyer, buyer, buyer, buyer, month×HS, month×HS, month×HS, month×HS, month×HS, month×HS, origin×dest.origin×dest.origin×dest.origin×dest.origin×dest.origin×dest.					

Notes: This table shows the results of estimating Equation (3.9). ***, ** and * indicate statistical significance at the 99, 95 and 90 percent level respectively. Columns 1-3 shows results for differentiated goods according to the conservative classification of Rauch [1999]. Columns 4-6 shows results for non-differentiated goods according to the conservative classification of Rauch [1999]. Columns 7-9 shows results for differentiated goods according to the liberal classification of Rauch [1999]. Columns 10-12 shows results for non-differentiated goods according to the liberal classification of Rauch [1999]. Good g is defined according to 6-digit HS classification. Sales trimmed by 4-digit HS code at 5 and 95 percent. Standard errors two-way clustered at the seller and 4-digit HS level. Standard errors in parentheses. The higher the Bhattacharyya coefficient, the culturally closer two firms are. Number of observations varies between specifications due to the dropping of observations separated by a fixed effect [Correia et al., 2019]. $\mathbb{I}^{court,first}$ indicates if the mean of the average number of days in between a court filing and first hearing for the origin-district court and the destination-district court is above the 75th percentile of the distribution. $\mathbb{I}^{court,decision}$ indicates if the mean of the average number of days in between a court filing and decision for the origin-district court and the destination-district court is above the 75th percentile of the distribution.

Table 3.5: Effect of cultural proximity on trade by number of varieties, intensive margin

	(1)	(2)	(3)	(4)
Dep. Variable	Log Sales	Log Sales	Log Sales	Log Sales
BC	0.111*** (0.040)	0.090** (0.040)	0.107*** (0.035)	0.097** (0.039)
$BC \times var_{\nu}^{sold}$	0.089 (0.126)			
$BC \times var_{\nu}^{bought}$		0.121 (0.084)		
$BC \times var_{\omega}^{sold}$			0.112** (0.051)	
$BC \times var_{\omega}^{bought}$				0.068 (0.043)
Obs.	32,843	32,843	32,843	32,843
Adj. R2	0.410	0.410	0.410	0.410
FE	Seller, buyer, origin $\times$ dest.	Seller, buyer, origin $\times$ dest.	Seller, buyer, origin $\times$ dest.	Seller, buyer, origin $\times$ dest.

	(5)	(6)	(7)	(8)
Dep. Variable	Log Transactions	Log Transactions	Log Transactions	Log Transactions
BC	0.056* (0.032)	0.030 (0.032)	0.056* (0.029)	0.042 (0.032)
$BC \times var_{\nu}^{sold}$	0.095 (0.105)			
$BC \times var_{\nu}^{bought}$		0.141** (0.067)		
$BC \times var_{\omega}^{sold}$			0.104** (0.042)	
$BC \times var_{\omega}^{bought}$				0.071** (0.036)
Obs.	32,843	32,843	32,843	32,843
Adj. R2	0.356	0.357	0.357	0.357
FE	Seller, buyer, origin $\times$ dest.	Seller, buyer, origin $\times$ dest.	Seller, buyer, origin $\times$ dest.	Seller, buyer, origin $\times$ dest.

Notes: This table shows the results of estimating Equation 3.10. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales and transactions refer to the total sales and total number of transactions for a given pair of seller and buyer, respectively. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are two-way clustered at the seller and buyer level. Standard errors in parentheses. The higher the Bhattacharyya coefficient, the culturally closer two firms are. $varieties_{\nu}^{sold}$ and $varieties_{\nu}^{bought}$ refer to the number of different HS codes at the 4-digit level sold and bought by the seller divided by 100, respectively. $varieties_{\omega}^{sold}$ and $varieties_{\omega}^{bought}$ refer to the number of different HS codes at the 4-digit level sold and bought by the buyer divided by 100, respectively.

Table 3.6: Cancellations

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable	Cancellation indicator (0/1)					
<i>BC</i>	-0.006* (0.003)	-0.006* (0.003)	-0.006* (0.003)	-0.005* (0.003)	-0.000 (0.002)	-0.005* (0.003)
Obs.	252,191	252,191	248,192	256,819	258,481	252,829
Adj. R2	0.102	0.102	0.110	0.102	0.0695	0.108
FE	Seller×HS, Seller×HS, Seller×HS, Seller×HS, Seller×HS, Seller×HS, buyer buyer, buyer, buyer, buyer, buyer, buyer, month month×HS origin×dest. month, month×HS, origin×destorigin×dest.					

Notes: This table shows the results of estimating Equation 3.11 at the transaction level. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Dependent variable is a dummy variable that is equal to 1 if there was at least 1 cancellation of a transaction involving good g for a given pair of seller and buyer. Good g is defined according to 6-digit HS classification. Sales are trimmed by 4-digit HS code at 5 and 95 percent. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are two-way clustered at the seller and 4-digit HS level. Standard errors are in parentheses. The higher the Bhattacharyya coefficient, the culturally closer firms are. The number of observations varies between specifications due to the dropping of observations separated by a fixed effect [Correia et al., 2019].

3.3.1.2 Preference-based mechanisms and discrimination.

Social hierarchies. To investigate the importance of vertical hierarchies and discrimination across cultural groups, we study whether there are asymmetric effects in transactions in which one firm is placed higher than the other based on the Varna-based hierarchy. This is one way of testing for preference-based discrimination across the social hierarchy. We define the Varna or religion for which a firm has the highest probability of belonging to.¹⁷ We do not find evidence that hierarchies (and preference-based discrimination) across social groups matter for our cultural proximity results.

We use of two different indicators: $\mathbb{I}_{\nu_H \omega_L}$ and $\mathbb{I}_{\nu_L \omega_H}$. The first one captures that the seller belongs to a higher hierarchy than the buyer. The second indicates the seller is placed below the buyer in the social hierarchy. We include these two indicators by interacting them with our measure of cultural proximity.

For the intensive margin, we estimate

$$\ln y(\nu, \omega) = \iota_\nu + \iota_\omega + \delta BC(\nu, \omega) + \eta \Xi(\nu, \omega) + \zeta_1 (BC(\nu, \omega) \times \mathbb{I}_{\nu_H \omega_L}) + \zeta_2 (BC(\nu, \omega) \times \mathbb{I}_{\nu_L \omega_H}) + \varepsilon(\nu, \omega), \quad (3.12)$$

¹⁷While the Varna-based hierarchy only relates to the Hindu religion, we also place other religions in this hierarchy based on their income levels.

Table 3.7: Effect of cultural proximity on trade by vertical hierarchies, intensive and extensive margins

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable	Log Sales	Log Trans- actions	Log Sales	Log Trans- actions	Trade Indicator	Trade Indicator
BC	0.099*** (0.034)	0.068** (0.028)	0.129*** (0.035)	0.079*** (0.029)	0.0010*** (0.0001)	0.0010*** (0.0001)
$BC \times \mathbb{I}_{\nu_H \omega_L}$	0.023 (0.113)	0.097 (0.091)	0.008 (0.116)	0.072 (0.092)	-0.0002 (0.0003)	-0.0003 (0.0003)
$BC \times \mathbb{I}_{\nu_L \omega_H}$	0.045 (0.128)	-0.076 (0.102)	-0.027 (0.129)	-0.123 (0.103)	-0.0002 (0.0002)	-0.0004 (0.0002)
Obs.	30,997	30,997	31,119	31,119	5,456,512	5,477,548
Adj. R2	0.418	0.360	0.412	0.357	0.614	0.0107
FE	Seller, buyer	Seller, buyer	Seller, buyer, origin×dest	Seller, buyer, origin×dest.	Seller, buyer	Seller, buyer, origin×dest.

Notes: Columns 1, 2, 3 and 4 show the results of estimating Equation 3.12. Columns 5 and 6 show the results of estimating Equation 3.13. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales and transactions refer to the total sales and total number of transactions for a given pair of seller and buyer, respectively. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are two-way clustered at the seller and buyer level. Standard errors are in parentheses. The higher the Bhattacharyya coefficient, the culturally closer the two firms are. The number of observations varies between specifications due to the dropping of observations separated by a fixed effect [Correia et al., 2019]. The subindex that accompanies ν denotes the hierarchical position of the seller, while the subindex that accompanies ω denotes the hierarchical position of the buyer. H denotes a higher position and L denotes a lower position. The baseline category is when both firms have the same hierarchical position.

while for the extensive margin we estimate

$$tr(\nu, \omega) = \iota_\nu + \iota_\omega + \delta BC(\nu, \omega) + \eta \Xi(\nu, \omega) + \zeta_1 (BC(\nu, \omega) \times \mathbb{I}_{\nu_H \omega_L}) + \zeta_2 (BC(\nu, \omega) \times \mathbb{I}_{\nu_L \omega_H}) + \varepsilon(\nu, \omega, t). \quad (3.13)$$

Table 3.7 presents the results for the intensive and extensive margins. The baseline category is that both firms belong to the same hierarchy. First, we find the baseline coefficient is very similar to those of Table 3.1. Second, we find there is no additional effect of cultural proximity when firms are placed differently in the hierarchy. We conclude that strong asymmetric effects caused by vertical discrimination across cultural groups are unlikely. The effect of cultural proximity is similar, whether or not the firms trading belong to the same or different hierarchies.

Age of firms. Preference-based discrimination is not profit maximizing, and is likely to lead to discriminating firms exiting the market. We follow Becker [1957] to analyze whether taste-based

discrimination is behind our main findings. If there is taste-based discrimination, then we should see that firms that sell to culturally close firms at lower prices are willing to forego profits because of their preferences. A consequence would be that these firms are more prone to go bankrupt.

For our empirical analysis, we leverage information on the establishment date from IndiaMART (the date on which a firm was established) and registration date from the tax authority (the data in which a firm obtained its permit to trade). If there is taste-based discrimination, then we should see older firms relying less on cultural proximity. This would mean that firms that had a preference for selling to firms culturally close to them eventually went bankrupt, while the survivors were those firms that did not show these preferences. Here, we estimate

$$\ln y(\nu, \omega) = \iota_\nu + \iota_\omega + \delta BC(\nu, \omega) + \eta \Xi(\nu, \omega) + \zeta(BC(\nu, \omega) \times \ln(\text{age}(\nu))) + \varepsilon(\nu, \omega), \quad (3.14)$$

where $\text{age}(\nu)$ is the age of seller ν .

Table 3.8 shows the results for a modified version of the intensive margin regressions according to Equation 3.1. If there was taste-based discrimination, then the interaction between the measure of cultural proximity and age should have a negative coefficient. However, we find weak evidence for taste-based discrimination, such that we cannot conclude this is the reason behind our results.

3.3.1.3 Cultural-group cartels

We explore whether our empirical patterns can be rationalized by a narrative of certain cultural-group cartels controlling a certain sector or market location, and excludes outsiders. This would imply different normative conclusions, as stabilizing cartels would lead to an efficiency loss. While not widely highlighted in the literature, certain case studies exist that may tangentially support these narratives. Such as the example of the Gounders community's specialized knitted garment production in Tirupur [Munshi, 2019].

For this analysis, we first measure the Herfindahl-Hirschman index (HHI) of cultural-group concentration, and test the importance of it in moderating the impact of cultural proximity. We measure the HHI across castes at the product level HHI^{good} , and then the location level $HHI^{district}$. Then, we assign each seller ν to their corresponding caste's HHI. The idea is to test if products or places with more market concentration (where cartels may survive and be important) have a greater positive impact of cultural proximity.

Furthermore, in order to simplify the interpretation of our results, we define the dummy variable

$$\mathbb{I}^{HHI}(\nu) = \begin{cases} 1 & HHI(\nu) \text{ in upper half} \\ 0 & \text{otherwise} \end{cases}, \quad (3.15)$$

Table 3.8: Effect of cultural proximity after controlling for establishment age of sellers, intensive margin

	(1)	(2)	(3)	(4)
Dep. Variable	Log Sales	Log Transactions	Log Sales	Log Transactions
<i>A) IndiaMART</i>				
<i>BC</i>	0.734** (0.355)	0.489* (0.296)	0.800** (0.371)	0.479 (0.311)
$BC \times \ln(\text{age}(\nu))$	-0.199* (0.111)	-0.124 (0.090)	-0.207* (0.112)	-0.122 (0.091)
Obs.	6,334	6,334	5,859	5,859
Adj. R2	0.428	0.303	0.387	0.237
<i>B) Tax Authority</i>				
<i>BC</i>	0.164 (0.115)	0.217** (0.094)	0.150 (0.116)	0.233** (0.097)
$BC \times \ln(\text{age}(\nu))$	-0.032 (0.050)	-0.076* (0.041)	-0.016 (0.050)	-0.082** (0.041)
Obs.	18,268	18,268	18,810	18,810
Adj. R2	0.406	0.333	0.403	0.332
FE	Seller, buyer	Seller, buyer	Seller, buyer, origin $\times$ dest.	Seller, buyer, origin $\times$ dest.

Notes: Columns 1, 2, 3 and 4 show the results of estimating Equation 3.14. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales and transactions refer to the total sales and total number of transactions for a given pair of seller and buyer, respectively. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are two-way clustered at the seller and buyer level. Standard errors are in parentheses. The higher the Bhattacharyya coefficient, the culturally closer two firms are. Number of observations varies between specifications due to the dropping of observations separated by a fixed effect [Correia et al., 2019]. In Panel A, the age of the seller comes from data reported in IndiaMART. In Panel B, the age of the seller comes from data reported by the tax authority.

Table 3.9: Effect of cultural proximity on trade interacted with market concentration, intensive margin

	(1)	(2)	(3)	(4)
Dep. Variable	Log Sales	Log Sales	Log Sales	Log Sales
BC	0.069** (0.032)	0.083*** (0.030)	0.062** (0.030)	0.078*** (0.027)
$BC \times \mathbb{I}^{HHI,good}$	0.006 (0.046)		0.010 (0.049)	
$BC \times \mathbb{I}^{HHI,district}$		-0.043 (0.059)		-0.049 (0.051)
Obs.	226,039	225,775	229,719	229,719
Adj. R2	0.853	0.853	0.854	0.854
FE	Seller×HS, Seller×HS, Seller×HS, Seller×HS, buyer, buyer, buyer, buyer, month×HS month×HS month×HS, month×HS, origin×dest. origin×dest.			

Notes: This table shows the results of estimating Equation (3.16). ***, ** and * indicate statistical significance at the 99, 95 and 90 percent level respectively. Sales refer to the sales of specific good g for a given pair of seller and buyer. Good g is defined according to 6-digit HS classification. Sales trimmed by 4-digit HS code at 5 and 95 percent. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors two-way clustered at the seller and 4-digit HS level. Standard errors in parentheses. The higher the Bhattacharyya coefficient, the culturally closer two firms are. Number of observations varies between specifications due to the dropping of observations separated by a fixed effect [Correia et al., 2019]. $\mathbb{I}^{HHI,good}$ indicates the caste of the seller is in the upper half of the good-based HHI distribution. $\mathbb{I}^{HHI,district}$ indicates the caste of the seller is in the upper half of the district-based HHI distribution.

that is, the dummy variable will be equal to 1 for those firms that belong to the castes with higher HHI.

Then, we estimate the following regression:

$$\ln n_g(\nu, \omega, t) = \iota_{\nu g} + \iota_{gt} + \iota_{\omega} + \delta BC(\nu, \omega) + \xi(BC(\nu, \omega) \times \mathbb{I}^{HHI}(\nu)) + \eta \Xi(\nu, \omega) + \epsilon_g(\nu, \omega). \quad (3.16)$$

Table 3.9 shows that market concentration has little effect on the impact of cultural proximity on sales. While we find little evidence of cartels driving our effects, we are unable to rule out that such forces may not be at play for certain products and locations.

3.3.1.4 Family ties and names

We further explore how family ties may facilitate trade. That is, we analyze how much of the cultural proximity picks up larger family links.

Here, since we have names of individuals, we can create a measure of how similar names are and see their relationship with trade. Indian names have various spellings (e.g., Bhattacharyya and

Bhataccharya), and so we require a measure of distance between names. We first define name distance measure $dist^{name}$ as follows

$$dist^{name}(\nu, \omega) = \frac{L(\nu, \omega)}{\max(\text{len}(\nu), \text{len}(\omega))},$$

where the denominator $L(\nu, \omega)$ is the Levenshtein distance, and the numerator is the maximum in between the length of the last name of seller ν and buyer ω .

The Levenshtein distance measures how many changes have to be made to the seller's last name in order to match the buyer's last name (i.e. changing a character, erasing a character, adding a character). Moreover, we divide this distance measure to normalize it to the number of characters contained in the last names of the seller and buyer. Thus, the larger the distance measure between names $dist^{name}$, the more different the last names of the seller and buyer are.

To see if our cultural proximity measure $BC(\omega, \nu)$ contains information that goes beyond firm owners sharing a last name, we modify Equation 3.1 to now include the name distance measure. Therefore, we estimate:

$$\ln y(\nu, \omega) = \iota_\nu + \iota_\omega + \delta BC(\nu, \omega) + \eta \Xi(\nu, \omega) + \xi dist^{name}(\nu, \omega) + \varepsilon(\nu, \omega), \quad (3.17)$$

Table 3.10 shows the results of estimating Equation 3.17. Names do matter, in that the closer the names are, the more the trade. Yet, even after controlling for name distance, the result for cultural proximity and trade remains almost identical. That is, even after considering how similar the names are, cultural proximity facilitates more trade. This suggests that the cultural proximity measure $BC(\omega, \nu)$ captures information beyond firms sharing the same last name.

3.3.2 Discussion of stylized facts

The stylized facts show that higher cultural proximity between a pair of firms favors trade in both the intensive and extensive margins, as well as lowers the price of the goods they trade. We discuss the possible mechanisms that may give rise to these findings.

Contracting frictions. In Section 3.3.1.1 we argue that contracting frictions could be the reason that drives the cultural proximity results. India is a country that suffers from a severe lack of contract enforcement. *A priori*, a buyer may not know if the seller will deliver the goods under the agreed conditions (delivery, quality, etc.). Likewise, *a priori*, the seller may not know if the buyer will pay under the agreed conditions. This means buyers and sellers incur contracting frictions to find suitable trading partners [Boehm and Oberfield, 2020]. Quantity-wise and matching-wise, this lowers trade as firms must pay a matching cost. Price-wise, this increases prices as the matching

Table 3.10: Effect of cultural proximity on trade, augmented with name distance, intensive margin

	(1)	(2)	(3)	(4)
Dep. Variable	Log Sales	Log Transactions	Log Sales	Log Transactions
<i>BC</i>	0.099** (0.039)	0.067** (0.031)	0.129*** (0.039)	0.083*** (0.031)
$\ln dist$	-0.020 (0.017)	-0.062*** (0.013)		
$dist^{name}$	-0.378*** (0.094)	-0.303*** (0.075)	-0.415*** (0.093)	-0.290*** (0.073)
Obs.	24,057	24,057	23,999	23,999
Adj. R2	0.412	0.364	0.404	0.362
FE	Seller, buyer	Seller, buyer	Seller, buyer, origin $\times$ dest.	Seller, buyer, origin $\times$ dest.

Notes: This table shows the results of estimating Equation (3.17). ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales and transactions refer to the total sales and total number of transactions for a given pair of seller and buyer, respectively. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors two-way clustered at the seller and buyer level. Standard errors in parentheses. The higher the Bhattacharyya coefficient, the culturally closer two firms are. Number of observations varies between specifications due to the dropping of observations separated by a fixed effect [Correia et al., 2019]. The larger the name distance $dist^{name}$, the more different the names of the firm owners are.

cost is passed down by the sellers to the aforementioned prices.

In this case, cultural proximity can work as a proxy for information and trust: culturally close firms may know and trust each other, and informally enforce contracts with social and reputational pressures. The higher the cultural proximity, the lower the contracting frictions. As a result, there would be more trade and lower prices, consistent with our previous findings. In Section 3.4.1, we present a simple theoretical framework in which cultural proximity affects contracting frictions and affects trade and prices. Our model suggests that if contracting frictions drive initial trade barriers, cultural proximity may reduce such frictions.

Preference-based mechanisms and discrimination. We argue the results are unlikely to emerge from buyers having an inherent preference for buying from sellers culturally close to them. This preference would be a demand shifter that is active for those sellers that are close in cultural terms. While this would certainly increase the quantity traded, it would increase the price of traded goods, a result inconsistent with our previous findings.

The stylized facts can arise from having sellers that show a preference for selling to culturally close buyers. This would imply increased supply for those buyers who are culturally close to the seller. However, we do not find conclusive evidence of this channel.

Discrimination from high-caste cultural groups against low-caste cultural groups may again reduce trade. Yet, we find there is no additional effect of cultural proximity when firms are placed

differently in the hierarchy. As such, we detect no asymmetric effects caused by vertical discrimination across cultural groups.

Cultural-group cartels. For the cultural-group cartels driving our results, while we cannot, conclusively rule out this narrative, we find little evidence in support of it.

We begin with the empirical assumption that cartels are more likely to have market power on a particular set of products or locations. It is challenging for a cartel to exist in a broad variety of industries and locations all over the country; as they may need substantial market power in many industries and regions. Yet, in Table 3.5 we show that, if anything, our results are stronger when more products are transacted – perhaps as more products imply more contracting frictions.

Furthermore, if cartels specialize in producing certain goods, in Appendix Table C.8 we find evidence that the effects of cultural proximity are not particularly stronger when firms sell the goods in which their own castes specialize in. Relatedly, even if the cartel is not focused on certain products, it may have a presence in a broader industry. But in controlling for seller-industry-by-buyer-industry fixed effects, our results are similarly robust, as Appendix Table C.9 shows.

In a similar vein, it is unlikely that these cartels emerge out of regional concentration and market power, as our regressions include origin-by-destination fixed effects, which implies we are looking within location pairs.

Lastly, our results are also inconsistent with the insight that cartels often consist of entrenched old firms. Table 3.8 shows that, for older firms, cultural proximity matters less, perhaps as they have smoothed out contracting frictions over time.

Family ties and names. In this part, we argue that cultural proximity encodes information that goes beyond firm owners sharing a name.

First of all, as our results in Section 3.3.1.4 show, firm owners having similar last names does matter for trade. The closer the two last names are, the more trade there will be. However, the results also show that after controlling for how similar names are, the main result of cultural proximity over trade remains.

This suggests that cultural proximity contains more information than just people sharing a name or having family ties. For instance, assume there are two firm owners that belong to the same caste (or even the same family) but have different last names. Our cultural proximity measure captures the fact the fact these two people from the same caste but with different names are culturally close, whereas the simple name distance measure would wrongly classify them as culturally distant.

3.3.3 Robustness

We examine alternative specifications and heterogeneity in responses that shed light on various other channels in Appendix C.2.

Correction for selection bias. Helpman et al. [2008] shows that the standard gravity equation estimations are biased as they do not account for selection issues. We follow their suggested correction in Appendix Tables C.3 and C.4. As the correction mentions, we need an excluded instrument that affects only the extensive margin (i.e. the matching cost) and not the intensive margin (i.e. the trade cost). We consider the participation of both seller and buyer in the IndiaMART online B2B platform, under the idea that online platforms should reduce their matching costs. The results show that the coefficients are downward biased if we do not account for the selection issues. Therefore, our main results in the paper provide a lower bound on the effect that cultural proximity has on the intensive margin of trade.

Alternative cultural proximity measure. As an alternative to the Bhattacharyya coefficient, we perform estimation exercises using a symmetric version of the Kullback and Leibler [1951] divergence. Tables C.5 and C.6 show our baseline findings are robust to this alternative cultural proximity measure.

Language. We test whether the results we find are driven by linguistic similarity. To do so, we follow the two linguistic distance measures from Kone et al. [2018]. Appendix Table C.7 shows that language does not affect the cultural proximity results already established.

Goods specialization. Cultural groups in India are, in many cases, defined by the production of specific goods [Munshi, 2019].¹⁸ Therefore, we analyze if the reason behind the cultural proximity results is cultural groups specializing in the production of certain goods and, given this, forming special bonds with their specific set of buyers. In Appendix Table C.8 we do not find evidence of good specialization driving the results. This means that cultural proximity matters for all types of goods: for those in which a cultural group specializes and for those in which a cultural group does not specialize too.

Industry pair linkages. In the production matrix of an economy, some sectors are more likely to trade with others because of the nature of their activities. The same cultural group may happen to participate in the same industry. In Appendix Table C.9, we present the results for the intensive margin after adding an industry of seller $\times$ industry of buyer fixed effect. We find that the result of there being more trade between culturally close firms prevails.

3.4 Quantitative Importance of Cultural Proximity

In this section, we perform a counterfactual analysis to quantify the importance of cultural proximity for trade and welfare. First, we describe the model. Then, we describe how we estimate and calibrate the model. Finally, we perform three counterfactual scenarios. In particular, we show the

¹⁸We can also understand this as certain cultural groups specializing in certain occupations.

importance of cultural proximity by implementing (i) social inclusion/mixing policies, (ii) social isolation policies, and (iii) reduction of contracting frictions.

3.4.1 Model

Summary. We build a quantitative inter-firm trade model and cultural differences between firm owners. The economy is comprised by a representative household and a continuum of firms. The representative household has preferences over all goods in the economy. Goods are differentiated and produced by the firms. Each good can be consumed by the household as a final good, or purchased as an intermediate input by other firms. Trade between the household and firms is riskless, but trade between firms is risky such that equilibrium price between firms is subject to a risk premium. Finally, we consider an endogenous firm network, so sellers optimally match with buyers, such that the profits of matching are larger than matching costs. Following our stylized facts, both premium and matching costs are a function of the cultural proximity between firms. For further details of the model, see Appendices C.3 and C.4.

Technology. There is a continuum of firms that operate under monopolistic competition and produce differentiated goods indexed by ω . Since each firm produces a unique good, ω denotes a firm or a good. We consider a roundabout production economy, so firms can purchase intermediate inputs from all firms in the economy. Each firm produces its differentiated good with hired labor from the representative household and intermediate inputs.

Firms are *ex ante* heterogeneous in their cultural endowments. This captures the idea that a firm owner is born within a cultural setup that he inherits and then uses to interact with other firm owners. To match with our empirical section, we consider that each firm receives a vector of cultural endowments ρ_ω , so we can construct a measure of cultural proximity between firms ν and ω as $BC(\nu, \omega) = \sqrt{\rho_\nu \rho_\omega}$.

Firms operate in three steps. First, sellers endogenously choose buyers (i.e. matching). In particular, a seller ν sells to buyer ω whenever the profits of doing so are larger than $\epsilon F(\nu, \omega)$, where $F(\nu, \omega)$ are pairwise fixed costs of matching, and ϵ are i.i.d. log normal errors. Guided by our stylized facts, cultural proximity $BC(\nu, \omega)$ between seller ν and buyer ω determines the pairwise matching costs $F(\nu, \omega)$. Sellers are looking for buyers, and vice-versa. Firms can meet costlessly, so all firms can meet each other before deciding who to match with. The purpose of a meeting is to infer the fixed costs the seller would incur if they were to trade with each buyer. Culture encodes information that is useful for sellers to decide which buyer to sell to [Allen et al., 2019, Balmaceda and Escobar, 2017].¹⁹ Guided by our stylized facts, we consider matching costs

¹⁹There could be other microfoundations motivating how culture influences matching costs (e.g. risk-sharing as in

to be a function of cultural proximity. In particular, we consider an exponential function of fixed matching costs to cultural proximity, such that

$$F(\nu, \omega) = \kappa + \exp(-\gamma BC(\nu, \omega)), \quad (3.18)$$

where $\gamma > 0$ measures the sensitivity of the pairwise matching cost to the cultural proximity $BC(\nu, \omega)$, and κ is a scaling constant. The intuition behind parameter γ is that it captures institutional features, like court quality. In this sense, Appendix Table C.2 suggests that parameter γ increases in magnitude when $\mathbb{I}^{court} = 1$ and firms face weak court quality. The lower the quality of courts is, the larger parameter γ is, and the more firms rely on cultural proximity to match.

Second, upon matching, seller ν and buyer ω set up the terms of the contract. There is a risk that the buyer withhold payment after the seller ships the goods. The seller's concerns around this issue can be monetized in the premium that the seller charges to the buyer. This premium gets passed to the unit price $p(\nu, \omega)$ that results from profit maximization $p(\nu, \omega) = \mu c(\nu) d(\nu, \omega)$, where $c(\nu)$ is the marginal cost, $\mu \equiv \frac{\sigma}{\sigma-1}$ is a markup, and $d(\nu, \omega) \geq 1$ is a risk premium. Guided by our stylized facts, we posit that cultural proximity $BC(\nu, \omega)$ between seller ν and buyer ω determines $d(\nu, \omega)$.²⁰ We then consider that the premium is

$$d(\nu, \omega) = \exp(\beta BC(\nu, \omega) + \epsilon(\nu, \omega)), \quad (3.19)$$

where $\beta < 0$ is a trade cost semi-elasticity, and $\epsilon(\nu, \omega)$ are unobservables. β depends on institutional features, like court quality.²¹

Third, firms maximize profits by selling to other firms and the household subject to the firms' and the household's demand. From the demand of intermediates and firm pricing, we can obtain the gravity equation as

$$\log(n(\nu, \omega)) = \iota_\nu + \iota_\omega + (1 - \sigma) \log(d(\nu, \omega)), \quad (3.20)$$

where ι_ν and ι_ω are seller and buyer fixed effects, also known as *multilateral resistance terms*. Here, the premium $d(\nu, \omega)$ enters the gravity equation as a trade cost. This gravity equation relates directly to Equation 3.1 that we estimate.

Ambrus et al. 2014, Bloch et al. 2008). These would generate similar frictions at the extensive and intensive margin.

²⁰This is similar to earlier work on how culture can be an informal institutional channel to solve hold-up problems. This would be consistent with other microfoundations on how culture influences trade costs due to reputation [Banerjee and Duflo, 2000, Chen and Wu, 2021] or loyalty [Board, 2011].

²¹Based on the result of Table 3.4, we know model parameter β increases in magnitude when $\mathbb{I}^{court} = 1$ and firms face bad courts. As such, β relates to how firms respond to court quality: the lower the quality of the courts, the larger the magnitude of β , and the more important cultural proximity is for trade.

Preferences. A representative household demands goods with from all firms and inelastically supplies labor to firms. The household maximizes utility $\left(\int_{\omega \in \Omega} y(\omega)^{\frac{\sigma-1}{\sigma}} d\omega\right)^{\frac{\sigma}{\sigma-1}}$ subject to its budget constraint $\int_{\omega \in \Omega} p(\omega) y(\omega) d\omega \leq Y$, where $y(\omega)$ is the household demand for good ω , $p(\omega)$ is the price the household pays for good ω , and Y is total income. This generates the demand function $x(\omega) = p(\omega)^{1-\sigma} p^{\sigma-1} Y$, where $x(\omega) \equiv p(\omega) y(\omega)$, and $p \equiv \left(\int_{\omega \in \Omega} p(\omega)^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$ is a CES price index. Finally, since firm ω sells its good to the household without risk, the price is $p(\omega) = \mu c(\omega)$, where μ is the markup and $c(\omega)$ is firm ω 's marginal cost.

3.4.2 Estimation and calibration

Here we explain how we estimate the key parameters of the model on cultural endowments, (intensive) trade costs, and seller matching costs. We also describe how we calibrate the remaining parameters of the model.

Cultural endowments ρ . For the cultural endowments, we assume each firm ν has a probability vector $\rho_\nu = [\rho_\nu(1), \dots, \rho_\nu(452)]$ of belonging to each of the 452 cultural groups, we observe in the data. We further assume the elements of ρ_ν are randomly drawn from a Dirichlet distribution, such that $\rho_\nu(1), \dots, \rho_\nu(452) \sim \mathcal{D}(\alpha_1, \dots, \alpha_{452})$, where $\alpha_1, \dots, \alpha_{452} > 0$ are concentration parameters.²² The probability density for the Dirichlet distribution is

$$\rho_\nu(1), \dots, \rho_\nu(452) \sim \mathcal{D}(\alpha_1, \dots, \alpha_{452}) = \frac{\Gamma(\sum_{x=1}^{452} \alpha_x)}{\prod_{x=1}^{452} \Gamma(\alpha_x)} \prod_{k=1}^{452} \rho_\nu(x)^{\alpha_x-1},$$

such that $\rho_\nu(x) \in [0, 1]$, $\sum_{x=1}^{452} \rho_\nu(x) = 1$, where $\Gamma(\cdot)$ is the gamma function and $\frac{\Gamma(\sum_{x=1}^{452} \alpha_x)}{\prod_{x=1}^{452} \Gamma(\alpha_x)}$ is a normalization constant. To ensure the theoretical Dirichlet distribution produces draws that are similar to the probabilities we see in the data, we estimate the vector $\alpha = [\alpha_1, \dots, \alpha_{452}]$ parameters by maximum likelihood.²³ Let $\boldsymbol{\rho} = \{\rho_1, \dots, \rho_N\}$, where N is the total number of firms. Then, the log-likelihood function is

$$\begin{aligned} \ln pr(\boldsymbol{\rho}|\alpha) &= \mathcal{N} \ln \Gamma(\sum_{x=1}^{452} \alpha_x) - \mathcal{N} \sum_{x=1}^{452} \ln \Gamma(\alpha_x) + \\ &\quad \mathcal{N} \sum_{x=1}^{452} (\alpha_x - 1) \left(\frac{1}{N} \sum_{\nu=1}^N \ln \rho_\nu(x) \right). \end{aligned} \quad (3.21)$$

Trade costs d . From Equation 3.19 we need an estimate for β . Our setup produces a clear empirical counterpart that we already estimated in the reduced-form section, conditional on high-

²²For a given x , the higher this parameter, the more disperse the realizations of $\rho_\nu(x)$ are across firms ν .

²³For this, we use the Matlab toolboxes `fastfit` and `lightspeed` by Tom Minka. We present the estimated parameters in Appendix Figure C.2.

dimensional fixed effects. So we obtain an estimate for this parameter by linking the theoretical gravity Equation 3.20 to the empirical gravity equation results (Column 1 from Table 3.1). Thus, we obtain $\beta = -0.03$.²⁴

Matching cost F . From Equation 3.18, we need an estimate for γ . We do this in two steps. First, using the extensive margin sample we run the following estimation

$$\ln \left[n \left(z, z' \right) \right] = \iota_z + \iota_{z'} + \delta BC \left(z, z' \right) + \eta \ln \left(dist \left(z, z' \right) \right) + \varepsilon \left(z, z' \right), \quad (3.22)$$

where we apply the inverse hyperbolic sine transformation to the dependent variable, so as to not lose the cases in which there is zero trade. With this, we recover

$$\ln \left[\widehat{n \left(z, z' \right)} \right] = \widehat{\iota}_z + \widehat{\iota}_{z'} + \widehat{\delta} BC \left(z, z' \right) + \widehat{\eta} \ln \left(dist \left(z, z' \right) \right),$$

where the hats denote estimated parameters and $\ln \left[\widehat{n \left(z, z' \right)} \right]$ are the predicted sales. This variable predicts what would be the sales for a pair of seller and buyer even in the case they did not actually trade in the data. Second, we combine and rearrange Equations (C.10) and (3.18), such that

$$l \left(z, z' \right) = \int 1 \left[\ln \left(\epsilon \left(z, z' \right) \right) < \ln \left[\widehat{n \left(z, z' \right)} \right] - \ln \left(\sigma \right) - \gamma BC \left(z, z' \right) \right] dH \left(\epsilon \left(z, z' \right) \right), \quad (3.23)$$

where we use the fact that $\pi \left(z, z' \right) = \frac{n \left(z, z' \right)}{\sigma}$ and replace $\ln \left[n \left(z, z' \right) \right]$ by its estimated counterpart $\ln \left[\widehat{n \left(z, z' \right)} \right]$.²⁵ We estimate this last equation with a probit (assuming $\epsilon \left(z, z' \right)$ is log-normally distributed). We find that $\gamma = -0.13$, and statistically different from 0.²⁶

Calibrated parameters and SMM. We calibrate the labor cost share $\alpha = 0.52$, the value reported for India for 2019 from the Penn World Tables [Feenstra et al., 2015]. This value also considers the informal sector, which plays a large role in India. For the markup we use $\mu = 1.34$, which is the median markup across all Indian sectors reported by De Loecker et al. [2016]. This markup implies an elasticity of substitution across suppliers $\sigma = 3.94$. Following Bernard et al. [2022] we normalize the total number of workers $L = 1$, take the nominal wage as the numeraire

²⁴Even though the wedge also appears in the price Equation C.11 of the model, we do not estimate this equation to identify β . The reason is that the price equation is not an equilibrium equation, while the gravity equation is. Also, for our simulations we add a constant to the trade cost, such that the minimum trade cost is equal to 1. Therefore, in our simulations we have $d \left(\nu, \omega \right) = \exp \left(-\beta + \beta BC \left(\nu, \omega \right) \right)$.

²⁵For these estimations we ignore the scaling constant κ that appears in Equation 3.18.

²⁶We present the results of the estimation in Appendix Table C.2. Also, for our simulations we add a constant to the matching cost, such that the minimum matching cost is equal to κ . Therefore, in our simulations we have $F \left(z, z' \right) = \kappa + \exp \left(-\gamma + \gamma BC \left(z, z' \right) \right)$.

so $w = 1$, and set the total number of firms $\mathcal{N} = 400$.

For the log-productivity distribution, we assume a mean $\mu_{\ln(z)} = 0$. The remaining parameters are (i) the standard deviation of the log-productivity distribution $\sigma_{\ln(z)}$ and (ii) the mean $\mu_{\ln(\epsilon)}$, (iii) the standard deviation $\sigma_{\ln(\epsilon)}$ of the link function noise distribution and (iv) the scaling constant for the pairwise matching cost κ . We estimate these four parameters so as to match targeted moments from the data, using a simulated method of moments (SMM). We explain this procedure below.

In Appendix C.5, we describe details on the targeted and untargeted moments we use. We show that, when it comes to the targeted moments, the model can very closely replicate the empirical ones. For the untargeted moments, the model gets reasonably close to the data.

3.4.3 Counterfactual analysis

We now present the results of various counterfactual exercises. First, we evaluate the effects of social mixing/inclusion and isolation policies, whereby we change the cultural proximity between firms (in our model terms, changing $BC(z, z')$). Second, we study the effects of a policy that reduces contracting frictions, such that firms rely less on cultural proximity when trading (in terms of our model, shrinking parameters β and γ).

To evaluate each scenario, we measure what happens to various model-based statistics. Welfare is measured by real wage, $\mathcal{W} = \frac{w}{P}$. To quantify the impact on aggregate productivity, we consider a sales-weighted average productivity measure such that $\mathcal{Z} = \left(\sum_{\nu=1}^N \phi_{\nu} z_{\nu}^{\sigma-1} \right)^{\frac{1}{\sigma-1}}$, where ϕ_{ν} represents the proportion of the sales of firm ν over the total sales of the economy.

To analyze the impact on the total economic activity, we measure total sales $\mathcal{S} = \sum_{\nu=1}^N S_{\nu}$, where S_{ν} are the total sales of firm ν . Additionally, we consider the average normalized intermediate sales $mean \left[\ln \left(\tilde{N}(\nu) / \mathcal{N}_b(\nu) \right) \right]$, where $\tilde{N}(\nu)$ are the total intermediate sales of seller ν , and the average normalized intermediate purchases $mean \left[\ln \left(N(\omega) / \mathcal{N}_s(\omega) \right) \right]$.

For the prices, we compare the changes in the aggregate price index P . Finally, to study how matching between firms is affected, we present the results for the average normalized number of buyers, $mean \left[\ln \left(\frac{\mathcal{N}_b(\nu)}{\mathcal{N}} \right) \right]$, and the average normalized number of sellers, $mean \left[\ln \left(\frac{\mathcal{N}_s(\omega)}{\mathcal{N}} \right) \right]$.²⁷

Social inclusion and social mixing policies. We analyze the effects of social inclusion/mixing policies. There is an important literature on the role of social cohesion for economic development [Alesina and Giuliano, 2015, Alesina and Reich, 2015, Bazzi et al., 2019, Depetris-Chauvin et al., 2020, Gradstein and Justman, 2019, Ritzen et al., 2000]. We tie our counterfactuals to the importance of implementing affirmative action policies with the intention of increasing cultural

²⁷In contrast to the previous sections, in this part we define the aggregate measures discretely. This is due to the simulations having a discrete number of firms, rather than a continuum.

proximity [Alan et al., 2021, Alesina et al., 2021], particularly for India [Khanna, 2020, Munshi, 2019]. For instance, affirmative actions programs may help incentivize students from different cultural groups to attend the same educative institutions. If these students then go on to become owners of the firms in the future, such policies may increase cultural proximity between firms, despite the fact the owners originally belonged to different cultural groups. Similarly, affirmative action in public sector jobs may also increase connections across caste lines, as individuals from different castes now work together.

To analyze the maximum potential of this policy within our theoretical framework, we propose case Counterfactual 1 (CF1), in which all the firms belong to the same cultural group. That is, we go from the baseline to $BC(z, z') = 1$ for all z, z' , which makes the firms become the closest possible in cultural terms. In this scenario, there are no contracting frictions, as firms know and/or trust each other, and so they pay the minimum trade and matching costs.

Table 3.11 shows how the model statistics change in each counterfactual to the baseline. In case CF1, we have that firms become the closest in cultural terms, so trade costs and matching costs go to their minimum possible. Aligned with our empirical facts, with lower trade costs, total sales increase by 2.76 percent, while the average intermediate sales and purchases go up by 1.52 percent and 1.15 percent, respectively. With the lower matching costs the average number of buyers also increases. Also, because there are lower trade and matching costs, aggregate prices fall by 1.73 percent, and welfare increases by 1.76 percent.

Besides welfare, another aggregate measure we analyze is average productivity, which falls by 0.13 percent. Yet, average productivity masks substantial compositional changes, as these results depend on whether the less productive firms are selling more or less to the baseline case. We show in Table 3.12 that, in case CF1, when trade and matching costs decrease, the less productive firms match more and sell more, which increases their weight in the aggregate and lowers average productivity.

Social isolation policies. Since the rise of democracy, efforts have been put in place by the Indian government to end the influence of the caste system in the modern economy [Iyer et al., 2013, Munshi, 2019]. What would have happened if sociopolitical forces perpetuated the social stratification of the caste system? To analyze the maximum impact of social isolation policies we propose case Counterfactual 2 (CF2), where we examine an extreme case in which each firm belongs to its own cultural group. Thus, we have a case where $BC(z, z') = 0$ for all z, z' and $z \neq z'$, which makes the firms the furthest possible in cultural terms. Under this scenario, firms incur the maximum contracting frictions, for which they pay the maximum trade cost and the maximum matching cost.

When all firms are the furthest in cultural terms, trade costs and matching costs are the highest.

Table 3.11: Effect of cultural proximity on aggregate outcomes (counterfactual scenarios)

	CF1: Social inclusion/mixing	CF2: Social isolation	CF3: Reducing contracting frictions
Welfare	1.76	-1.45	0.87
Ave. productivity	-0.13	0.10	-0.06
Total sales	2.76	-2.23	1.37
Ave. normalized intermediate sales	1.52	-1.20	0.76
Ave. normalized intermediate purchases	1.15	-0.94	0.57
Ave. normalized number of buyers	1.07	-0.87	0.53
Ave. normalized number of sellers	1.00	-0.82	0.50
Agg. price index	-1.73	1.47	-0.87

Notes: We present the percentage gains or losses to the baseline scenario. CF1 is a case where all the firms belong to the same cultural group. This is, we go from the baseline to $BC(z, z') = 1$ for all z, z' , which makes the firms to become the closest possible in cultural terms. In this scenario, there are no contracting frictions, as firms know and/or trust each other, and so they pay the minimum trade and matching costs. CF2 is a case where each firm belongs to its own cultural group. Thus, we have a case where $BC(z, z') = 0$ for all z, z' and $z \neq z'$, which makes the firms the furthest possible in cultural terms. Under this scenario, firms incur the maximum contracting frictions, for which they pay the maximum trade cost and the maximum matching cost. CF3 is a scenario where trade and matching costs become less sensitive to cultural proximity. In this case parameters β and γ shrink by 50 percent.

Table 3.11 shows that in case CF2 total sales fall by 2.23 percent, and prices increase by 1.47 percent. Average intermediate sales and purchases also fall. There are also fewer matches, and the average number of buyers and sellers falls by at least 0.82 percentage points. As a result, welfare falls by 1.45 percent. Average productivity increases by 0.10 percent, relative to the baseline. Table 3.12 shows that with social isolation, every firm loses in terms of sales. However, the firms that lose the most are the least productive, which shrinks their weight in the aggregate and, thus, drives average productivity up.

Reducing contracting frictions. Now we turn to study which would be the effect of reducing contracting frictions. A policy that improves the quality of courts would reduce the contracting frictions firms face. In terms of our framework, this means that the trade cost and the matching cost become less sensitive to our measure of cultural proximity. Thus, in the Counterfactual 3 (CF3) we analyze a case where parameters β_2 and γ shrink by 50 percent. This captures how

Table 3.12: Change in sales by productivity quartiles

	CF1: Social inclusion/mixing	CF2: Social isolation	CF3: Reducing contracting frictions
1st quartile (most productive)	2.73	-2.21	1.35
2nd quartile	2.91	-2.35	1.44
3rd quartile	2.91	-2.31	1.44
4th quartile (least productive)	2.86	-2.32	1.42

Notes: We aggregate the sales of all firms that belong to a productivity quartile and calculate their percentage variation to the baseline. CF1 is a case where all the firms belong to the same cultural group. This is, we go from the baseline to $BC(z, z') = 1$ for all z, z' , which makes the firms to become the closest possible in cultural terms. In this scenario, there are no contracting frictions, as firms know and/or trust each other, and so they pay the minimum trade and matching costs. CF2 is a case where each firm belongs to its own cultural group. Thus, we have a case where $BC(z, z') = 0$ for all z, z' and $z \neq z'$, which makes the firms the furthest possible in cultural terms. Under this scenario, firms incur the maximum contracting frictions, for which they pay the maximum trade cost and the maximum matching cost. CF3 is a scenario where trade and matching costs become less sensitive to cultural proximity. In this case parameters β and γ shrink by 50 percent.

reducing contracting frictions affects aggregate outcomes via the channel of trade becoming less reliant on cultural proximity.²⁸

Table 3.11 shows that after reducing contracting frictions in case CF3, total sales rise by 1.37 percent, and prices fall by 0.87 percent. Average intermediate sales and purchases increase. The number of matches also increases, with the average number of buyers and sellers rising by about 0.5 percent. Welfare increases by 0.87 percent, and average productivity falls by 0.06 percent. In Table 3.12 we show that in reducing contracting frictions, all firms gain in terms of sales to the baseline. Nonetheless, it is the less productive firms that gain the most, such that their weight in the aggregate increases. This drives the average productivity down.

3.5 Conclusions

We shed light on how cultural proximity shapes the formation of production networks and its implications for welfare. We first provide empirical evidence on the role of cultural proximity for inter-firm trade by leveraging a new dataset of firm-to-firm transactions from a large Indian state, along with data on firm owner names and their cultural proximity derived from India's caste and religious system.

We report three new stylized facts. First, culturally closer firms report higher sales between

²⁸Reducing contracting frictions may affect aggregate outcomes through other channels as well, such as more investments in differentiated products, and more trade across longer distances.

them, on the intensive margin. Second, culturally closer firms are more likely to trade with each other, on the extensive margin. Third, firms that are culturally further apart report higher unit prices in their transactions.

We explore various mechanisms and find evidence most consistent with the importance of alleviating contracting frictions. We do not find sufficient evidence that hierarchies (and preference-based discrimination) matter, or that linguistic distance and the specialization in certain goods matter for our results. We show evidence that suggests that the effect we find of cultural proximity on trade is stronger for differentiated goods, which often rely on either formal or informal contract enforcement [Nunn, 2007, Rauch, 1999]. We also find that the importance of cultural proximity is elevated in regions with poor court quality (and so worse contract enforcement). Indeed, consistent with our narrative, the importance of court quality is only seen in trades of differentiated products, rather than homogeneous goods. We understand these results as consistent with the possibility that cultural proximity is an informal mechanism that substitutes formal contract enforcement [Munshi, 2014, 2019].

We then build a quantitative general equilibrium model of firm-to-firm trade and cultural proximity. We introduce our measure of cultural proximity as a wedge that affects trade and matching costs, and estimate the key parameters of the model: the semi-elasticity of the trade cost to cultural proximity and the semi-elasticity of matching cost to cultural proximity. The model generates estimable specifications, that we take directly to the data. We use the estimated parameters to quantify the implications for welfare and other model-based statistics of implementing different policies. Welfare increases by 1.76 percent under social inclusion policies, falls by 1.45 percent under social isolation, and increases by 0.87 percent when reducing contracting frictions makes firms less reliant on cultural proximity.

In contexts like India, cultural and social networks may be used informally to overcome the lack of formal institutions that uphold contracts. Our paper is among the first to establish the consequences of these cultural ties in the context of trade. We study how social relationships influence firm-level decisions and quantify their importance for welfare. Our results have strong implications for policy. Promoting social inclusion and mixing via diversity-friendly policies can help facilitate matches and trade, with substantial implications for aggregate output and welfare. Furthermore, investing in reducing contracting frictions will allow firms to not have to rely on cultural ties, facilitating matches with more productive and low-cost suppliers, and once again improving economic well-being.

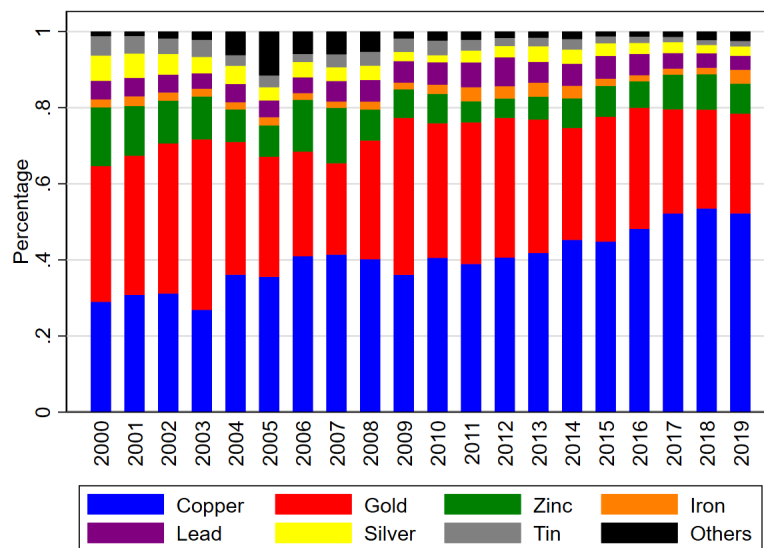
APPENDIX A

Appendix to Chapter 1

A.1 Empirical Appendix

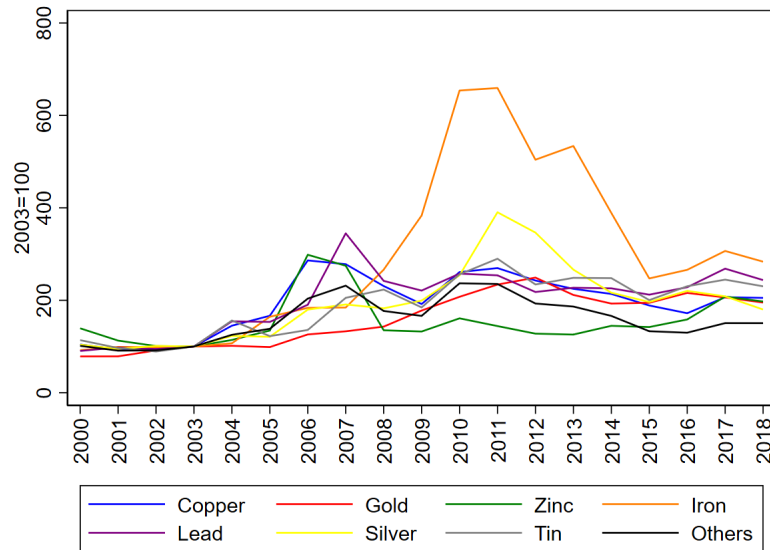
A.1.1 Mining Sector

Figure A.1: Mining exports by mineral



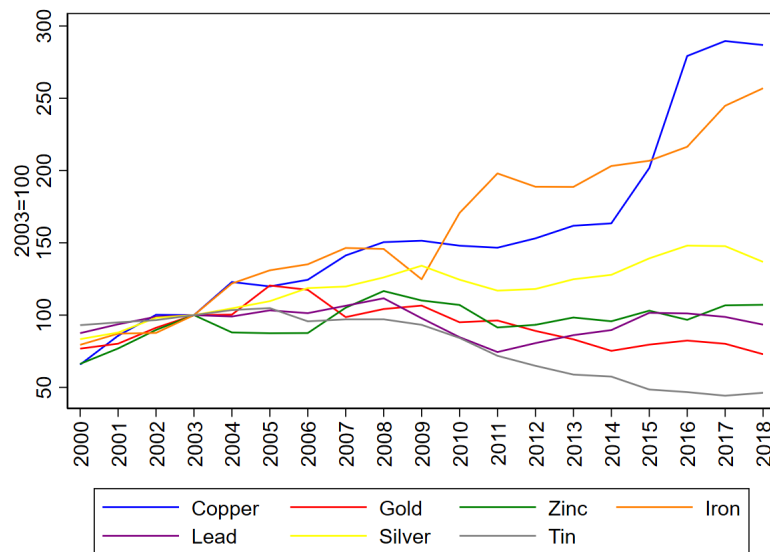
Note: Figure shows evolution of the shares of each mining commodity over total mining commodity exports.

Figure A.2: Mining prices



Note: Figure shows evolution of mining commodity prices. We normalize the dollar prices for the commodities and then divide them by the Peruvian imports price index.

Figure A.3: Mining output



Note: Figure shows evolution of mining physical output by commodity.

Table A.1: Mining production ranking of Peru, 2011

Mineral	Share of total world output	Largest competitors
Copper	8%	CHL: 33%, CHN: 8%, USA: 7%
Gold	6%	CHN: 14%, AUS: 10%, USA: 9%
Zinc	10%	CHN: 34%, AUS: 12%, USA: 10%
Lead	5%	CHN: 50%, AUS: 13%, USA: 7%
Tin	12%	CHN: 49%, IDN: 17%, BOL: 8%
Silver	15%	MEX: 18%, CHN: 16%, AUS: 7%
Iron	<1%	CHN: 45%, AUS: 17%, BRA: 13%

Note: Figure shows share of Peruvian output over total world output for each of the listed mining commodities.

Source: United States Geological Survey.

A.1.2 Sample Statistics

Table A.2: Bank statistics

	Mean	Std. dev.	Min.	Max.
Assets (bill. USD)	7.09	8.40	0.22	32.00
Firm deposits (bill. USD)	1.26	1.64	0.01	5.92
Foreign wholesale funding (bill. USD)	0.85	1.14	-0.04	5.37
E_{bt}	0.13	0.10	0.00	0.30

Note: Table shows information for the 9 banks considered in the analysis. For those banks that merged during this period, we consider them one entity and add their accounts. Additionally, we do not consider banks that are specialized in (i) serving foreign-owned firms only or (ii) credit card and consumer loans only. Foreign wholesale funding is defined as liabilities minus total deposits and equity. E_{bt} is the bank-level weighted commodity price exposure index from Equation 1.1.

Table A.3: Loan statistics

	Mean	Std. dev.	Min.	Max.
Loans (thousands USD)	283.05	594.41	0.00	6134.41
Number of bank relationships per firm	2.68	0.98	2.00	9.00

Note: Table shows information the on outstanding loan amounts and number of bank relationships per firm.

Table A.4: Firm statistics

	Mean	Std. dev.	Min.	Max.
Wage bill (thousands USD)	2029.56	5154.77	0.00	111591.40
Sales (thousands USD)	12049.26	20580.52	49.31	282293.70
e_{it}	0.03	0.07	0.00	0.30

Note: Table shows information for approximately 4,800 firms we consider in the analysis. e_{it} is the firm-level weighted commodity price exposure index from Equation 1.2.

A.2 Model Appendix

A.2.1 Microfoundation for the bank cost function

In our model, domestic banks obtain foreign wholesale funding from foreign lenders, for which they must pay a premium that depends on their reputation and the amount of foreign wholesale funding demanded. In this section, we provide a possible microfoundation for this cost, based on that of Drechsel and Tenreyro [2018].

We focus on the foreign lender, which we assume is risk neutral. The cost of the foreign lender for obtaining the funds is $(1 + i^*) N$, where i^* is the interest rate prevailing in foreign markets, and N is the amount of wholesale funding demanded by the domestic bank. Meanwhile, the expected benefit for the foreign lender is $\rho (1 + r^*) N + (1 - \rho) S$, where r^* is the interest rate charged to the domestic bank, S is the amount of collateral that can be seized in case of default, and ρ is the probability of repayment. We further assume that $S < (1 + i^*) N$, which means that the amount that can be recovered in case of default is lower than the actual value of the debt plus interests.

The breakeven condition for the foreign lender is

$$(1 + i^*) N = \rho (1 + r^*) N + (1 - \rho) S.$$

We can solve for the interest rate charged by the foreign lender to the domestic bank

$$1 + r^* = \frac{1 + i^*}{\rho} - \frac{(1 - \rho) S}{\rho N}.$$

From this expression, we have that

$$\frac{\partial (1 + r^*)}{\partial N} > 0,$$

which implies that domestic banks must pay a higher interest rate when they demand larger amounts of foreign wholesale funding.

From our result, we also have that

$$\frac{\partial (1 + r^*)}{\partial S} < 0, \frac{\partial (1 + r^*)}{\partial \rho} < 0,$$

which relates to the interest rate charged to the domestic bank depending on its reputation. If domestic banks possess high collateral or a high probability of repayment (i.e., they have good reputation in the market), they will be charged lower interest rates for the foreign wholesale funding.

Therefore, the interest rate charged by a foreign lender to a domestic bank is increasing in the amount of foreign wholesale funding demanded, decreasing in the amount of collateral, and decreasing in the probability of repayment. Although this derivation does not result in the exact functional form for the premium used in the main text, it clearly explains the intuition behind it.

A.2.2 Model Summary

Final consumption good bundle

$$c = \left[(\Lambda^{NT})^{\frac{1}{\gamma}} (c^N)^{\frac{\gamma-1}{\gamma}} + (1 - \Lambda^{NT})^{\frac{1}{\gamma}} (c^T)^{\frac{\gamma-1}{\gamma}} \right]^{\frac{\gamma}{\gamma-1}}$$

Tradable consumption good bundle

$$c^T = \left[(\Lambda^{XM})^{\frac{1}{\chi}} (c^X)^{\frac{\chi-1}{\chi}} + (1 - \Lambda^{XM})^{\frac{1}{\chi}} (c^M)^{\frac{\chi-1}{\chi}} \right]^{\frac{\chi}{\chi-1}}$$

Non-tradable consumption good bundle

$$c^N = \left[\sum_i \Lambda_i^{\frac{1}{\sigma}} c_i^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}$$

Non-tradable good consumption

$$c^N = \Lambda^{NT} \left(\frac{p^N}{p} \right)^{-\gamma} c$$

Tradable good consumption

$$c^T = (1 - \Lambda^{NT}) \left(\frac{p^T}{p} \right)^{-\gamma} c$$

Exportable good consumption

$$c^X = \Lambda^{XM} \left(\frac{p^X}{p^T} \right)^{-\chi} c^T$$

Importable good consumption

$$c^M = (1 - \Lambda^{XM}) \left(\frac{1}{p^T} \right)^{-\chi} c^T$$

Exportable good production

$$y^X = Ah^X$$

Exportable sector labor supply

$$(h^X)^\psi = \frac{w^X}{p}$$

Exportable good labor demand

$$w^X = Ap^X$$

Non-tradable sector labor supply

$$(h^N)^\psi = \frac{w^N}{p}$$

Non-tradable intermediate good consumption, for all i

$$c_i = \Lambda_i \left(\frac{p_i}{p^N} \right)^{-\sigma} c^N$$

Non-tradable intermediate good production, for all i

$$y_i = A h_i$$

Non-tradable intermediate good labor demand, for all i

$$p_i = \frac{\sigma}{\sigma - 1} \left[\frac{w^N}{A} (1 + \theta r_i^L) \right]$$

Loan bundle, for all i

$$L_i = \left[\sum_b s_{ib}^{\frac{1}{\varepsilon}} L_{ib}^{\frac{\varepsilon-1}{\varepsilon}} \right]^{\frac{\varepsilon}{\varepsilon-1}}$$

Loan demand, for all i

$$L_i = \theta w^N h_i$$

Bank interest rate, for all b

$$1 + r_b^L = \frac{\varepsilon}{\varepsilon - 1} \left[1 + r + \phi \ln \left(\frac{\mathcal{B}_b}{D_b} \right) \right]$$

Deposits, for all b

$$\ln(D_b) = \alpha_b + \beta \omega_b^{CO} \ln(p^{CO})$$

Total bank loans, for all b

$$L_b = \sum_i L_{ib}$$

Bank balance sheet, for all b

$$L_b = D_b + K_b + N_b$$

Individual bank loan demand, for all i and b

$$L_{ib} = s_{ib} \left(\frac{1 + r_b^L}{1 + r_i^L} \right)^{-\varepsilon} L_i$$

Total labor

$$h = h^X + h^N$$

Total labor in non-tradable sector

$$h^N = \sum_i h_i$$

Non-tradable good market equilibrium

$$c^N = y^N$$

Non-tradable intermediate good market equilibrium, for all i

$$c_i = y_i$$

Real exchange rate

$$rer = \frac{1}{p}$$

Total non-commodity output

$$Y^{NCO} = y^X + y^N$$

Real GDP, where overline denotes constant base year prices

$$\overline{p}GDP = \overline{p}c + \overline{p^{CO}}y^{CO} + \overline{p^X}(y^X - c^X) - \overline{p^M}c^M$$

Balance of payments

$$p^{CO}y^{CO} + p^X(y^X - c^X) - c^M = \sum_b \left[r + \phi \ln \left(\frac{\mathcal{B}_b}{D_b} \right) \right] N_b$$

A.2.3 Hat Algebra Model Summary

Final consumption good bundle

$$\widehat{c} = \left[\Lambda^{NT} \widehat{c}^N^{\frac{\gamma-1}{\gamma}} + (1 - \Lambda^{NT}) \widehat{c}^T^{\frac{\gamma-1}{\gamma}} \right]^{\frac{\gamma}{\gamma-1}}$$

Tradable consumption good bundle

$$\widehat{c}^T = \left[\Lambda^{XM} \widehat{c}^X^{\frac{\chi-1}{\chi}} + (1 - \Lambda^{XM}) \widehat{c}^M^{\frac{\chi-1}{\chi}} \right]^{\frac{\chi}{\chi-1}}$$

Non-tradable consumption good bundle

$$\widehat{c}^N = \left[\sum_i \Lambda_i \widehat{c}_i^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}$$

Non-tradable good consumption

$$\widehat{c}^N = \left(\frac{\widehat{p}^N}{\widehat{p}} \right)^{-\gamma} \widehat{c}$$

Tradable good consumption

$$\widehat{c}^T = \left(\frac{\widehat{p}^T}{\widehat{p}} \right)^{-\gamma} \widehat{c}$$

Exportable good consumption

$$\widehat{c}^X = \left(\frac{\widehat{p}^X}{\widehat{p}^T} \right)^{-\chi} \widehat{c}^T$$

Importable good consumption

$$\widehat{c}^M = \left(\frac{1}{\widehat{p}^T} \right)^{-\chi} \widehat{c}^T$$

Exportable good production

$$\widehat{y}^X = \widehat{A} \widehat{h}^X$$

Exportable sector labor supply

$$\widehat{h}^X{}^\psi = \frac{\widehat{w}^X}{\widehat{p}}$$

Exportable good labor demand

$$\widehat{w}^X = \widehat{A} \widehat{p}^X$$

Non-tradable sector labor supply

$$\widehat{h^N}^\psi = \frac{\widehat{w^N}}{\widehat{p}}$$

Non-tradable intermediate good consumption, for all i

$$\widehat{c}_i = \left(\frac{\widehat{p}_i}{\widehat{p^N}} \right)^{-\sigma} \widehat{c^N}$$

Non-tradable intermediate good production, for all i

$$\widehat{y}_i = \widehat{A} \widehat{h}_i$$

Non-tradable intermediate good labor demand, for all i

$$\widehat{w^N} \left(1 + \theta \widehat{r}_i^L \right) = \widehat{A} \widehat{p}_i$$

Loan bundle, for all i

$$\widehat{L}_i = \left[\sum_b s_{ib} \left(\widehat{L}_{ib} \right)^{\frac{\varepsilon-1}{\varepsilon}} \right]^{\frac{\varepsilon}{\varepsilon-1}}$$

Loan demand, for all i

$$\widehat{L}_i = \widehat{w^N} \widehat{h}_i$$

Bank interest rate, for all b

$$1 + \widehat{r}_b^L = 1 + \widehat{r} - \phi \ln \left(\widehat{D}_b \right)$$

Deposits, for all b

$$\ln \left(\widehat{D}_b \right) = \beta \omega_b^{CO} \ln \left(\widehat{p^{CO}} \right)$$

Total bank loans, for all b

$$\widehat{L}_b = \sum_i \omega_{ib}^L \widehat{L}_{ib}$$

Bank balance sheet, for all b

$$\widehat{L}_b = \Omega_b^D \widehat{D}_b + \Omega_b^K \widehat{K}_b + \Omega_b^N \widehat{N}_b + \Omega_b^O$$

Individual bank loan demand, for all i and b

$$\widehat{L}_{ib} = \left(\frac{1 + \widehat{r}_b^L}{1 + \widehat{r}_i^L} \right)^{-\varepsilon} \widehat{L}_i$$

Total labor

$$\widehat{h} = \Omega_X^h \widehat{h}^X + (1 - \Omega_X^h) \widehat{h}^N$$

Total labor in non-tradable sector

$$\widehat{h}^N = \sum_i \Lambda_i^h \widehat{h}_i$$

Non-tradable good market equilibrium

$$\widehat{c}^N = \widehat{y}^N$$

Non-tradable intermediate good market equilibrium, for all i

$$\widehat{c}_i = \widehat{y}_i$$

Real exchange rate

$$\widehat{rer} = \frac{1}{\widehat{p}}$$

Total non-commodity output

$$\widehat{Y}^{NCO} = \Omega_X^{NCO} \widehat{y}_X + (1 - \Omega_X^{NCO}) \widehat{y}_N$$

Real GDP

$$\widehat{GDP} = \Omega_C \widehat{c} + \Omega_X \eta_X + \Omega_X (1 - \eta_X) [\Omega_Y^X \widehat{y}_X - (\Omega_Y^X - 1) \widehat{c}_X] - \Omega_M \widehat{c}_M$$

Balance of payments

$$\begin{aligned} & \eta_X \widehat{p}^{CO} + (1 - \eta_X) \widehat{p}^X [\Omega_Y^X \widehat{y}^X - (\Omega_Y^X - 1) \widehat{c}^X] - \widehat{c}^M \\ &= \sum_b \omega_b^N \left\{ (1 + \eta^N) \left[1 + \widehat{r} - \phi \ln(\widehat{D}_b) \right] \widehat{N}_b - \widehat{N}_b \right\} \end{aligned}$$

A.2.4 Ratios

Table A.5: Ratios and other weights

Trade balance ratios		
η^{CO}	Mining commodity exports / Exports	0.683
η^N	Interests / Foreign wholesale funding	0.02
ω_b^N	Foreign wholesale funding over total	By bank
GDP , output, ratios		
Ω^C	(Consumption + Investment + Gov. exp.) / GDP	1
Ω^X	Exports / GDP	0.28
Ω^M	Imports / GDP	0.28
Ω_X^{NCO}	Exportable sector output / Non-commodity output	0.3
Ω_X^h	Exportable sector labor / Non-commodity labor	0.42
Ω_Y^X	Exportable goods output / Non-commodity exports	1.4
Λ_i^h	Labor in NT int. firms / Total NT labor	By firm
Bank balance sheet ratios		
Ω_b^N	Wholesale funding / Firm loans	By bank
Ω_b^D	Firm deposits / Firm loans	By bank
Ω_b^K	Equity / Firm loans	By bank
Ω_b^O	Other liabilities / Firm loans	By bank
ω_{ib}^L	Individual loans / Total loans, by bank	By firm and bank

Note: All ratios calculated using deflated data and expressed in domestic currency. Imports and exports consider traded goods only.

A.2.5 From Model to Regressions

Loan-level regression. We start from the equilibrium conditions for loans expressed in natural logarithms. We have the loan demand

$$\ln(L_{ib}) = -\varepsilon r_b^L + \varepsilon r_i^L + \ln(s_{ib}) + \ln(L_i),$$

and the interest rate set by banks

$$r_b^L = \ln\left(\frac{\varepsilon}{\varepsilon - 1}\right) + r + \phi \ln(\mathcal{B}_b) - \phi \ln(D_b),$$

where we require $\frac{\mathcal{B}_b}{D_b} > 1$ and $\phi \ln\left(\frac{\mathcal{B}_b}{D_b}\right) \approx 0$ for this approximation to hold and to obtain a positive interest rate.

Replace to obtain

$$\ln(L_{ib}) = -\varepsilon \ln\left(\frac{\varepsilon}{\varepsilon - 1}\right) - \varepsilon r + \varepsilon r_i^L + \ln(s_{ib}) + \ln(L_i) - \varepsilon \phi \ln(\mathcal{B}_b) + \varepsilon \phi \ln(D_b).$$

Replace $\ln(D_b) = \alpha_b + \beta \sum_m \omega_{bm}^{CO} \ln(p_m^{CO})$, add time subindices and collect terms to find

$$\begin{aligned} \ln(L_{ibt}) = & \underbrace{-\varepsilon \ln\left(\frac{\varepsilon}{\varepsilon - 1}\right)}_{\text{constant}} + \underbrace{\ln(s_{ib}) + \varepsilon \phi \alpha_b - \varepsilon \phi \ln(\mathcal{B}_b)}_{\alpha_{ib}} \\ & + \underbrace{\varepsilon r_{it}^L + \ln(L_{it}) - \varepsilon r_t}_{\alpha_{it}} + \underbrace{\varepsilon \phi \beta}_{\zeta} \sum_m \omega_{bm}^{CO} \ln(p_m^{CO}). \end{aligned}$$

Add a residual, which we interpret as the classical measurement error, to obtain

$$\ln(L_{ibt}) = \alpha_{ib} + \alpha_{it} + \zeta \sum_m \omega_{bm}^{CO} \ln(p_{mt}^{CO}) + u_{ibt},$$

$$\zeta = \varepsilon \phi \beta.$$

This expression corresponds with Equation 1.4. Note that in this derivation we assumed there is more than one commodity exporter, so we obtain the case we have in the data.

Firm-level regression. We begin with the equilibrium conditions of the non-tradable intermediate firms expressed in natural logarithms. We have the demand faced by non-tradable intermediate firms

$$\ln(y_i) = -\sigma \ln(p_i) + \sigma \ln(p^N) + \ln(y^N),$$

the price set by non-tradable intermediate firms

$$\ln(p_i) = \ln\left(\frac{\sigma}{\sigma-1}\right) + \ln(w^N) - \ln(A) + \theta r_i^L.$$

the interest rate faced by non-tradable intermediate firms

$$r_i^L = \sum_b s_{ib} r_b^L,$$

and the interest rate set by banks

$$r_b^L = \ln\left(\frac{\varepsilon}{\varepsilon-1}\right) + r + \phi \ln(\mathcal{B}_b) - \phi \ln(D_b).$$

Replace to obtain

$$\begin{aligned} \ln(p_i y_i) &= \sigma \ln(p^N) + \ln(y^N) - (\sigma-1) \ln\left(\frac{\sigma}{\sigma-1}\right) \\ &\quad - (\sigma-1) \ln(w^N) + (\sigma-1) \ln(A) - (\sigma-1) \theta r_i^L. \end{aligned}$$

Replace interest rate, $\ln(D_b) = \alpha_b + \beta \sum_m \omega_{bm}^{CO} \ln(p_m^{CO})$, add time subindices and collect terms to find

$$\begin{aligned} \ln(p_{it} y_{it}) &= \underbrace{-(\sigma-1) \ln\left(\frac{\sigma}{\sigma-1}\right) - (\sigma-1) \theta \ln\left(\frac{\varepsilon}{\varepsilon-1}\right) - (\sigma-1) \theta \phi \sum_b s_{ib} \ln(\mathcal{B}_b)}_{\text{constant}} \\ &\quad + \underbrace{\sigma \ln(p_t^N) + \ln(y_t^N) - (\sigma-1) \ln(w_t^N) - (\sigma-1) \theta r_t + (\sigma-1) \ln(A_t)}_{\alpha_t} \\ &\quad + \underbrace{(\sigma-1) \theta \phi \beta}_{\kappa} \sum_b s_{ib} \sum_m \omega_{bm}^{CO} \ln(p_{mt}^{CO}). \end{aligned}$$

Add a residual, which we interpret as the classical measurement error, to obtain

$$\begin{aligned} \ln(p_{it} y_{it}) &= \alpha_i + \alpha_t + \kappa \sum_b s_{ib} \sum_m \omega_{bm}^{CO} \ln(p_{mt}^{CO}) + u_{it}, \\ \kappa &= (\sigma-1) \theta \phi \beta. \end{aligned}$$

This expression corresponds with Equation 1.5. As in the previous case, for this derivation we assumed there is more than one commodity exporter, so we recover the case we have in the data.

A.2.6 Data-Based Moments and Shocks

Targeted and untargeted moments. We set the moments to match their observed changes in the data between 2003 and 2011.

- $\widehat{GDP}$: We use the real GDP series.
- $\widehat{c}$: We use the sum of the real consumption, investment and government expenditure series.
- $\widehat{y^{NCO}}$: We use the real non-mining GDP series.
- $\widehat{c^M}$: We use the real imported goods series.
- $\widehat{rer}$: We use the real exchange rate series. A decrease means a real appreciation of the domestic consumption basket with respect to the foreign consumption basket. We use the bilateral series with respect to the US. We do not use the multilateral real exchange rate (i.e., an average real exchange rate with respect to different trade partners) as it considers weights that shift over time.
- $\widehat{L_b}$: We use the aggregate firm loans issued by banks, deflated by the GDP deflator.
- $\widehat{D_b}$: We use the aggregate firm deposits held in banks, deflated by the GDP deflator.
- $\widehat{N_b}$: We consider a simplified bank balance sheet. On the asset side we have firm loans L_b . On the liabilities side we have firm deposits D_b , equity K_b and foreign wholesale funding N_b . However, N_b is not properly measured in the data. We calculate $N_b = L_b - D_b - K_b$ as a proxy for foreign wholesale funding. All involved series are deflated by the GDP deflator.
- $\widehat{r_i^L}$: We use the average loan interest rate series for both domestic currency and US dollars. Then, we calculate a weighted interest rate using the respective weights of domestic currency-denominated and US dollar-denominated loans over total loans.

Shocks. We set the shocks to match their observed changes in the data between 2003 and 2011.

- $\widehat{K_b}$: We use the aggregate bank equity series, deflated by the GDP deflator.
- $\widehat{r}$: We calculate the change in the sum of US federal funds rate and change in Peruvian EMBI minus Peruvian inflation rate.

A.2.7 Complete Simulation Results

Table A.6: Main simulation, complete results

	(1)	(2)	(3)
Variable	Baseline: Fin. mech.	CF: No fin. mech. ($\beta = 0$)	Baseline minus CF
$\widehat{GDP}$	1.65	1.59	0.06
$\widehat{c}$	2.00	1.89	0.11
$\widehat{y^{NCO}}$	1.66	1.61	0.05
$\widehat{c^M}$	2.15	2.00	0.14
$\widehat{rer}$	0.87	0.89	-0.02
Ave. $\widehat{L}_b$	2.46	2.23	0.23
Ave. $\widehat{D}_b$	2.49	1.00	1.49
Ave. $\widehat{N}_b$	4.24	6.65	-2.41
Ave. $\widehat{r}_b^L$	-0.08	-0.02	-0.06
$\widehat{c^N}$	1.90	1.81	0.10
$\widehat{c^T}$	2.15	2.00	0.14
$\widehat{c^X}$	2.15	2.00	0.14
$\widehat{y^N}$	1.90	1.81	0.10
$\widehat{y^X}$	1.09	1.14	-0.05
$\widehat{p}$	1.15	1.13	0.02
$\widehat{p^N}$	1.27	1.23	0.04
$\widehat{p^T}$	1.00	1.00	0.00
$\widehat{h}$	1.38	1.35	0.03
$\widehat{h^N}$	1.68	1.60	0.09
$\widehat{h^X}$	0.97	1.01	-0.04
$\widehat{w^N}$	1.46	1.39	0.06
$\widehat{w^X}$	1.13	1.13	0.00

Note: Column 1 shows moments from baseline simulation with financial propagation mechanism. Column 2 shows moments from counterfactual simulation with no financial propagation mechanism. Column 3 shows the difference between baseline and counterfactual. All moments are expressed in percentage variations between initial and final states. Averages for the bank-related model variables are computed using loan-based weights of each bank in total loans. These weights are calibrated according to observed data.

Table A.7: Commodity price shock-only simulations, complete results

	(1)	(2)	(3)
Variable	Baseline: Fin. mech.	CF: No fin. mech. ($\beta = 0$)	Baseline minus CF
$\widehat{GDP}$	1.32	1.25	0.07
$\widehat{c}$	1.65	1.51	0.14
$\widehat{y^{NCO}}$	1.23	1.17	0.06
$\widehat{c^M}$	1.87	1.68	0.19
$\widehat{rer}$	0.78	0.81	-0.03
Ave. $\widehat{L_b}$	2.35	2.02	0.33
Ave. $\widehat{D_b}$	2.49	1.00	1.49
Ave. $\widehat{N_b}$	6.13	7.79	-1.66
Ave. $\widehat{r_b^L}$	-0.06	0.00	-0.06
$\widehat{c^N}$	1.51	1.40	0.11
$\widehat{c^T}$	1.87	1.68	0.19
$\widehat{c^X}$	1.87	1.68	0.19
$\widehat{y^N}$	1.51	1.40	0.11
$\widehat{y^X}$	0.57	0.63	-0.05
$\widehat{p}$	1.29	1.24	0.05
$\widehat{p^N}$	1.54	1.44	0.10
$\widehat{p^T}$	1.00	1.00	0.00
$\widehat{h}$	1.12	1.08	0.04
$\widehat{h^N}$	1.51	1.40	0.11
$\widehat{h^X}$	0.57	0.63	-0.05
$\widehat{w^N}$	1.55	1.44	0.11
$\widehat{w^X}$	1.00	1.00	0.00

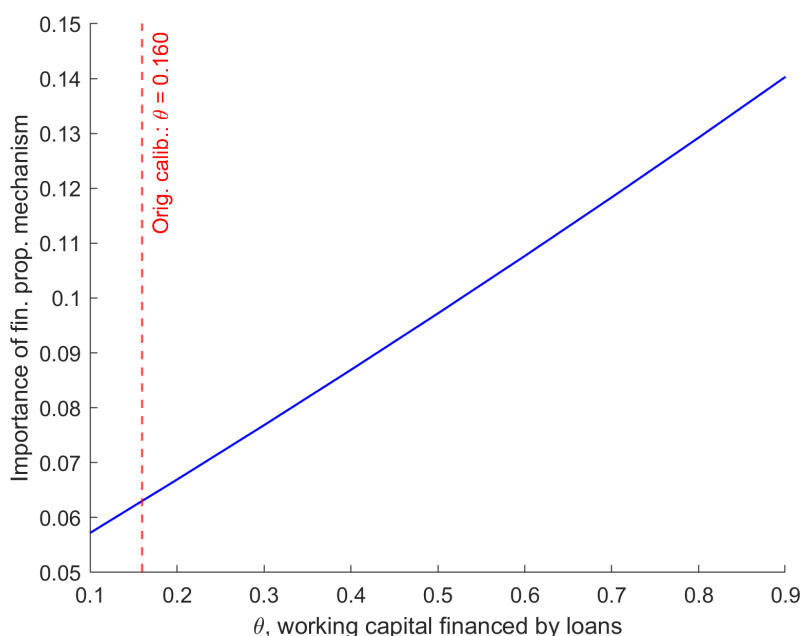
Note: Column 1 shows moments from baseline simulation with financial propagation mechanism. Column 2 shows moments from counterfactual simulation with no financial propagation mechanism. Column 3 shows the difference between baseline and counterfactual. All moments are expressed in percentage variations between initial and final states. Averages for the bank-related model variables are computed using loan-based weights of each bank in total loans. These weights are calibrated according to observed data.

A.2.8 Model Sensitivity to Different Parameterizations

In this section, we explore the sensitivity of the model results to variations in key bank-related and commodity-related parameters. We keep the original calibration and shocks, only changing the numerical values of selected parameters. We focus on examining the importance of the financial propagation mechanism, that is, the difference in GDP growth when the mechanism is turned on and off.

θ , **working capital financed by loans.** In Figure A.4 we see that the importance of the financial propagation mechanism grows as θ increases. This is because a higher θ means firms require to take higher loans, making the financial propagation mechanism becomes more important overall.

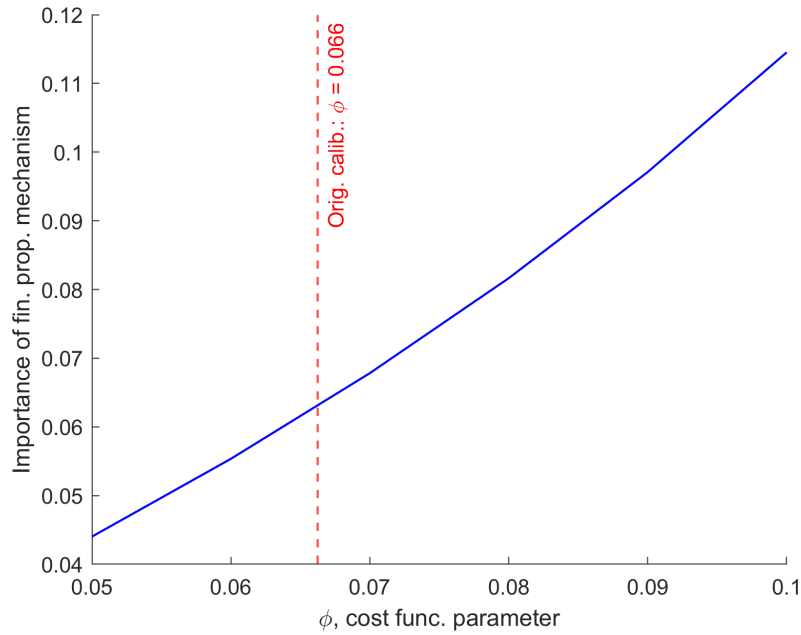
Figure A.4: Financial propagation mechanism under alt. parameters: θ



Note: Figure shows the importance of the financial mechanism in the theoretical model under different values for θ . The importance is defined as the difference in GDP growth between the baseline case and the counterfactual case with no financial propagation mechanism. Vertical line denotes original calibration.

ϕ , **bank cost function parameter.** Figure A.5 shows that the importance of the importance of the financial propagation mechanism grows as ϕ increases. With a larger ϕ , bank costs become more responsive to the amount of deposits held, intensifying the role of the financial propagation mechanism in the economy.

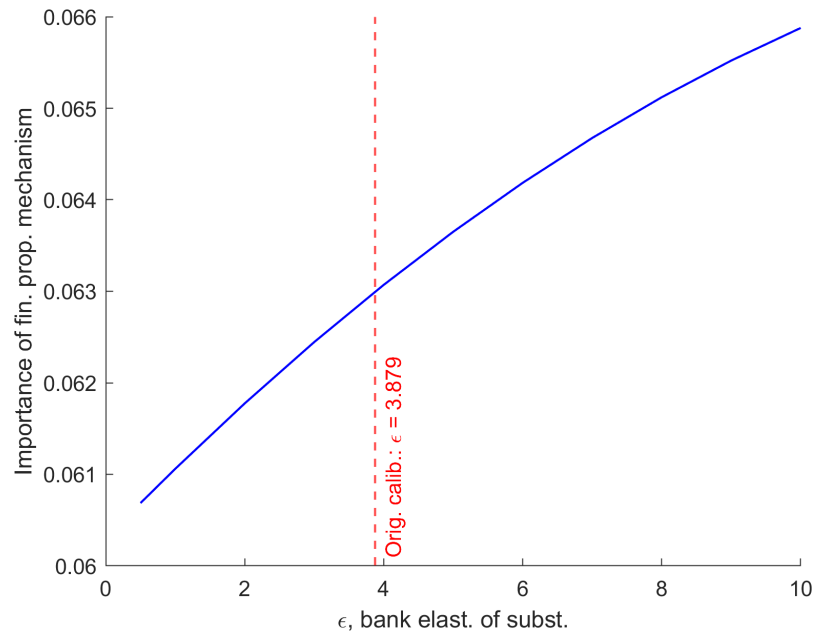
Figure A.5: Financial propagation mechanism under alt. parameters: ϕ



Note: Figure shows the importance of the financial mechanism in the theoretical model under different values for ϕ . The importance is defined as the difference in GDP growth between the baseline case and the counterfactual case with no financial propagation mechanism. Vertical line denotes original calibration.

ε , **bank elasticity of substitution**. Figure A.6 displays that the importance of the importance of the financial propagation mechanism grows as ε increases. A larger ε implies the demand for loans becomes more sensitive to interest rate changes. This way, the role the financial propagation mechanism plays in the model economy expands.

Figure A.6: Financial propagation mechanism under alt. parameters: ε

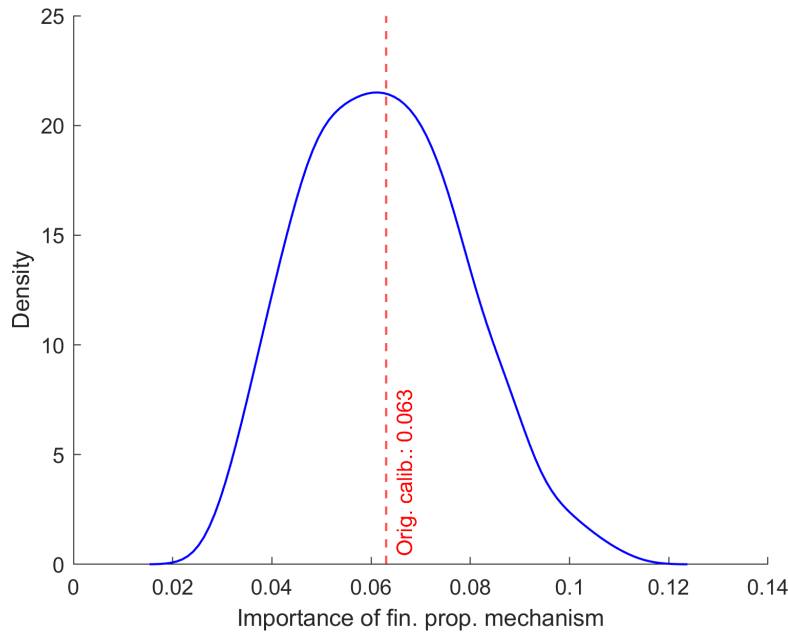


Note: Figure shows the importance of the financial mechanism in the theoretical model under different values for ε . The importance is defined as the difference in GDP growth between the baseline case and the counterfactual case with no financial propagation mechanism. Vertical line denotes original calibration.

Estimated coefficients. The bank block calibration relies on estimated coefficients β , ζ , and κ . To account for the uncertainty around their estimation, we perform the following exercise based on Huo et al. [2024]. We draw random values of β , ζ , and κ from a normally distributed range around their point estimates. We further assume these distributions are independent of each other. Then, we simulate the model and calculate the importance of the financial propagation mechanism.

Figure A.7 presents the density of the importance of the financial propagation mechanism after repeating the exercise 1000 times. While the range of results can vary from 0.0294 to 0.1097, our result under the original calibration is close to the mean (0.0625) and median (0.0619) of the distribution. Therefore, we find our main result is a central value within the range.

Figure A.7: Financial propagation mechanism under alt. parameters: Estimated coefficients



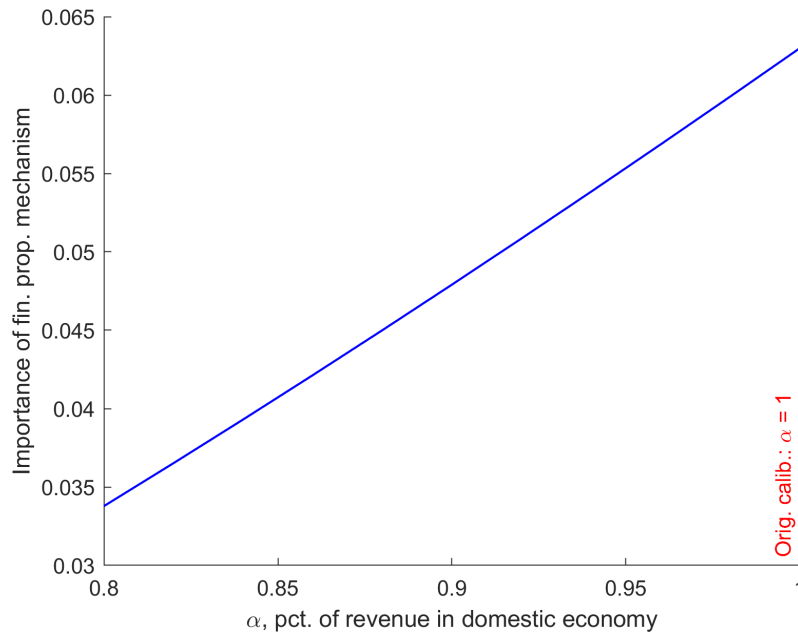
Note: Figure shows the importance of the financial mechanism in the theoretical model under different values for of estimated coefficients β , ζ , and κ . The importance is defined as the difference in GDP growth between the baseline case and the counterfactual case with no financial propagation mechanism. Vertical line denotes original calibration.

α , **percentage of mining commodity firm's revenue in domestic economy.** In this section, we assume a more general form of the balance of payments equation

$$\alpha p^{CO} y^{CO} + p^X (y^X - c^X) - c^M = \sum_b \left[r + \phi \ln \left(\frac{\mathcal{B}_b}{D_b} \right) \right] N_b,$$

where parameter α determines how much of the revenues of the mining commodity firm stay in the domestic economy, and $1 - \alpha$ determines the amount of dividends paid to foreign investors. In other words, parameter α governs the strength of the wealth channel. In the main simulations, we assume $\alpha = 1$.¹ Figure A.8 shows how the importance of the financial propagation mechanism changes when we vary this parameter. As α increases, more wealth remains in the domestic economy, which in turn increases the demand for loans and the importance the financial propagation mechanism has.

Figure A.8: Financial propagation mechanism under alt. parameters: α



Note: Figure shows the importance of the financial mechanism in the theoretical model under different values for α . The importance is defined as the difference in GDP growth between the baseline case and the counterfactual case with no financial propagation mechanism. Vertical line denotes original calibration.

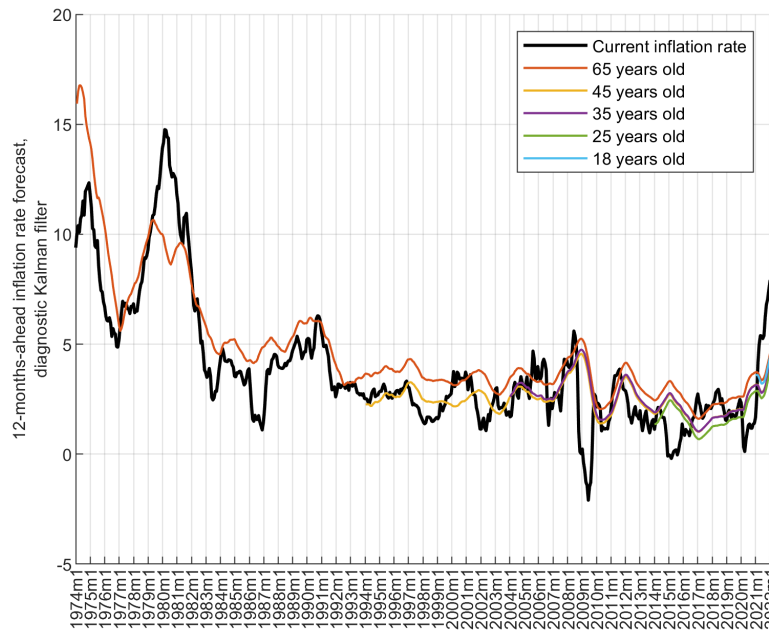
¹According to the 2007 economic census, the dividends paid to foreign owners by mining commodity firms were equivalent to 4 percent of their revenues.

APPENDIX B

Appendix to Chapter 2

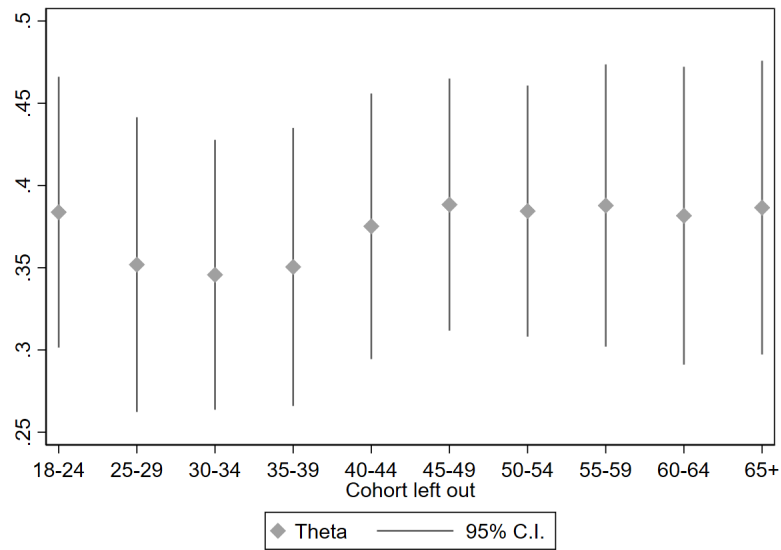
B.1 Additional figures and tables

Figure B.1: History-based Kalman-filter-based inflation forecasts by cohort, full sample



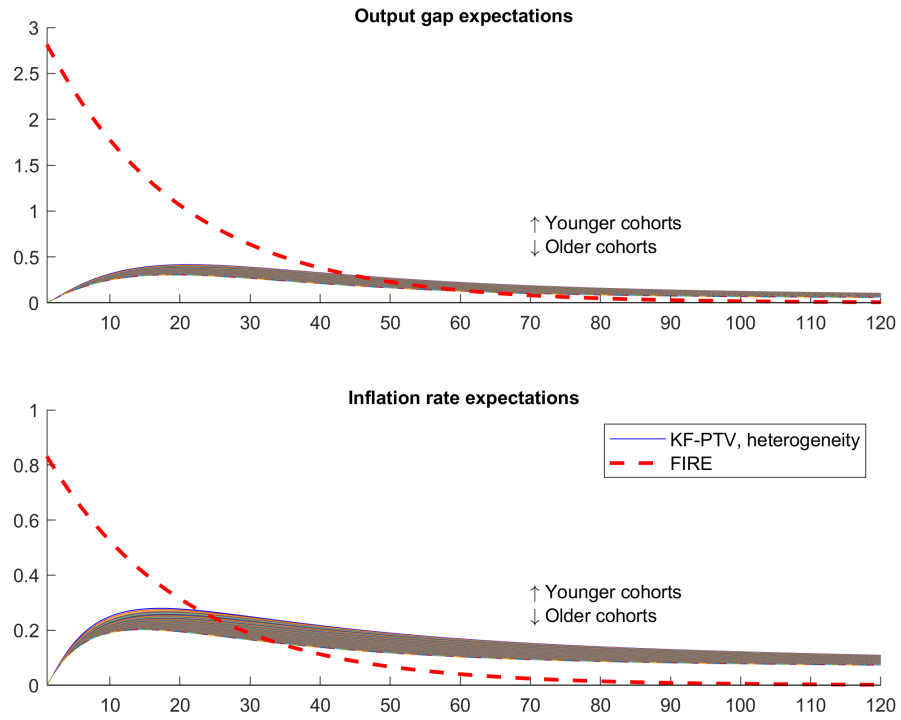
Note: Figure shows forecasts for selected cohorts according to the Kalman-filter-augmented expectations and considering the estimate for θ from Column 1 of Table 2.2. Selected cohorts differentiated by their age in 2021. We further assume that each cohort starts forecasting when they become 18 years old. The current inflation rate measure is based on the monthly YoY percentage variation of the CPI.

Figure B.2: Parameter estimation, stability by cohorts



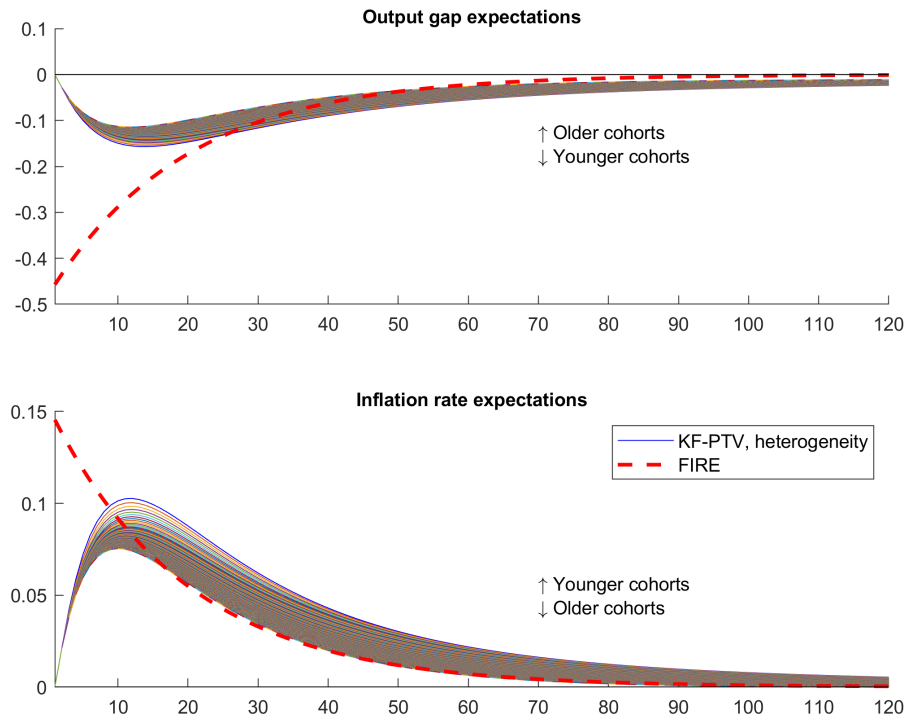
Note: Figure shows the coefficient resulting from estimating Equation 2.6, but leaving one cohort out of the sample. The cohort left out of the estimation changes across the horizontal axis.

Figure B.3: Impulse response functions, history-based expectations by cohort, taste shock



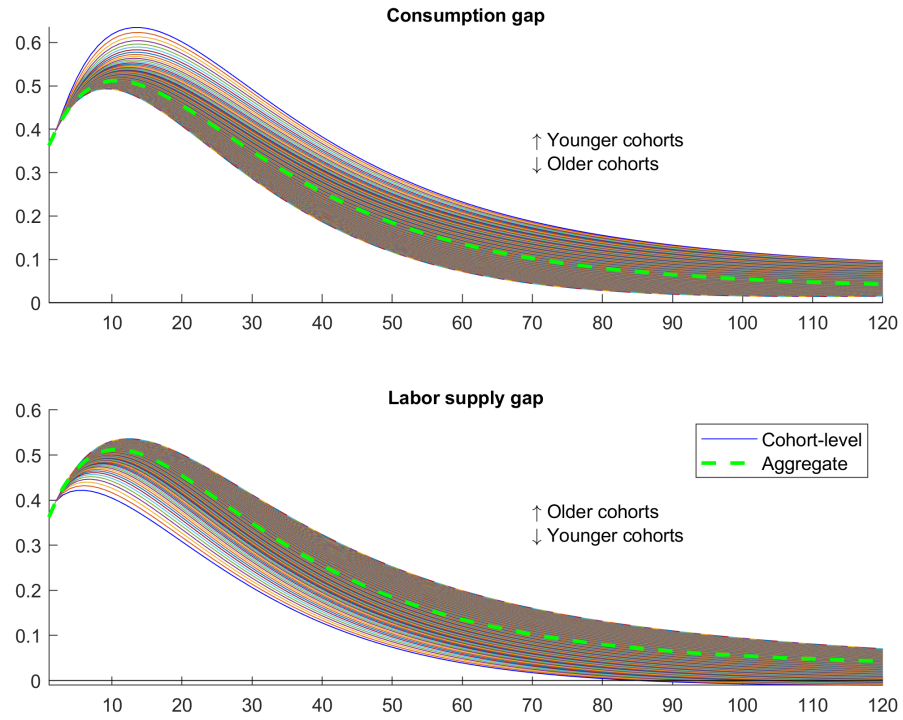
Note: Figure shows the heterogeneous expectations generated by the history-based Kalman filter. Cohorts denote age at the time of the shock. The solid lines represent different cohorts in the history-based Kalman filter model. The dashed red line is the full information rational expectations model. Horizontal axis denotes months after the shock.

Figure B.4: Impulse response functions, history-based expectations by cohort, cost shock



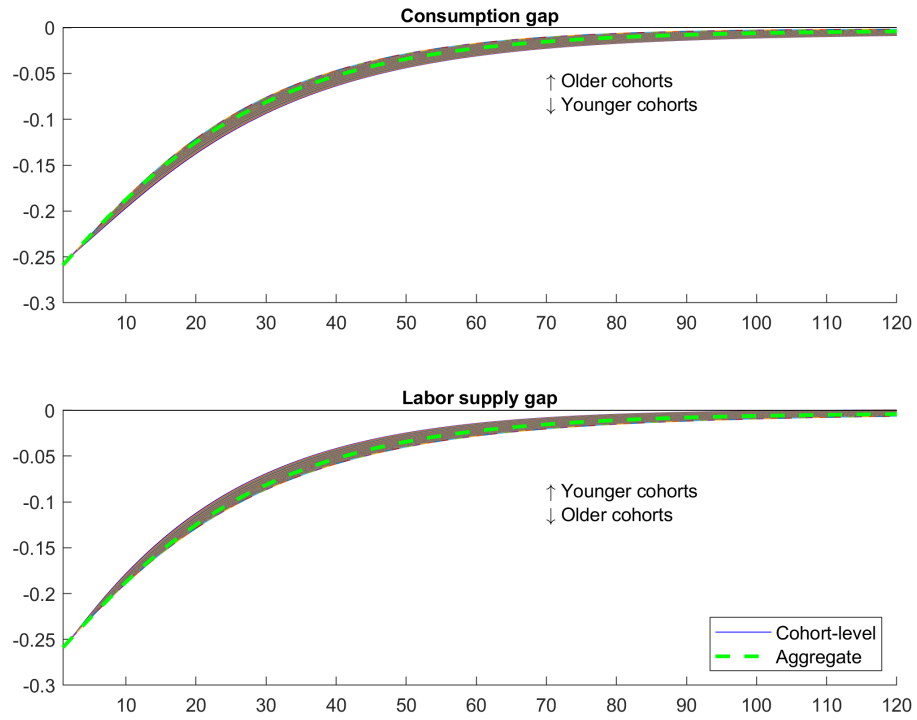
Note: Figure shows the heterogeneous expectations generated by the history-based Kalman filter. Cohorts denote age at the time of the shock. The solid lines represent different cohorts in the history-based Kalman filter model. The dashed red line is the full information rational expectations model. Horizontal axis denotes months after the shock.

Figure B.5: Impulse response functions, consumption and labor supply by cohort, history-based KF with heterogeneity, taste shock



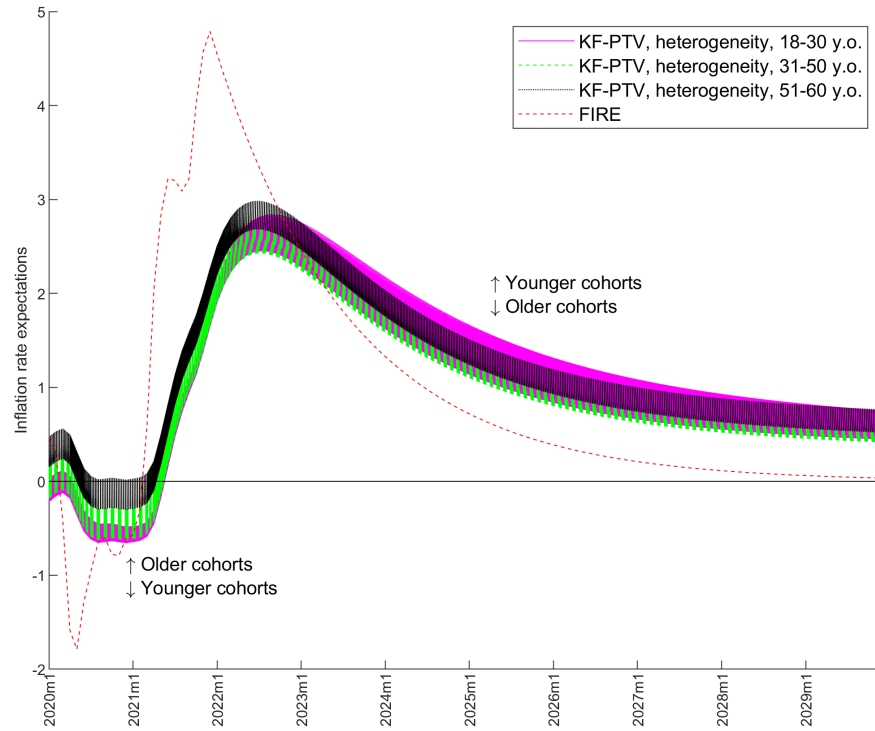
Note: Figure shows the heterogeneous consumption and labor supply gaps generated in the case where expectations follow the history-based Kalman filter. Cohorts denote age at the time of the shock. The solid lines represent different cohorts in the history-based Kalman filter model. The dashed red line is the aggregate. Horizontal axis denotes months after the shock.

Figure B.6: Impulse response functions, consumption and labor supply by cohort, history-based KF with heterogeneity, cost shock



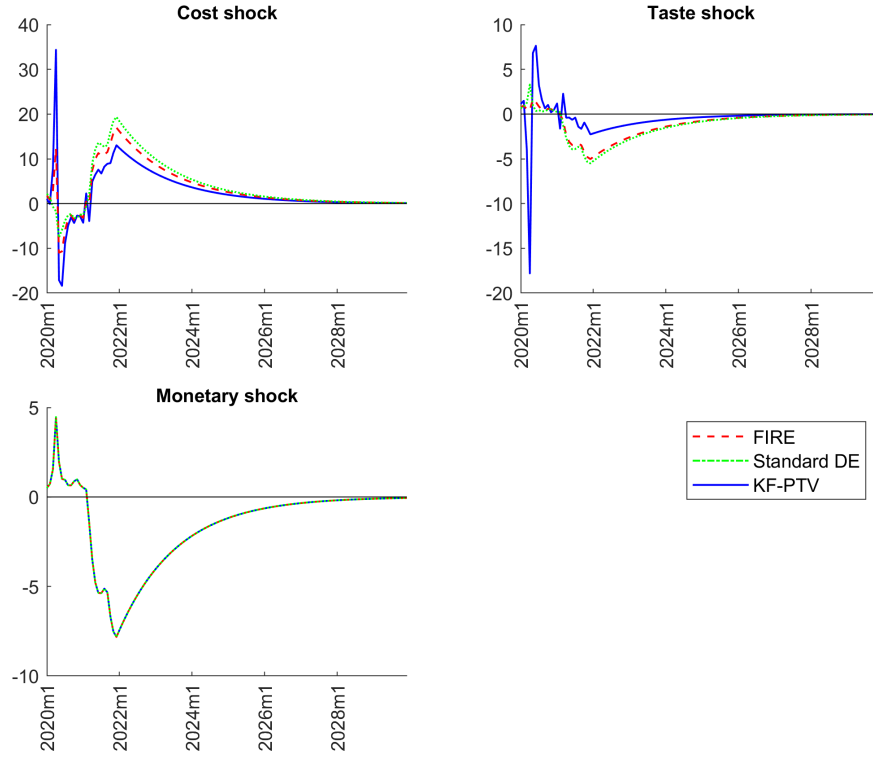
Note: Figure shows the heterogeneous consumption and labor supply gaps generated in the case where expectations follow the history-based Kalman filter. Cohorts denote age at the time of the shock. The solid lines represent different cohorts in the history-based Kalman filter model. The dashed red line is the aggregate. Horizontal axis denotes months after the shock.

Figure B.7: Impulse response functions, inflation rate history-based expectations by cohort, forecast



Note: Figure shows the heterogeneous expectations generated by the history-based Kalman filter and the data (up to December 2021). Cohorts denote age in 2021. Horizontal axis denotes months.

Figure B.8: Impulse response functions, shocks, forecast



Note: Figure shows the paths shocks follow according to the model and the data (up to December 2021). The red dashed line shows the results for the case of the full information rational expectations model (FIRE), the green dotted line shows the results of a standard diagnostic expectations operator and the solid blue line shows the history-based Kalman filter model. For the standard diagnostic expectations case we assume that agents use the expectations operator $\mathbb{E}_t^{\theta, s}[X_{t+h}] = \mathbb{E}_t[X_{t+h}] + \zeta(\mathbb{E}_t[X_{t+h}] - \mathbb{E}_{t-3}[X_{t+h}])$ with $\zeta > 0$. Horizontal axis denotes months.

Table B.1: Model calibration

Parameter	Value	Parameter	Value
β	0.9967	χ_y	0.125
η	1	ρ^{cost}	0.9
ϕ	0.9167	ρ^{taste}	0.9
σ	1	λ	0.001
ε	9	K	0.1751
χ_π	1.5	θ	-0.374

Note: Table shows the parameters used for the model. We follow a standard monthly calibration.

Table B.2: Optimal Taylor rule parameters, cost shock

	χ_π^*	χ_y^*
FIRE	119.20	0.74
History-based KF-PTV	140.79	0.47

Note: Table shows the Taylor rule parameters that minimize objective function $\mathbb{E}_t [\pi_t^2 + \vartheta y_t^2]$ when an unexpected cost shock hits the economy.

Table B.3: Optimal Taylor rule parameters, taste shock

	χ_π^*	χ_y^*
FIRE	8.73	28.10
History-based KF-PTV	9.18	27.19

Note: Table shows the Taylor rule parameters that minimize objective function $\mathbb{E}_t [\pi_t^2 + \vartheta y_t^2]$ when an unexpected taste shock hits the economy.

B.2 Normality of history-based expectations

Let $f(\pi_{t+1}|\mathcal{I}_t)$ be the true distribution of future inflation conditional on the currently available information set $\mathcal{I}_t$. We assume this behaves as

$$f(\pi_{t+1}|\mathcal{I}_t) \sim N(\mathbb{E}_t^{KF}[\pi_{t+1}], \sigma_\pi^2),$$

where $\mathbb{E}_t^{KF}[\pi_{t+1}]$ is the expectation computed according to the standard Kalman filter. On the other hand, let $f(\pi_{t+1}|\mathcal{I}_{i,t}^{ref})$ be the distribution of the inflation rate conditional on the referential information set associated with cohort i . Under our current assumptions, this distribution behaves as

$$f(\pi_{t+1}|\mathcal{I}_{i,t}^{ref}) \sim N(\mathbb{E}_{i,t}^{ref}[\pi_{t+1}], (t - k_i)^{-1} \sigma_\pi^2),$$

where

$$\mathbb{E}_{i,t}^{ref}[\pi_{t+1}] = \frac{\sum_{j=1}^{t-k_i} \mathbb{E}_{i,t-j}^{KF}[\pi_{t+1}]}{(t - k_i)}.$$

Given these two elements, we define the distribution as

$$f_{i,t}^\theta(\pi_{t+1}) = f(\pi_{t+1}|\mathcal{I}_t) D_{i,t}^\theta(\pi_{t+1}),$$

where

$$D_{i,t}^\theta(\pi_{t+1}) = \left[\frac{f(\pi_{t+1}|\mathcal{I}_t)}{f(\pi_{t+1}|\mathcal{I}_{i,t}^{ref})^{\frac{1}{t-k_i}}} \right]^\theta Z_{i,t},$$

where $Z_{i,t}^{-1} = \int f_{i,t}^\theta(\pi_{t+1}) d\pi_{t+1}$ is a term that ensures that $f_{i,t}^\theta(\pi_{t+1})$ integrates to one. Therefore, under some algebraic procedure, the probability density function of the distribution can be written as

$$f_{i,t}^\theta(\pi_{t+1}) = \frac{\left[\frac{1}{\sigma_\pi \sqrt{2\pi}} \exp \left\{ -\frac{(\pi_{t+1} - \mathbb{E}_t^{KF}[\pi_{t+1}])^2}{2\sigma_\pi^2} \right\} \right]^{(1+\theta)} Z_{i,t}}{\left[\frac{1}{\sigma_\pi \sqrt{2\pi}} \exp \left\{ -\frac{(\pi_{t+1} - \mathbb{E}_{i,t}^{ref}[\pi_{t+1}])^2}{2\sigma_\pi^2} \right\} \right]^\theta},$$

And we can make the following approximation

$$f_{i,t}^{\theta}(\pi_{t+1}) \approx \frac{1}{\sigma_{\pi} \sqrt{2\pi}} \exp \left\{ -\frac{(\pi_{t+1} - \mathbb{E}_{i,t}^{\theta}[\pi_{t+1}])^2}{2\sigma_{\pi}^2} \right\} Z_{i,t},$$

where

$$\mathbb{E}_{i,t}^{\theta}[\pi_{t+1}] = \mathbb{E}_t^{KF}[\pi_{t+1}] + \theta \left(\mathbb{E}_t^{KF}[\pi_{t+1}] - \mathbb{E}_{i,t}^{ref}[\pi_{t+1}] \right).$$

Thus, we conclude that

$$f_{i,t}^{\theta}(\pi_{t+1}) \sim N \left(\mathbb{E}_{i,t}^{\theta}[\pi_{t+1}], \sigma_{\pi}^2 \right).$$

B.3 Monthly inflation as a random walk

Throughout the paper we consider the monthly inflation to be a random walk process. The reason behind is that, with monthly inflation data, we cannot reject the hypothesis of unit root. To further expand on this, in Table B.4 we present the results of AR(1) regressions on the monthly US inflation rate from January 1960 to March 2022. Column 1 does not consider a constant while Column 2 does.

Table B.4: AR(1) regression for monthly inflation

	(1)	(2)
π_{t-1}	0.999*** (0.993 - 1.005)	0.992*** (0.982 - 1.002)
Observations	747	747
R-squared	0.993	0.981

Note: Table shows the results of AR(1) regressions with the monthly inflation rate in the US from January 1960 to March 2022. Column 1 does not consider a constant while Column 2 does. The current inflation rate measure is based on the monthly YoY percentage variation of the CPI. 95 percent confidence intervals in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

We see that in both specifications the autorregressive coefficient is very close to 1. Furthermore, the value of 1 falls within the 95 percent confidence intervals. Also, the F-test in which the null hypothesis is that the autorregressive coefficient is equal to 1 gives a p-value of 0.66 when we do not consider a constant and a p-value of 0.12 when we consider a constant, such that we cannot reject the null hypothesis in any of the two cases. Lastly, an augmented Dickey-Fuller test on the monthly inflation gives a p-value of 0.52, so that we cannot reject the null hypothesis of a unit root. These findings are in line with those of Pivetta and Reis [2007], who find that there is a very high persistence in the quarterly inflation rate in the US.

Given that we take the inflation rate to be a random walk process, we now turn to the calibration of the Kalman filter of section 2.4.1. There, we defined the agents assume that inflation behaves as

$$\pi_{t+1} = \pi_t + \varepsilon_t,$$

where ε_t is a normally independent and identically distributed inflation shock.

Agents also receive a signal s_t given by

$$s_t = \pi_{t+1} + v_t,$$

where v_t is a normally independent and identically distributed signal noise.

Furthermore, we assume that the signal is $s_t = \pi_{t-1}^{food}$ where π_{t-1}^{food} denotes the inflation in period $t - 1$ of the food component of the CPI. By manipulating the equations we get

$$\varepsilon_t = \pi_t - \pi_{t+1},$$

$$v_t = s_t - \pi_{t+1} = \pi_{t-1}^{food} - \pi_{t+1}.$$

Then, armed with the US monthly inflation and food inflation series from January 1960 to March 2022 we calculate the variances σ_ε^2 , σ_v^2 and the covariance $\sigma_{\varepsilon v}$.

B.4 Constant Kalman gain across cohorts

In Section 2.3 we argued that cohorts do not react differently to inflation news. Here, we provide further empirical evidence that supports this fact in a structurally-based fashion. We start from the standard Kalman filter specification to obtain

$$\mathbb{E}_{i,t} [\pi_{t+1}] = \mathbb{E}_{i,t-1} [\pi_{t+1}] + \kappa (s_t - \mathbb{E}_{i,t-1} [\pi_{t+1}]),$$

where $\mathbb{E}_{i,t} [\pi_{t+1}]$ is the inflation expectation of cohort i , s_t is the signal and κ is the Kalman gain.

In order to uncover whether cohorts react differently to the signal they are receiving, we estimate κ by cohorts as

$$\mathbb{E}_{m,i,t}^{SCE} [\pi_{t+1}] = \mathbb{E}_{m,i,t-1}^{SCE} [\pi_{t+1}] + \sum_{i \neq 1941}^{2002} \kappa_i (s_t - \mathbb{E}_{m,i,t-1}^{SCE} [\pi_{t+1}]) \times \mathbb{I}_i, \quad (\text{B.1})$$

where we use the inflation expectations coming from agent m of cohort i in period t from the Survey of Consumer Expectations as the empirical counterpart for the theoretical inflation expectations¹. $\mathbb{I}_i$ serves as an indicator variable for each cohort, which is defined by the date of birth. If different cohorts react differently to the signal they are receiving, then we should obtain different coefficients across cohorts.

Panel (a) of Figure B.9 shows the results. We find that there are no systematically significant differences across cohorts when it comes to the Kalman gain. We interpret this as evidence that supports our fact of cohorts not reacting differently to the inflation signals they receive.

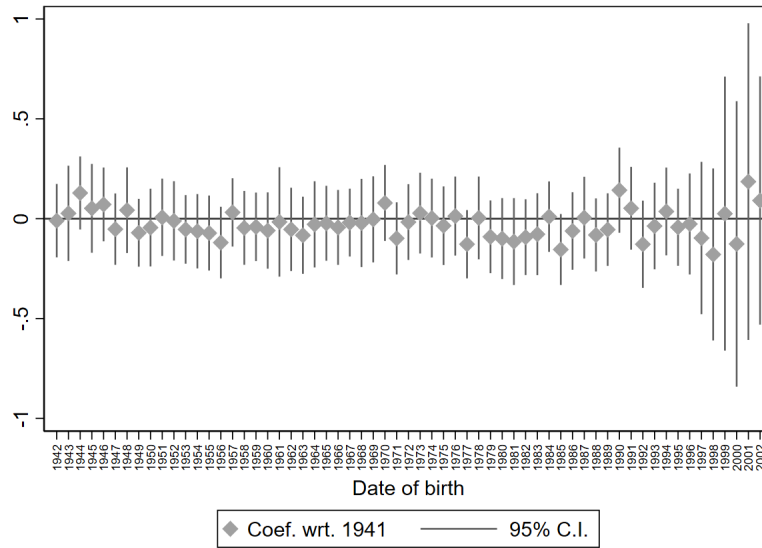
An alternative way of defining cohorts is by the age of the agents at the moment of the survey. Under this definition we estimate

$$\mathbb{E}_{m,i,t}^{SCE} [\pi_{t+1}] = \mathbb{E}_{m,i,t-1}^{SCE} [\pi_{t+1}] + \sum_{i \neq 18}^{80} \kappa_i (s_t - \mathbb{E}_{m,i,t-1}^{SCE} [\pi_{t+1}]) \times \mathbb{I}_i, \quad (\text{B.2})$$

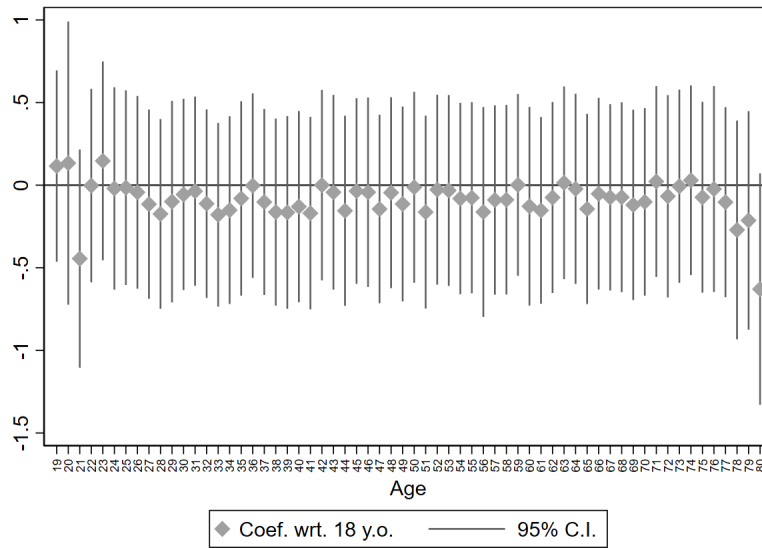
Panel (b) of Figure B.9 presents the results. As before, we find there are no systematic differences in the Kalman gain across cohorts when defined by current age.

¹In order to obtain $\mathbb{E}_{m,i,t-1}^{SCE} [\pi_{t+1}]$ we assume agents believe inflation behaves as a random walk.

Figure B.9: Constant Kalman gain across cohorts



(a) Cohorts by date of birth



(b) Cohorts by age

Note: Figure shows the coefficient resulting from estimating Equations B.1 and B.2. The base cohort in Panel (a) is those born in 1941. The base cohort in Panel (b) is those who were 18 years old at the moment of the survey. We control for period and individual fixed effects.

B.5 History-based Kalman filter with AR(1) assumption

In this section we repeat the forecasting exercise from Section 2.4 but replacing the random walk assumption with an AR(1) specification. Therefore, agents assume that inflation behaves as

$$\pi_{t+1} = \rho_\pi \pi_t + \varepsilon_t,$$

where the coefficient $\rho_\pi \in [0, 1]$ captures the mean-reversion of the inflation variable. Here, we assume the inflation rate has been properly demeaned.

As before, we assume that the signal is given by

$$s_t = \zeta \pi_{t+1} + v_t.$$

The forecasted value of the inflation variable is

$$\mathbb{E}_{i,t}^{KF} [\pi_{t+1}] = (1 - \zeta K) \mathbb{E}_{i,t-1}^{KF} [\pi_{t+1}] + K s_t,$$

where the difference now lies in the fact that agents use the AR(1) assumption to forecast the inflation rate such that

$$\mathbb{E}_{i,t}^{KF} [\pi_{t+h}] = \rho_\pi^{h-1} \mathbb{E}_{i,t}^{KF} [\pi_{t+1}].$$

In this section we assume $\zeta = 1$, $\rho = 0.99$, $\sigma_\varepsilon = 0.15$, $\sigma_\nu = 4.09$ and $\sigma_{\varepsilon\nu} = -0.06$.² This gives $K = 0.175$.

Table B.5 presents the result of the parameter estimation. In this case, $\theta = -0.612$. Finally, Figure B.10 presents the comparison of the forecasts with the AR(1) assumption and the observed forecasts in the data. We find that, as with the random walk process, this version of the forecast based on an AR(1) assumption provides a good fit to the data.

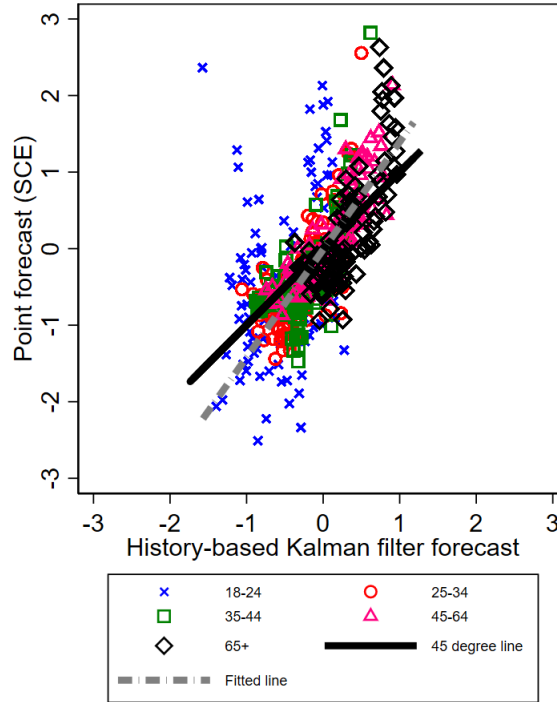
²We calibrate the variances and covariance according to Appendix B.3.

Table B.5: Parameter estimation, AR(1)

(1)	
$\mathbb{E}_{i,t}^{ref} [\pi_{t+12}]$	0.612*** (0.067)
Time FE	Yes
Observations	83,991
R-squared	0.100

Note: Table shows results of Regression 2.6. $\mathbb{E}_{i,t}^{ref} [\pi_{t+12}]$ is the reference constructed for a respondent of age i as explained in the main text. We use population weights. Column 1 has only a time fixed effect as an additional control. Robust standard errors in parentheses. Standard errors clustered by age. Dependent variable trimmed at 10 percent and 90 percent in each period. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure B.10: Observed inflation forecasts and history-based Kalman filter forecasts, AR(1)



Note: Figure shows binned scatterplot across history-based Kalman filter forecasts (x-axis) and point forecast inflation expectations according to the Survey of Consumer Expectations (SCE) of the Federal Reserve Bank of New York (y-axis). Variables demeaned by the intercept. Data go from June 2013 to December 2021. SCE variable trimmed at 10 percent and 90 percent in each period.

B.6 History-based Kalman filter with alternative signals

Throughout the paper we use the lagged inflation rate of the food component of the CPI as a signal of the non-observed aggregate inflation variable. This is based on previous papers study how consumers form their inflation expectations based on shopping experiences and food prices [D’Acunto et al., 2021, D’Acunto and Weber, 2022]. In this section we consider two alternative components of the CPI as the signals consumers use: food at home and dairy.

Table B.6 presents the estimation of the parameter according to each signal. In all cases we find a positive weight on the reference. Moreover, the parameter is close to the baseline from Table 2.2.

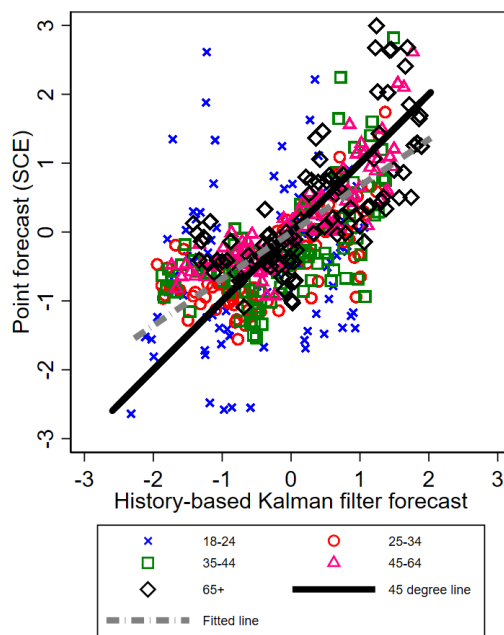
Figure B.11 shows how the history-based Kalman filter forecasts with alternative signals fare against the observed inflation forecasts. Using the food at home and dairy CPI components as a signal provides a decent fit to the observed data.

Table B.6: Main parameter estimation, alternative signals

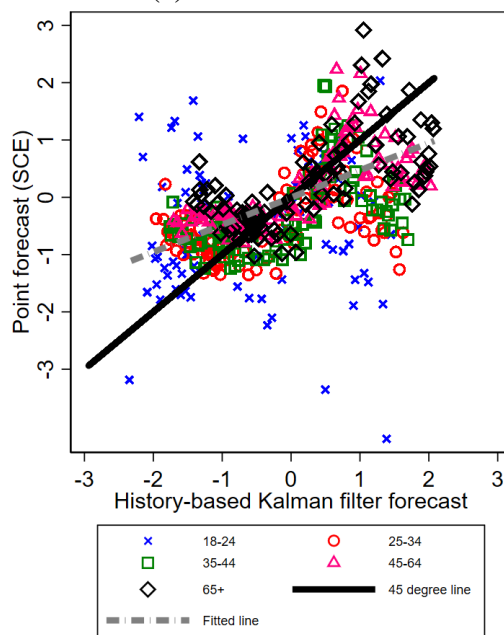
	Food at home	Dairy
$\mathbb{E}_{i,t}^{ref} [\pi_{t+12}]$	0.359*** (0.038)	0.380*** (0.043)
Observations	83,991	83,991
R-squared	0.101	0.100

Note: Table shows results of Regression 2.6 using alternative CPI components as the signal. $\mathbb{E}_{i,t}^{ref} [\pi_{t+12}]$ is the reference constructed for a respondent of age i as explained in the main text. We use population weights. Columns have only a time fixed effect as an additional control. Robust standard errors in parentheses. Standard errors clustered by age. Dependent variable trimmed at 10 percent and 90 percent in each period. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure B.11: Observed inflation forecasts and history-based Kalman filter forecasts, alternative signals



(a) Food at home



(b) Dairy

Note: Figure shows binned scatterplot across history-based Kalman filter forecasts (x-axis) and point forecast inflation expectations according to the Survey of Consumer Expectations (SCE) of the Federal Reserve Bank of New York (y-axis) using alternative CPI components as the signal. Variables demeaned by the intercept. Data go from June 2013 to December 2021. SCE variable trimmed at 10 percent and 90 percent in each period.

B.7 External validity: European data

We check the external validity of our results using data from the Consumer Expectations Survey (CES) of the European Central Bank. It contains monthly data between April 2020 and September 2022 for six countries: Belgium, France, Germany, Italy, the Netherlands, and Spain.³

We see that in Europe, lifetime experiences with the inflation rate are also heterogeneous across cohorts, as we show in Figure B.13. Moreover, we see that by 2020 the youngest cohorts had not been exposed to high inflation rates, but this changes after the high inflation rate episode of 2021 and 2022. After this, the youngest cohorts are the ones that show the highest lifetime average for the inflation rate, even larger than that of the people who experienced the high inflation rates of the 80s.

In Figure B.14 we relate the two previous facts and find that in Europe, similar to the US, the larger the inflation rate individuals have experienced in their lifetimes, the higher their inflation expectations.

Table B.7 shows that in Europe, as happened in the US, after controlling for the average lifetime inflation rate, younger generations do not react more strongly to inflation news than older cohorts.⁴

We now turn to the history-based Kalman filter of Section 2.4.1. Table B.8 shows the parameters that go into the Kalman filter calibration after using European inflation rate and food inflation rate. Then, we estimate the parameter according to Equation 2.6.⁵ In Table B.9 in our baseline specification of Column 1 we find $\theta^{eur} = -0.156$. With this parameter, in Figure B.15 we plot inflation expectations according to our history-based Kalman filter, across cohorts and in each of the six countries in our sample. We find that the oldest cohorts have the highest inflation expectations before 2021. Then, after 2021 the youngest cohorts start catching up with the oldest ones and even surpass them in some countries.

Table B.9 also shows additional specifications of the estimation of Equation 2.6. We find that after controlling for cohort and country fixed effects, we still find that agents weight positively their references when forming their expectations. These additional specifications also tell us that the heterogeneity in expectations across cohorts is not due to people of different ages or from

³There is a relevant difference between the data sets of US and Europe. In the former we have the exact age of the respondents. In the latter we do not have detailed information on the age of the respondents, as they are classified in 4 age groups: 18-34, 35-49, 50-70 and 71+.

⁴We confirm the finding with a F-test where the null hypothesis is that all of the interactions are jointly equal to zero. The test gives a p-value of 0.35, so we cannot reject the null hypothesis.

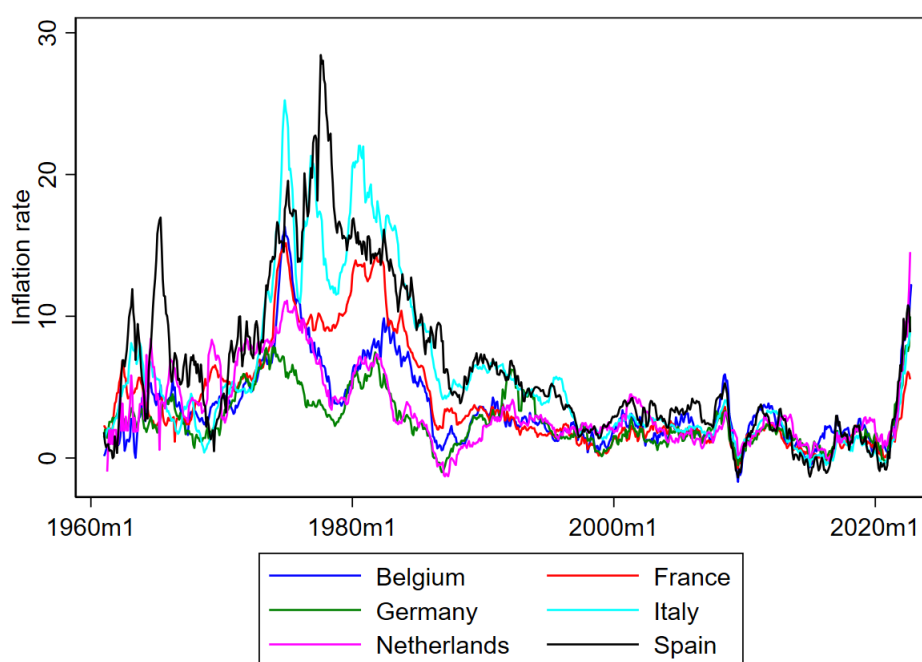
⁵Because we do not know the exact age of the respondents, we do not know which are the exact lifetime average inflation rates they have experienced. Therefore, for this estimation, we assume that every agent in cohort 18-34 has the lifetime average inflation rate of a 25-year-old, every agent in cohort 35-49 has the lifetime average inflation rate of a 35-year-old, every agent in cohort 50-70 has the lifetime average inflation rate of a 50-year-old and every agent in cohort 71+ has the lifetime average inflation rate of a 71-year-old. On the signals used, because the series on the inflation rate of the food component of the CPI have varying starting dates in the different countries, we replace the missing values with the observed inflation rate in order to make the starting dates of all countries uniform.

different countries facing different consumption bundles or having different preferences, but to the proposed anchoring-to-the-past mechanism. Thus, it is past experiences that define expectations, not the age or the geographic location per se.⁶

Lastly, in Figure B.16 we compare the inflation expectations generated by our history-based Kalman filter to the survey data. We see that we have a decent fit to the data.

We conclude that our findings from the main text are also valid for Europe. We find evidence that supports the claim that (i) inflation expectations are also heterogeneous in Europe and (ii) can also be modeled by a history-based Kalman filter with positive weight on their historical reference.

Figure B.12: Inflation rate, Europe

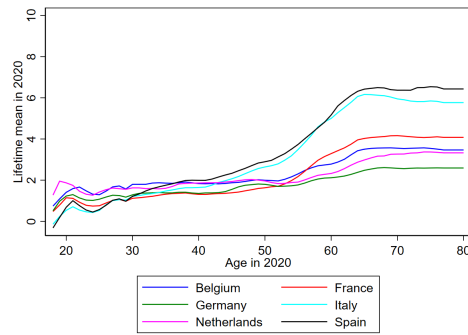


Note: The current inflation rate measure is based on the monthly YoY percentage variation of the CPI.

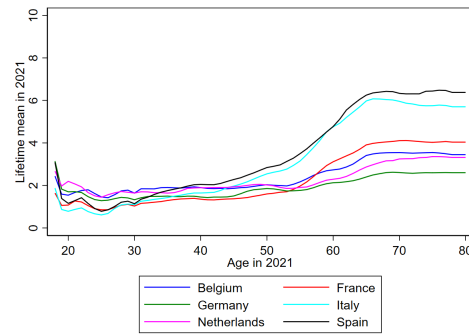
Source: FRED.

⁶See Hajdini et al. [2022a] for a further discussion on this.

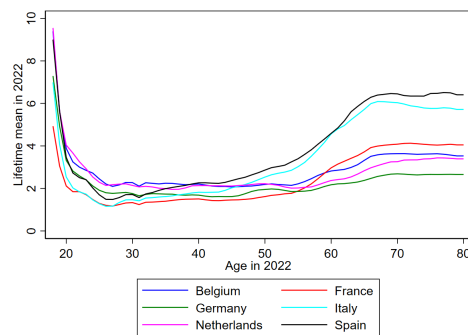
Figure B.13: Lifetime average inflation rate among respondents, Europe



(a) 2020



(b) 2021

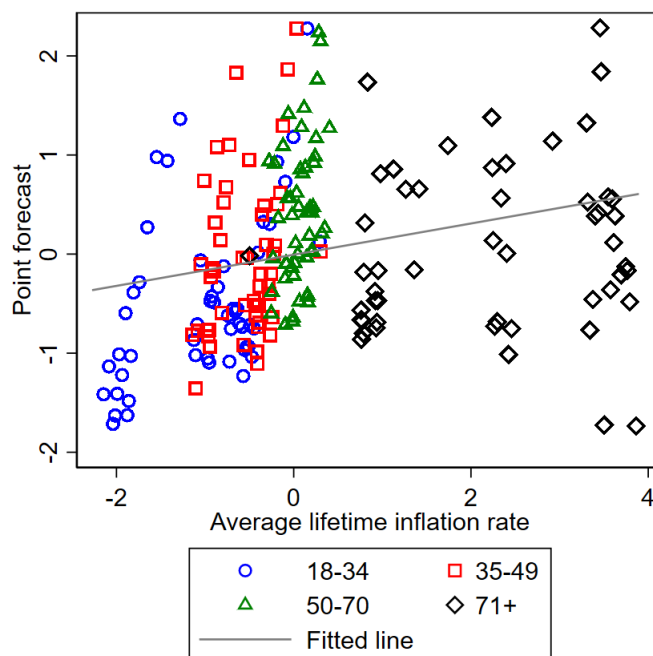


(c) 2022

Note: Figure shows the mean of the monthly YoY inflation rate that people of the age shown in 2020, 2021, and 2022 have experienced in their lifetimes, starting when they were age 18. The current inflation rate measure is based on the monthly YoY percentage variation of the CPI.

Source: FRED.

Figure B.14: Inflation point forecast and average lifetime inflation rate, Europe



Note: Figure shows binned scatterplot across lifetime average inflation rate bins. Variables residualized by respondent gender and commuting zone. Data go from April 2020 to September 2022. Ages correspond to the interviewee's age at the time of the survey. The average lifetime inflation rate measure is based on the monthly YoY percentage variation of the CPI.

Source: Consumer Expectations Survey.

Table B.7: Effects of current and experienced inflation rates on inflation expectations, Europe

Dep. var.: Inflation expectations	(1)	(2)	(3)	(4)
Average lifetime inflation rate	0.276* (0.132)	0.244** (0.100)	0.301* (0.149)	0.252* (0.129)
Current inflation	0.351*** (0.033)	0.275*** (0.057)		
Cohort 35-49		0.155 (0.090)		
Cohort 50-70		0.203* (0.112)		
Cohort 71+		-0.401 (0.348)		
Current inflation $\times$ 35-49		0.074 (0.074)		
Current inflation $\times$ 50-70		0.122 (0.069)		
Current inflation $\times$ 71+		0.124 (0.083)		
Time FE	No	No	Yes	Yes
Controls	No	No	No	Yes
Observations	294,232	294,232	294,232	294,232
R-squared	0.128	0.135	0.152	0.164

Note: Table shows regressions where the dependent variable is inflation expectations according to the Consumer Expectations Survey (CES) of the European Central Bank. Column 1 shows controls for the average lifetime inflation of respondents of a given age at each period in time and the last inflation measure. Column 2 follows Column 1 but adds cohort fixed effects and the interaction of those cohort fixed effects with the current inflation. Column 3 follows Column 1 but adds time fixed effects and, hence, omits the current inflation variable. Column 4 follows Column 1 but adds time fixed effects and demographic controls. The demographic controls are income, gender, educational level, and country. The current inflation rate measure is based on the monthly YoY percentage variation of the CPI. Robust standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by age. The dependent variable is trimmed, dropping the lower and upper 10 percent of answers in each period.

Table B.8: History-based Kalman filter parameters, Europe

	σ_ε^2	σ_v^2	$\sigma_{\varepsilon v}$	K
Belgium	0.16	2.51	-0.28	0.24
France	0.09	2.63	-0.14	0.17
Germany	0.12	3.13	-0.15	0.18
Italy	0.23	3.38	-0.38	0.24
Netherlands	0.13	3.54	-0.18	0.18
Spain	0.45	5.83	-0.65	0.26

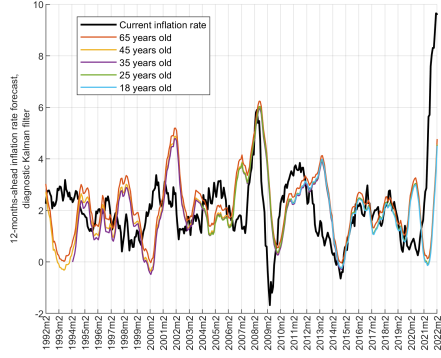
Note: We obtain this calibration for each country following the steps outlined in Appendix B.3. The data for these calculations goes from January 1971 to October 2022.

Table B.9: Parameter estimation, Europe

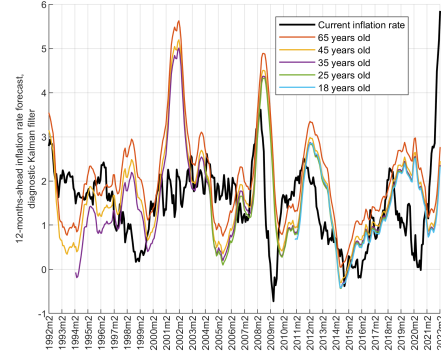
	(1)	(2)	(3)	(4)
$\mathbb{E}_{i,t}^{ref} [\pi_{t+12}]$	0.156*** (0.011)	0.208*** (0.016)	0.094*** (0.011)	0.060*** (0.018)
Time FE	Yes	Yes	Yes	Yes
Controls	No	Cohort	Country	Cohort, country
Observations	271,311	271,311	271,311	271,311
R-squared	0.122	0.130	0.132	0.140

Note: Table shows results of Regression 2.6, but changing the dependent variable for inflation expectations according to the Consumer Expectations Survey (CES) of the European Central Bank. We use population weights. The independent variable $\mathbb{E}_{i,t}^{ref} [\pi_{t+12}]$ is the reference constructed for a respondent of age i as explained in the main text. Column 1 considers a time fixed effect as a control. Column 2 has time and cohort fixed effects. Column 3 has time and country fixed effects. Column 4 has time, cohort, and country fixed effects. Robust standard errors in parentheses. Dependent variable trimmed at 10 percent and 90 percent in each period. *** p<0.01, ** p<0.05, * p<0.1.

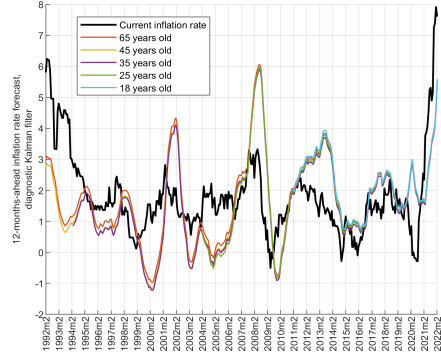
Figure B.15: History-based Kalman-filter-based inflation forecasts by cohort, Europe



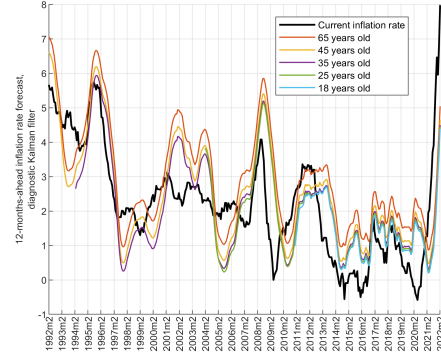
(a) Belgium



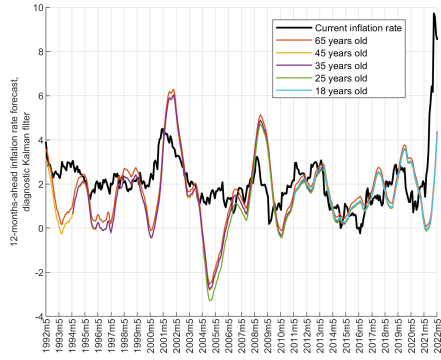
(b) France



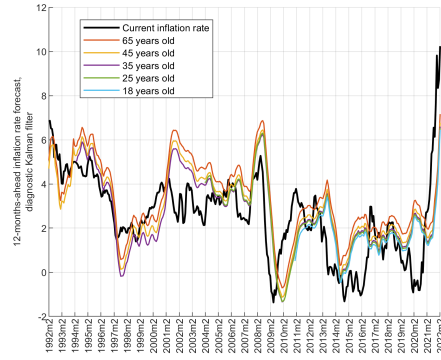
(c) Germany



(d) Italy



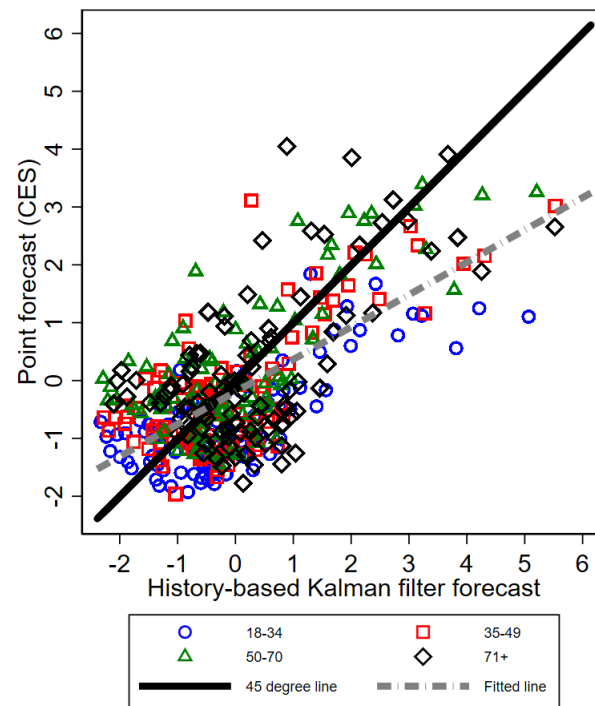
(e) Netherlands



(f) Spain

Note: Figure shows forecasts for selected cohorts according to the Kalman-filter-augmented expectations and considering the estimate for θ from Column 1 of Table B.9. Selected cohorts differentiated by their age in 2021. We further assume that each cohort starts forecasting when they become 18 years old. The current inflation rate measure is based on the monthly YoY percentage variation of the CPI.

Figure B.16: Observed inflation forecasts and history-based Kalman filter forecasts, Europe



Note: Figure shows binned scatterplot across history-based Kalman filter forecasts (x-axis) and point forecasts of inflation expectations according to the Consumer Expectations Survey (CES) of the European Central Bank (y-axis). Variables demeaned by the intercept. Data go from April 2020 to September 2022. SCE variable trimmed at 10 percent and 90 percent in each period.

B.8 Derivations for firm block

B.8.1 Final good producer

The final good producer operates in a perfectly competitive market. It produces the final good Y_t using a CES basket composed by a continuum of intermediate goods $Y_t(j)$ with $j \in [0, 1]$. Therefore, the optimization problem of the firm is

$$\max_{\{Y_t(j)\}_{j \in [0,1]}} P_t Y_t - \int_{j \in [0,1]} P_t(j) Y_t(j) dj,$$

subject to

$$Y_t = \left[\int_{j \in [0,1]} Y_t(j)^{\frac{\varepsilon-1}{\varepsilon}} dj \right]^{\frac{\varepsilon}{\varepsilon-1}},$$

where $P_t(j)$ is the price of intermediate good j , and $\varepsilon > 1$ is the elasticity of substitution in the CES basket. As a solution, we obtain the corresponding downward-sloping demand curve $\forall j \in [0, 1]$

$$Y_t(j) = \left(\frac{P_t(j)}{P_t} \right)^{-\varepsilon} Y_t,$$

where the price of this production basket is given by

$$P_t = \left[\int_{j \in [0,1]} P_t(j)^{1-\varepsilon} dj \right]^{\frac{1}{1-\varepsilon}},$$

B.8.2 Intermediate good producers

There is a continuum of monopolistically competitive firms indexed by $j \in [0, 1]$. Each firm $j \in [0, 1]$ produces the intermediate good $Y_t(j)$ with a production function given by

$$Y_t(j) = A_t L_t(j),$$

where A_t is a process that represents technology, and $L_t(j)$ is the labor employed by firm j . The intermediate good producer will pay a nominal wage w_t to workers.

We assume that intermediate good producers face price rigidities à la Calvo, where the probability of price change corresponds to $(1 - \phi) \in (0, 1)$. Therefore, a reoptimizing intermediate

good producer $j \in [0, 1]$ solves

$$\max_{\{P_t(j)\}} \mathbb{E}_t \left[\sum_{k \geq t} \phi^{k-t} Q_{t,k} (P_t(j) Y_{k|t}(j) - \Psi_k(Y_{k|t}(j))) \right],$$

where $\Psi_t(\cdot)$ is a well-defined cost function. Here, $Q_{t,k} = \beta^{k-t} (C_k/C_t)^{-\sigma} (P_t/P_k)$ with $k \geq t$ is the household's stochastic discount factor used by firms. We assume that intermediate goods producers use the stochastic discount factor associated with the aggregate consumption level. This optimization is subject to the following sequence of demand for $k \geq t$:

$$Y_{k|t}(j) = \left(\frac{P_t(j)}{P_k} \right)^{-\epsilon} Y_k.$$

The first-order condition of this problem leads to

$$Y_{t|t}(j) \left[1 - \mu MC_{t|t}(j) \left(\frac{P_t}{P_t(j)} \right) \right] + \mathbb{E}_t \left[\sum_{k > t} \phi^{k-t} Q_{t,k} Y_{k|t}(j) \left[1 - \mu MC_{k|t}(j) \left(\frac{P_k}{P_k(j)} \right) \right] \right] = 0,$$

where $MC_{k|t}(j) = \Psi'_k(Y_{k|t}(j)) / P_k$ denotes real marginal costs for $k \geq t$, and $\mu = \epsilon/(\epsilon-1) > 1$ is the desired frictionless mark-up. Under the assumption of symmetry across all firms, the solution implies $P_t(j) = P_t \forall j \in [0, 1]$. Therefore, the previous first-order condition reduces to

$$Y_{t|t}(j) (1 - \mu MC_{t|t}(j)) + \mathbb{E}_t \left[\sum_{k > t} \phi^{k-t} Q_{t,k} Y_{k|t}(j) (1 - \mu MC_{k|t}(j)) \right] = 0.$$

Regarding the cost structure of this model, i.e. the $\Psi_t(\cdot)$ function, we assume that the only cost of the firms are labor costs

$$\Psi_k(Y_{k|t}(j)) = \frac{W_k Y_{k|t}(j)}{A_k},$$

which implies $MC_{k|t}(j) = W_k / (A_k P_k) = MC_k \forall j \in [0, 1]$ and

$$\mathcal{M}_t = P_t MC_t = \frac{W_t}{A_t}.$$

Using the aggregate market clearing condition in the market of goods, i.e. $Y_t = C_t \forall t$, the optimal price P_t^* must satisfy:

$$P_t^* = \mu \frac{\mathbb{E}_t \left[\sum_{k \geq t} \phi^{k-t} Q_{t,k} Y_{k|t}(j) \mathcal{M}_k \right]}{\mathbb{E}_t \left[\sum_{k \geq t} \phi^{k-t} Q_{t,k} Y_{k|t}(j) \right]} = \mu \frac{\mathbb{E}_t \left[\sum_{k \geq t} (\phi\beta)^{k-t} P_k^\varepsilon C_k^{1-\sigma} MC_k \right]}{\mathbb{E}_t \left[\sum_{k \geq t} (\phi\beta)^{k-t} P_k^{\varepsilon-1} C_k^{1-\sigma} \right]}.$$

A log-linearization of the previous equation leads to the following condition for the optimal price

$$p_t^* = \mu + (1 - \phi\beta) \mathbb{E}_t \left[\sum_{k \geq t} (\phi\beta)^{k-t} (p_k + mc_k) \right].$$

Finally, and following the standard procedure in the literature, we can derive the New Keynesian Phillips curve. Under our framework, this equations corresponds to

$$\pi_t = \beta \mathbb{E}_t [\pi_{t+1}] + \frac{(1 - \phi)(1 - \phi\beta)}{\phi} (\sigma + \eta) y_t.$$

B.9 Heterogeneous cohorts

In the baseline exercises we assumed a history-based Kalman filter operator where history varies by cohort, as in Equation 2.3. In this section we analyze a variation of such operator, where we assume that history is fixed across all cohorts. We define this alternative history-based Kalman filter operator as

$$E_t^{\theta,alt} [X_{t+h}] = E_t^{KF} [X_{t+h}] + \theta \left(E_t^{KF} [X_{t+h}] - \sum_{j=1}^J \frac{E_{t-j}^{KF} [X_{t+h}]}{J} \right),$$

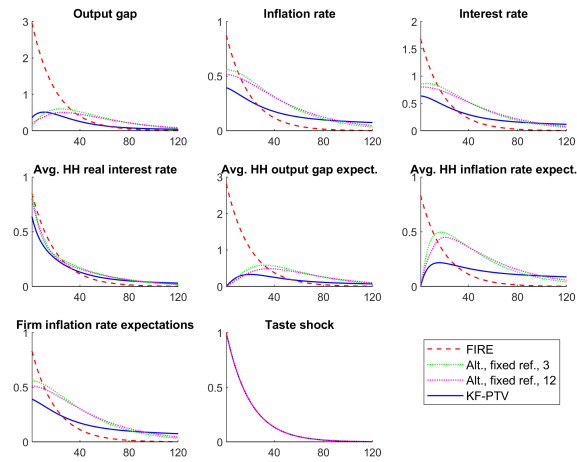
where agents remember what occurred in the last J periods. As in the baseline case, we set $\theta = -0.374$.

Figure B.17 shows the impulse response functions to a cost shock and a taste shock according to our model of Section 2.5. Besides the responses coming from the FIRE and history-based Kalman filter cases with varying history across cohorts, we also consider the two cases for the alternative history-based Kalman filter operator: fixing the history of all cohorts to the last 3 periods and fixing the reference of all cohorts to the last 12 periods.

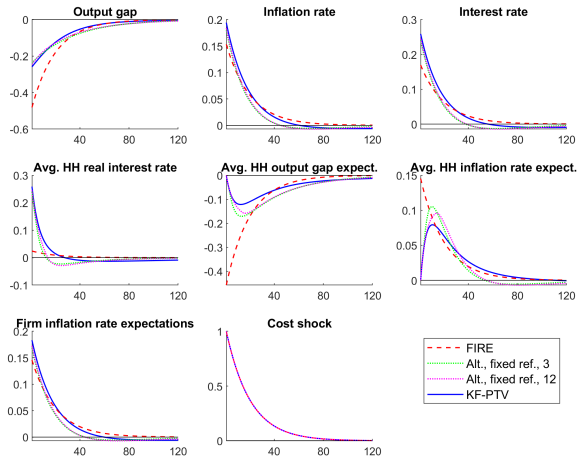
We see that the three cases follow the same pattern when compared to the FIRE case. However, the expectations under the fixed reference cases show a stronger reaction to the shocks. This is because the history span in these two alternative fixed reference cases is shorter than in the full-fledged history-based case. While the latter remembers and is pegged to the steady state for longer, the former cases do not.

It could be argued that if we choose a sufficiently long history with the alternative fixed-reference Kalman filter operator, then we could closely replicate the results coming from the full-fledged history-based case with different cohorts. However, we prefer the full-fledged history-based case with different cohorts as it picks up the richness of the data and introduces it in the model: there are different cohorts living at the same time, each has had different life experiences and each has different beliefs.

Figure B.17: Impulse response functions, comparison with alternative history-based Kalman filter operator



(a) Taste shock



(b) Cost shock

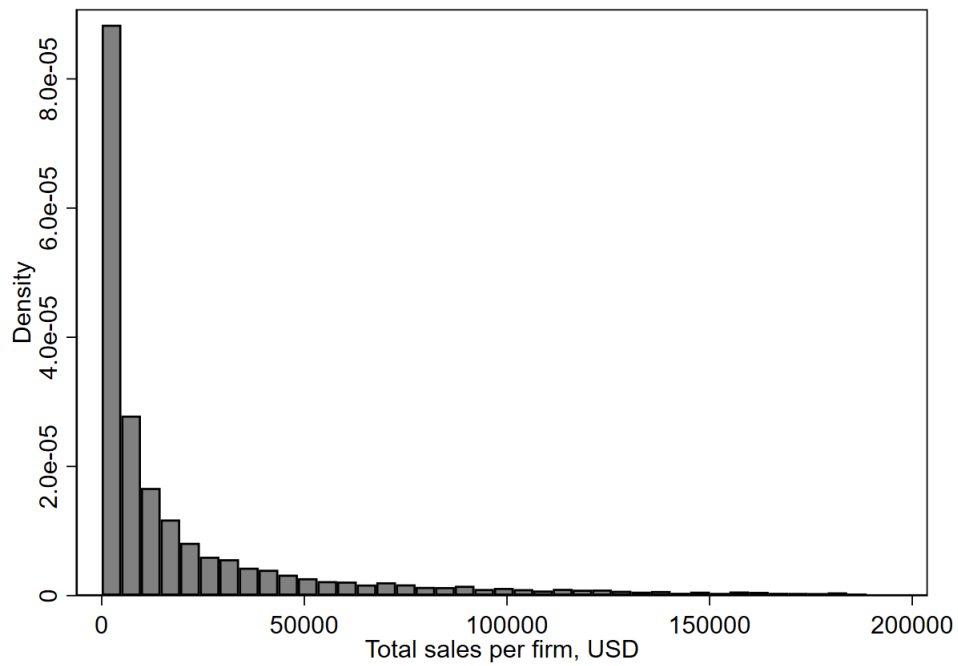
Note: Figure shows impulse response functions for a selected group of variables after the mentioned shocks. The red dashed line shows the results for the case of the full information rational expectations model (FIRE), the green and magenta dotted lines show the results of history-based Kalman filter model with fixed reference and the solid blue line shows the history-based Kalman filter model where the reference varies by cohort. Horizontal axis denotes months after the shock.

APPENDIX C

Appendix to Chapter 3

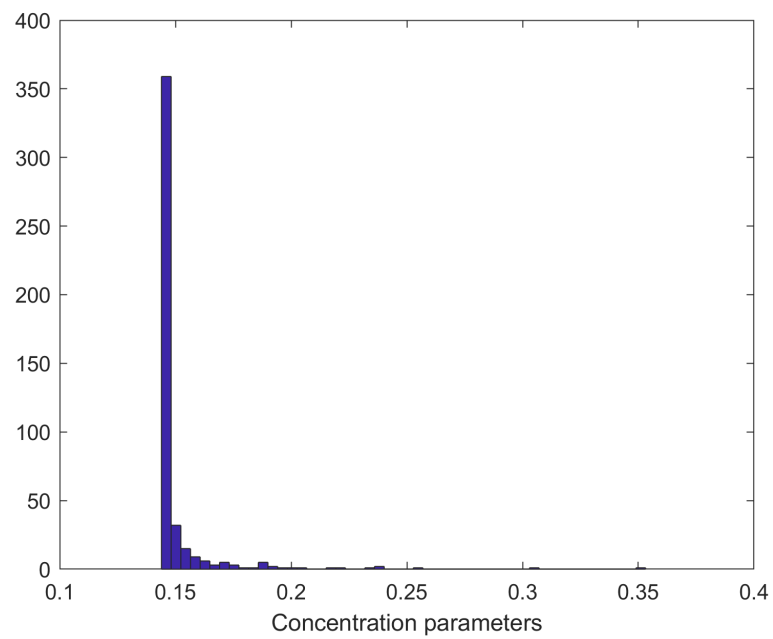
C.1 Additional figures and tables

Figure C.1: Distribution of total sales per firm



Notes: Figure shows distribution of total sales per firm up to the 90th percentile. Values expressed in US dollars.

Figure C.2: Histogram of estimated concentration parameters for Dirichlet distribution



Notes: Estimated concentration parameters for a Dirichlet distribution according to the maximum likelihood estimation from Equation 3.21.

Table C.1: Data summary

Panel A: Aggregate statistics	
Number of firms	21,497
Number of sellers	10,010
Number of buyers	16,403
Transactions	823,923
Total sales	1.2 billion USD
Average sales per transaction	1,434 USD

Panel B: Statistics per firm						
	Average	Std. Dev.	Min.	Median	Max.	IQR
Sales per seller	118,069	698,015	0	7,972	28,300,000	35,411
Transactions per seller	82	834	1	6	60391	23
Matches per seller	4	11	1	2	322	3
Types of varieties sold per seller	3	6	1	1	100	2
Purchases per buyer	72,052	450,260	0	5,805	17,400,000	25,151
Transactions per buyer	50	276	1	7	16,059	26
Matches per buyer	3	4	1	1	147	2
Types of varieties bought per buyer	5	8	1	2	211	5

Notes: Sales and purchases expressed in USD.

Table C.2: Estimation for matching cost

	(1)	(2)
	1st Stage	2nd Stage
Dep. Variable	Sales (Hyperbolic Inverse Sine)	Trade Indicator
BC	0.013*** (0.001)	0.131*** (0.008)
$\text{iht}[\widehat{n(z, z')}]$		8.340*** (0.024)
Obs.	5,606,627	5,606,627
Adj. R2	0.595	-
Pseudo R2	-	0.453
FE	Seller, buyer	-

Notes: Column 1 shows the results of estimating Equation 3.22. Column 2 shows the results of estimating Equation 3.23. We winsorize $\ln \widehat{n(z, z')}$ at 1 percent and 99 percent. Sample only contains in-state firms. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales refer to the total sales for a given pair of seller and buyer, respectively. The trade indicator is equal to 1 if there was at least 1 recorded transaction for a given pair of seller and buyer, and it is equal to 0 otherwise. Standard errors are clustered at the seller and buyer level in Column 1. Standard errors are in parentheses. The higher the Bhattacharyya coefficient, the culturally closer two firms are.

C.2 Robustness

C.2.1 Correction for selection bias

Helpman et al. [2008] suggest that the traditional gravity equation estimations, which we use for our intensive margin regressions, are biased because of selection issues. Thus, in this section, we follow their proposed correction for selection bias.

In the first stage, we estimate the following linear probability model:

$$tr(\nu, \omega) = \iota_\nu + \iota_\omega + \delta BC(\nu, \omega) + \eta \Xi(\nu, \omega) + \gamma B2B(\nu, \omega) + \epsilon(\nu, \omega), \quad (C.1)$$

where we follow the nomenclature and the in-state sample from our extensive margin regressions. Here, we need an excluded instrument that affects only the extensive margin (i.e. the matching cost) and not the intensive margin (i.e. the trade cost). Thus, we consider the indicator variable $B2B(\nu, \omega)$ that equals 1 when both seller ν and buyer ω are in IndiaMART and equals 0 otherwise. As mentioned in Section 3.2.2, IndiaMART is the largest e-commerce platform for business-to-business (B2B) transactions in India.¹ Thus, the idea here is that it is easier for both firms to match if they take part in this platform.

We present the results of this first stage in Table C.3. As before, the closer the firms are in cultural terms, the more likely it is they will trade. Additionally, if both firms participate in IndiaMART, the more likely the trade.

After the estimation, we recover the predicted probability of trading $\hat{tr}(\nu, \omega)$, with which we calculate the latent variable

$$\hat{\zeta}(\nu, \omega) = \Phi^{-1}(\hat{tr}(\nu, \omega)),$$

where $\Phi^{-1}(\cdot)$ is the inverse CDF of the standard normal distribution. Following Heckman [1979], we obtain the inverse Mills ratio

$$\Upsilon(\hat{\zeta}) = \frac{\phi(\hat{\zeta}(\nu, \omega))}{\Phi(\hat{\zeta}(\nu, \omega))},$$

where $\phi(\cdot)$ is the PDF of the standard normal distribution, and $\Phi(\cdot)$ is the CDF of the standard normal distribution.

For the second stage, we estimate

$$\ln y(\nu, \omega) = \iota_\nu + \iota_\omega + \delta BC(\nu, \omega) + \eta \Xi(\nu, \omega) + v \Upsilon(\hat{\zeta}) + \epsilon(\nu, \omega), \quad (C.2)$$

¹In 2019, there were between 5 and 6 million registered firms in IndiaMART (<https://www.indiamart.com>), which represented all firm size groups and all geographic regions in India.

where $y(\nu, \omega)$ is the total positive sales of seller ν to buyer ω and the term $\Upsilon(\hat{\zeta})$ accounts for selection bias.

We present the second stage results in Columns 3 and 4 of Table C.4. We must note that for computational reasons, we work with only an in-state sample, our results are not directly comparable to the baseline results from Table 3.1. Therefore, Columns 1 and 2 present the results with the in-state sample but without the correction for selection bias.

We find that not considering the correction for selection biases the coefficients downwards. This way, we conclude that the main results related to the intensive margin in the paper represent a lower bound of the effect of cultural proximity on trade.

Table C.3: Correction for selection bias, first stage

Dep. Variable	Trade Indicator
<i>BC</i>	0.0010*** (0.0001)
<i>B2B</i>	0.0016*** (0.0003)
Obs.	5,628,290
Adj. R2	0.0106
FE	Seller, buyer, origin×dest.

Notes: Table shows the results of estimating Equation C.1. The sample contains only in-state firms. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. The trade indicator is equal to 1 if there was at least 1 recorded transaction for a given pair of seller and buyer, and it is equal to 0 otherwise. Standard errors are two-way clustered at the seller and buyer level. Standard errors are in parentheses. The higher the Bhattacharyya coefficient, the culturally closer two firms are.

Table C.4: Correction for selection bias, second stage

	(1)	(2)	(3)	(4)
Dep. Variable	Log Sales	Log Transactions	Log Sales	Log Transactions
<i>BC</i>	0.148*** (0.035)	0.095*** (0.029)	0.223*** (0.074)	0.132** (0.055)
Inv. Mills rat.			0.503 (0.421)	0.246 (0.298)
Obs.	26,238	26,238	26,238	26,238
Adj. R2	0.392	0.360	0.392	0.360
FE	Seller, buyer, origin×dest.	Seller, buyer, origin×dest.	Seller, buyer, origin×dest.	Seller, buyer, origin×dest.

Notes: Columns 1, 2, 3 and 4 show the results of estimating Equation C.2. Columns 1 and 2 do not consider the correction for selection bias term. The sample contains only in-state firms. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales and transactions refer to the total sales and total number of transactions for a given pair of seller and buyer, respectively. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are two-way clustered at the seller and buyer level. Standard errors are in parentheses. The higher the Bhattacharyya coefficient, the culturally closer two firms are. Number of observations varies between specifications due to the dropping of observations separated by a fixed effect [Correia et al., 2019].

C.2.2 Alternative cultural proximity measure

In this section, we present an alternative measure of cultural proximity to that of the Bhattacharyya coefficient. Define the standard discrete distribution-based Kullback and Leibler [1951] divergence as

$$KL(\nu||\omega) = \sum_{x=1}^X \rho_{\nu}(x) \log \left(\frac{\rho_{\nu}(x)}{\rho_{\omega}(x)} \right).$$

We have that $KL(\nu||\omega) \geq 0$, where $KL(\nu||\omega) = 0$ when sellers and buyers have exactly equal probability distributions, while it will be higher the more different the two probability distributions are.² Intuitively, we can see this measure as the expected difference between two probability distributions. However, this proximity measure is not symmetric; that is, $KL(\nu||\omega) \neq KL(\omega||\nu)$. Consider our previous example where we record a transaction between a seller ν and a buyer with distribution ω , from which we calculate $KL(\nu||\omega)$. If, in a second transaction, the roles of the firms

²This interpretation diverts from the standard use the Kullback-Leibler has in information theory, where a higher divergence means a higher information loss.

revert, then the Kullback-Leibler divergence would be $KL(\omega||\nu)$, implying the cultural proximity between the two firms has changed, when it should not change. To convert this measure into a symmetric one, we define

$$KL_{sym}(\nu||\omega) = KL(\nu||\omega) + KL(\omega||\nu) = KL_{sym}(\omega||\nu).$$

Notice this similarity measure needs $\rho_\nu(x) > 0$ and $\rho_\omega(x) > 0$ for all x . However, the probability of a firm belonging to a certain cultural group may be zero. In those cases, we replace that probability of zero for a probability $\varepsilon \rightarrow 0^+$ such that KL_{sym} is well-defined. Tables C.5 and C.6 show the regression results for the intensive margin, extensive margin, and unit prices. In this case, the higher the Kullback-Leibler divergence, the more culturally different the buyer from the seller. The results confirm the findings from the main text.

Table C.5: Effect of cultural proximity on trade, intensive and extensive margins, Kullback-Leibler

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable	Log Sales	Log Transactions	Log Sales	Log Transactions	Trade Indicator	Trade Indicator
KL_{sym}	-0.004*** (0.001)	-0.003** (0.001)	-0.005*** (0.002)	-0.003** (0.001)	-0.00004*** (0.00000)	-0.00004*** (0.00000)
Log dist.	-0.023 (0.015)	-0.065*** (0.011)			0.00007 (0.00005)	
Obs.	32,678	32,678	32,843	32,843	5,606,627	5,628,290
Adj. R2	0.415	0.359	0.410	0.356	0.617	0.0106
FE	Seller, buyer	Seller, buyer	Seller, buyer, origin $\times$ dest.	Seller, buyer, origin $\times$ dest.	Seller, buyer	Seller, buyer, origin $\times$ dest.

Notes: Columns 1, 2, 3 and 4 show the results of estimating a modified version of Equation 3.1. Columns 5 and 6 show the results of estimating a modified version of Equation 3.2. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales and transactions refer to the total sales and total number of transactions for a given pair of seller and buyer, respectively. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are two-way clustered at the seller and buyer level. Standard errors are in parentheses. A higher Kullback-Leibler divergence means two firms are socially farther away. Number of observations varies between specifications due to the dropping of observations separated by a fixed effect [Correia et al., 2019].

Table C.6: Effect of cultural proximity on prices, Kullback-Leibler

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable	Log Prices	Log Prices	Log Prices	Log Prices	Log Prices	Log Prices
KL_{sym}	0.003** (0.001)	0.003** (0.001)	0.003** (0.001)	0.002* (0.001)	0.002** (0.001)	0.001 (0.001)
Log dist.	0.023 (0.016)	0.023 (0.016)	0.028* (0.017)			
Obs.	230,744	230,744	226,645	235,001	236,617	230,900
Adj. R2	0.932	0.932	0.935	0.933	0.925	0.936
FE	Seller×HS, Seller×HS, Seller×HS, Seller×HS, Seller×HS, Seller×HS, buyer buyer, buyer, buyer, buyer, buyer, month month×HS origin×dest. month, month×HS, origin×dest.origin×dest.					

Notes: This table shows the results of estimating a modified version of Equation 3.3. Good g is defined according to 6-digit HS classification. Prices trimmed by 4-digit HS code at 5 and 95 percent. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Prices refer to unit prices of goods sold. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are multi-way clustered at the seller, 4-digit HS and origin-destination level. Standard errors are in parentheses. A higher Kullback-Leibler divergence means two firms are culturally farther away. Number of observations varies between specifications due to the dropping of observations separated by a fixed effect [Correia et al., 2019].

C.2.3 Language

In this section, we check if the results we find are driven by language similarity. To do so, we follow the two language similarity measures from Kone et al. [2018]. Define ϑ_i^l as the share of people with mother tongue l in district i . Then, the common language measure between districts i and j is

$$commlang_{ij} = \sum_l \vartheta_i^l \vartheta_j^l. \quad (C.3)$$

We can also define a language overlap measure, defined as

$$overlang_{ij} = \sum_l \min \{ \vartheta_i^l, \vartheta_j^l \}. \quad (C.4)$$

In both cases, the larger the measures, the less likely it should be for people in these districts to face communication barriers. Then, we estimate the following intensive margin regression that considers language measures:

$$\ln y(\nu, \omega) = \iota_\nu + \iota_\omega + \delta BC(\nu, \omega) + \eta \Xi(\nu, \omega) + \zeta lang + \varepsilon(\nu, \omega), \quad (C.5)$$

where $lang$ is either $commlang_{ij}$ or $overlang_{ij}$.

Table C.7 presents the results. We find that none of the measures is statistically significant. This suggests that the cultural proximity result is not driven by firms sharing the same language.

Table C.7: Effect of cultural proximity and language on trade, intensive margin

	(1)	(2)	(3)	(4)
Dep. Variable	Log Sales	Log Transactions	Log Sales	Log Transactions
<i>BC</i>	0.108*** (0.033)	0.068** (0.028)	0.108*** (0.033)	0.068** (0.028)
<i>commlang</i>	-0.322 (0.389)	-0.126 (0.305)		
<i>overlang</i>			-0.419 (0.406)	-0.061 (0.324)
Obs.	30,703	30,703	30,703	30,703
Adj. R2	0.409	0.357	0.409	0.357
FE	Seller, buyer	Seller, buyer	Seller, buyer	Seller, buyer

Notes: This table shows the results of estimating Equation C.5. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales and transactions refer to the total sales and total number of transactions for a given pair of seller and buyer, respectively. Standard errors two-way clustered at the seller and buyer level. Standard errors are in parentheses. *commlang* and *overlang* are language distance measures according to Equations C.3 and C.4.

C.2.4 Goods specialization

The cultural groups in India are, in many cases, defined by the production of specific goods [Munshi, 2019].³ In this section, we study whether the cultural proximity results come from the fact that cultural groups specialize in the production and consumption of certain goods. If this were true, then the cultural proximity results would be due to different cultural groups specializing in certain goods, and therefore forming specific links with other cultural groups that either produce their inputs or consume their goods.

Here, we first must know which are the goods in which each cultural group specializes in selling and buying within our dataset. For this, we take the following steps:

1. We assign each firm to a unique cultural group. We do this by assigning each firm to the cultural group to which it has the highest probability of belonging according to its vector of probabilities ρ .
2. We determine which is the most important 4-digit HS code in terms of sales and purchases for each cultural group.
3. We create a dummy variable $\mathbb{I}_g^{spec,seller}$ that equals 1 when a firm sells the good in which its cultural group specializes in selling. Similarly, we define dummy variable $\mathbb{I}_g^{spec,buyer}$ that equals 1 when a firm buys the good in which its cultural group specializes in buying.

Working with a version of our dataset at the transaction level, we estimate the regression

$$\ln n_g(\nu, \omega, t) = \iota_{\nu g} + \iota_{gt} + \iota_{\omega} + \delta BC(\nu, \omega) + \xi(BC(\nu, \omega) \times \mathbb{I}_g^{spec}) + \eta \Xi(\nu, \omega) + \epsilon_g(\nu, \omega). \quad (C.6)$$

Table C.8 presents the results for the sales. First, if the cultural proximity results were only driven by cultural groups producing specific specialized goods, then we would expect the term on cultural proximity to be close to zero, and the term on the interactions to be statistically different from zero. However, we find that cultural proximity matters for all types of goods: for those in which a cultural group specializes and for those in which a cultural group does not specialize.

Second, in Column 2, we find that the coefficient on the interaction term is positive and statistically significant. Nevertheless, we lose this statistical significance after controlling for additional variables in Column 4. This could point to cultural proximity mattering more for those goods in which cultural groups specialize in buying, but the result is not conclusive.

All in all, the results suggest that specialization does not play a role in the determination of the effect of cultural proximity on trade.

³We can also understand this as certain cultural groups specializing in certain occupations.

Table C.8: Effect of cultural proximity on trade by good specialization, intensive margin

	(1)	(2)	(3)	(4)
Dep. Variable	Log Sales	Log Sales	Log Sales	Log Sales
BC	0.072*** (0.026)	0.071*** (0.025)	0.064*** (0.023)	0.064*** (0.023)
$BC \times \mathbb{I}_g^{spec,seller}$	-0.016 (0.160)		0.135 (0.304)	
$BC \times \mathbb{I}_g^{spec,buyer}$		0.152*** (0.008)		0.185 (0.118)
Obs.	226,039	226,039	229,719	229,719
Adj. R2	0.853	0.853	0.854	0.854
FE	Seller×HS, Seller×HS, Seller×HS, Seller×HS, buyer, buyer, buyer, buyer, month×HS month×HS month×HS, month×HS, origin×dest. origin×dest.			

Notes: This table shows the results of estimating Equation C.6. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales refer to the sales of specific good g for a given pair of seller and buyer. Good g is defined according to 6-digit HS classification. Sales were trimmed by 4-digit HS code at 5 and 95 percent. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are two-way clustered at the seller and 4-digit HS level. Standard errors are in parentheses. The higher the Bhattacharyya coefficient, the culturally closer the two firms are. The number of observations varies between specifications due to the dropping of observations separated by a fixed effect [Correia et al., 2019]. $\mathbb{I}_g^{spec,seller}$ indicates the good g is the good in which the seller's cultural group specializes in selling. $\mathbb{I}_g^{spec,buyer}$ indicates the good g is the good in which the buyer's cultural group specializes in buying.

C.2.5 Industry pair linkages

In this section, we revise the intensive margin regressions after considering that there are pairs of industries that are bound to trade more than other pairs. For instance, perhaps certain castes happen to specialize in certain industries, and these industries are more likely to trade with each other. For this analysis, we add a selling firm's industry $\times$ buying firm's industry fixed effect to the intensive margin estimation. The sectors are based on the 4-digit HS code of the good with the highest sales for each firm. Therefore, we estimate

$$\ln y(\nu, \omega) = \iota_\nu + \iota_\omega + \delta BC(\nu, \omega) + \eta \Xi(\nu, \omega) + (\iota_\nu^{ind} \times \iota_\omega^{ind}) + \varepsilon(\nu, \omega), \quad (C.7)$$

where ι_ν^{ind} and ι_ω^{ind} denote industry fixed effects for the seller ν and the buyer ω .

Table C.9 presents the results. When compared to the results in Table 3.1, we find that, while the effect of cultural proximity is slightly higher, the main message prevails.

Table C.9: Effect of cultural proximity after controlling for industries, intensive margin

	(1)	(2)	(3)	(4)
Dep. Variable	Log Sales	Log Transactions	Log Sales	Log Transactions
<i>BC</i>	0.105** (0.052)	0.089** (0.043)	0.145*** (0.055)	0.104** (0.045)
Obs.	16,194	16,194	16,229	16,229
Adj. R2	0.414	0.326	0.395	0.308
FE	Seller, buyer, seller ind. $\times$ buyer ind.	Seller, buyer, seller ind. $\times$ buyer ind.	Seller, buyer, origin $\times$ dest., seller ind. $\times$ buyer ind.	Seller, buyer, origin $\times$ dest., seller ind. $\times$ buyer ind.

Notes: Columns 1, 2, 3 and 4 show the results of estimating Equation C.7. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales and transactions refer to the total sales and total number of transactions for a given pair of seller and buyer, respectively. Origin-destination fixed effect considers the district of the seller and the buyer. An industry is classified according to the 4-digit HS classification of the most sold good by each firm. Standard errors are two-way clustered at the seller and buyer level. Standard errors are in parentheses. The higher the Bhattacharyya coefficient, the culturally closer two firms are. Number of observations varies between specifications due to the dropping of observations separated by a fixed effect [Correia et al., 2019].

C.3 Model

C.3.1 Technology

There is a continuum of firms that operate under monopolistic competition and produce differentiated goods indexed by ω . Since each firm produces a unique good, ω denotes a firm or a good. We consider a roundabout production economy, so firms can purchase intermediate inputs from all firms in the economy. Each firm produces its differentiated good with hired labor from the representative household and intermediate inputs.

Firms operate in three steps. First, sellers endogenously choose buyers (i.e. matching). In particular, a seller ν sells to buyer ω whenever the profits of doing so are larger than the fixed costs of matching $\epsilon F(\nu, \omega)$. Guided by our stylized facts, we argue that cultural proximity $BC(\nu, \omega)$ between seller ν and buyer ω determines the pairwise matching costs $F(\nu, \omega)$. Indeed, previous work describes how culture encodes information that is useful for sellers to decide which buyer to sell to [Allen et al., 2019, Balmaceda and Escobar, 2017].⁴

Second, upon matching, due to the risk of buyers withholding payment, sellers charge a premium to buyers (i.e. terms of trade). In particular, the price that seller ν charges to buyer ω includes a premium $d(\nu, \omega) \geq 1$, which is a function of different factors that influence the hold-up process. Guided by our stylized facts, we posit that cultural proximity $BC(\nu, \omega)$ between seller ν and buyer ω determines $d(\nu, \omega)$. This is similar to earlier work on how culture can be an informal institutional channel to solve hold-up problems.⁵ Third, sellers and buyers trade.

Step 1: Matching. In the first step, we endogenize sellers selecting which buyers to trade with by laying out the maximization problem of firms, and how cultural proximity influences it. A seller ν matches with buyer ω whenever the seller's profits of doing so $\pi(\nu, \omega)$ are larger than the fixed costs of matching $\epsilon F(\nu, \omega)$, where $F(\nu, \omega)$ are pairwise fixed costs of matching, and ϵ are i.i.d. log normal errors with mean $\mu_{\ln(\epsilon)}$ and standard deviation $\sigma_{\ln(\epsilon)}$. It can be shown that profits are proportional to the value of intermediate sales $n(\nu, \omega)$ from seller ν to buyer ω , such that

$$\pi(\nu, \omega) = \frac{n(\nu, \omega)}{\sigma}. \quad (\text{C.8})$$

Sellers are looking for buyers, and vice-versa. Firms can meet costlessly, so all firms can meet each other before deciding who to match with. The purpose of a meeting is to infer the fixed costs the seller would incur if they were to trade with each buyer. Here we rationalize how sellers infer

⁴There could be other microfoundations motivating how culture influences matching costs (e.g. risk-sharing as in Ambrus et al. 2014, Bloch et al. 2008. These would generate similar frictions at the extensive and intensive margin.

⁵Again, this would be consistent with other microfoundations on how culture influences trade costs due to reputation [Banerjee and Duflo, 2000, Chen and Wu, 2021] or loyalty [Board, 2011].

the costs of matching arising from cultural proximity. We assume that each firm owner has a name and that sellers and buyers exchange names upon meeting. Then, each name has a probability mapping to different cultural groups. We assume that this mapping is public information. So, for example, we assume that all firm owners know that the surname *Shah* is associated with either of the three groups: Jain, Muslim (Faquir), or Hindu (Vaishnav baniya), with probabilities reflecting the empirical distribution.

For each cultural group j , we consider the matching function

$$M^j(\nu, \omega) = (\rho_\nu(j))^\varphi (\rho_\omega(j))^{1-\varphi},$$

where $\rho_\nu(j)$ is the probability that seller ν belongs to cultural group j , $\rho_\omega(j)$ is the probability that buyer ω belongs to cultural group j , φ is the weight of ν to determine their proximity. For simplicity, we assume $\varphi = \frac{1}{2}$. The expected proximity between ν and ω is

$$\bar{M}(\nu, \omega) = \frac{1}{\mathcal{X}} \sum_{j=1}^{\mathcal{X}} M^j(\nu, \omega) \propto BC(\nu, \omega).$$

That is, the expected proximity between seller ν and buyer ω is proportional to the Bhattacharyya coefficient we use in the empirical part of the paper to measure cultural proximity between firms. Finally, following our stylized facts, the fixed costs of matching are a function of cultural proximity such that $F(\nu, \omega) = f(\bar{M}(\nu, \omega))$, where $f(\cdot)$ is an exponential function for simplicity. Our modeling decision allows for the fact that a firm belonging to a cultural group encodes information that other firms use to determine the cost of matching. Then, we consider

$$F(\nu, \omega) = \kappa + \exp(-\gamma BC(\nu, \omega)), \quad (\text{C.9})$$

where $\gamma > 0$ measures the sensitivity of the pairwise matching cost to the cultural proximity $BC(\nu, \omega)$, and κ is a scaling constant.

The intuition behind parameter γ is that it encases court quality, such that $\gamma = \gamma(\mathbb{I}^{court})$, where $\mathbb{I}^{court}$ is an indicator of low court quality. In this sense, Table C.2 suggests that parameter γ increases in magnitude when $\mathbb{I}^{court} = 1$ and firms face weak court quality. The lower the quality of courts is, the larger parameter γ is, and the more firms rely on cultural proximity to match.

We now describe how firms match. Before the formation of the network, firms are characterized by $\lambda = (z, \rho)$, where z is productivity, and ρ is the vector of probabilities of firm λ belonging to each cultural group. After firms meet each other and infer their cultural proximity, firms now only

differ in their productivity z . So the share of seller-buyer pairs (z, z') is

$$l(z, z') = \int \mathbb{I} \left[\ln(\pi(z, z')) - \ln(F(z, z')) - \ln(\epsilon(z, z')) > 0 \right] dH(\epsilon(z, z')), \quad (\text{C.10})$$

where $l(z, z')$ is called the *link function*. From Equations 3.18 and C.10, we see that the higher the cultural proximity, the lower the matching cost and the larger the probability of matching. This relates to Stylized Fact 2.

Step 2: Terms of the contract. After matching, seller ν and buyer ω set up the terms of the contract. We first consider the possibility that the buyer may withhold payment after the seller ships the goods. The seller's concerns around this issue can be monetized in the premium that the seller charges to the buyer. This premium gets passed to the unit price $p(\nu, \omega)$ that results from profit maximization

$$p(\nu, \omega) = \mu c(\nu) d(\nu, \omega), \quad (\text{C.11})$$

where $c(\nu)$ is the marginal cost and $\mu \equiv \frac{\sigma}{\sigma-1}$ is a markup.

We now turn to explain the premium $d(\nu, \omega)$. Time is continuous within our one-period static model. Seller ν considers the time T that buyer ω withholds payment for the first time is a random variable. T follows an exponential distribution with intensity rate $\delta(\nu, \omega)$. Then, the probability seller ν waits for more than t units of time until buyer ω withholds payment for the first time is

$$\begin{aligned} P(T > t) &= 1 - P(T \leq t), \\ &= 1 - (1 - \exp(-\delta(\nu, \omega)t)), \\ &= \exp(-\delta(\nu, \omega)t). \end{aligned}$$

Seller ν cares that buyer ω never withholds payment. Then, the probability $\bar{p}$ that seller ν waits more than a unit of time until buyer ω withholds payment is

$$\begin{aligned} \bar{p} &= P(T > 1), \\ &= \exp(-\delta(\nu, \omega)). \end{aligned}$$

Then, the expected number of times buyer ω withholds payment is $\frac{1}{\bar{p}}$. We posit that the premium $d(\nu, \omega)$ is proportional to the expected number of times buyer ω withholds payment to seller ν . The intuition is that seller ν will charge a higher premium if there is a higher hold-up risk from buyer

ω . For simplicity,

$$\begin{aligned} d(\nu, \omega) &= \frac{1}{\bar{p}} \geq 1, \\ &= \exp(\delta(\nu, \omega)). \end{aligned}$$

Finally, we allow $\delta(\nu, \omega)$ to include a set of covariates that affects the terms of trade. For example, inputs can spoil or get lost more easily with larger distances, which increases the probability of buyer ω not paying seller ν since the goods did not arrive as agreed.

More importantly, guided by our stylized facts, we allow for the fact that, upon matching, cultural proximity $BC(\nu, \omega)$ also influences $\delta(\nu, \omega)$. In particular, we consider

$$d(\nu, \omega) = \exp(\beta BC(\nu, \omega) + \epsilon(\nu, \omega)), \quad (\text{C.12})$$

where $\beta < 0$ is a trade cost semi-elasticity, and $\epsilon(\nu, \omega)$ are unobservables.

From Equation C.11, we have that the higher the cultural proximity, the lower the prices, which relates to Stylized Fact 3. Likewise, from Equation 3.20, we see that the higher the cultural proximity, the higher the intermediate sales, which relates to Stylized Fact 1.

The economic intuition of parameter β is that it incorporates information on court quality, such that $\beta = \beta(\mathbb{I}^{court})$. Based on the result of Table 3.4, we know model parameter β increases in magnitude when $\mathbb{I}^{court} = 1$ and firms face bad courts. Therefore, parameter β relates to how firms respond to court quality: the lower the quality of the courts, the larger the magnitude of parameter β , and the more important cultural proximity is for trade.

Step 3: Trade. Finally, we model trade. Each firm has a technology

$$y(\omega) = \kappa_\alpha z(\omega) l(\omega)^\alpha m(\omega)^{1-\alpha}, \quad (\text{C.13})$$

where $y(\omega)$ is output, $\kappa_\alpha \equiv \frac{1}{\alpha^\alpha(1-\alpha)^{1-\alpha}}$ is a normalization constant, $z(\omega)$ is firm-level productivity, $l(\omega)$ is labor, and $m(\omega)$ are intermediate inputs from other firms. In turn, the intermediate inputs are defined as a CES composite so

$$m(\omega) = \left(\int_{\nu \in \Omega(\omega)} m(\nu, \omega)^{\frac{\sigma-1}{\sigma}} d\nu \right)^{\frac{\sigma}{\sigma-1}},$$

where $m(\nu, \omega)$ is quantity of inputs from seller ν to buyer ω , $\sigma > 1$ is the elasticity of substitution across intermediates, and $\Omega(\omega)$ is the endogenous set of suppliers of buyer ω . By cost minimiza-

tion, we get

$$c(\omega) = \frac{P(\omega)^{1-\alpha}}{z(\omega)}, \quad (\text{C.14})$$

where $P(\omega) \equiv \left(\int_{\nu \in \Omega(\omega)} p(\nu, \omega)^{1-\sigma} d\nu \right)^{\frac{1}{1-\sigma}}$ is a CES price index across prices of intermediates, and labor is the numeraire good, so $w = 1$. Profit maximization subject to demand generates constant markup pricing such that the unit price is $p(\nu, \omega) = \mu c(\nu) d(\nu, \omega)$, as stated before.

We now derive the demand for intermediates, so

$$n(\nu, \omega) = p(\nu, \omega)^{1-\sigma} P(\omega)^{\sigma-1} N(\omega), \quad (\text{C.15})$$

where $N(\omega) = \int_{\nu \in \Omega(\omega)} n(\nu, \omega) d\nu$ is the total intermediate purchases by buyer ω and $n(\nu, \omega) \equiv p(\nu, \omega) m(\nu, \omega)$ is the value of purchases from seller ν to buyer ω . From Equation C.15 we can obtain the gravity equation as

$$\log(n(\nu, \omega)) = \iota_\nu + \iota_\omega + (1 - \sigma) \log(d(\nu, \omega)), \quad (\text{C.16})$$

where ι_ν and ι_ω are seller and buyer fixed effects. Here, the premium $d(\nu, \omega)$ enters the gravity equation as a trade cost. This gravity equation relates directly to Equation 3.1 that we estimate.

C.3.2 Equilibrium given extensive margin

Conditional on the extensive margin (i.e. sellers matching with buyers), firms only differ in productivity z . Based on the price index of all of the goods acquired by firm z' , we get

$$P(z')^{1-\sigma} = \mu^{1-\sigma} \int P(z)^{(1-\alpha)(1-\sigma)} z^{\sigma-1} d(z, z')^{1-\sigma} l(z, z') dG(z), \quad (\text{C.17})$$

where $l(z, z')$ is the share of sellers of productivity z that sell to buyers with productivity z' , also called the *link function*. Now, total sales of firm z is the sum of sales to households plus intermediates, so

$$S(z) = \left[\mu^{1-\sigma} P(z)^{(1-\alpha)(1-\sigma)} z^{\sigma-1} \right] \times \left[\frac{Y}{p^{1-\sigma}} + \left(\frac{1-\alpha}{\mu} \right) \left(\int \left[d(z, z')^{1-\sigma} P(z')^{\sigma-1} S(z') \right] l(z, z') dG(z') \right) \right]. \quad (\text{C.18})$$

C.3.3 Preferences

A representative household demands goods from firms with CES $\sigma > 1$, the elasticity of substitution between goods, and inelastically supplies labor to firms. The household maximizes utility

$$\max_{\{y(\omega)\}} \left(\int_{\omega \in \Omega} y(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}}, \text{ s.t. } \int_{\omega \in \Omega} p(\omega) y(\omega) d\omega \leq Y,$$

where $y(\omega)$ is the household demand for good ω , $p(\omega)$ is the price the household pays for good ω , Ω is the set of goods in the economy, and Y is total income. This generates the demand for good ω

$$x(\omega) = p(\omega)^{1-\sigma} p^{\sigma-1} Y, \quad (\text{C.19})$$

where $x(\omega) \equiv p(\omega) y(\omega)$, and $p \equiv \left(\int_{\omega \in \Omega} p(\omega)^{1-\sigma} d\omega \right)^{\frac{1}{1-\sigma}}$ is a CES price index.

C.4 Model derivations

C.4.1 Technology

A unique variety ω is produced by a single firm which minimizes its unit cost of production subject to its production technology, so

$$\begin{aligned} \min_{\{m(\nu, \omega)\}} \quad & \int_{\nu \in \Omega(\omega)} m(\nu, \omega) p(\nu, \omega) d\nu + w l(\omega), \text{ s.t. } \\ & y(\omega) = \kappa_\alpha z(\omega) l(\omega)^\alpha m(\omega)^{1-\alpha}, \\ & m(\omega) = \left(\int_{\nu \in \Omega(\omega)} m(\nu, \omega)^{\frac{\sigma-1}{\sigma}} d\nu \right)^{\frac{\sigma}{\sigma-1}}, \\ & y(\omega) = 1. \end{aligned}$$

Merge the first and third constraints, such that

$$\begin{aligned} y(\omega) &= \kappa_\alpha z(\omega) l(\omega)^\alpha m(\omega)^{1-\alpha}, \\ 1 &= \kappa_\alpha z(\omega) l(\omega)^\alpha m(\omega)^{1-\alpha}, \\ l(\omega)^\alpha &= \frac{1}{\kappa_\alpha z(\omega) m(\omega)^{1-\alpha}}, \\ &= \kappa_\alpha^{-1} z(\omega)^{-1} m(\omega)^{\alpha-1}, \\ l(\omega) &= \kappa_\alpha^{-\frac{1}{\alpha}} z(\omega)^{-\frac{1}{\alpha}} m(\omega)^{\frac{\alpha-1}{\alpha}}. \end{aligned}$$

Rewrite the minimization problem, such that

$$\begin{aligned} \min_{\{m(\nu, \omega)\}} & \int_{\nu \in \Omega(\omega)} m(\nu, \omega) p(\nu, \omega) d\nu + w l(\omega), \\ & \int_{\nu \in \Omega(\omega)} m(\nu, \omega) p(\nu, \omega) d\nu + \kappa_\alpha^{-\frac{1}{\alpha}} w z(\omega)^{-\frac{1}{\alpha}} m(\omega)^{\frac{\alpha-1}{\alpha}}, \\ & \int_{\nu \in \Omega(\omega)} m(\nu, \omega) p(\nu, \omega) d\nu + \kappa_\alpha^{-\frac{1}{\alpha}} w z(\omega)^{-\frac{1}{\alpha}} \left(\int_{\nu \in \Omega(\omega)} m(\nu, \omega)^{\frac{\sigma-1}{\sigma}} d\nu \right)^{\frac{\sigma}{\sigma-1} \frac{\alpha-1}{\alpha}}. \end{aligned}$$

The first order condition with respect to $m(\nu, \omega)$ is

$$\begin{aligned} 0 &= p(\nu, \omega) + \kappa_\alpha^{-\frac{1}{\alpha}} w z(\omega)^{-\frac{1}{\alpha}} \left(\frac{\sigma}{\sigma-1} \frac{\alpha-1}{\alpha} \right) (\dots)^{\frac{\sigma}{\sigma-1} \frac{\alpha-1}{\alpha} - 1} \left(\frac{\sigma-1}{\sigma} \right) m(\nu, \omega)^{\frac{\sigma-1}{\sigma} - 1}, \\ p(\nu, \omega) &= \kappa_\alpha^{-\frac{1}{\alpha}} \left(\frac{1-\alpha}{\alpha} \right) w z(\omega)^{-\frac{1}{\alpha}} (\dots)^{\frac{\sigma}{\sigma-1} \frac{\alpha-1}{\alpha} - 1} m(\nu, \omega)^{-\frac{1}{\sigma}}, \\ m(\nu, \omega)^{\frac{1}{\sigma}} &= \frac{\kappa_\alpha^{-\frac{1}{\alpha}} \left(\frac{1-\alpha}{\alpha} \right) w z(\omega)^{-\frac{1}{\alpha}} (\dots)^{\frac{\sigma}{\sigma-1} \frac{\alpha-1}{\alpha} - 1}}{p(\nu, \omega)}, \\ m(\nu, \omega) &= \frac{\kappa_\alpha^{-\frac{\sigma}{\alpha}} \left(\frac{1-\alpha}{\alpha} \right)^\sigma w^\sigma z(\omega)^{-\frac{\sigma}{\alpha}} (\dots)^{\sigma \left(\frac{\sigma}{\sigma-1} \frac{\alpha-1}{\alpha} - 1 \right)}}{p(\nu, \omega)^\sigma}. \end{aligned}$$

Now, the first order condition with respect to $m(\nu, \omega)$ is

$$m(\nu, \omega) = \frac{\kappa_\alpha^{-\frac{\sigma}{\alpha}} \left(\frac{1-\alpha}{\alpha} \right)^\sigma w^\sigma z(\omega)^{-\frac{\sigma}{\alpha}} (\dots)^{\sigma \left(\frac{\sigma}{\sigma-1} \frac{\alpha-1}{\alpha} - 1 \right)}}{p(\nu', \omega)^\sigma}.$$

We divide both first order conditions, such that

$$\begin{aligned} \frac{m(\nu, \omega)}{m(\nu', \omega)} &= \frac{\frac{\kappa_\alpha^{-\frac{\sigma}{\alpha}} \left(\frac{1-\alpha}{\alpha} \right)^\sigma w^\sigma z(\omega)^{-\frac{\sigma}{\alpha}} (\dots)^{\sigma \left(\frac{\sigma}{\sigma-1} \frac{\alpha-1}{\alpha} - 1 \right)}}{p(\nu, \omega)^\sigma}}{\frac{\kappa_\alpha^{-\frac{\sigma}{\alpha}} \left(\frac{1-\alpha}{\alpha} \right)^\sigma w^\sigma z(\omega)^{-\frac{\sigma}{\alpha}} (\dots)^{\sigma \left(\frac{\sigma}{\sigma-1} \frac{\alpha-1}{\alpha} - 1 \right)}}{p(\nu', \omega)^\sigma}}, \\ &= \frac{\frac{z(\omega)^{-\frac{\sigma}{\alpha}}}{p(\nu, \omega)^\sigma}}{\frac{z(\omega)^{-\frac{\sigma}{\alpha}}}{p(\nu', \omega)^\sigma}}, \\ &= \frac{p(\nu', \omega)^\sigma}{p(\nu, \omega)^\sigma}, \\ m(\nu', \omega) &= \frac{p(\nu, \omega)^\sigma m(\nu, \omega)}{p(\nu', \omega)^\sigma}. \end{aligned}$$

We plug this expression back into the expression for the composite of intermediates, so

$$\begin{aligned}
m(\omega) &= \left(\int_{\nu' \in \Omega(\omega)} m(\nu', \omega)^{\frac{\sigma-1}{\sigma}} d\nu' \right)^{\frac{\sigma}{\sigma-1}}, \\
&= \left(\int_{\nu' \in \Omega(\omega)} \left(\frac{p(\nu', \omega)^\sigma m(\nu', \omega)}{p(\nu', \omega)^\sigma} \right)^{\frac{\sigma-1}{\sigma}} d\nu' \right)^{\frac{\sigma}{\sigma-1}}, \\
&= p(\nu, \omega)^\sigma m(\nu, \omega) \underbrace{\left(\int_{\nu' \in \Omega(\omega)} p(\nu', \omega)^{1-\sigma} d\nu' \right)^{\frac{\sigma}{\sigma-1}}}_{= (P(\omega)^{1-\sigma})^{\frac{\sigma}{\sigma-1}}}, \\
&= p(\nu, \omega)^\sigma m(\nu, \omega) (P(\omega)^{1-\sigma})^{\frac{\sigma}{\sigma-1}}, \\
&= p(\nu, \omega)^\sigma m(\nu, \omega) P(\omega)^{-\sigma}, \\
&= m(\omega) p(\nu, \omega)^{-\sigma} P(\omega)^\sigma, \\
p(\nu, \omega) m(\nu, \omega) &= m(\omega) p(\nu, \omega)^{1-\sigma} P(\omega)^\sigma, \\
n(\nu, \omega) &= P(\omega) m(\omega) p(\nu, \omega)^{1-\sigma} P(\omega)^{\sigma-1}, \\
&= N(\omega) p(\nu, \omega)^{1-\sigma} P(\omega)^{\sigma-1},
\end{aligned}$$

which is the demand of firm ω from variety ν , where $P(\omega)^{1-\sigma} = \int_{\nu \in \Omega(\omega)} p(\nu, \omega)^{1-\sigma} d\nu$ is the price index faced by firm ω , $n(\nu, \omega) = p(\nu, \omega) m(\nu, \omega)$ is the expenditure of ω on variety ν , and $N(\omega) = P(\omega) m(\omega)$ is the total expenditure of firm ω .

The expression for the unit cost of production is

$$\begin{aligned}
c(\omega) &= \frac{w^\alpha P(\omega)^{1-\alpha}}{z(\omega)}, \\
&= \frac{P(\omega)^{1-\alpha}}{z(\omega)},
\end{aligned}$$

where wages $w = 1$ is the numeraire price.

Now, firms engage in monopolistic competition since they produce a unique variety. In particular, firm ν maximizes profits by selling its good to buyers ω subject to the demand for its intermediate, so

$$\begin{aligned}
\max_{\{p(\nu, \omega)\}} \quad & \int_{\omega \in \Omega(\nu)} (p(\nu, \omega) - d(\nu, \omega) c(\nu)) m(\nu, \omega), \text{ s.t.} \\
& m(\nu, \omega) = m(\omega) p(\nu, \omega)^{-\sigma} P(\omega)^\sigma,
\end{aligned}$$

where $d(\nu, \omega)$ is the iceberg cost of firm ν selling to ω . Rewrite the profit function $\pi(\nu, \omega)$, such

that

$$\begin{aligned}
\pi(\nu, \omega) &= (p(\nu, \omega) - d(\nu, \omega) c(\nu)) m(\nu, \omega), \\
&= p(\nu, \omega) m(\nu, \omega) - d(\nu, \omega) c(\nu) m(\nu, \omega), \\
&= p(\nu, \omega) m(\omega) p(\nu, \omega)^{-\sigma} P(\omega)^\sigma - d(\nu, \omega) c(\nu) m(\omega) p(\nu, \omega)^{-\sigma} P(\omega)^\sigma, \\
&= m(\omega) p(\nu, \omega)^{1-\sigma} P(\omega)^\sigma - d(\nu, \omega) c(\nu) m(\omega) p(\nu, \omega)^{-\sigma} P(\omega)^\sigma.
\end{aligned}$$

The first order condition is

$$\begin{aligned}
[p(\nu, \omega)] : (1 - \sigma) m(\omega) p(\nu, \omega)^{-\sigma} P(\omega)^\sigma \\
- (-\sigma) d(\nu, \omega) c(\nu) m(\omega) p(\nu, \omega)^{-\sigma-1} P(\omega)^\sigma &= 0, \\
(\sigma - 1) m(\omega) p(\nu, \omega)^{-\sigma} P(\omega)^\sigma &= \sigma d(\nu, \omega) c(\nu) m(\omega) p(\nu, \omega)^{-\sigma-1} P(\omega)^\sigma, \\
(\sigma - 1) &= \sigma d(\nu, \omega) c(\nu) p(\nu, \omega)^{-1}, \\
p(\nu, \omega) &= \left(\frac{\sigma}{\sigma - 1} \right) c(\nu) d(\nu, \omega), \\
&= \mu c(\nu) d(\nu, \omega),
\end{aligned}$$

where $\mu = \frac{\sigma}{\sigma-1}$ is the markup. Finally, firm ω maximizes profits subject to the household demand for its good:

$$\begin{aligned}
\max_{\{p(\omega)\}} (p(\omega) - c(\omega)) y(\omega), \text{ s.t.} \\
y(\omega) = p(\omega)^{-\sigma} p^{\sigma-1} Y.
\end{aligned}$$

Rewriting profits:

$$\begin{aligned}
&= (p(\omega) - c(\omega)) (p(\omega)^{-\sigma} p^{\sigma-1} Y), \\
&= p(\omega)^{1-\sigma} p^{\sigma-1} Y - c(\omega) p(\omega)^{-\sigma} p^{\sigma-1} Y.
\end{aligned}$$

Then, the first order condition is:

$$\begin{aligned}
[p(\omega)] : (1 - \sigma) p(\omega)^{-\sigma} p^{\sigma-1} Y - (-\sigma) c(\omega) p(\omega)^{-\sigma-1} p^{\sigma-1} Y &= 0, \\
(1 - \sigma) - (-\sigma) c(\omega) p(\omega)^{-1} &= 0, \\
p(\omega) &= \left(\frac{\sigma}{\sigma - 1} \right) c(\omega), \\
p(\omega) &= \mu c(\omega),
\end{aligned}$$

where μ is the markup.

C.4.2 Preferences

A representative household maximizes its utility subject to its budget constraint, so

$$\max_{\{y(\omega)\}} \left(\int_{\omega \in \Omega} y(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}}, \text{ s.t. } \int_{\omega \in \Omega} p(\omega) y(\omega) d\omega \leq Y.$$

The first order condition with respect to firm ω is

$$\begin{aligned} [y(\omega)] : \left(\frac{\sigma}{\sigma-1} \right) (\dots)^{\frac{\sigma}{\sigma-1}-1} \left(\frac{\sigma-1}{\sigma} \right) y(\omega)^{\frac{\sigma-1}{\sigma}-1} &= \lambda p(\omega), \\ \lambda p(\omega) &= (\dots)^{\frac{\sigma}{\sigma-1}-1} y(\omega)^{-\frac{1}{\sigma}}, \end{aligned}$$

where λ is the Lagrangian multiplier of the budget constraint, and $(\dots)$ is an aggregate term we do not write down since it will cancel out during the derivation. Now, the first order condition with respect to another firm ω' is

$$\lambda p(\omega') = (\dots)^{\frac{\sigma}{\sigma-1}-1} y(\omega')^{-\frac{1}{\sigma}}.$$

We then divide both first-order conditions, such that

$$\begin{aligned} \frac{\lambda p(\omega)}{\lambda p(\omega')} &= \frac{(\dots)^{\frac{\sigma}{\sigma-1}-1} y(\omega)^{-\frac{1}{\sigma}}}{(\dots)^{\frac{\sigma}{\sigma-1}-1} y(\omega')^{-\frac{1}{\sigma}}}, \\ \frac{p(\omega)}{p(\omega')} &= \frac{y(\omega)^{-\frac{1}{\sigma}}}{y(\omega')^{-\frac{1}{\sigma}}}, \\ &= \frac{y(\omega')^{\frac{1}{\sigma}}}{y(\omega)^{\frac{1}{\sigma}}}, \\ y(\omega')^{\frac{1}{\sigma}} &= y(\omega)^{\frac{1}{\sigma}} \frac{p(\omega)}{p(\omega')}, \\ y(\omega') &= y(\omega) \left(\frac{p(\omega)}{p(\omega')} \right)^{\sigma}. \end{aligned}$$

We plug this demand back into the budget constraint, which holds with equality, so

$$\begin{aligned}
Y &= \int_{\omega' \in \Omega} p(\omega') y(\omega') d\omega, \\
&= \int_{\omega' \in \Omega} p(\omega') \left[y(\omega) \left(\frac{p(\omega)}{p(\omega')} \right)^\sigma \right] d\omega, \\
&= y(\omega) p(\omega)^\sigma \underbrace{\int_{\omega' \in \Omega} p(\omega')^{1-\sigma} d\omega}_{=p^{1-\sigma}}, \\
&= y(\omega) p(\omega)^\sigma p^{1-\sigma}, \\
&= (p(\omega) y(\omega)) p(\omega)^{\sigma-1} p^{1-\sigma}, \\
&= x(\omega) p(\omega)^{\sigma-1} p^{1-\sigma}, \\
x(\omega) &= p(\omega)^{1-\sigma} p^{\sigma-1} Y,
\end{aligned}$$

which is the demand function for the unique variety of firm ω , where $p^{1-\sigma} = \int_{\omega \in \Omega} p(\omega)^{1-\sigma} d\omega$ is a CES aggregate price index, and $x(\omega) = p(\omega) y(\omega)$ is the expenditure on a variety ω .

C.4.3 Gravity of intermediates

By plugging the pricing equation in the demand of firm ω for intermediates from firm ν , we derive the firm-level gravity equation

$$\begin{aligned}
n(\nu, \omega) &= p(\nu, \omega)^{1-\sigma} P(\omega)^{\sigma-1} N(\omega), \\
&= (\mu c(\nu) d(\nu, \omega))^{1-\sigma} P(\omega)^{\sigma-1} N(\omega), \\
&= \mu^{1-\sigma} d(\nu, \omega)^{1-\sigma} c(\nu)^{1-\sigma} P(\omega)^{\sigma-1} N(\omega), \\
\log(n(\nu, \omega)) &= \log(\mu^{1-\sigma} d(\nu, \omega)^{1-\sigma} c(\nu)^{1-\sigma} P(\omega)^{\sigma-1} N(\omega)), \\
&= \underbrace{\log(\mu^{1-\sigma})}_{\iota} + \underbrace{\log(c(\nu)^{1-\sigma})}_{\iota_\nu} + \underbrace{\log(P(\omega)^{\sigma-1} N(\omega))}_{\iota_\omega} + \log(d(\nu, \omega)^{1-\sigma}), \\
&= \iota + \iota_\nu + \iota_\omega + (1 - \sigma) \log(d(\nu, \omega)),
\end{aligned}$$

where ι is an intercept, ι_ν are seller fixed effects, and ι_ω are buyer fixed effects.

C.4.4 Equilibrium given the extensive margin

In this section, we derive the expression for the equilibrium objects given the structure of the production network. We first derive the recursive expression for prices, and then for total sales.

Recursive expression for prices. Consider the expression for the CES price index, so

$$\begin{aligned}
P(\omega)^{1-\sigma} &= \int_{\nu \in \Omega(\omega)} p(\nu, \omega)^{1-\sigma} d\nu, \\
P(z')^{1-\sigma} &= \int p(z, z')^{1-\sigma} l(z, z') dG(z), \\
&= \int \left(\left(\frac{\sigma}{\sigma-1} \right) c(z) d(z, z') \right)^{1-\sigma} l(z, z') dG(z), \\
&= \mu^{1-\sigma} \int (c(z) d(z, z'))^{1-\sigma} dG(z), \\
&= \mu^{1-\sigma} \int \left(\frac{P(z)^{1-\alpha}}{z} d(z, z') \right)^{1-\sigma} l(z, z') dG(z), \\
&= \mu^{1-\sigma} \int \left(\frac{P(z)^{1-\alpha}}{z} d(z, z') \right)^{1-\sigma} l(z, z') dG(z), \\
&= \mu^{1-\sigma} \int P(z)^{(1-\alpha)(1-\sigma)} z^{\sigma-1} d(z, z')^{1-\sigma} l(z, z') dG(z).
\end{aligned}$$

That is, the price index for firms of productivity z' can be expressed as a function of all other price indexes of firms z . This forms a system of equations we can solve.

Total sales. Consider the expression for total sales (i.e. sales to the household and firms), so

$$\begin{aligned}
S(\nu) &= x(\nu) + \int_{\omega \in \Omega(\nu)} n(\nu, \omega) d\omega, \\
S(z) &= x(z) + \int n(z, z') l(z, z') dG(z'), \\
&= p(z)^{1-\sigma} p^{\sigma-1} Y \\
&\quad + \int \left[\left(\frac{\sigma}{\sigma-1} \right)^{1-\sigma} d(z, z')^{1-\sigma} c(z)^{1-\sigma} P(z')^{\sigma-1} N(z') \right] l(z, z') dG(z'), \\
&= p(z)^{1-\sigma} p^{\sigma-1} Y \\
&\quad + \int \left[\left(\frac{\sigma}{\sigma-1} \right)^{1-\sigma} d(z, z')^{1-\sigma} \left[\frac{P(z)^{1-\alpha}}{z} \right]^{1-\sigma} P(z')^{\sigma-1} N(z') \right] l(z, z') dG(z'), \\
&= p(z)^{1-\sigma} p^{\sigma-1} Y \\
&\quad + \left[\left(\frac{\sigma}{\sigma-1} \right)^{1-\sigma} \left(\frac{P(z)^{1-\alpha}}{z} \right)^{1-\sigma} \right] \int \left[d(z, z')^{1-\sigma} P(z')^{\sigma-1} N(z') \right] l(z, z') dG(z'), \\
&= \frac{p(z)^{1-\sigma} Y}{p^{1-\sigma}} \\
&\quad + \left[\mu^{1-\sigma} P(z)^{(1-\alpha)(1-\sigma)} z^{\sigma-1} \right] \int \left[d(z, z')^{1-\sigma} P(z')^{\sigma-1} N(z') \right] l(z, z') dG(z'), \\
&= \frac{[\mu c(z)]^{1-\sigma} Y}{p^{1-\sigma}} \\
&\quad + \left[\mu^{1-\sigma} P(z)^{(1-\alpha)(1-\sigma)} z^{\sigma-1} \right] \int \left[d(z, z')^{1-\sigma} P(z')^{\sigma-1} \left(\frac{(1-\alpha) S(z')}{\mu} \right) \right] l(z, z') dG(z'), \\
&= \frac{\left[\mu \frac{P(z)^{1-\alpha}}{z} \right]^{1-\sigma} Y}{p^{1-\sigma}} \\
&\quad + \left[\mu^{1-\sigma} P(z)^{(1-\alpha)(1-\sigma)} z^{\sigma-1} \right] \left[\frac{1-\alpha}{\mu} \right] \int \left[d(z, z')^{1-\sigma} P(z')^{\sigma-1} S(z') \right] l(z, z') dG(z'), \\
&= \left[\mu^{1-\sigma} P(z)^{(1-\alpha)(1-\sigma)} z^{\sigma-1} \right] \\
&\quad \left[\frac{Y}{p^{1-\sigma}} + \left(\frac{1-\alpha}{\mu} \right) \left(\int \left[d(z, z')^{1-\sigma} P(z')^{\sigma-1} S(z') \right] l(z, z') dG(z') \right) \right],
\end{aligned}$$

where we use the fact that $N(z') = \frac{(1-\alpha)S(z')}{\mu}$. Given prices $P(z)$, this forms a system of equations for sales we can solve.

C.5 Targeted and untargeted moments

Since the link function noise distribution affects how firms match between them, to identify the parameters related to this distribution we must target moments that are related to the extensive margin.

First, we target the mean of the log-normalized number of buyers $\ln \left(\frac{\mathcal{N}_b(\nu)}{\mathcal{N}} \right)$, where $\mathcal{N}_b(\nu)$ is the number of buyers a seller ν has; and the mean of the log-normalized number of sellers $\ln \left(\frac{\mathcal{N}_s(\omega)}{\mathcal{N}} \right)$, where $\mathcal{N}_s(\omega)$ is the number of sellers a buyer ω has. Because these two moments are related to magnitude of the matching, they should inform us about the mean of the link function noise distribution $\mu_{\ln(\epsilon)}$ and the scaling constant for the pairwise matching cost κ .

Second, this being mostly a seller-oriented model, to identify the standard deviation of the link function noise distribution $\sigma_{\ln(\epsilon)}$ we target the variance of the log-normalized number of buyers $\ln \left(\frac{\mathcal{N}_b(\nu)}{\mathcal{N}} \right)$. Lastly, to identify the standard deviation of the log-productivity distribution, we must choose a moment that is related to the variance of the intensive margin. Thus, we target the variance of the log-normalized intermediate sales $\ln \left(\frac{\tilde{N}(\nu)}{\mathcal{N}_b(\nu)} \right)$, where $\tilde{N}(\nu)$ is the total intermediate sales a seller ν makes.

The first untargeted moment we consider is the variance of the log-normalized number of sellers $\ln \left(\frac{\mathcal{N}_s(\omega)}{\mathcal{N}} \right)$. The second untargeted moment we examine is the variance of the log-normalized intermediate purchases $\ln \left(\frac{N(\omega)}{\mathcal{N}_s(\omega)} \right)$.

The exact definition of the targeted and untargeted moments, as well as the construction of their empirical counterparts are as follows:

Normalized number of buyers and sellers

Data. In our dataset, for each firm i , we calculate the number of firms it sold to and the number of firms it bought from. Then, to normalize this measure, we divide this number by the total number of firms in our sample. Thus, for a specific firm i , we can understand this measure as the share of firms this specific firm i is connected to, both as a buyer and a seller.

Model. For this part, we start with the link function matrix, where each element $l(z, z')$ represents the pairwise probability that seller z will match with buyer z' . For each seller z , we take the average $l(z, z')$ across all the possible buyers. This represents the proportion of firms that seller z will match to the total number of firms. We multiply this number by the total number of firms $\mathcal{N}$ to obtain the number of buyers for each seller z . We follow a similar procedure to calculate the number of sellers each buyer z' has.

Normalized intermediate sales and purchases

Data. In our dataset, for each firm i , we calculate the total sales to other firms and the total purchases from other firms. In the case of the sellers, we normalize this measure by dividing the total sales of firm i by the total number of buyers this firm has. We follow a similar procedure with

the buyers to calculate the normalized intermediate purchases.

Model. We use the intermediate sales matrix, where each element $n(z, z')$ represents the total sales of intermediate goods from seller z to buyer z' . We sum all the sales for each seller z and divide this number by the number of buyers it has. Thus, we obtain the normalized intermediate sales for a given seller. For the normalized intermediate purchases, we follow a similar procedure with the buyers.

Goodness of fit. After our matching procedure, we find the parameters $\sigma_{\ln(z)} = 0.88$, $\mu_{\ln(\epsilon)} = 64.30$, $\sigma_{\ln(\epsilon)} = 10.85$ and $\kappa = 14.80$. Table C.10 in Appendix C.5 shows how the model-based moments fare against their empirical counterparts. When it comes to the targeted moments, the model can very closely replicate the empirical ones. For the untargeted moments, the model gets reasonably close to the data.

Table C.10: Targeted and untargeted moments

Targeted Moments		
	Data	Model
$mean [\ln (\mathcal{N}_b (\nu) / \mathcal{N})]$	-9.24	-9.48
$var [\ln (\mathcal{N}_b (\nu) / \mathcal{N})]$	0.98	0.89
$var \left[\ln \left(\tilde{N} (\nu) / \mathcal{N}_b (\nu) \right) \right]$	2.82	2.82
$mean [\ln (\mathcal{N}_s (\omega) / \mathcal{N})]$	-9.39	-9.14
Untargeted Moments		
	Data	Model
$var [\ln (\mathcal{N}_s (\omega) / \mathcal{N})]$	0.60	0.16
$var [\ln (N (\omega) / \mathcal{N}_s (\omega))]$	2.73	0.56

Notes: The targeted moments are the mean of the log-normalized number of buyers $mean [\ln (\mathcal{N}_b (\nu) / \mathcal{N})]$, the variance of the log-normalized number of buyers $var [\ln (\mathcal{N}_b (\nu) / \mathcal{N})]$ and the variance of the log-normalized intermediate sales $var \left[\ln \left(\tilde{N} (\nu) / \mathcal{N}_b (\nu) \right) \right]$, where $\tilde{N} (\nu)$ are the total intermediate sales of seller ν . The untargeted moments are the mean of the log-normalized number of sellers $mean [\ln (\mathcal{N}_s (\omega) / \mathcal{N})]$, the variance of the log-normalized number of sellers $var [\ln (\mathcal{N}_s (\omega) / \mathcal{N})]$ and the variance of the log-normalized intermediate purchases $var [\ln (N (\omega) / \mathcal{N}_s (\omega))]$.

BIBLIOGRAPHY

- K. Adam. Learning and equilibrium selection in a monetary overlapping generations model with sticky prices. *Review of Economic Studies*, 70(4):887–907, 2003. URL <https://doi.org/10.1111/1467-937X.00271>.
- H. Afrouzi. Strategic inattention, inflation dynamics, and the non-neutrality of money. 2020. URL <http://dx.doi.org/10.2139/ssrn.3576296>.
- H. Afrouzi, S. Y. Kwon, A. Landier, Y. Ma, and D. Thesmar. Overreaction in expectations: Evidence and theory. *Quarterly Journal of Economics*, 2023.
- I. Agarwal, R. Duttagupta, and A. F. Presbitero. Commodity prices and bank lending. *Economic inquiry*, 58(2):953–979, 2020.
- S. Alan, C. Baysan, M. Gumren, and E. Kubilay. Building social cohesion in ethnically mixed schools: An intervention on perspective taking. *The Quarterly Journal of Economics*, 136(4): 2147–2194, 2021.
- A. Alesina and P. Giuliano. Culture and institutions. *Journal of economic literature*, 53(4):898–944, 2015.
- A. Alesina, P. Giuliano, and B. Reich. Nation-building and education. *The Economic Journal*, 131 (638):2273–2303, 2021.
- A. F. Alesina and B. Reich. Nation building. 2015.
- S. N. Ali and D. A. Miller. Ostracism and forgiveness. *American Economic Review*, 106(8): 2329–48, 2016.
- H. Allcott and D. Keniston. Dutch disease or agglomeration? the local economic effects of natural resource booms in modern america. *The Review of Economic Studies*, 85(2):695–731, 2018.
- T. Allen, K. Bilir, Z. Chen, and C. Tonetti. Knowledge diffusion through networks. *Working Paper*, 2019.
- A. Ambrus, M. Mobius, and A. Szeidl. Consumption risk-sharing in social networks. *American Economic Review*, 104(1):149–82, 2014.
- J. Andrés, O. Arce, and C. Thomas. Banking competition, collateral constraints, and optimal monetary policy. *Journal of Money, Credit and Banking*, 45(s2):87–125, 2013.

- G.-M. Angeletos, Z. Huo, and K. A. Sastry. Imperfect macroeconomic expectations: Evidence and theory. *NBER Macroeconomics Annual*, 35(1):1–86, 2021.
- E. Ash, S. Asher, A. Bhowmick, S. Bhupatiraju, D. Chen, T. Devi, C. Goessmann, P. Novosad, and B. Siddiqi. In-group bias in the Indian judiciary: Evidence from 5 million criminal cases. *Working Paper*, 2021.
- F. Balmaceda and J. F. Escobar. Trust in cohesive communities. *Journal of Economic Theory*, 170: 289–318, 2017.
- S. Bandyopadhyay, C. C. Coughlin, and H. J. Wall. Ethnic networks and US exports. *Review of International Economics*, 16(1):199–213, 2008.
- A. V. Banerjee and E. Duflo. Reputation effects and the limits of contracting: A study of the Indian software industry. *Quarterly Journal of Economics*, 115(3):989–1017, 2000.
- S. Bazzi, A. Gaduh, A. D. Rothenberg, and M. Wong. Unity in diversity? how intergroup contact can foster nation building. *American Economic Review*, 109(11):3978–4025, 2019.
- G. S. Becker. *The economics of discrimination*. University of Chicago Press, 1957.
- F. Benguria, F. Saffie, and S. Urzua. Transmission of Commodity Price Super-Cycles. *Review of Economic Studies*, pages 1–33, 08 2023.
- D. J. Benjamin. Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics: Applications and Foundations 1*, 2:69–186, 2019.
- A. B. Bernard, E. Dhyne, G. Magerman, K. Manova, and A. Moxnes. The origins of firm heterogeneity: A production network approach. *Journal of Political Economy*, 130(7):1765–1804, 2022.
- S. Bhagavatula, M. Bhalla, M. Goel, and B. Vissa. Cultural diversity on corporate boards and firm outcomes. *Working Paper*, 2018.
- A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, 1943.
- F. Bianchi, C. L. Ilut, and H. Saijo. Implications of diagnostic expectations: Theory and applications. 2021. URL <https://doi.org/10.3386/w28604>.
- O. J. Blanchard. Debt, deficits, and finite horizons. *Journal of Political Economy*, 93(2):223–247, 1985. URL <https://doi.org/10.1086/261297>.
- O. J. Blanchard, J.-P. L’Huillier, and G. Lorenzoni. News, noise, and fluctuations: An empirical exploration. *American Economic Review*, 103(7):3045–70, 2013. URL <https://doi.org/10.1257/aer.103.7.3045>.
- F. Bloch, G. Genicot, and D. Ray. Informal insurance in social networks. *Journal of Economic Theory*, 143(1):36–58, 2008.

- S. Board. Relational contracts and the value of loyalty. *American Economic Review*, 101(7): 3349–3367, 2011.
- J. Boehm and E. Oberfield. Misallocation in the market for inputs: Enforcement and the organization of production. *Quarterly Journal of Economics*, 135(4):2007–2058, 2020.
- J. Boken, L. Gadenne, T. Nandi, and M. Santamaria. Community networks and trade. *Working Paper*, 2022.
- P. Bordalo, N. Gennaioli, and A. Shleifer. Diagnostic expectations and credit cycles. *Journal of Finance*, 73(1):199–227, 2018. URL <https://doi.org/10.1111/jofi.12586>.
- P. Bordalo, N. Gennaioli, R. L. Porta, and A. Shleifer. Diagnostic expectations and stock returns. *Journal of Finance*, 74(6):2839–2874, 2019. URL <https://doi.org/10.1111/jofi.12833>.
- P. Bordalo, N. Gennaioli, Y. Ma, and A. Shleifer. Overreaction in macroeconomic expectations. *American Economic Review*, 110(9):2748–82, 2020. URL <https://doi.org/10.1257/aer.20181219>.
- P. Bordalo, J. J. Conlon, N. Gennaioli, S. Y. Kwon, and A. Shleifer. Memory and probability. *The Quarterly Journal of Economics*, 138(1):265–311, 2023.
- K. Borusyak, P. Hull, and X. Jaravel. Quasi-experimental shift-share research designs. *Review of Economic Studies*, 89(1):181–213, 2022.
- K. Borusyak, P. Hull, and X. Jaravel. A practical guide to shift-share instruments. December 2024. doi: 10.3386/w33236. URL <http://www.nber.org/papers/w33236>.
- W. A. Branch and B. McGough. Dynamic predictor selection in a new keynesian model with heterogeneous expectations. *Journal of Economic Dynamics and Control*, 34(8):1492–1508, 2010. URL <https://doi.org/10.1016/j.jedc.2010.03.012>.
- P. Bustos, G. Garber, and J. Ponticelli. Capital accumulation and structural transformation. *The Quarterly Journal of Economics*, 135(2):1037–1094, 2020.
- C. Campos, M. McMain, and M. Pedemonte. Understanding which prices affect inflation expectations. *Economic Commentary*, (2022-06), 2022. URL <https://www.doi.org/10.26509/frbc-ec-202206>.
- B. Candia, O. Coibion, and Y. Gorodnichenko. The macroeconomic expectations of firms. 2022. URL <https://doi.org/10.3386/w30042>.
- C. D. Carroll. Macroeconomic expectations of households and professional forecasters. *Quarterly Journal of Economics*, 118(1):269–298, 2003. URL <https://doi.org/10.1162/00335530360535207>.
- M. X. Chen and M. Wu. The value of reputation in trade: Evidence from Alibaba. *Review of Economics and Statistics*, 103(5):857–873, 2021.

- Y.-W. Cheung. Exchange rate risk premiums. *Journal of International Money and Finance*, 12(2): 182–194, 1993. URL [https://doi.org/10.1016/0261-5606\(93\)90023-5](https://doi.org/10.1016/0261-5606(93)90023-5).
- G. Chodorow-Reich. The employment effects of credit market disruptions: Firm-level evidence from the 2008–9 financial crisis. *The Quarterly Journal of Economics*, 129(1):1–59, 2014.
- O. Coibion and Y. Gorodnichenko. Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, 105(8):2644–78, 2015. URL <https://www.doi.org/10.1257/aer.20110306>.
- O. Coibion, Y. Gorodnichenko, and S. Kumar. How do firms form their expectations? New survey evidence. *American Economic Review*, 108(9):2671–2713, 2018. URL <https://doi.org/10.1257/aer.20151299>.
- O. Coibion, Y. Gorodnichenko, and T. Ropele. Inflation expectations and firm decisions: New causal evidence. *Quarterly Journal of Economics*, 135(1):165–219, 2020. URL <https://doi.org/10.1093/qje/qjz029>.
- O. Coibion, D. Georgarakos, Y. Gorodnichenko, and M. Van Rooij. How does consumption respond to news about inflation? Field evidence from a randomized control trial. *American Economic Journal: Macroeconomics*, 15(3):109–152, 2023. URL <https://doi.org/10.1257/mac.20200445>.
- P. Collin-Dufresne, M. Johannes, and L. A. Lochstoer. Asset pricing when ‘this time is different’. *Review of Financial Studies*, 30(2):505–535, 2017. URL <https://doi.org/10.1093/rfs/hhw084>.
- L. W. Cong, H. Gao, J. Ponticelli, and X. Yang. Credit allocation under economic stimulus: Evidence from china. *The Review of Financial Studies*, 32(9):3412–3460, 2019.
- W. M. Corden and J. P. Neary. Booming sector and de-industrialisation in a small open economy. *Economic Journal*, 92(368):825–848, 1982. ISSN 00130133, 14680297. URL <http://www.jstor.org/stable/2232670>.
- S. Correia, P. Guimarães, and T. Zylkin. ppmlhdfc: Fast Poisson estimation with high-dimensional fixed effects. *Working Paper*, 2019.
- F. D’Acunto and M. Weber. Memory and beliefs: Evidence from the field. 2022.
- F. D’Acunto, U. Malmendier, J. Ospina, and M. Weber. Exposure to grocery prices and inflation expectations. *Journal of Political Economy*, 129(5):1615–1639, 2021. URL <https://doi.org/10.1086/713192>.
- F. D’Acunto, U. Malmendier, and M. Weber. What do the data tell us about inflation expectations? 2022. URL <https://doi.org/10.3386/w29825>.
- P. De Grauwe. Animal spirits and monetary policy. *Economic Theory*, 47(2):423–457, 2011. URL <https://www.jstor.org/stable/41485845>.

- J. De Loecker, P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik. Prices, markups, and trade reform. *Econometrica*, 84(2):445–510, 2016.
- H. Degryse, O. De Jonghe, S. Jakovljević, K. Mulier, and G. Schepens. Identifying credit supply shocks with bank-firm data: Methods and applications. *Journal of Financial Intermediation*, 40: 100813, 2019.
- R. Dekle, J. Eaton, and S. Kortum. Global rebalancing with gravity: Measuring the burden of adjustment. 2008.
- E. Depetris-Chauvin, R. Durante, and F. Campante. Building nations through shared experiences: Evidence from african football. *American Economic Review*, 110(5):1572–1602, 2020.
- A. M. Dietrich. Consumption categories, household attention, and inflation expectations: Implications for optimal monetary policy. 2022.
- T. Drechsel and S. Tenreyro. Commodity booms and busts in emerging economies. *Journal of International Economics*, 112:200–218, 2018.
- W. Edwards. Conservatism in human information processing. *Formal representation of human judgment*, 1968.
- G. W. Evans and S. Honkapohja. *Learning and Expectations in Macroeconomics*. Princeton University Press, 2001. URL <https://doi.org/10.1515/9781400824267>.
- E. Faia, M. Mayer, and V. Pezone. The value of firm networks: a natural experiment on board connections. 2021.
- S. Federico, F. Hassan, and V. Rappoport. Trade shocks and credit reallocation. 2023.
- R. C. Feenstra, R. Inklaar, and M. P. Timmer. The next generation of the Penn World Table. *American Economic Review*, 105(10):3150–82, 2015.
- A. Fernández, S. Schmitt-Grohé, and M. Uribe. How important is the commodity supercycle? 2023.
- R. Fisman, D. Paravisini, and V. Vig. Cultural proximity and loan outcomes. *American Economic Review*, 107(2):457–92, 2017.
- X. Gabaix. A behavioral New Keynesian model. *American Economic Review*, 110(8):2271–2327, 2020. URL <https://doi.org/10.1257/aer.20162005>.
- J. Gali. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications*. Princeton University Press, 2015.
- J. Gali. Monetary policy and bubbles in a new keynesian model with overlapping generations. *American Economic Journal: Macroeconomics*, 13(2):121–67, 2021. URL <https://www.aeaweb.org/articles?id=10.1257/mac.20180427>.

- J. García-Cicco and E. Kawamura. Dealing with the dutch disease: Fiscal rules and macro-prudential policies. *Journal of International Money and Finance*, 55:205–239, 2015.
- L. Gáti. Monetary policy & anchored expectations: An endogenous gain learning model. 2020.
- A. Gerali, S. Neri, L. Sessa, and F. M. Signoretti. Credit and banking in a dsge model of the euro area. *Journal of Money, Credit and Banking*, 42:107–141, 2010.
- E. P. Gilje. Does local access to finance matter? evidence from us oil and natural gas shale booms. *Management Science*, 65(1):1–18, 2019.
- E. P. Gilje, E. Loutskina, and P. E. Strahan. Exporting liquidity: Branch banking and financial integration. *The Journal of Finance*, 71(3):1159–1184, 2016.
- S. S. Goraya. How does caste affect entrepreneurship? Birth versus worth. *Journal of Monetary Economics*, 135:116–133, 2023.
- M. Gradstein and M. Justman. Cultural interaction and economic development: An overview. *European Journal of Political Economy*, 59:243–251, 2019.
- J. Greenwood, Z. Hercowitz, and G. W. Huffman. Investment, capacity utilization, and the real business cycle. *The American Economic Review*, pages 402–417, 1988.
- L. Guiso, P. Sapienza, and L. Zingales. Cultural biases in economic exchange? *Quarterly Journal of Economics*, 124(3):1095–1131, 2009.
- Á. Gutiérrez-Daza. Business cycles when consumers learn by shopping. 2022.
- I. Hajdini, E. S. Knotek II, J. Leer, M. Pedemonte, R. Rich, and R. Schoenle. Indirect consumer inflation expectations: Theory and evidence. *Federal Reserve Bank of Cleveland Working Paper*, 2022a. URL <https://doi.org/10.26509/frbc-wp-202235>.
- I. Hajdini, E. S. Knotek II, J. Leer, M. Pedemonte, R. W. Rich, and R. S. Schoenle. Low passthrough from inflation expectations to income growth expectations: Why people dislike inflation. Working Paper 22-21, Federal Reserve Bank of Cleveland, June 2022b. URL <https://doi.org/10.26509/frbc-wp-202221>.
- I. Hajdini, E. S. Knotek II, M. Pedemonte, R. Rich, J. Leer, and R. Schoenle. Indirect consumer inflation expectations. *Economic Commentary*, (2022-03), 2022c. URL <https://doi.org/10.26509/frbc-ec-202203>.
- J. J. Heckman. Sample selection bias as a specification error. *Econometrica*, pages 153–161, 1979.
- E. Helpman, M. Melitz, and Y. Rubinstein. Estimating trade flows: Trading partners and trading volumes. *Quarterly Journal of Economics*, 123(2):441–487, 2008.
- Z. Huo, A. Levchenko, and N. Pandalai-Nayar. International comovement in the global production network. *Review of Economic Studies*, page rdae033, 2024.

- L. Iyer, T. Khanna, and A. Varshney. Caste and entrepreneurship in india. *Economic and Political Weekly*, pages 52–60, 2013.
- R. Iyer, J.-L. Peydró, S. da Rocha-Lopes, and A. Schoar. Interbank liquidity crunch and the firm credit crunch: Evidence from the 2007–2009 crisis. *The Review of Financial Studies*, 27(1): 347–372, 2014.
- J. Jayaratne and D. P. Morgan. Capital market frictions and deposit constraints at banks. *Journal of Money, Credit and Banking*, pages 74–92, 2000.
- U. Jermann and V. Quadrini. Macroeconomic effects of financial shocks. *American Economic Review*, 102(1):238–271, 2012.
- R. C. Jump, C. Hommes, and P. Levine. Learning, heterogeneity, and complexity in the New Keynesian model. *Journal of Economic Behavior & Organization*, 166:446–470, 2019. URL <https://doi.org/10.1016/j.jebo.2019.07.014>.
- D. Kahneman and A. Tversky. Subjective probability: A judgment of representativeness. *Cognitive psychology*, 3(3):430–454, 1972.
- G. Kaplan, B. Moll, and G. L. Violante. Monetary policy according to HANK. *American Economic Review*, 108(3):697–743, 2018. URL <https://doi.org/10.1257/aer.20160042>.
- A. K. Kashyap and J. C. Stein. What do a million observations on banks say about the transmission of monetary policy? *American Economic Review*, 90(3):407–428, 2000.
- G. Khanna. Does affirmative action incentivize schooling? evidence from india. *Review of Economics and Statistics*, 102(2):219–233, 2020.
- A. I. Khwaja and A. Mian. Tracing the impact of bank liquidity shocks: Evidence from an emerging market. *American Economic Review*, 98(4):1413–1442, 2008.
- Z. L. Kone, M. Y. Liu, A. Mattoo, C. Ozden, and S. Sharma. Internal borders and migration in india. *Journal of Economic Geography*, 18(4):729–759, 2018.
- M. A. Kose. Explaining business cycles in small open economies: How much do world prices matter? *Journal of International Economics*, 56(2):299–327, 2002.
- T. Kuchler and B. Zafar. Personal experiences and expectations about aggregate outcomes. *Journal of Finance*, 74(5):2491–2542, 2019. URL <https://doi.org/10.1111/jofi.12819>.
- S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951.
- J.-P. L’Huillier, S. R. Singh, and D. Yoo. Diagnostic expectations and macroeconomic volatility. 2021.
- J. Liao. Over/under-reaction and judgment noise in expectations formation. 2023.

- R. Macchiavello and A. Morjaria. The value of relationships: Evidence from a supply shock to Kenyan rose exports. *American Economic Review*, 105(9):2911–45, 2015.
- U. Malmendier. Exposure, experience, and expertise: Why personal histories matter in economics. *Journal of the European Economic Association*, 19:2857–2894, 2021. URL <https://doi.org/10.1093/jeea/jvab045>.
- U. Malmendier and S. Nagel. Learning from inflation experiences. *Quarterly Journal of Economics*, 131(1):53–87, 2016. URL <https://doi.org/10.1093/qje/qjv037>.
- U. Malmendier and J. A. Wachter. Memory of past experiences and economic decisions. 2022.
- U. Malmendier, S. Nagel, and Z. Yan. The making of hawks and doves. *Journal of Monetary Economics*, 117:19–42, 2021. URL <https://doi.org/10.1016/j.jmoneco.2020.04.002>.
- N. G. Mankiw and R. Reis. Sticky information versus sticky prices: A proposal to replace the New Keynesian Phillips curve. *Quarterly Journal of Economics*, 117(4):1295–1328, 2002. URL <https://doi.org/10.1162/003355302320935034>.
- N. G. Mankiw and R. Reis. Pervasive stickiness. *American Economic Review*, 96(2):164–169, 2006. URL <https://doi.org/10.1257/000282806777211937>.
- N. G. Mankiw and R. Reis. Imperfect information and aggregate supply. *Handbook of Monetary Economics*, 3:183–229, 2011. URL <https://doi.org/10.1016/B978-0-444-53238-1.00005-3>.
- A. Marcet and T. J. Sargent. Convergence of least squares learning mechanisms in self-referential linear stochastic models. *Journal of Economic Theory*, 48(2):337–368, 1989. URL [https://doi.org/10.1016/0022-0531\(89\)90032-X](https://doi.org/10.1016/0022-0531(89)90032-X).
- E. G. Mendoza. The terms of trade, the real exchange rate, and economic fluctuations. *International Economic Review*, pages 101–137, 1995.
- J. M. Morelli, P. Ottonello, and D. J. Perez. Global banks and systemic debt crises. *Econometrica*, 90(2):749–798, 2022.
- K. Munshi. Community networks and the process of development. *Journal of Economic Perspectives*, 28(4):49–76, 2014.
- K. Munshi. Caste and the Indian economy. *Journal of Economic Literature*, 57(4):781–834, 2019.
- K. Munshi and M. Rosenzweig. Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review*, 106(1):46–98, 2016.
- P. A. Neumeyer and F. Perri. Business cycles in emerging economies: the role of interest rates. *Journal of Monetary Economics*, 52(2):345–380, 2005.
- N. Nunn. Relationship-specificity, incomplete contracts, and the pattern of trade. *Quarterly Journal of Economics*, 122(2):569–600, 2007.

- D. Paravisini. Local bank financial constraints and firm access to external finance. *The Journal of Finance*, 63(5):2161–2193, 2008.
- D. Paravisini, V. Rappoport, and P. Schnabl. Specialization in bank lending: Evidence from exporting firms. *Journal of Finance*, 78(4):2049–2085, 2023.
- J. Peek and E. S. Rosengren. Unnatural selection: Perverse incentives and the misallocation of credit in japan. *American Economic Review*, 95(4):1144–1166, 2005.
- O. Pfäuti and F. Seyrich. A behavioral heterogeneous agent New Keynesian model. 2023.
- F. Pivetta and R. Reis. The persistence of inflation in the United States. *Journal of Economic Dynamics and Control*, 31(4):1326–1358, 2007. URL <https://doi.org/10.1016/j.jedc.2006.05.001>.
- J. E. Rauch. Trade and search: Social capital, sogo shosha, and spillovers. *Working Paper*, 1996.
- J. E. Rauch. Networks versus markets in international trade. *Journal of international Economics*, 48(1):7–35, 1999.
- J. E. Rauch and A. Casella. Overcoming informational barriers to international resource allocation: Prices and ties. *Economic Journal*, 113(484):21–42, 2003.
- J. E. Rauch and V. Trindade. Ethnic Chinese networks in international trade. *Review of Economics and Statistics*, 84(1):116–130, 2002.
- B. D. Richman. How community institutions create economic advantage: Jewish diamond merchants in New York. *Law & Social Inquiry*, 31(2):383–420, 2006.
- J. Ritzen, W. Easterly, and M. J. Woolcock. *On” good” politicians and” bad” policies: Social cohesion, institutions, and growth*, volume 2448. World Bank Publications, 2000.
- C. Roth and J. Wohlfart. How do expectations about the macroeconomy affect personal expectations and behavior? *Review of Economics and Statistics*, 102(4):731–748, 2020. URL https://doi.org/10.1162/rest_a_00867.
- I. Salle, Y. Gorodnichenko, and O. Coibion. Lifetime memories of inflation: Evidence from surveys and the lab. Technical report, National Bureau of Economic Research, 2024.
- W. Salter. Internal and external balance: the role of price and expenditure effects. *Economic Record*, 35(71):226–238, 1959.
- S. Schmitt-Grohé and M. Uribe. How important are terms-of-trade shocks? *International Economic Review*, 59(1):85–111, 2018.
- P. Schnabl. The international transmission of bank liquidity shocks: Evidence from an emerging market. *Journal of Finance*, 67(3):897–932, 2012.
- A. Schoar, R. K. Iyer, and S. Kumar. Importance of ethnic networks for business transactions of the small enterprises. *Working Paper*, 2008.

- S. Shousha. Macroeconomic effects of commodity booms and busts: The role of financial frictions. 2016.
- A. Silva, P. Caraiani, J. Miranda-Pinto, and J. Olaya-Agudelo. Commodity prices and production networks in small open economies. 2024.
- M. Startz. The value of face-to-face: Search and contracting problems in Nigerian trade. *Working Paper*, 2016.
- T. W. Swan. Longer-run problems of the balance of payments. *Australian Economy*, pages 384–95, 1963.
- A. Tversky and D. Kahneman. Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2):207–232, 1973.
- M. Ulate. Going negative at the zero lower bound: The effects of negative nominal interest rates. *American Economic Review*, 111(1):1–40, January 2021.
- M. Uribe and S. Schmitt-Grohé. *Open economy macroeconomics*. Princeton University Press, 2017.
- Y. Wang, T. M. Whited, Y. Wu, and K. Xiao. Bank market power and monetary policy transmission: Evidence from a structural estimation. *Journal of Finance*, 77(4):2093–2141, 2022.
- M. Weber, B. Candia, T. Ropele, R. Lluberas, S. Frache, B. H. Meyer, S. Kumar, Y. Gorodnichenko, D. Georgarakos, O. Coibion, et al. Tell me something i don’t already know: Learning in low and high-inflation settings. 2023.
- T. M. Whited, Y. Wu, and K. Xiao. Will central bank digital currency disintermediate banks? 2022.
- M. Woodford. Imperfect common knowledge and the effects of monetary policy. 2001. URL <https://doi.org/10.3386/w8673>.
- M. E. Yaari. Uncertain lifetime, life insurance, and the theory of the consumer. *Review of Economic Studies*, 32(2):137–150, 1965. URL <https://doi.org/10.2307/2296058>.
- Y. Zhou. Inter-firm linkages, ethnic networks, and territorial agglomeration: Chinese computer firms in Los Angeles. *Papers in Regional Science*, 75(3):265, 1996.