DOCUMENTOS DE TRABAJO

From Invoices to Economic Monitoring Using Machine Learning

Emiliano Luttini
Matías Pizarro
Dagoberto Quevedo
Marco Rojas

N° 1088 Agosto 2026
BANCO CENTRAL DE CHILE

Documentos de Trabajo del Banco Central de Chile
Working Papers of the Central Bank of Chile
Agustinas 1180, Santiago, Chile
Teléfono: (56-2) 3882475; Fax: (56-2) 38822311

Documento de Trabajo N° 1088

Working Paper N° 1088

From Invoices to Economic Monitoring Using Machine Learning*

Emiliano Luttini
World Bank

Matías Pizzaro
Central Bank of Chile

Dagoberto Quevedo
Central Bank of Chile

Marco Rojas
Central Bank of Chile

Resumen

Utilizando el universo de facturas electrónicas en Chile, desarrollamos una metodología para clasificar transacciones de firmas en categorías de producto utilizando un algoritmo de machine learning entrenado sobre un corpus etiquetado de descripciones de productos. La clasificación resultante nos permite construir medidas de gastos en maquinaria y equipo y de consumo de bienes, las cuales siguen de cerca sus contrapartes de cuentas nacionales, y a su vez reflejan los efectos de shocks sobre el consumo y sus implicancias para medidas del índice de precios al consumidor. Usando el COVID-19 como caso de estudio, documentamos reasignación hacia bienes durables, mostramos que las restricciones de movilidad y políticas de apoyo al ingreso familiar no explican la totalidad de estos cambios, y cuantificamos las implicancias de reasignación del gastos sobre la inflación medida en este contexto.

Abstract

Using the universe of electronic invoices in Chile, we develop a methodology to classify firms' transactions into product categories based on a machine learning algo-rithm trained on a labeled corpus of product descriptions. The resulting classification allows us to construct measures of machinery and equipment outlays and goods con-sumption that closely track their national accounts counterparts, while also tracing the effects of shocks on consumption and their implications for consumer price index mea-surement. Using the COVID-19 pandemic as a case study, we document a reallocation toward durable goods, show that mobility restrictions and income support policies do not fully account for these shifts, and quantify the implications of expenditure reallocation for measured inflation.

* We would like to thank Elias Albagli for his support throughout the project. We also thank seminar participants at the Central Bank of Chile, "Nontraditional Data, Machine Learning, and Natural Language Processing in Macroeconomics" at the Federal Reserve Board, "3rd Irving Fisher Committee Workshop on Data Science in Central Banking" at Banca d'Italia, and "Economic Data Science Society Conference" at Banque de France for thoughtful comments and discussions. We also thank Michele Lenza for his detailed comments. All errors are our own. Disclaimers are found at the end of this document.
E-mail addresses: eluttini@worldbank.org, mpizarro@bcentral.cl, dquevedo@bcentral.cl, mrojaso@bcentral.cl.

"... In an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it." Simon (1971).

1 Introduction

As Herbert Simon noted in 1971, living in an information-rich world translates into a dearth of analytical capabilities. Since this quote, the abundance of information has become even more dramatic, as has the dearth he pointed out. This information-rich world features large databases and text data that have become increasingly prevalent in monitoring and analyzing the economy. Private companies and public bureaus, collecting thousands of transactions, have become vital actors in supplying this information with the potential to complement traditional data sources used by statistical agencies that allow policymakers to understand the economy. The massiveness of the data is such that practitioners must use new techniques to exploit this information. More prominently, from the field of computer science, machine learning is used to perform tasks that otherwise would require the job of countless human beings. Machine learning can be potentially valuable in interpreting information embedded in tax documents; however, trying to leverage this potential has yet to receive attention from the literature.

In this article, we use natural language processing techniques to classify text descriptions from electronic tax documents into standardized product categories. We draw on the universe of electronic tax documents covering business-to-business transactions in Chile from 2014 to 2022—over two billion documents containing ten billion free-text product descriptions—and apply a supervised machine learning classifier to map these descriptions into categories useful for economic analysis. The resulting classification allows us to construct machinery and equipment outlays and goods consumption aggregate series that closely track their national accounts counterparts, are available within days after the end of each month, and are not subject to ex-post revisions. More broadly, our work illustrates how tax information combined with text analytics can help document the full impact of shocks affecting consumption, allowing us to study how expenditure patterns shift across product categories and how these shifts feed into CPI measurement. In the context of the COVID-19 pandemic, we document a reallocation of expenditure toward durable and away from non-durable goods, study the effects of stay-at-home orders and pension savings withdrawals on consumption, and construct a flexible-weight consumer price index. We show that, while fixed- and flexible-weight measures diverge during specific episodes—notably at the onset of the pandemic and during the subsequent high-inflation period—these differences are not a first-order concern over the pandemic period as a whole.

We employ a supervised multi-class text classification model to identify the products involved in firms' transactions. A key challenge is that invoice descriptions have not been previously labeled using standard economic classifications, and the language used in invoicing is highly idiosyncratic and domain-specific, making pre-trained language models ill-suited for the task. We address this by constructing a labeled corpus of text descriptions through a targeted manual labeling strategy, prioritizing descriptions based on a score that balances revenue, frequency, and monthly occurrence across economic sectors. We train the FastText model (Joulin et al., 2016) using this corpus, mapping descriptions into three complementary classification schemes—the Central Product Classification, the Unique Product Classification, and the Classification of Individual Consumption by Purpose—each suited to a different economic application. We use a series of metrics to assess the goodness of fit of the algorithm and find strong performance. We then apply our trained classifier to the universe of electronic invoice descriptions to learn about the nature of firms' transactions in the Chilean economy.

The classification exercise allows us to construct aggregate series for machinery and equipment outlays and goods consumption that closely track their national accounts counterparts. We do so by mapping firms' purchases into product categories classified as gross fixed capital formation or consumption under the Broad Economic Categories framework used in national accounts. To approximate final goods consumption, we use purchases by firms in the retail sector, assuming that products reaching retailers turn over rapidly to final consumers. Our constructed series replicate the main trends, seasonal patterns, and cyclical movements of the official data. We also show that because of legal time constraints to report the electronic tax documents, the production of these indices can be available approximately ten days after the month ends.

Our work illustrates how tax information combined with text analytics can help document the effects of shocks on consumption by tracing how expenditure patterns shift across product categories, which we study in the context of the COVID-19 pandemic. While statistical agencies typically rely on household surveys to measure such patterns, our classification provides more timely information. Using data from 2018 to 2022, we document a marked reallocation of expenditure beginning in the second half of 2020, as households shifted spending toward durable goods and away from non-durables, consistent with evidence from other countries. This reallocation was driven mainly by increased spending on durable Entertainment and Clothing and General Merchandise goods, while non-durable Transport expenditure contracted sharply and became the largest negative contributor. A natural interpretation is that the distinctive conditions of the pandemic led households to limit exposure to social activities and redirect spending toward goods consumed at home.

To further understand the drivers of these consumption patterns, we study the role of key policy responses implemented during the pandemic. We begin by examining stay-at-home orders, which imposed mobility constraints across municipalities during 2020 and 2021. Using an event-study design, we find that lockdowns compressed spending broadly across consumption categories. However, the net effect of entering and exiting lockdowns points to a reallocation toward durable Entertainment goods and non-durable Clothing and General Merchandise, consistent with households shifting consumption toward goods that can be used at home while limiting exposure to social activities. While this evidence is consistent with households reallocating consumption toward home-based durable goods, consistent with limiting exposure to the virus, the presence of mobility constraints more broadly affected the cross-section of consumption categories. We then study the role of income support policies through the pension withdrawal programs enacted during the same period. Focusing on the first withdrawal, we find that these liquidity injections led to a sustained increase in consumption, with effects that build over time. The response is broad-based, driven primarily by Food and Clothing and General Merchandise, while among durable goods, Entertainment exhibits the largest increase.

Taken together, the evidence suggests a limited role of stay-at-home orders and pension withdrawals in rationalizing the change in consumption patterns over the period, and points to the distinctive conditions of the pandemic period itself as the main determinant of the aggregate shifts in consumption.

Finally, we examine the implications of these changes for inflation measurement. Using price information from electronic tax documents, we construct a flexible-weight consumer price index and compare it to its fixed-weight counterpart. Although the two measures diverge during specific episodes—particularly at the onset of the pandemic and during the subsequent inflation surge—we find that these differences are not quantitatively large over the period as a whole. This suggests that, despite substantial shifts in expenditure patterns, the resulting bias in standard cost-of-living measures is limited.

Related literature. This work contributes to the growing body of papers that use text analytics for classifying corpus of text. Hansen and Ash (2023) summarizes the recent developments in the field applied to economics analysis, emphasizing the complementary role of human knowledge and machine learning as we do here. To our knowledge Hoberg and Phillips (2016) is the closest to our work regarding the classification problem they face, as they classify US firms and their competitors into industries using the forms filed by public companies that describe their business and products. We contribute to this approach by categorizing into standardized categories the goods and services firms exchange.

Our article also contributes to the literature that provides strategies and sources to complement traditional aggregate statistics. Ehrlich et al. (2019) discuss the challenges of traditional methods of collecting data from businesses and households, and to circumvent them, they propose using alternative datasets arising from the digitization of market transactions. Chetty et al. (2023) build a high-frequency dataset that tracks economic activity in the US using data from private companies. Buda et al. (2022) use transaction-level data from the universe of Spanish bank current account retail accounts to produce an aggregate that successfully reproduces consumption dynamics from national accounts. Our paper is the first to implement such a strategy to build goods consumption and machinery and equipment purchases aggregates using invoice data.

Our work relates to the literature studying the effects of the COVID-19 pandemic and stay-at-home orders on the economy. Chetty et al. (2023) show how the COVID-19 pandemic affected consumers' spending across ZIP codes with different median household incomes and provide evidence that consumption decreased by the most in affluent ZIP codes at the onset of the pandemic. Alexander and Karger (2021) and Cox et al. (2020) show that spending reductions were abrupt in sectors that require face-to-face interaction. Goolsbee and Syverson (2021) estimate the causal effect of stay-at-home orders in the US on the traffic in individual businesses, showing that mobility constraints can explain a drop of about ten percent in foot traffic at individual businesses (see also Alexander and Karger (2021)). We add to this literature by documenting how stay-at-home orders distinctly affect the consumption of different types of goods and characterize the aggregate consumption pattern during different stages of the pandemic, showing that households contracted the consumption of durable goods at the onset of the pandemic, favoring the consumption of perishable ones. In contrast, the opposite pattern holds when deploying income support programs at later stages.

2 Data

2.1 Electronic tax documents

Our data is drawn from the electronic tax documents that firms in Chile generate every time they have transactions with each other, and which must be reported to the Chilean Internal Tax Service (SII for its acronym in Spanish). The documents include regular electronic invoices—which are the overwhelming majority—, tax-exempt invoices, export invoices, and credit notes, among others. The system originates from a comprehensive tax overhaul, which made it compulsory for firms to adopt a system of electronic tax documents. The motivation of this system was to simplify the administrative burden associated with issuing paper invoices and improve the inspection and audit by the tax authority. The system was implemented in three stages: For firms operating in urban settings, the electronic invoice was binding from November 2014, for large firms, August 2016, for small and medium enterprises, and February 2017, for microenterprises. In each electronic tax document, we observe firm identifiers for the seller and buyer and their economic sectors. The document contains information on the items sold, such as price, quantity, and a free-text description of the good or service traded, among other variables. This text data is unstructured and is the crucial element we use to classify items into standardized categories.[1]

We use data between 2015 and 2022, summarized in Table 1, for the universe of documents. The first column shows the number of documents that amount to over 4 billion, with an average of 44 million monthly. We primarily rely on regular electronic invoices that amount to about 27 million documents. For this relevant sample, the fourth and fifth columns show the total number of items and descriptions, which are around 10 and 1 billion, respectively. It is over these text descriptions that we proceed to classify according to the strategy outlined in the next section. Then, we use this information to produce economic information to keep track of the economy in real time.

[1]Illustrative examples of electronic invoices can be found on the Chilean Internal Tax Service website.

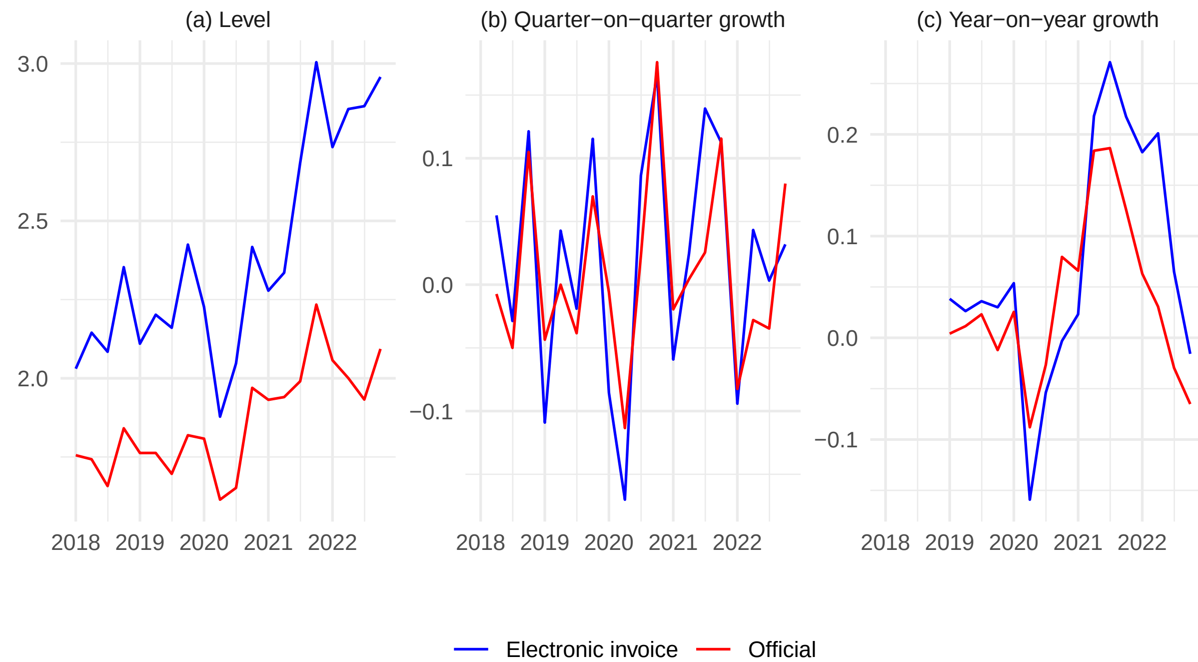
Table 1: Summary statistics of electronic tax documents and electronic invoices (in millions)

	Sellers	Buyers	Documents	Transactions
	Electronic Tax Documents			
Monthly average	0.54	5.10	44.19	185.95
Annual average	6.47	61.16	530.28	2,231.35
Total	51.78	489.25	4,242.22	17,850.84
	Electronic Invoices			
Monthly average	0.30	1.67	26.95	111.14
Annual average	3.56	20.04	323.43	1,333.67
Total	28.46	160.30	2,587.47	10,669.35

Notes: The data cover 2015–2022 and report monthly and yearly averages and totals for sellers, buyers, documents and transactions (in millions). Both electronic tax documents and electronic invoices exclude canceled documents. Electronic invoices includes only electronic invoices documents.

The data contains information about aggregate economic activity. Figure 1 compares official GDP with the total volume of sales recorded in electronic invoices. Panel (a) focuses on levels, while panels (b) and (c) compare quarterly and annual growth rates. Aggregate activity arising from electronic invoices is an order of magnitude larger than GDP because it double counts value added along the supply chain, whereas GDP measures only the value added at each stage. Despite this difference in levels, the series comove closely, indicating that electronic invoice data capture the main dynamics of aggregate output.

Figure 1: Gross domestic product: Official Statistics and Electronic Invoice Aggregate



Notes: Panel (a) reports chained-volume indices for gross domestic product from official national accounts data and electronic invoices. Panels (b) and (c) report quarter and annual growth rates, respectively.

Real-time availability. One advantage of this data is that it is readily available. According to the SII rules, firms have up to ten working days to report these documents. As a result, virtually all business-to-business transactions are observable within a few days after the end of the month. This is a significant improvement compared to other administrative data that relies on tax forms subject to significant delays relative to the period in which economic activity occurs.

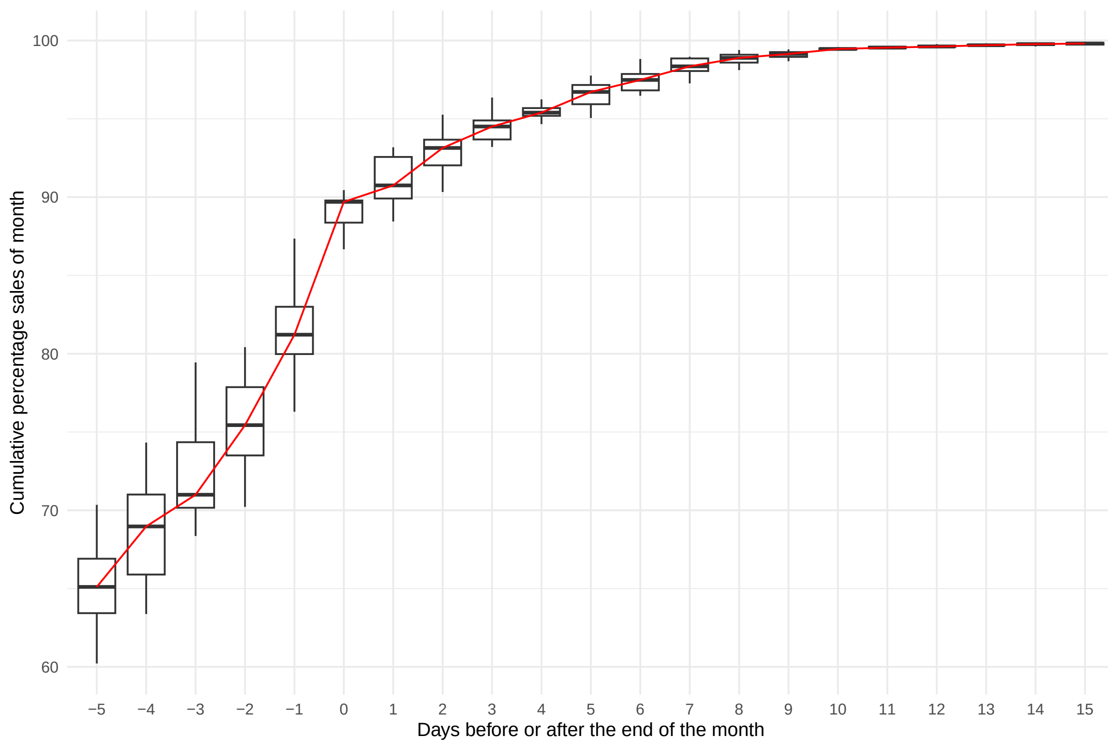
We study the practical implications of this 10-day reporting requirement. To do so, we exploit two dates recorded in each invoice: the transaction date and reporting date. This information allows us to quantify reporting delays and assess the timeliness of the data. We analyze the percentage of transactions reported after the month ends relative to the total amount of transactions:

$$s_{t,d} = \frac{r_{t,d}}{r_{t,\infty}},$$

where $r_{t,d}$ is the sales of month t reported d days before or after the end of such month, so $d \in \{\dots, -2, -1, 0, 1, \dots\}$, and $r_{t\infty}$ is the sales using all the information for that month.

Figure 2 shows the cumulative percentage of total sales as a function of the days before or after the month ends. Five days before the end of the month, the median value for $r_{t,-5}$ indicates a coverage of 65 percent of the total sales, and two weeks after, nearly all the information from the previous month has already been processed. Although there is some dispersion in the data, especially at horizons farther from month-end, the interquartile ranges remain relatively concentrated around the median. In this context, the ability to construct indicators that closely track official statistics is particularly valuable, because such indicators can be made available promptly and with only limited ex post revisions, complementing official national accounts data, which are revised as additional information becomes available.

Figure 2: Electronic invoices effective transaction date versus actual data availability



Notes: This figure plots the share of total sales received by the Bank by days relative to month-end ($d = 0$). Each box shows the median, the interquartile range, and the minimum–maximum values for each day. The plot includes transactions from 2023.

3 Methodology

This section describes how we construct the labeled corpus of free-text product descriptions, preprocess the text, and implement the classification algorithm used to assign standardized product categories to previously unclassified descriptions.

3.1 Constructing the corpus

In our setting, converting the free-text product descriptions into standardized categories suitable for economic analysis poses two challenges. First, invoice descriptions have not been previously labeled using standard economic classifications, so no off-the-shelf training data are available. Second, the language used in invoicing is highly idiosyncratic and domain-specific, making pre-trained language models ill-suited for the task. We therefore develop a classification strategy tailored to the structure and content of Chilean electronic invoice data, mapping descriptions into three product classifications: i) the Central Product Classification (CPC), which provides a broad and fine taxonomy of goods; ii) the Unique Product Classification (UPC), a Chilean adaptation of the CPC designed to better reflect the structure of goods production in the Chilean economy; iii) the Classification of Individual Consumption by Purpose (COICOP). The first two are better suited to analyze more general aspects of the economy, while the latter zooms in on consumption goods; hence, in the application section we will use the classifiers accordingly. We turn to discussing how the training corpus is assembled.[2]

3.1.1 Sample selection

We construct a training set from free-text product descriptions that are manually assigned to one of K classes, $\mathcal{K} = \{1, \ldots, K\}$. Let S denote the set of economic sectors and s its index. For each sector s, let $G(s)$ be the set of distinct free-text descriptions observed in that sector, and $G = \bigcup_{s \in S} G(s)$ the universe of descriptions.

Prior to manual labeling, we rank descriptions by an importance score that balances revenue, frequency, and monthly occurrence probability. For description i in sector s we

[2]For the CPC we have 1,766 categories, the UPC we have 284 categories, and the COICOP we have 283 categories.

define

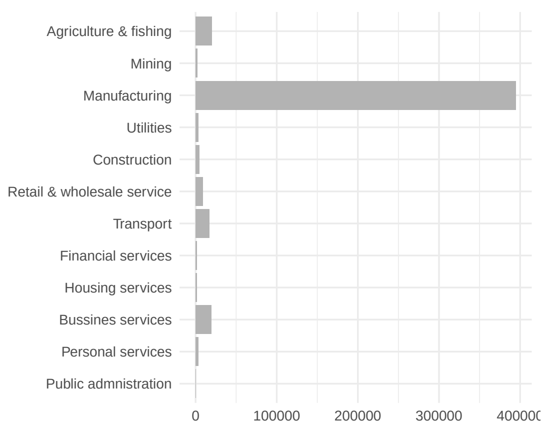
$$\gamma_{i,s} \equiv \frac{1}{3}\left(\frac{m_i}{\sum_{j\in G(s)} m_j}\right) + \frac{1}{3}\left(\frac{e_i}{\sum_{j\in G(s)} e_j}\right) + \frac{1}{3}\,\rho_{i,s},$$

where m_i is the total revenue associated with description i, e_i is its frequency over time, and $\rho_{i,s}$ is the monthly probability that description i occurs in sector s based on electronic invoice transaction lines.[3] The equal weights (1/3) reflect a pragmatic balance across economic relevance, coverage and intensity. Within each sector s, we select the top 1,000–2,000 $\gamma_{i,s}$ for manual labeling.[4] Let the resulting labeled corpus be

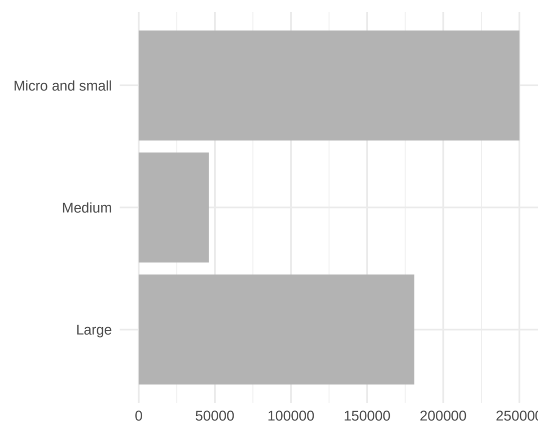
$$\mathcal{C} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N},$$

where each $\mathbf{x}_i$ is a token sequence $\mathbf{x}_i = (w_1, \dots, w_T)$ representing a (raw) free-text description and $y_i \in \mathcal{K}$ is its class label or the true label of observation i.

Figure 3: Training sample composition



(a) UPC Division



(b) Firm sale size

Notes: Composition of the training sample by UPC division and firm sales size, measured by the number of free-text descriptions.

[3]Unlike e_i, which aggregates occurrences, $\rho_{i,s}$ is the probability that description i appears at least once in a given month.

[4]In addition, the training sample is iteratively updated using two complementary criteria. First, categories with low quality metric scores after cross-validation are prioritized for additional manual labeling. Second, and more distinctively, we evaluate classification quality through an economic application: for the CPI and COICOP classifiers, we construct product-level price indices and compare them against official time series published by the National Statistics Office, assessing fit using R-squared. Products are then ranked by goodness of fit, and manual labeling efforts are concentrated on the poorest-performing categories.

Figure 3 shows the sectoral and firm size composition of the training sample. By sector, manufacturing accounts for the largest share of labeled descriptions, reflecting the greater product variety and specificity of invoicing language in that sector. By firm size, the sample is skewed toward large firms and micro and small enterprises, with medium-sized firms representing a lower proportion.

3.1.2 Text preprocessing and auxiliary features

Free-text descriptions are unstructured and often contain non-informative tokens for classification. We therefore apply a preprocessing function to clean and normalize the text. Let

$$\phi : \mathcal{C} \to \mathcal{D},$$

be a function mapping each raw document $\mathbf{x}_i$ to a normalized document $\phi(\mathbf{x}_i)$. Following Gentzkow et al. (2019); Manning et al. (2008), we clean the text data as follows: (i) lowercasing; (ii) transliterating Spanish diacritics to ASCII (e.g., á→a, ñ→n); (iii) removing characters outside $[a\text{–}z]$ (dropping digits and punctuation); (iv) removing stopwords; (v) stemming (e.g., "bottles"/"bottle" → "bottl"; in Spanish, "computadoras" → "computador"). We adopt stemming for simplicity and speed; (vi) removing tokens shorter than two characters; and (vii) normalizing whitespace.[5]

We augment each document with the seller's sector as an auxiliary feature, $\phi(\mathbf{x}_i) \cup \{\langle s_i \rangle\}$. A useful illustration of why sector information improves classification involves maintenance services. Consider the description "air conditioner": a retailer using this label is almost certainly selling the physical unit, whereas a firm in the business services sector is more likely invoicing for its installation or repair. Without sector information, the algorithm cannot resolve this ambiguity and will tend to misclassify one of the two cases.

3.2 Classification approach

We employ a supervised multi-class text classification strategy to assign each electronic-invoice entry to one of K product categories. Specifically, we use the FastText classification algorithm of Joulin et al. (2016), which jointly learns subword and word n-gram embeddings and a linear classifier through a single optimization problem. This approach is particularly well-suited to our setting: invoice descriptions are short and noisy, frequently containing misspellings, abbreviations, and seller-specific conventions, and the classifier

[5]In our setting, lemmatization delivered similar accuracy at higher computational cost.

must scale to millions of unlabeled entries at inference time.

3.2.1 Subword and word n-gram features

To handle misspellings, neologisms, and morphological variation, FastText represents each token through its character n-grams (Bojanowski et al., 2016; Joulin et al., 2016). Let $\langle w \rangle$ denote a token w padded with boundary markers. For $n \in [n_{\min}, n_{\max}]$, the set of character n-grams of w is

$$\mathcal{G}(w) = \bigcup_{n=n_{\min}}^{n_{\max}} \{\langle w \rangle[j : j+n-1] \mid 1 \leq j \leq |\langle w \rangle| - n + 1\},$$

where $|\langle w \rangle|$ denotes the number of characters in $\langle w \rangle$. The whole-word token $\langle w \rangle$ is also included as a feature. For example, with $w =$ `lotion` and $n_{\min} = n_{\max} = 3$, the character n-grams are {<lo, lot, oti, tio, ion, on>} together with the whole-word token <lotion>.

To capture local word order, we also include word n-grams up to length $n_g \in \mathbb{N}$ (Wang and Manning, 2012). For a tokenized document $\mathbf{x} = (w_1, \ldots, w_T)$, the set of word n-grams is

$$\mathcal{F}(\mathbf{x}) = \bigcup_{n=1}^{n_g} \left\{(w_j, \ldots, w_{j+n-1}) \,\middle|\, 1 \leq j \leq T - n + 1\right\}.$$

For instance, with $n_g = 2$ and the description `hand lotion extra dry`, the bigrams are {`hand lotion`, `lotion extra`, `extra dry`}.

Finally, the document's active feature multiset combines subword and word features is:

$$\mathcal{S}(\mathbf{x}) \;=\; \left(\bigcup_{w \in \mathbf{x}} \mathcal{G}(w)\right) \cup \mathcal{F}(\mathbf{x}).$$

3.2.2 Document representation, training, and inference

To map features to embedding vectors efficiently, FastText uses the hashing trick (Weinberger et al., 2009), implementing it via the Fowler–Noll–Vo hash function (Fowler et al., 2024): a fixed hash function $h : \mathcal{S}(\mathbf{x}) \to \{1, \ldots, B\}$, so that all features mapped to the same bucket share a common embedding $\mathbf{u}_{h(w)} \in \mathbb{R}^d$. This eliminates the need for an explicit vocabulary dictionary and keeps memory requirements constant regardless of how many distinct features are observed.

The document embedding is then the average of the embeddings of all active features:

$$\mathbf{z}_{\mathbf{x}} = \frac{1}{|\mathcal{S}(\mathbf{x})|} \sum_{w \in \mathcal{S}(\mathbf{x})} \mathbf{u}_{h(w)} \in \mathbb{R}^{d}.$$

Averaging over character and word n-gram embeddings renders the representation order-invariant at the global level while word n-grams retain sensitivity to local word order.

A linear classifier maps the document embedding to scores over the K classes:

$$s(\mathbf{x}) = \mathbf{W}\mathbf{z}_{\mathbf{x}} + \mathbf{b}, \qquad \mathbf{W} \in \mathbb{R}^{K \times d}, \quad \mathbf{b} \in \mathbb{R}^{K},$$

where the k-th row $\mathbf{w}_k^{\top}$ produces the score for class k, and b_k is a class-specific intercept. Class probabilities are obtained via the softmax transformation,

$$p(k \mid \mathbf{x}) = \frac{e^{s_k(\mathbf{x})}}{\sum_{j=1}^{K} e^{s_j(\mathbf{x})}},$$

and the model is trained by minimizing the negative log-likelihood over the labeled training set:

$$\min_{\theta} \sum_{(\mathbf{x},y) \in \mathcal{D}} \mathcal{L}(\mathbf{x}, y), \qquad \mathcal{L}(\mathbf{x}, y) = -\log p(y \mid \mathbf{x}), \tag{1}$$

where $\theta = \{\mathbf{W}, \mathbf{b}, \mathbf{u}_1, \ldots, \mathbf{u}_B\}$. Optimization proceeds via stochastic gradient descent with a linearly decaying learning rate, trained over E epochs using asynchronous multi-threaded updates on CPU (Bottou, 2010; Niu et al., 2011), which allows the classifier to scale to the full invoice dataset without specialized hardware.

At inference, the predicted category for a new document $\mathbf{x}$ is

$$\hat{y}(\mathbf{x}) = \arg\max_{k \in \mathcal{K}} p(k \mid \mathbf{x}).$$

Where needed, we return the top-t predictions with their associated probabilities to support human-in-the-loop validation of ambiguous entries. Appendix A describes the computational implementation and the hyperparameter search strategy.

3.3 Validation

To assess how well our classification model performs, we evaluate the model's predictions using cross-validated performance metrics. We rely on repeated stratified q-fold cross-validation, where the original training sample is divided into q equal parts, using $q-1$ folds for training and the remaining one for evaluation. In our case, we use $q=5$ folds and repeat this procedure 10 times to reduce sampling variability.

For each classifier, repetition, fold q, and class k, we compute standard classification metrics. First, Precision is defined as

$$P^{qk} = \frac{T_p^{qk}}{T_p^{qk} + F_p^{qk}},$$

where T_p^{qk} denotes true positives and F_p^{qk} false positives. Precision measures the proportion of instances predicted as class k that are correctly classified.

Second, Recall is defined as

$$R^{qk} = \frac{T_p^{qk}}{T_p^{qk} + F_n^{qk}},$$

where F_n^{qk} represents false negatives. Recall measures the proportion of actual instances of class k that are correctly identified by the classifier.

Finally, we compute the F1-score, which combines Precision and Recall through their harmonic mean:

$$F1^{qk} = 2\frac{P^{qk} \times R^{qk}}{P^{qk} + R^{qk}}.$$

The F1-score ranges from 0 (the model fails to correctly classify any instance in class k) to 1 (perfect classification). Table 2 reports aggregate, class-weighted averages of Precision, Recall, and F1-score for the three classifiers, where weights correspond to the number of observations in class k in the training sample.

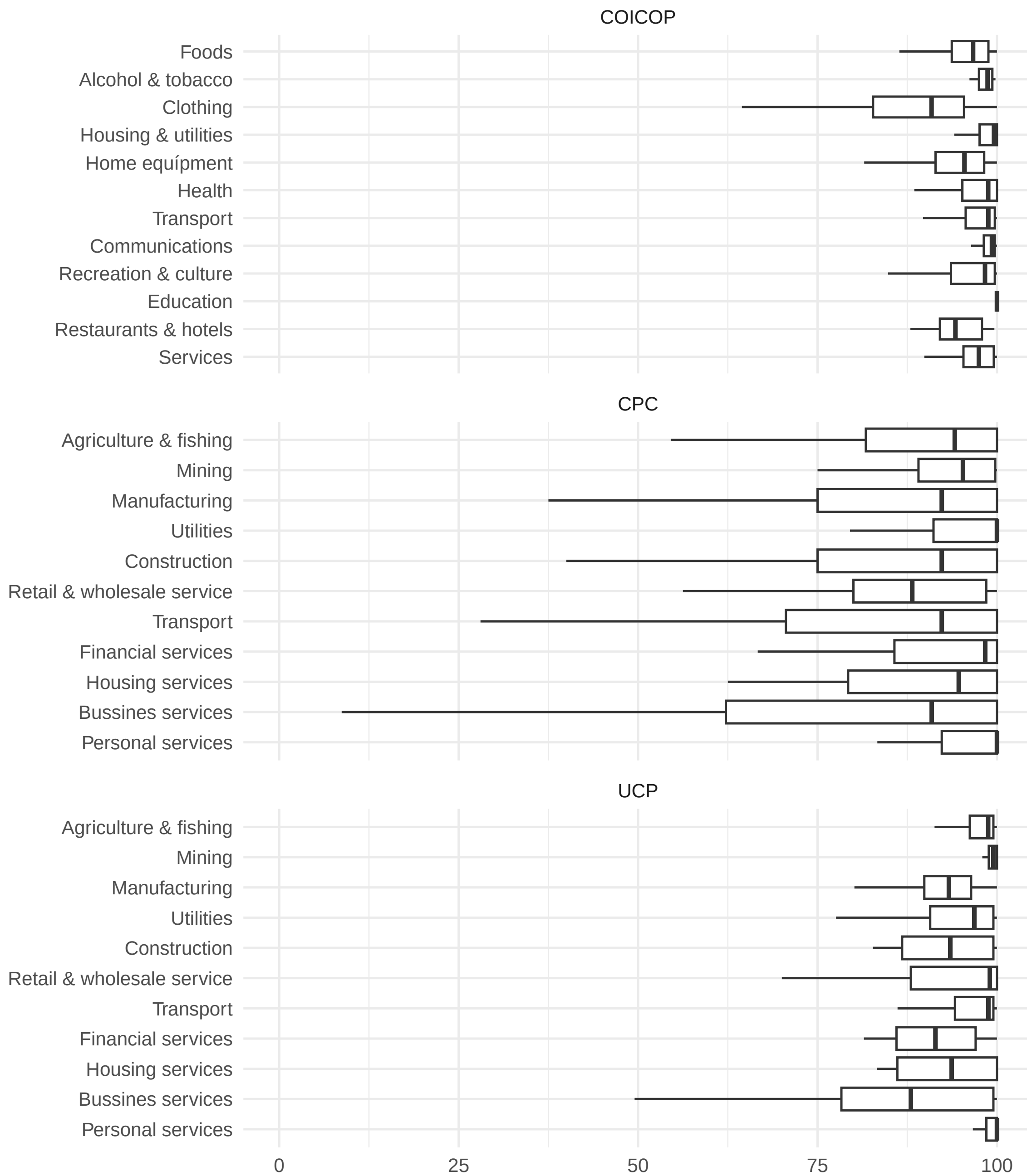
Table 2: Quality metrics by classifier and feature model

Classifier	Precision	Recall	F1
COICOP	0.947	0.946	0.946
CPC	0.913	0.908	0.909
UPC	0.936	0.935	0.935

Notes: Metrics are computed using q-fold cross-validation. Reported values correspond to class-weighted averages, where weights are given by the number of observations in each class in the training sample.

Figure 4 shows the distribution of classification performance across divisions for each of the three classifiers. All three exhibit generally high F1 scores, with relatively tight distributions in the case of COICOP and UPC, and somewhat greater dispersion for the CPC classifier. This is expected given that CPC is a broader classification scheme with a larger number of categories, which naturally introduces more confusion between adjacent product classes.

Figure 4: Distribution of classification performance (F1 score) by classifier



Notes: Boxplots show the distribution of F1 scores for each classifier across divisions. Metrics are computed using q-fold cross-validation.

4 Applications

The product classifications constructed in the previous section allow us both to build real-time measures of economic activity and to trace how shocks affect consumption. We construct Machinery and Equipment and Goods Consumption series that closely track their national accounts counterparts and are available within days of month-end. We then use the product-level information embedded in the classifications to study expenditure reallocation during the COVID-19 pandemic, examine the role of lockdowns and income support policies in shaping these shifts, and compare flexible- and fixed-weight consumer price indices.

4.1 Machinery and Equipment and Consumption tracking

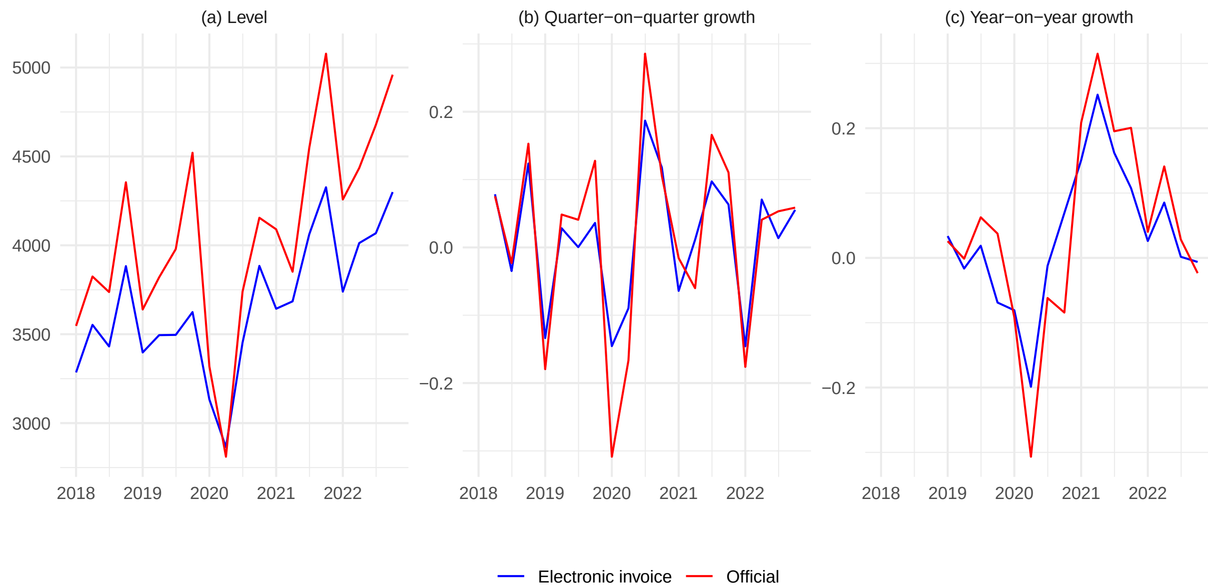
We leverage our classification algorithm to construct an aggregate series that mimics the official machinery and equipment outlays series from National Accounts. We build the series using the mapping between firms' purchases and their associated CPC and UPC categories, which together provide broad coverage of the products identified as gross fixed capital formation under the Broad Economic Categories classification. To ensure consistency, we retain only goods for which both classifiers agree that the item is a capital good. After cleaning the data we aggregate firms' purchases of capital goods to obtain our measure of machinery and equipment outlays at quarterly frequency.[6]

Each panel in Figure 5 presents our constructed machinery and equipment outlays series together with the corresponding national accounts counterpart. Panels (a), (b), and (c) show the levels, quarter-on-quarter growth, and year-on-year growth of such series, respectively. Our level measure is consistently below the official statistics series, as we do not include direct imports. Despite this, quarter-on-quarter and year-on-year growth strongly correlate with their national accounts counterparts, thus capturing the official statistics' trend, seasonal pattern, and cyclical upswing. For instance, our estimates capture the collapse of the outlays at the beginning of the COVID-19 pandemic. The correlation for quarter and annual variations of the constructed machinery and equipment outlays and national accounts series are 0.93 and 0.90, respectively. Interestingly, we also compute the correlation coefficient between the first national accounts release of machinery and

[6]Before aggregation, (i) intra-firm transactions are removed; (ii) purchases of capital goods from the retail sector are excluded to avoid double counting; (iii) transactions exceeding USD $110,000,000$ are dropped; and (iv) invoices modified by credit notes are adjusted accordingly.

equipment outlays and our reported measure, finding weaker correlations than with the final release: 0.69 and 0.54, respectively. This pattern suggests that the electronic-invoice measure aligns more closely with the revised official statistics.

Figure 5: Machinery and equipment outlays: Official statistics and electronic invoice aggregate

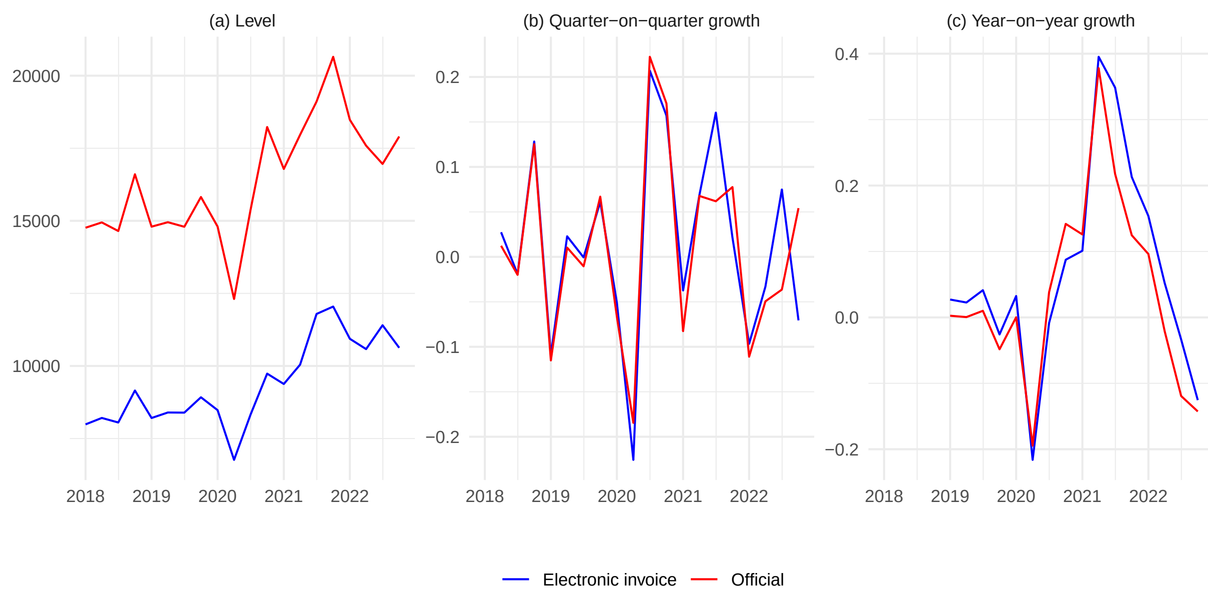


Notes: Panel (a) reports chained-volume indices for machinery and equipment outlays from official national accounts data and electronic invoices (EI, right axis). Panels (b) and (c) report quarter-on-quarter and year-on-year log differences, respectively. The EI series excludes direct imports and retail-sector purchases.

Similar to machinery and equipment outlays, we also construct a series for goods consumption. Considering that we use electronic invoices that originate in business-to-business transactions, we make assumptions to obtain an aggregate that conceptually resembles that of goods consumption. We proxy household consumption of final goods through the input purchases of firms in the Retail and Restaurants and Hotels sectors. The assumption is that once products reach retailers, they have a fast turnover and reach final consumers within a quarter so as to avoid generating large deviations between retail purchases and final consumption. We further restrict attention to items whose product codes map to consumption goods under the Broad Economic Categories classification and its correspondence to COICOP. Aggregating all the items' values over the quarter, we obtain our measure of aggregate consumption.

Each panel in Figure 6 presents our constructed goods consumption series together with its official counterparts. Panels (a), (b), and (c) show the levels, quarter-on-quarter growth and year-on-year growth of such series, respectively. Our estimated level of goods consumption series underestimates the official counterpart, which could be explained by variations in the markup retailers charge consumers. Despite this, the movements in the consumption measure computed from electronic invoices strongly correlate with official statistics. For quarter-on-quarter and year-on-year growth, the correlation coefficient is 0.95 and 0.94, respectively. We also considered the correlation coefficient between the first national accounts release of goods consumption and our reported measure, finding values similar to those with the final release: 0.95 and 0.97, respectively. The difference with the earlier evidence presented in the case of the machinery and equipment outlays indicator is explained by the greater revision machinery and equipment outlays national accounts data is subject to.

Figure 6: Goods consumption: Official statistics and electronic invoice aggregate



Notes: Panel (a) reports chained-volume indices for goods consumption from official national accounts data and electronic invoices (EI, right axis). Panels (b) and (c) report quarter-on-quarter and year-on-year log differences, respectively. The EI series proxies final goods consumption using retail-sector purchases and excluding services.

4.2 Consumption patterns over time

We leverage our classification to document the reallocation of expenditure across durable and non-durable goods during the COVID-19 pandemic, a period of unusually large and rapid changes in consumption patterns. Using data from 2018 to 2023, we first characterize the aggregate shift toward durable goods that emerges in the second half of 2020, and then decompose this reallocation across broad product categories. We subsequently examine whether two policies happening over this period may have shaped these patterns: the stay-at-home orders imposed by the Chilean government at the municipal level, and pension withdrawals enacted to provide liquidity support to households.

4.2.1 Durable and Non-durable consumption

We turn to the analysis of consumption patterns in recent years, with a particular focus on the period during and after the COVID-19 pandemic. This is especially relevant for our application, as official statistics on consumption behavior are updated only every five years, limiting the ability to study interim changes. Figure 7 displays aggregate quarterly consumption shares for durable and non-durable goods between 2018 and 2022, using both our constructed series and the official data from GDP expenditure in personal consumption.[7] Beginning in the second half of 2020, both series exhibit a marked increase in the share of durable goods accompanied by a decline in non-durable consumption. This pattern has also been documented in other countries (Tauber and Zandweghe, 2021). Although our estimates indicate a relatively higher share of durables than in the official data, the evolution of durable and non-durable shares is highly correlated across the two measures.

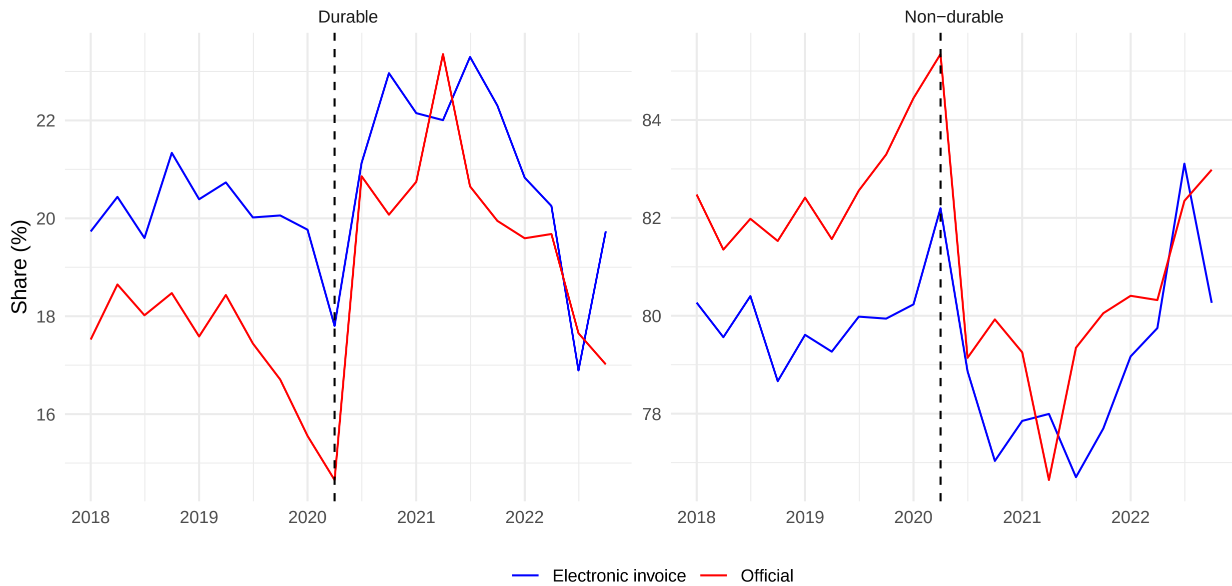
We investigate which product categories account for the rise in durable and the decline in non-durable expenditure shares during the first year following the onset of the pandemic. To this end, we decompose expenditure into four broad categories—Food, Clothing and General Merchandise, Transport, and Entertainment—splitting each into its durable and non-durable components. For each category-durability j, we estimate the following regression:

$$\omega_{jt} = \lambda_j + \beta_j D_t + \varepsilon_{jt}, \tag{2}$$

where ω_{jt} denotes the expenditure share of each category-durability j in quarter t, with $\sum_{j \in J} \omega_{jt} = 1$, and D_t is an indicator equal to one for the period 2020Q3–2021Q2 and zero otherwise. The category fixed effect λ_j absorbs the average expenditure share of each category over the full sample, so that $\widehat{\beta}_j$ measures the average deviation of category j's

[7] Appendix Figure B.1 further reports quarterly growth rates of these variables.

Figure 7: Consumption shares: Durable and non-Durable



Notes: This figure reports consumption shares for durable and nondurable goods using retail-sector purchases from electronic invoices (EI) and the corresponding official national accounts data.

share during the pandemic period relative to the pre-pandemic baseline.

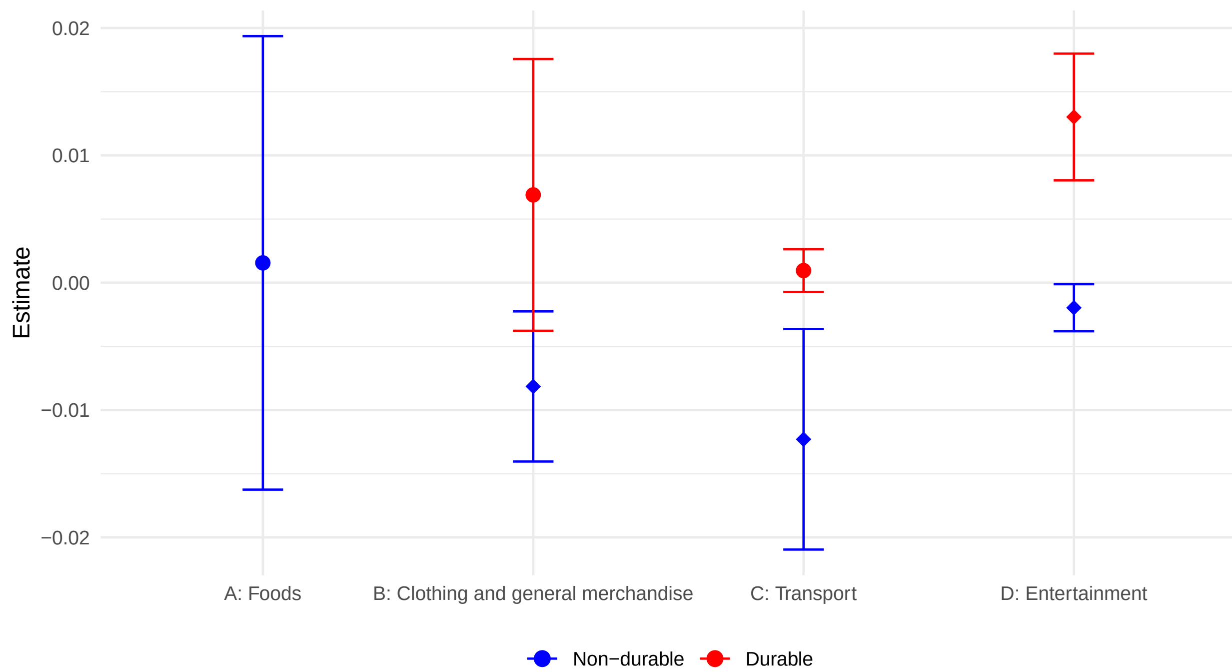
Figure 8 reports $\widehat{\beta}_j$ for each category. The results reveal a broad reallocation of expenditure toward durable goods, driven primarily by gains in Clothing and General Merchandise and Entertainment, offsetting declines in their non-durable counterparts. Particularly striking is the sharp contraction in non-durable Transport expenditure, which emerges as the single largest negative contributor. On net, the pandemic period was characterized by a marked shift away from Transport goods and toward Entertainment goods, reflecting the substitution from mobility-related spending toward home-based consumption during this period.

4.2.2 The effects of stay-at-home orders over consumption

We zoom in on how the imposition of stay-at-home orders affected consumption differently over the period. We describe the mobility constraints the Chilean government established during the period and use an event-study research design to quantify their effect on different consumption categories.

Background information and methodology. On March 18, 2020, the President of Chile

Figure 8: Expenditure changes during COVID-19



Notes: This figure shows changes in expenditure during COVID-19. The circles report the β_j coefficients from the regression defined in Equation (2) for durable and non-durable consumption categories. The range bars denote the 90% confidence interval.

declared a state of national catastrophe. On March 22, a nationwide curfew was introduced, prohibiting movement between 10:00 p.m. and 5:00 a.m. On March 25, the government announced a system of *Rotating Lockdowns*, under which the national authority imposed targeted lockdowns at the municipal level based on local epidemiological conditions. The order was binary: a municipality was either under lockdown or not.

On July 19, 2020, the government introduced a five-step plan to manage the gradual reopening of economic activity. Step 1 imposed strict mobility restrictions, requiring government permits for essential activities such as grocery shopping or commuting, and suspending all non-essential face-to-face activities.[8] Step 2 relaxed confinement on weekdays but maintained mandatory lockdown on weekends, holidays, and for individuals aged 75 and above; non-essential activities remained prohibited. Steps 3 through 5 lifted lockdown requirements for the general population, differing only in the maximum permitted capacity for non-essential activities. For our empirical analysis, we focus on whether a municipality was subject to a full lockdown or not—that is, whether it was under the

[8]These included the closure of schools and restaurants and the prohibition of gatherings exceeding 50 people.

Rotating Lockdown system or Step 1 of the five-step plan, or neither.

We rely on a two-way fixed-effects event-study design to study the effect of these policies on consumption. To do so, we first remove weekly shocks common to all firms operating in the country. We then exploit within-municipality variation over time, combined with cross-municipality differences in exposure to the lockdown policy.[9] Our analysis covers the period from January 2020 to July 2021. We implement this approach by regressing the logarithm of input purchases of retailers, which is our proxy for consumption, on municipality and week fixed effects, as well as event-time indicators capturing periods before and after the entering and exiting stay-at-home orders. Following Alexander and Karger (2021), we estimate the following regression model:

$$\log X_{c,t} = \lambda_c + \lambda_t + \sum_{s \in S} \alpha_s \mathbb{1}^{\text{Event}}_{c,t+s} + \epsilon_{c,t}, \tag{3}$$

where $X_{c,t}$ is consumption in municipality c during week t, $S = \{-4, \ldots, 0, \ldots, 4\}$, $Event$ can either be *Enter* or *Exit*, and $\mathbb{1}^{\text{Event}}_{c,t+s}$ takes the value of one if municipality c is under $Event$ during week $t + s$, and zero otherwise. The coefficients of interest, α_s, trace the dynamic response of average consumption at the municipality level around the introduction of stay-at-home orders. The error term $\epsilon_{c,t}$ captures idiosyncratic shocks to consumption, and standard errors are clustered at the municipality level to allow for arbitrary serial correlation within municipalities.

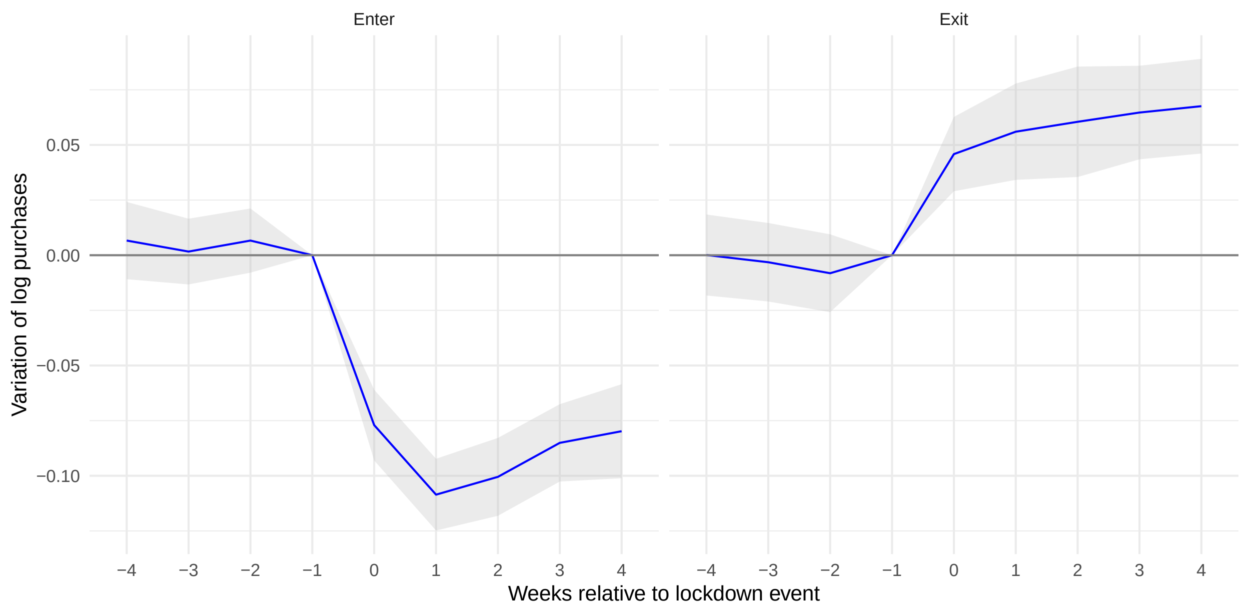
Effects of lockdown and opening. We describe the consumption adjustment when the Chilean national authority imposes stay-at-home orders. Panel *Enter* of Figure 9 shows estimates of the event study parameters. In the weeks before the lockdown imposition, no statistical differences existed between the treatment and control municipalities, which supports the use of non-lockdown municipalities as a control group.

Upon entering into lockdown, consumption contracts sharply, declining by approximately 8 log points on impact (where log point coefficients are scaled by 100 for readability), and deepening to nearly 11 log points one week later. Although consumption begins to recover two weeks after the onset of lockdown, it remains significantly below pre-lockdown levels, with consumption still between 8 and 9 log points lower four weeks into the restriction period. These magnitudes are reported in column (1) of Appendix Table B.1, which summarizes the average effect across all consumption categories.

[9]Data on lockdowns, the Rotating Lockdown policy, and the Step by Step We Take Care of Ourselves plan comes from the Chilean Ministry of Science.

Panel *Exit* of Figure 9 reports estimates around the lifting of stay-at-home orders. In the weeks prior to reopening, consumption in lockdown municipalities remains modestly below that of non-lockdown municipalities. Following the removal of mobility restrictions, consumption increases by roughly 5 log points in the first week after reopening. In the subsequent weeks consumption keeps recovering, although it does not fully revert to pre-lockdown levels within the event window.

Figure 9: Lockdown effect on consumption.



Notes: Variation of log consumption before or during lockdown. This figure shows $\sum_{s \in S} \alpha_s$ from regression defined in Equation 3. The gray area denotes the 90% confidence interval.

We can further understand the changes consumption goods experience when we zoom in on disaggregated categories. Figure 10 presents the estimates for i) foods, ii) Entertainment, iii) Clothing and general merchandise, iv) Transportation.

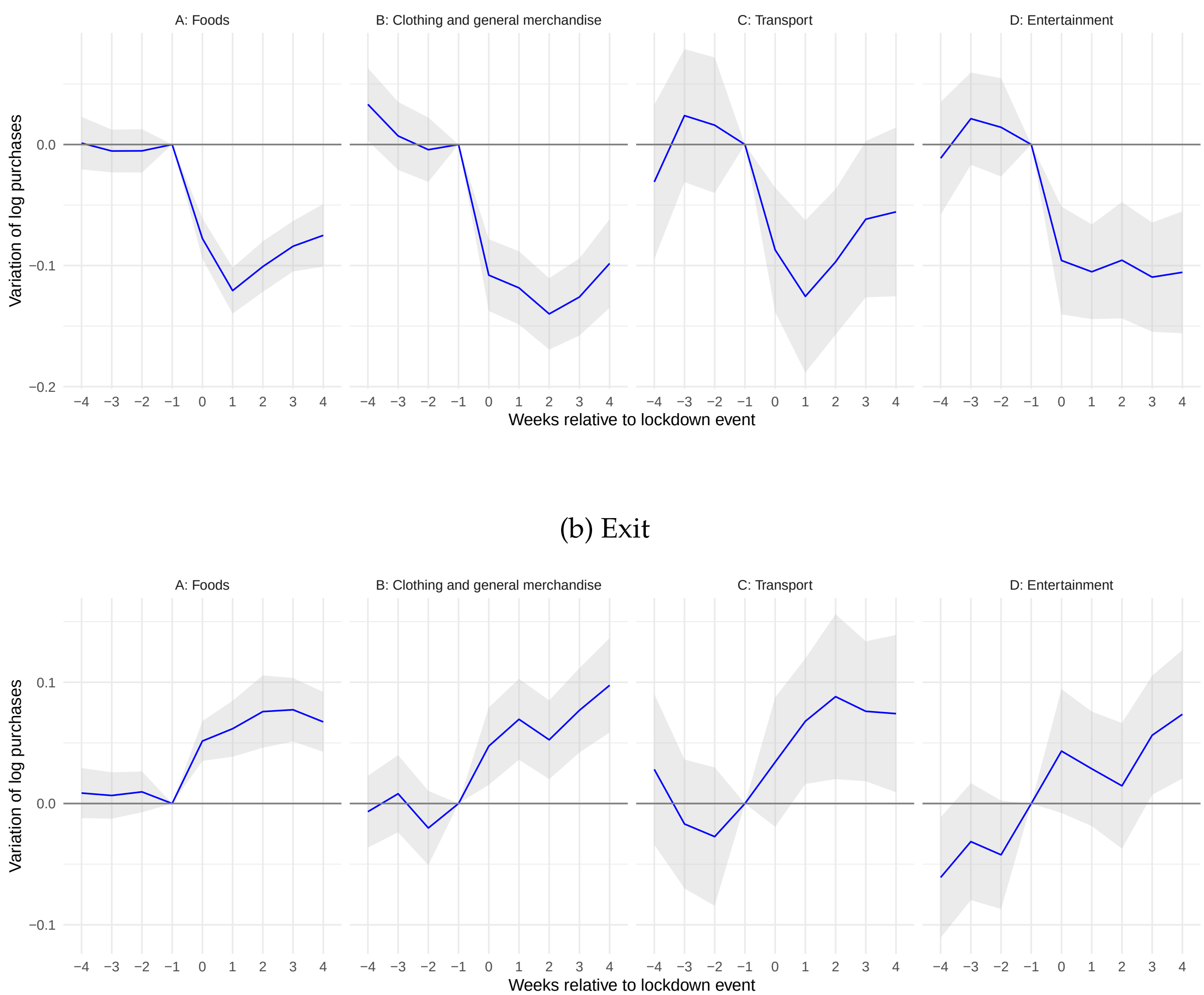
At the onset of lockdowns, food consumption declines by approximately 8 log points on impact, deepening to 12 log points one week later and recovering gradually thereafter, remaining about 8 log points below baseline four weeks into the lockdown. Clothing and general merchandise experiences the sharpest contraction, falling by roughly 11 log points on impact and by 14 log points one week later, remaining approximately 10 log points below baseline after four weeks. Entertainment follows a similar pattern, declining by about 10 log points on impact and remaining persistently depressed throughout the restriction period. Transport, in contrast, displays a somewhat more muted and less precisely estimated contraction, declining by roughly 9 log points on impact and recovering more rapidly, with the coefficient no longer statistically significant four weeks into the lockdown.

Opening the economy shows substantial heterogeneity across goods categories. Food is the category that reacts more quickly to reopening, rising by approximately 5 log points upon impact, but remaining relatively flat afterwards, with the coefficient stabilizing at around 7 log points. The other categories present a somewhat different pattern. Clothing and general merchandise and transport initially recover less upon reopening, with estimates of roughly 5 and 3 log points, respectively, but tend to present a more robust response in the weeks after, with clothing reaching approximately 10 log points and transport approximately 8 log points. Entertainment follows a similar trajectory, displaying a modest initial rebound of roughly 4 log points that builds to approximately 7 log points. Yet, estimates are not precise across all categories, and in general we cannot reject that the responses are statistically the same across goods.

Finally, Table 3 shows further disaggregation for durable and non-durable categories and estimated effects four weeks after entering or exiting the lockdown. Durable Transport and Entertainment and non-durable Clothing and general merchandise goods are the only categories in which the net effect between week four after entry and week four after exit of stay-at-home orders seems to have increased their overall participation in consumption, at the expense of other categories. Durable Transport also exhibits a positive net effect as Entertainment, yet estimates for this category are generally imprecise. Appendix Tables B.2 and B.3 report the full set of estimates.

Figure 10: Lockdown and divisions of consumption goods.

(a) Enter



Notes: Variation of log consumption before or during lockdown by product division. This figure shows $\sum_{s \in S} \alpha_s$ from regression defined in Equation 3. The gray area denotes the 90% confidence interval. Standard errors are clustered at the municipality level.

Table 3: Lockdown and divisions of consumption goods: Net effect at week 4

	Aggregate	Foods	Clothing and general merchandise	Transport	Entertainment
			(a) Non-durable		
Enter ($s = 4$)	-0.078***	-0.075***	-0.084***	-0.064*	-0.101***
	(0.011)	(0.013)	(0.021)	(0.030)	(0.028)
Exit ($s = 4$)	0.069***	0.067***	0.113***	0.044	0.071*
	(0.011)	(0.013)	(0.020)	(0.031)	(0.029)
			(b) Durable		
Enter ($s = 4$)	-0.107**		-0.154***	-0.087	-0.077.
	(0.034)		(0.032)	(0.106)	(0.044)
Exit ($s = 4$)	0.086*		0.084*	0.202.	0.120**
	(0.039)		(0.034)	(0.111)	(0.045)

Notes: This table reports event-study estimates of the effect of stay-at-home orders on consumption by product division at week $s = 4$. Panel (a) reports non-durable goods and panel (b) reports durable goods. Each column corresponds to a separate regression of the Equation 3. The omitted category is the week immediately preceding the event ($s = -1$). All regressions include municipality and week fixed effects. Standard errors, reported in parentheses, are clustered at the municipality level. Coefficients are reported in log points and can be interpreted approximately as percent changes. Statistical significance is denoted by *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$, and . $p < 0.15$.

4.2.3 The effects of pension withdrawals over consumption

Chile passed three pieces of legislation that allowed pension withdrawals during the COVID-19 pandemic where affiliates could withdraw up to ten percent of their accumulated mandatory contributions in the private pension system. The policies were enacted on July 30, 2020, December 10, 2020, and April 28, 2021, and applied on a one-time basis for each episode. Withdrawals were subject to a minimum and a maximum amount, and if ten percent of an affiliate's pension balance was below the minimum threshold, the individual was allowed to withdraw the full available balance.[10]

The withdrawals display three pronounced spikes that closely correspond to the enactment of the relevant laws. Appendix Figure B.2 shows the monthly evolution of pension withdrawals, expressed in millions of UF (Unidad de Fomento, a Chilean inflation-indexed unit of account). In each episode, withdrawals peak in the first month following authorization and decline rapidly thereafter. This front-loaded response is consistent with affiliates requesting funds shortly after becoming eligible, suggesting that withdrawals

[10]The minimum and maximum were the equivalent of about 1,250 and 5,300 USD. Withdrawals were generally tax-exempt, with the exception of the second episode, and affiliates had up to 365 days to request their funds.

were driven primarily by immediate liquidity needs. Following each spike, withdrawal activity falls to near-zero levels, indicating limited take-up beyond the initial months of each policy episode. We next examine the evolution of consumption during and up to 2 months after each episode.

We focus on the effect of the first episode, as this was largely unanticipated at the time of implementation. Anticipation was substantially more pronounced for the second and third withdrawals. Consistent with this interpretation, in Appendix Figure B.3 we show that Google Trends data for keywords related to pension withdrawals show a sharp increase in search intensity around the first withdrawal episode, followed by a monotonic decline for the second and third episodes, suggesting a diminishing novel component over time. This matters for our application: when withdrawals are anticipated, retailers may build up inventories ahead of the policy, so the estimated effect on retail purchases can carry the wrong sign.

We estimate the dynamic effects of pension fund withdrawals on local economic activity using an event-study design with continuous treatment intensity. For the first withdrawal, we specifically estimate the following:

$$\log X_{c,t} = \lambda_c + \lambda_t + \sum_{s \in S} \alpha_s D_{t+s} \psi_c + \epsilon_{c,t}, \tag{4}$$

where $X_{c,t}$ is total log consumption in municipality c during month t, $S = \{0, 1, 2\}$ and D_{t+s} takes the value of 1 in the month that is s periods from the treatment.

The treatment dynamics are mediated by an intensity variable that varies across municipalities and is captured by $\psi_c = \log(1 + \text{withdrawal}_c/\text{income}_c)$, where withdrawal_c is the amount withdrawn in municipality c during the first withdrawal window and income_c is average labor income in 2019. The specification includes municipality fixed effects, which absorb all time-invariant differences across municipalities, and time fixed effects, which control for aggregate shocks common to all municipalities, including macroeconomic conditions and seasonality.

The coefficients of interest, α_s, trace out the dynamic response of consumption following a withdrawal. Each coefficient measures the percent change in month s associated with a one-percent increase in pension withdrawals relative to monthly labor income.

Figure 11 reports the estimated dynamic effects of the first pension withdrawal. Consumption rises following the start of the policy, with effects that build over time. The point estimates rise from the month of withdrawal initiation and remain elevated for one to two months thereafter. The gradual build-up in retailers' purchasing response is consistent with a setting in which retailers initially satisfy the increase in final demand by drawing down existing inventories before placing new orders, though we cannot directly confirm this mechanism from our data, which measures input purchases rather than inventory levels or final sales. This pattern contrasts with evidence from household-level debit and credit card data, which typically shows stronger consumption responses on impact (Chetty et al., 2023; Cox et al., 2020; Hadzi-Vaskov et al., 2025). In our setting, the results suggest that while retailers respond to the liquidity shock from pension withdrawals, their adjustments lag households' spending decisions. Although the point estimate in the first month is smaller than in subsequent months, the estimates are not statistically distinguishable from one another.[11]

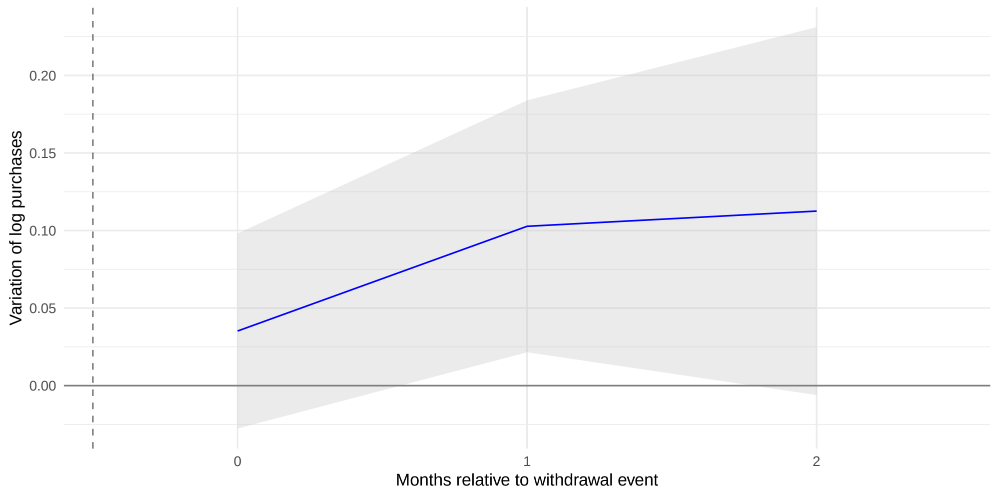
We next examine heterogeneity in the response to pension fund withdrawals across broad categories of consumption goods. Figure 12 reports estimates by product-category using the same specification as in Equation 4.[12]

The positive response in consumption appears to be driven primarily by Food and Clothing and general merchandise, for which point estimates increase following the withdrawal and remain elevated in subsequent months. In contrast, estimates for transportation and entertainment goods are imprecise and do not exhibit clear dynamic patterns. Interestingly, when we zoom in on four broad categories—Food, Clothing and General Merchandise, Transport, and Entertainment—splitting each into its durable and nondurable components, we find a positive response for the Entertainment category. Table 4 shows the results of replicating the analysis separately for durable and nondurable components. Among durables, Entertainment exhibits the largest point estimates. Among nondurables, estimated responses are generally smaller: Food shows a statistically significant increase, while Clothing and General Merchandise and Transportation are positive but imprecisely estimated.

[11] Results for the estimated dynamic effects of the second and third pension withdrawal are reported and discussed in Appendix B.

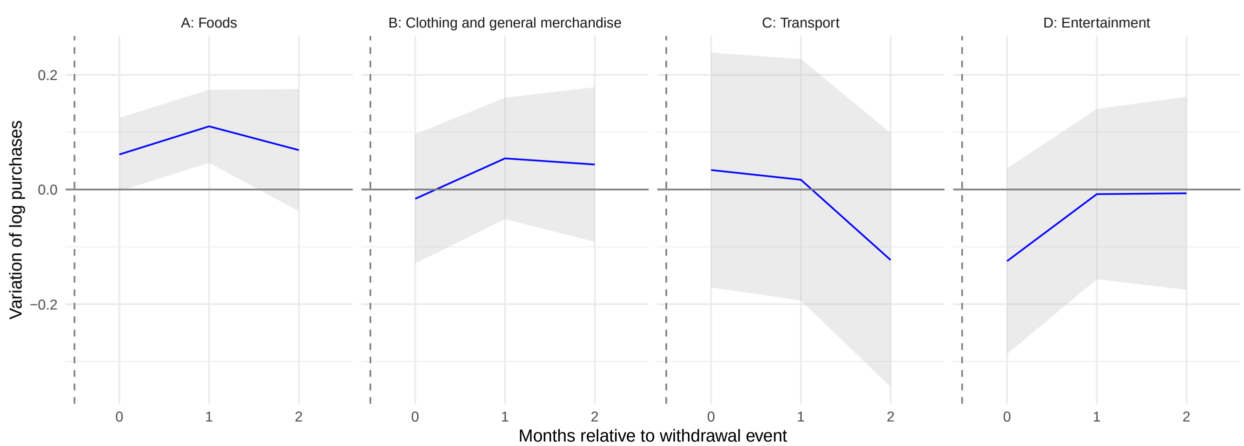
[12] Appendix Table B.4 reports the coefficients and their standard deviations.

Figure 11: First withdrawals effect on consumption.



Notes: Variation of log consumption after the first withdrawal. This figure shows $\sum_{s \in S} \alpha_s$ from regression defined in Equation 4. The gray area denotes the 90% confidence interval. Standard errors are clustered at the municipality level.

Figure 12: First withdrawal and divisions of consumption goods.



Notes: Variation of log consumption after first withdrawal by product division. This figure shows $\sum_{s \in S} \alpha_s$ from regression defined in Equation 4. The gray area denotes the 90% confidence interval. Standard errors are clustered at the municipality level.

Table 4: First Withdrawal and divisions of consumption goods

s	All	Foods	Clothing and general merchandise	Transport	Entertainment
			(a) Non-durable		
0	0.035	0.061.	-0.016	0.034	-0.125
	(0.032)	(0.033)	(0.057)	(0.104)	(0.082)
1	0.103*	0.110***	0.054	0.017	-0.008
	(0.041)	(0.032)	(0.054)	(0.107)	(0.075)
2	0.113.	0.069	0.044	-0.123	-0.007
	(0.060)	(0.054)	(0.069)	(0.113)	(0.086)
			(b) Durable		
0	-0.056		-0.070	0.163	0.116
	(0.081)		(0.068)	(0.230)	(0.116)
1	-0.019		0.001	0.072	0.206.
	(0.086)		(0.073)	(0.253)	(0.118)
2	-0.069		0.035	-0.137	0.213
	(0.106)		(0.095)	(0.240)	(0.139)

Notes: Variation difference of log consumption after the first withdrawal by division product. Panel (a) reports non-durable goods and panel (b) reports durable goods. This table shows $\sum_{s\in S}\alpha_s$ from regression defined in Equation 4. Standard errors are clustered at the municipality level. Statistical significance is denoted by *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$, and . $p < 0.15$.

4.3 Implications for price indices

We examine how the shifts in consumption patterns documented in the previous section affect the measurement of the cost of living. Combining price data from electronic invoices with our classification methodology, we construct price indices for a broad sample of products and compare a fixed-weight CPI measure against a time-varying weight alternative.

4.3.1 Methodology

We construct a price index analogous to the official CPI, updating consumption weights to reflect the changes in expenditure shares documented in Section 4.2.

The construction of the Consumer Price Index follows the methodology employed by the National Statistics Institute (INE for its acronym in Spanish) (Ministerio de Economía, 2020a). The first issue we address is the transactional nature of our data, as firms may sell the same variety at multiple prices within a month. To obtain a single representative quote for each period, we compute the median price across all transactions for a variety v, sold by firm f, during month t. We define variety v using the text description in the invoice, which —using our classification procedure— has been classified as a COICOP product i. The relevant price we use is denoted by $P_{v(i),f,t}$, and yields a sample of prices comparable to the one used by INE, which quotes prices for a given variety once a month.

Two additional challenges are worth noting. First, for services, electronic invoices do not always report the price per unit of the service or per hour, but instead provide only the total charge for the service provided. As a result, electronic invoices do not offer a reliable measure for tracking the price of services. Second, our dataset consists of business-to-business transactions, which prevents us from observing business-to-consumer transactions in sectors such as Education and Health services. Consequently, we cannot compute price indices for those sectors. Due to these limitations, we can track 227 of 303 products in the CPI basket, representing 55.0% of the total weights in the CPI. Further filtering of the data, summarized in Appendix C, leaves us tracking 157 products representing 36.4% of the consumption basket.

INE's methodology relies on constructing price indices for common varieties at the product level i in a recurring form. First, the set of firms selling a particular variety is denoted by M_{vt}, and their prices are averaged across as follows,

$$\Delta P_{v(i),t} = \left(\prod_{f \in M_{vt}} \frac{P_{v(i),f,t}}{P_{v(i),f,t-1}} \right)^{\frac{1}{|M_{vt}|}},$$

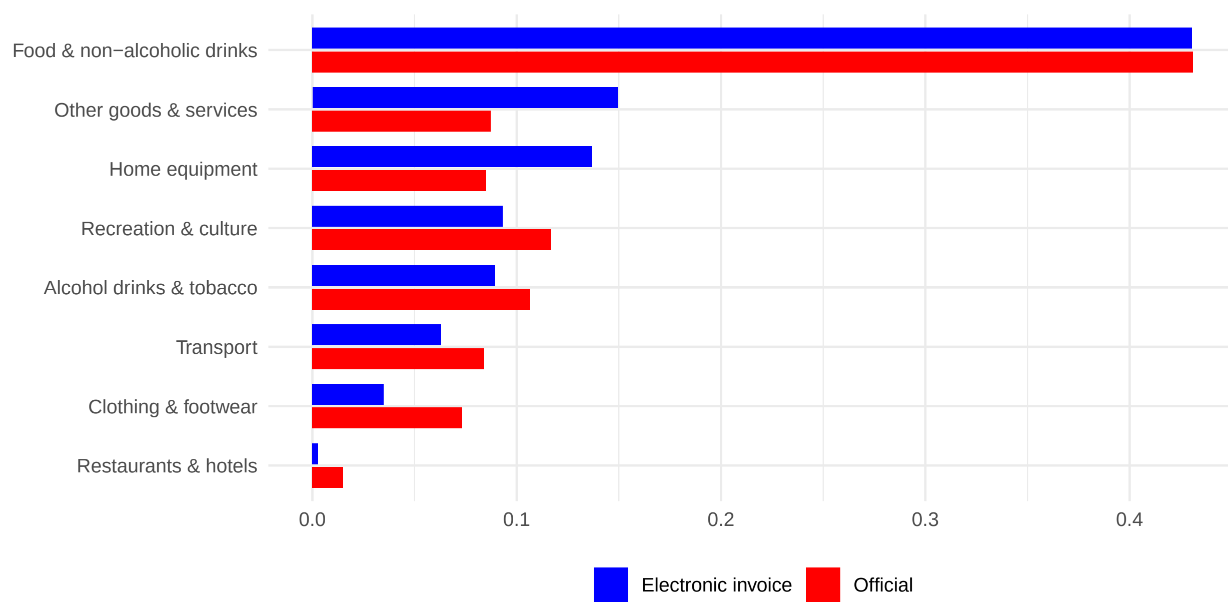
then, we use the set of varieties belonging to product i, denoted by N_{it}, to obtain a product-specific price index,

$$\Delta P_{i,t} = \left(\prod_{v \in N_{it}} \Delta P_{v(i),t} \right)^{\frac{1}{|N_{it}|}}.$$

This element is continuously updated using $P_{i,t} \equiv P_{i,t-1} \Delta P_{i,t}$, where $P_{i,0} = 100$. The official CPI is constructed using these price indices together with weights for each product.

We examine the implications of allowing the consumption weights to vary over time. To do so, we update the weights using changes in retailers' product-level purchasing shares, following the methodology described in Section 4.2. Figure 13 compares the 2018 division expenditure shares derived from this approach, based on electronic invoices, with the (normalized) official weights used by the INE to construct the CPI. The strong similarity supports the use of these documents as a reliable source for updating the consumption basket underlying the CPI, as well as for documenting the evolving pattern of consumption expenditures during the pandemic. The most salient difference comes from the "Restaurants & Hotels" categories, where our data cannot capture an equivalent relative level of expenditure. As discussed above, the nature of electronic invoices leads to an under-representation of business-to-consumer transactions, particularly within service-oriented divisions.

Figure 13: Consumption weights by division



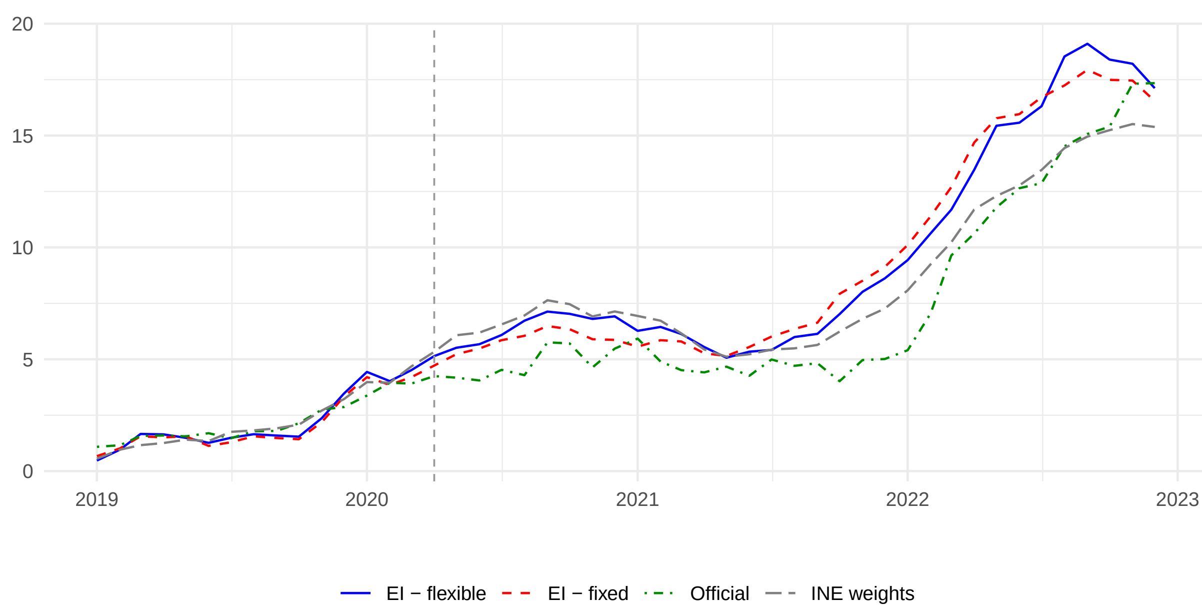
Notes: The figure compares expenditures shares for CPI product divisions in 2018. Blue bars are computed using purchases in electronic invoices and red bars are the official (normalized) weights.

4.3.2 Results

Figure 14 presents four annual inflation series considering the 157 products we are able to track. The blue line corresponds to electronic invoice price indices with fixed weights computed from electronic invoice expenditure shares. The dashed red line uses the same prices, but allows weights to vary over time. The dot-dashed green line replaces these with the official INE weights, while the dashed gray line corresponds to the official CPI measure, incorporating both INE prices and weights. The constructed CPI series closely tracks the official CPI, with a correlation of 0.97.

Regardless of the overall similarity across the series, notable differences emerge around the onset of the pandemic. Prior to this period (2019m1–2020m3), all indices display similar dynamics and magnitudes. Differences become evident during the first months of the COVID-19 outbreak and persist until August 2021. During this period, the constructed inflation series deviate from the official series by roughly 2–3 percentage points, with the gap originating between March and July 2020. Stay-at-home orders implemented during this period (see Section 4.2.2), likely affected the collection of prices by the INE. In particular, the share of imputed prices rose sharply from an average of about 9 percent of products to 43 percent in March 2020 (Ministerio de Economía, 2020b). Having access to a larger and more readily available set of observed prices is therefore an advantage in this context, in line with the arguments made in Diewert and Fox (2020) regarding the importance of continuous consumption and price collection. Accordingly, the administrative dataset used in this paper is less susceptible to distortions arising from missing price observations.

Figure 14: Annual inflation



Notes: The figure compares annual CPI measures using data from electronic invoices and from official statistics over 157 products. The solid blue line uses both varying weights and prices from EI (electronic invoice), the dashed red line uses both 2018 fixed weights and prices from EI, the dot-dashed green line uses official weights and prices from EI, and the dashed gray line is the official CPI inflation for the set of products tracked.

How the time-varying-weight price index differs from the fixed-weight price index is informative for our understanding of the welfare costs of shocks taking place during this period. Figure 14 shows that up to 2020m6, fixed and flexible weight indices track each other closely, but thereafter a gap emerges. Between 2020m7 and 2021m3, inflation measured using varying weights is higher than that implied by fixed weights, and also exceeds the official inflation rate. A similar difference is documented by Cavallo (2024) for both the US and Chile. This divergence highlights the consumption shift towards product categories whose prices were rapidly rising at the beginning of the pandemic, most notably food.[13]

Figure 14 shows that between 2021m5 and 2022m7 this gap reverses as inflation measured with fixed weights can exceed flexible weight inflation by roughly 1 percentage point. This period coincides with a phase of accelerating inflation, during which headline annual CPI inflation rose from 3.5 percent to 13.1 percent, almost coinciding with the peak of 14.1 percent recorded a month later. This reversal points to consumers substituting away from rapidly increasing prices, and underscoring the substitution bias present in measured headline inflation, which intensifies during episodes of high inflation.

Our results speak to biases in measured inflation and align with recent studies such as Christiano et al. (2025), who examine the implications of using true price indices in workhorse macroeconomic models, which have flexible weights, as opposed to using fixed weights used in official inflation measures. Their estimates of substitution bias are consistent with our findings for the same period. More broadly, our results connect to theoretical contributions by Feenstra (1994) and Redding and Weinstein (2019), who develop price aggregation frameworks that account for endogenous changes in product availability and consumer preferences. These mechanisms are particularly relevant in the context of the COVID-19 pandemic, during which some products became unavailable while new varieties simultaneously emerged.

5 Conclusions

Using the universe of electronic invoices in Chile—whose free-text product descriptions are highly idiosyncratic and domain-specific—we develop a methodology to construct a labeled corpus of standardized product categories and apply a machine learning algorithm to identify the products involved in firms' transactions. The resulting classification allows us to construct machinery and equipment outlays and goods consumption aggregate series that closely track their national accounts counterparts without ex-post revision.

[13] Appendix Figure B.5 shows how during these first months, Food prices were rising, while other like Transport even had negative inflation rates.

Our work illustrates how tax information combined with text analytics can help document the impact of shocks affecting consumption, allowing us to study how expenditure patterns shift across product categories and how these shifts feed into CPI measurement. During the COVID-19 pandemic, durable expenditure accounted for much of the recovery phase, while stay-at-home orders compressed spending broadly across non-durable categories. We also show that unanticipated pension withdrawals generated sizable and persistent increases in consumption, with substantial heterogeneity across categories. Finally, we show that flexible-weight and fixed-weight consumer price indices diverged during specific episodes—notably at the onset of the pandemic and during the subsequent inflation surge.

Beyond the applications developed here, the results from the classification algorithm can be applied to other economic research questions, including the measurement of price rigidities at the product level, the estimation of substitution bias in inflation measures, the analysis of market concentration in narrowly defined markets, and production function estimation. Several of these applications are already being pursued by researchers at the Central Bank of Chile.

Our results carry a broader message for the production of official statistics. Naturally occurring data, like electronic invoice systems, contain sufficient information to construct economic aggregates that closely track their national accounts counterparts in real time. Realizing this potential requires formal collaboration between tax agencies and national statistical offices that could expand the set of high-frequency indicators available to policymakers. Our exercise offers a proof of concept for what such collaboration can deliver.

References

Alexander, Diane and Ezra Karger, "Do Stay-at-Home Orders Cause People to Stay at Home? Effects of Stay-at-Home Orders on Consumer Behavior," *The Review of Economics and Statistics*, 09 2021, pp. 1–25.

Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomás Mikolov, "Enriching Word Vectors with Subword Information," *CoRR*, 2016, *abs/1607.04606.*

Bottou, Léon, "Large-Scale Machine Learning with Stochastic Gradient Descent," in "Proceedings of COMPSTAT 2010" Springer 2010, pp. 177–186.

Buda, G., V. M. Carvalho, S. Hansen, J. V. R. Mora, Ã . Ortiz, and T. Rodrigo, "National Accounts in a World of Naturally Occurring Data: A Proof of Concept for Consumption," Cambridge Working Papers in Economics 2244, Faculty of Economics, University of Cambridge July 2022.

Cavallo, Alberto, "Inflation with Covid Consumption Baskets," *IMF Economic Review*, June 2024, *72* (2), 902–917.

Chetty, Raj, John N Friedman, Michael Stepner, and Opportunity Insights Team, "The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data*," *The Quarterly Journal of Economics*, 10 2023, *139* (2), 829–889.

Christiano, Lawrence, Martin S. Eichenbaum, and Benjamin K. Johannsen, "Substitution Bias and Fixed-Weight Price Indices in Time-Dependent Pricing Models," *American Economic Review: Insights*, Apr 2025.

Cox, Natalie, Peter Ganong, Pascal Noel, Joseph Vavra, Arlene Wong, Diana Farrell, Fiona Greig, and Erica Deadman, "Initial Impacts of the Pandemic on Consumer Behavior: Evidence from Linked Income, Spending, and Savings Data," *Brookings Papers on Economic Activity*, 2020, pp. 35–69.

Diewert, W. Erwin and Kevin J Fox, "Measuring Real Consumption and CPI Bias under Lockdown Conditions," Working Paper 27144, National Bureau of Economic Research May 2020.

Ehrlich, Gabriel, John Haltiwanger, Ron Jarmin, David Johnson, and Matthew D Shapiro, "Re-engineering Key National Economic Indicators," Working Paper 26116, National Bureau of Economic Research July 2019.

Feenstra, Robert C., "New Product Varieties and the Measurement of International Prices," *The American Economic Review*, 1994, *84* (1), 157–177.

Fowler, Glenn, Landon Curt Noll, Kiem-Phong Vo, Donald E. Eastlake 3rd, and Tony Hansen, "The FNV Non-Cryptographic Hash Algorithm," Internet-Draft draft-eastlake-fnv-21, IETF January 2024. Work in Progress.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, "Text as Data," *Journal of Economic Literature*, September 2019, *57* (3), 535–574.

Goolsbee, Austan and Chad Syverson, "Fear, lockdown, and diversion: Comparing drivers of pandemic economic decline 2020," *Journal of Public Economics*, 2021, *193*, 104311.

Hadzi-Vaskov, Metodij, Emiliano Luttini, and Mr Luca A Ricci, "Micro-Evidence on the Consumption Impact of Income-Support Policies During COVID-19," Technical Report, International Monetary Fund 2025.

Hansen, Stephen and Elliott Ash, "Text Algorithms in Economics," CEPR Discussion Papers 18125, C.E.P.R. Discussion Papers April 2023.

Hoberg, Gerard and Gordon Phillips, "Text-Based Network Industries and Endogenous Product Differentiation," *Journal of Political Economy*, 2016, *124* (5), 1423–1465.

Joulin, Armand, Edouard Grave, Piotr Bojanowski, and Tomás Mikolov, "Bag of Tricks for Efficient Text Classification," *CoRR*, 2016, *abs/1607.01759.*

Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze, *Introduction to Information Retrieval*, USA: Cambridge University Press, 2008.

Ministerio de Economía, "Manual Metodologico del IPC," Working Paper, Instituto Nacional de Estadísticas 2020.

—, "Separata Técnica IPC," Working Paper, Instituto Nacional de Estadísticas May 2020.

Niu, Feng, Benjamin Recht, Christopher Ré, and Stephen J. Wright, "HOGWILD!: A Lock-Free Approach to Parallelizing Stochastic Gradient Descent," Technical Report, University of Wisconsin–Madison 2011. arXiv:1106.5730.

Redding, Stephen J and David E Weinstein, "Measuring Aggregate Price Indices with Taste Shocks: Theory and Evidence for CES Preferences*," *The Quarterly Journal of Economics*, 09 2019, *135* (1), 503–560.

Simon, Herbert A., "Designing organizations for an information rich world," in Martin Greenberger, ed., *Computers, communications, and the public interest*, 1971, pp. 37–72.

Tauber, Kristen and Willem Van Zandweghe, "Why Has Durable Goods Spending Been So Strong during the COVID-19 Pandemic?," *Economic Commentary*, July 2021, *2021* (16), 1–6.

Wang, Sida and Christopher D. Manning, "Baselines and bigrams: simple, good sentiment and topic classification," in "Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers - Volume 2" ACL '12 Association for Computational Linguistics USA 2012, p. 90–94.

Weinberger, Kilian, Anirban Dasgupta, John Langford, Alex Smola, and Josh Attenberg, "Feature hashing for large scale multitask learning," in "Proceedings of the 26th Annual International Conference on Machine Learning" ICML '09 Association for Computing Machinery New York, NY, USA 2009, p. 1113–1120.

Appendices

A Computational implementation of FastText

The FastText library was obtained from the official GitHub repository and compiled from source. We use the *autotune* option to perform a time-bounded hyperparameter search, described below, this search ran for approximately 12 hours and selected the configuration that maximized validation performance to train the final model. The final fitted model is then applied to unlabeled electronic invoice entries out of sample. The implementation was executed on a system running Windows Server 2022, equipped with an Intel Xeon Gold 6126 (2×2.60 GHz) processor and 1.5 TB of RAM. The implementation used Python 3.11.x, and the `scikit-learn` package was employed to compute all quality metrics.

A.1 Autotune: Time-bounded hyperparameter search

FastText's *autotune* performs a sequential stochastic search over the hyperparameter space under time-bounded hyperparameter search Q, iteratively proposing, training, and evaluating candidate configurations while concentrating probability mass around the best-performing regions. Let $\mathcal{D}_{\text{tr}}$ be the training set and $\mathcal{D}_{\text{val}}$ the validation set. Hyperparameters live in a mixed domain

$$\mathcal{H} \subseteq \mathbb{R}^{d_c} \times \mathbb{Z}^{d_i} \times \mathcal{C}^{d_d},$$

covering learning rate η, embedding dimension d, number of epochs E, word n-grams n_g, subword range $[n_{\min}, n_{\max}]$, bucket size B, minimum word count t, etc. For $h \in \mathcal{H}$, define the trained model

$$M(h) = \text{train}(\mathcal{D}_{\text{tr}}, h).$$

The validation score J is the micro-averaged F_1. For each validation instance i with ground-truth label $y_i \in \mathcal{K}$ and predicted label $\hat{y}_i$, we compute

$$P = \frac{\sum_i \mathbf{1}\{\hat{y}_i = y_i\}}{\sum_i 1}, \qquad R = \frac{\sum_i \mathbf{1}\{\hat{y}_i = y_i\}}{\sum_i 1}, \qquad F_1 = \frac{2PR}{P + R}.$$

The optimization problem is

$$\max_{h \in \mathcal{H}} f(h) \equiv J(M(h), \mathcal{D}_{\text{val}}) \quad \text{s.t.} \quad \mathsf{T} \leq Q, \tag{5}$$

where T denotes the elapsed wall-clock time.

Optimization protocol. Autotune runs a sequential stochastic search under the time-bound Q. Let $p_t(h)$ denote a proposal distribution over $\mathcal{H}$ at iteration t, and let

$$h_t^\star \in \arg \max_{h \in \{h_1, \ldots, h_t\}} f(h) \tag{6}$$

be the current incumbent. While $\mathsf{T} < Q$:

1. *Propose:* sample $h_{t+1} \sim p_t$.
2. *Train:* fit $M_{t+1} \leftarrow \text{train}(\mathcal{D}_{\text{tr}}, h_{t+1})$.

3. *Evaluate:* compute $s_{t+1} \leftarrow f(h_{t+1}) = J(M_{t+1}, \mathcal{D}_{\text{val}})$.
4. *Select:* set $h^{\star}_{t+1} \leftarrow \arg\max\{f(h^{\star}_t), s_{t+1}\}$.
5. *Adapt:* update p_{t+1} to concentrate probability mass near promising neighborhoods of $h^{\star}_{t+1}$.

When the budget elapses, autotune finishes the current training run and returns a model retrained on $\mathcal{D}_{\text{tr}}$ with the best hyperparameters found. The table A.1 shows the optimal hyperparameter configurations by classifier, which were used to train the final model and generate predictions for all electronic invoice documents.

Table A.1: Set of hyperparameters configurations by classifier

Parameter	CPC	UPC	COICOP
Epoch E	81	100	100
Learning rate η	0.05	0.05	0.05
Dimension d	871	100	233
Word n-grams	2	2	4
Bucket B	9,590,022	2,000,000	6,132,768
n_{min}	0	0	3
n_{max}	0	0	6

Notes: Each classifier was trained independently using the fastText supervised learning framework. Hyperparameters were tuned via FastText's *autotune*. E denotes the number of training epochs; η the learning rate; d the embedding dimension; B the number of buckets for the hashing trick; and n_{min}, n_{max} the minimum and maximum character n-gram lengths, respectively. A value of zero for n_{min} and n_{max} indicates that character n-grams were not used.

B Additional Tables and Figures

This appendix presents supplementary figures and tables that complement the main results discussed in Section 4. Figure B.1 reports the quarterly growth rates of durable and non-durable goods consumption, comparing electronic invoice aggregates with official national accounts data. Tables B.1 through B.3 provide the full set of event-study estimates for the effects of stay-at-home orders on consumption, broken down by product division for the aggregate, durable, and non-durable cases respectively. Figure B.2 documents the monthly evolution of pension withdrawals across the three withdrawal episodes. Table B.4 reports the consumption response to the first pension withdrawal episode disaggregated across broad product categories, complementing the aggregate results shown in Figure 12 in the main text.

Figure B.4 presents the corresponding estimates for the second and third pension withdrawal episodes. Unlike the first one, anticipation was more pronounced, so the results must be interpreted with caution: to the extent that retailers stocked up ahead of the policy, our estimates likely understate the true consumption response. For the second pension withdrawal, the qualitative pattern nonetheless mirrors that of the first withdrawal. The third withdrawal episode, where anticipation was even more pronounced, tells a different story. As the policy became anticipated, many retailers appear to have made purchases prior to implementation. Consequently, municipalities more exposed to the withdrawal saw retail purchases slow down relative to less exposed ones, producing estimates opposite in sign to those found in the first two episodes.

Figure B.5 disaggregates the annual inflation comparison from Section 4.3 across four broad product divisions—Food, Clothing and General Merchandise, Transport, and Entertainment—allowing for a more granular assessment of where fixed-weight and flexible-weight price indices diverge over the pandemic period.

Table B.1: Lockdown and divisions of consumption goods

s	All	Foods	Clothing and general merchandise	Transport	Entertainment
	(a) Enter				
-4	0.007	0.001	-0.007	-0.031	-0.011
	(0.009)	(0.011)	(0.015)	(0.032)	(0.024)
-3	0.002	-0.005	0.008	0.024	0.021
	(0.008)	(0.009)	(0.016)	(0.028)	(0.019)
-2	0.007	-0.005	-0.020	0.016	0.014
	(0.007)	(0.009)	(0.016)	(0.028)	(0.021)
0	-0.077***	-0.078***	0.047**	-0.087***	-0.096***
	(0.008)	(0.009)	(0.016)	(0.026)	(0.023)
1	-0.109***	-0.121***	0.069***	-0.125***	-0.105***
	(0.008)	(0.010)	(0.017)	(0.032)	(0.020)
2	-0.100***	-0.101***	0.053**	-0.097**	-0.096***
	(0.009)	(0.011)	(0.017)	(0.031)	(0.024)
3	-0.085***	-0.084***	0.077***	-0.062.	-0.110***
	(0.009)	(0.011)	(0.018)	(0.033)	(0.023)
4	-0.080***	-0.075***	0.098***	-0.056	-0.106***
	(0.011)	(0.013)	(0.020)	(0.035)	(0.026)
	(b) Exit				
-4	0.000	0.009	0.033*	0.028	-0.061*
	(0.009)	(0.010)	(0.015)	(0.032)	(0.025)
-3	-0.003	0.007	0.007	-0.017	-0.031
	(0.009)	(0.010)	(0.014)	(0.027)	(0.025)
-2	-0.008	0.010	-0.004	-0.027	-0.042.
	(0.009)	(0.009)	(0.014)	(0.029)	(0.023)
0	0.046***	0.052***	-0.108***	0.034	0.043.
	(0.009)	(0.008)	(0.015)	(0.027)	(0.026)
1	0.056***	0.062***	-0.118***	0.068*	0.029
	(0.011)	(0.012)	(0.015)	(0.026)	(0.024)
2	0.060***	0.076***	-0.140***	0.088*	0.015
	(0.013)	(0.015)	(0.015)	(0.035)	(0.026)
3	0.065***	0.077***	-0.126***	0.076**	0.056*
	(0.011)	(0.013)	(0.016)	(0.029)	(0.025)
4	0.068***	0.067***	-0.098***	0.074*	0.074**
	(0.011)	(0.013)	(0.019)	(0.033)	(0.027)

Notes: This table reports event-study estimates of the effect of stay-at-home orders on consumption by product division. Each column corresponds to a separate regression of the Equation 3. Event time s is defined relative to the week in which a municipality enters (Panel a) or exits (Panel b) a lockdown, with $s = 0$ denoting the week of policy implementation or removal. The omitted category is the week immediately preceding the event ($s = -1$), so all coefficients are interpreted relative to that baseline. All regressions include municipality and week fixed effects. Standard errors, reported in parentheses, are clustered at the municipality level. Coefficients are reported in log points and can be interpreted approximately as percent changes. Statistical significance is denoted by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table B.2: Lockdown and divisions of consumption goods: durable

s	All	Clothing and general merchandise	Transport	Entertainment
	(a) Enter			
-4	-0.012	0.026	-0.043	0.047
	(0.029)	(0.030)	(0.094)	(0.042)
-3	-0.006	-0.031	0.097	0.077*
	(0.027)	(0.028)	(0.083)	(0.035)
-2	-0.021	-0.036	0.003	0.072*
	(0.029)	(0.028)	(0.083)	(0.036)
0	-0.158***	-0.157***	-0.216*	-0.104**
	(0.028)	(0.027)	(0.087)	(0.039)
1	-0.103**	-0.153***	-0.142	-0.076*
	(0.031)	(0.028)	(0.097)	(0.038)
2	-0.146***	-0.174***	-0.150	-0.094*
	(0.031)	(0.029)	(0.100)	(0.044)
3	-0.102**	-0.168***	-0.144	-0.030
	(0.033)	(0.033)	(0.102)	(0.044)
4	-0.107**	-0.154***	-0.087	-0.077.
	(0.034)	(0.032)	(0.106)	(0.044)
	(b) Exit			
-4	0.009	-0.025	0.133	-0.021
	(0.035)	(0.029)	(0.109)	(0.043)
-3	0.008	0.005	0.044	-0.053
	(0.030)	(0.029)	(0.092)	(0.041)
-2	-0.034	-0.020	-0.014	0.006
	(0.033)	(0.028)	(0.088)	(0.040)
0	0.059.	0.052.	0.132	0.079*
	(0.031)	(0.027)	(0.081)	(0.040)
1	0.050	0.050	0.163	0.035
	(0.035)	(0.032)	(0.099)	(0.041)
2	0.036	0.060*	0.113	0.033
	(0.034)	(0.029)	(0.093)	(0.041)
3	0.070*	0.080**	0.142	0.092*
	(0.035)	(0.031)	(0.100)	(0.045)
4	0.086*	0.084*	0.202.	0.120**
	(0.039)	(0.034)	(0.111)	(0.045)

Notes: This table reports event-study estimates of the effect of stay-at-home orders on consumption by product division. Each column corresponds to a separate regression of the Equation 3. Event time s is defined relative to the week in which a municipality enters (Panel a) or exits (Panel b) a lockdown, with $s = 0$ denoting the week of policy implementation or removal. The omitted category is the week immediately preceding the event ($s = -1$), so all coefficients are interpreted relative to that baseline. All regressions include municipality and week fixed effects. Standard errors, reported in parentheses, are clustered at the municipality level. Coefficients are reported in log points and can be interpreted approximately as percent changes. Statistical significance is denoted by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table B.3: Lockdown and divisions of consumption goods: non-durable

s	All	Foods	Clothing and general merchandise	Transport	Entertainment
	(a) Enter				
-4	0.009	0.001	0.040*	-0.016	-0.017
	(0.009)	(0.011)	(0.018)	(0.027)	(0.024)
-3	0.001	-0.005	0.020	0.012	0.007
	(0.008)	(0.009)	(0.016)	(0.023)	(0.022)
-2	0.007	-0.005	0.010	0.021	-0.006
	(0.007)	(0.009)	(0.013)	(0.025)	(0.022)
0	-0.071***	-0.078***	-0.095***	-0.060*	-0.084***
	(0.008)	(0.009)	(0.017)	(0.024)	(0.025)
1	-0.115***	-0.121***	-0.114***	-0.117***	-0.128***
	(0.008)	(0.010)	(0.016)	(0.024)	(0.022)
2	-0.101***	-0.101***	-0.126***	-0.084**	-0.106***
	(0.009)	(0.011)	(0.016)	(0.028)	(0.026)
3	-0.089***	-0.084***	-0.121***	-0.058*	-0.136***
	(0.009)	(0.011)	(0.017)	(0.026)	(0.024)
4	-0.078***	-0.075***	-0.084***	-0.064*	-0.101***
	(0.011)	(0.013)	(0.021)	(0.030)	(0.028)
	(b) Exit				
-4	-0.002	0.009	-0.004	0.005	-0.061*
	(0.009)	(0.010)	(0.017)	(0.026)	(0.026)
-3	-0.003	0.007	0.008	-0.029	-0.019
	(0.009)	(0.010)	(0.017)	(0.024)	(0.025)
-2	-0.006	0.010	-0.018	-0.027	-0.057*
	(0.008)	(0.009)	(0.016)	(0.027)	(0.023)
0	0.046***	0.052***	0.052**	0.024	0.033
	(0.007)	(0.008)	(0.017)	(0.024)	(0.024)
1	0.060***	0.062***	0.078***	0.059*	0.028
	(0.010)	(0.012)	(0.017)	(0.024)	(0.023)
2	0.063***	0.076***	0.056**	0.061*	0.025
	(0.013)	(0.015)	(0.017)	(0.029)	(0.025)
3	0.068***	0.077***	0.086***	0.058*	0.054*
	(0.011)	(0.013)	(0.018)	(0.026)	(0.025)
4	0.069***	0.067***	0.113***	0.044	0.071*
	(0.011)	(0.013)	(0.020)	(0.031)	(0.029)

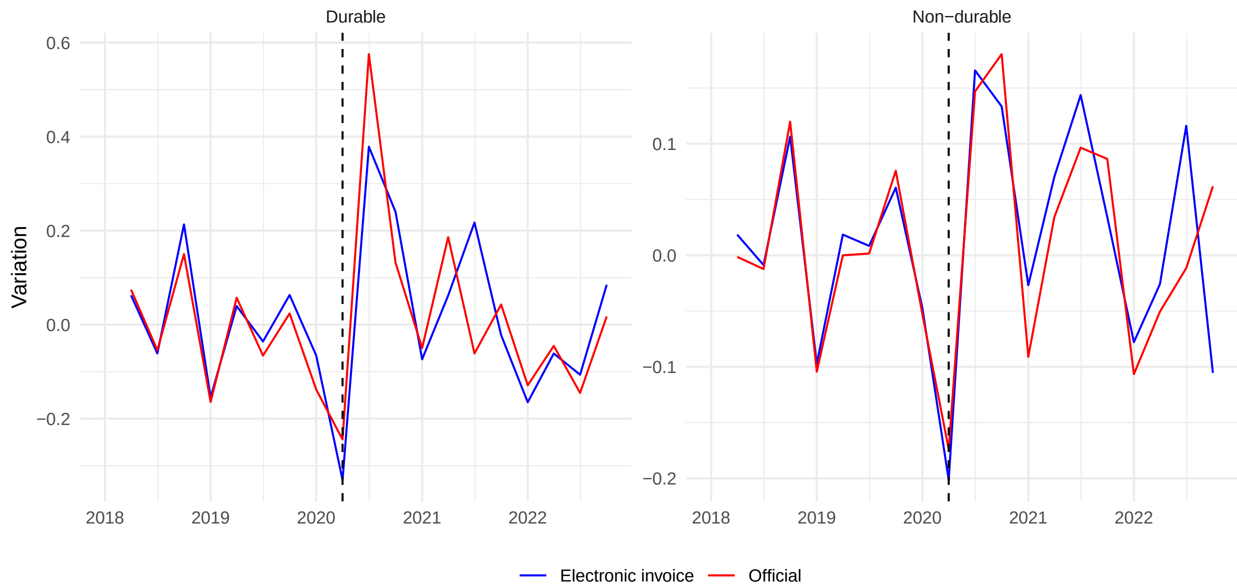
Notes: This table reports event-study estimates of the effect of stay-at-home orders on consumption by product division. Each column corresponds to a separate regression of the Equation 3. Event time s is defined relative to the week in which a municipality enters (Panel a) or exits (Panel b) a lockdown, with $s = 0$ denoting the week of policy implementation or removal. The omitted category is the week immediately preceding the event ($s = -1$), so all coefficients are interpreted relative to that baseline. All regressions include municipality and week fixed effects. Standard errors, reported in parentheses, are clustered at the municipality level. Coefficients are reported in log points and can be interpreted approximately as percent changes. Statistical significance is denoted by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table B.4: Withdrawals and divisions of consumption goods.

s	All	Foods	Clothing and general merchandise	Transport	Entertainment
	First withdrawal				
0	0.035	0.061.	-0.016	0.034	-0.125
	(0.032)	(0.033)	(0.057)	(0.104)	(0.082)
1	0.103*	0.110***	0.054	0.017	-0.008
	(0.041)	(0.032)	(0.054)	(0.107)	(0.075)
2	0.113.	0.069	0.044	-0.123	-0.007
	(0.060)	(0.054)	(0.069)	(0.113)	(0.086)

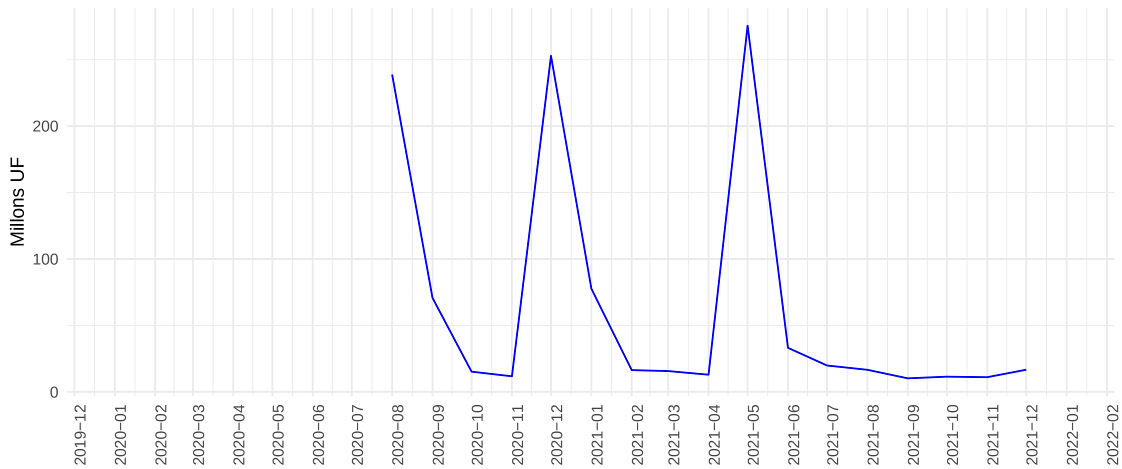
Notes: Variation difference of log consumption after withdrawals by division product. This table shows $\sum_{s \in S} \alpha_s$ from regression defined in Equation 4. Standard errors are clustered at the municipality level.

Figure B.1: Durable and non-durable goods consumption growth: Official statistics and electronic invoice aggregate



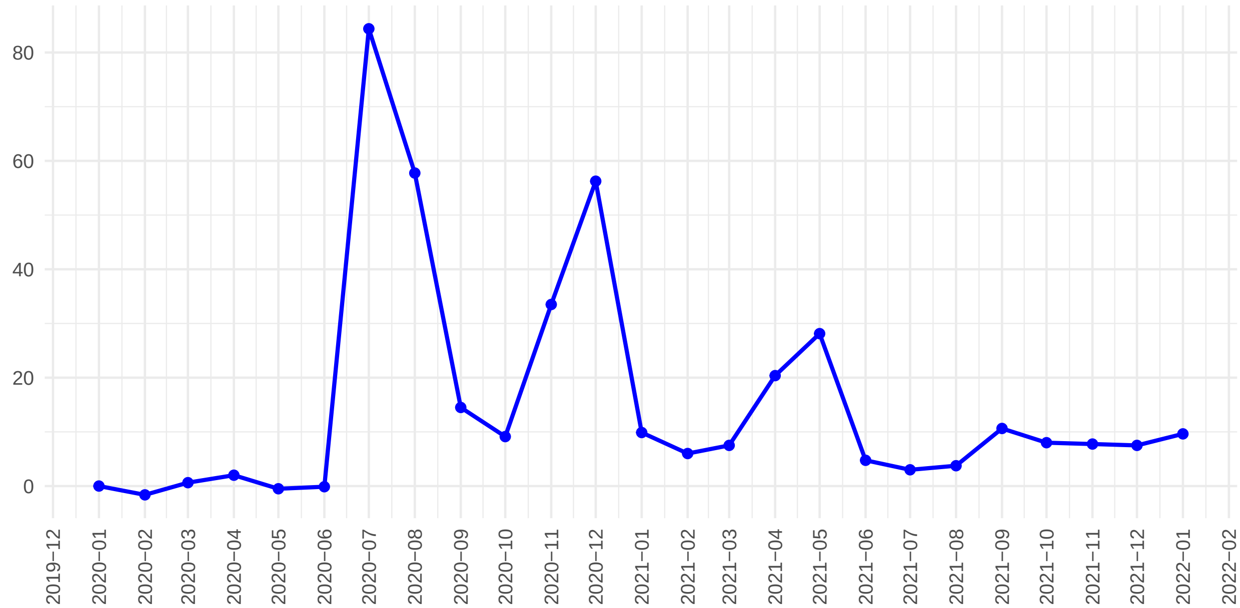
Notes: This figure reports quarterly variations in consumption for durable and nondurable goods, comparing retail-sector purchases from electronic invoices with the corresponding official national accounts data.

Figure B.2: Evolution of withdrawals by month, expressed in millions of UF.



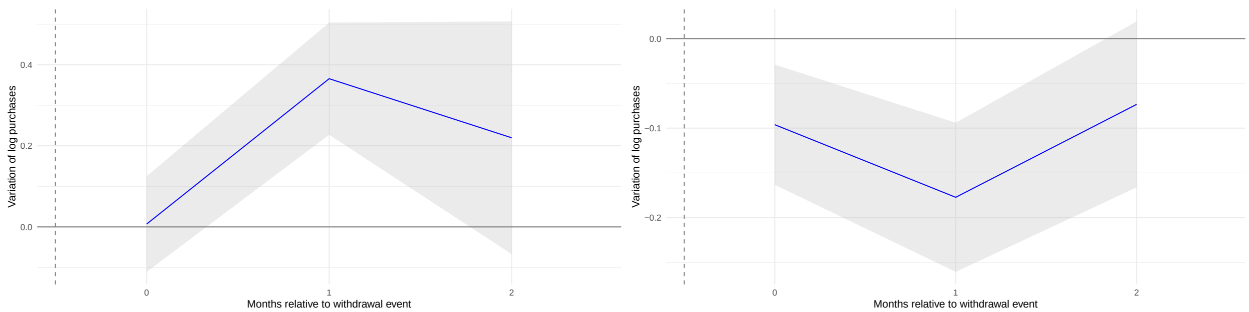
Notes: This figure reports time series aggregates of pensions withdrawals.

Figure B.3: Google Trends indices for pension withdrawal-related keywords



Notes: The figure displays Google Trends indices for keywords related to pension withdrawals. The series corresponds to the simple average across all keywords, covering the period from January 2020 to January 2021.

Figure B.4: Subsequent pension withdrawal and Retail and Restaurant and Hotels consumption

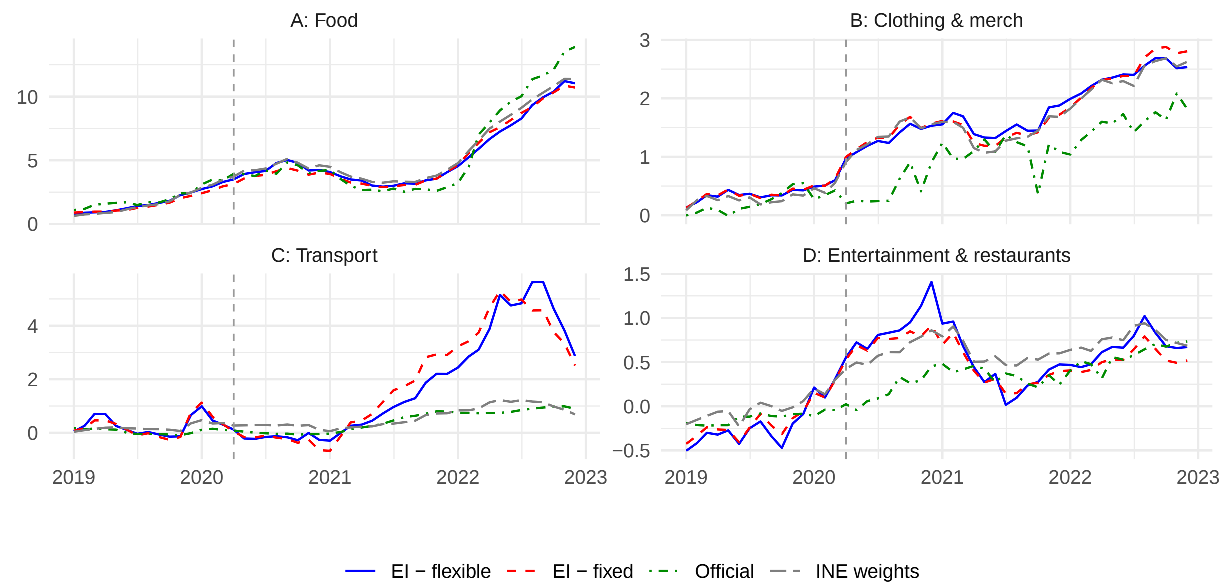


(a) Second pension withdrawal

(b) Third pension withdrawal

Notes: Variation of log consumption after second withdrawal. This figure shows $\sum_{s \in S} \alpha_s$ from regression defined in Equation 4. The gray area denotes the 90% confidence interval. Standard errors are clustered at the municipality level.

Figure B.5: Annual Inflation by macro division



Notes: The figure compares annual CPI measures using data from electronic invoices and from official statistics over 157 products, separated by four macro divisions. Each solid blue line uses both varying weights and prices from EI (electronic invoice), the dashed red line uses both 2018 fixed weights and prices from EI, the dot-dashed green line uses official weights and prices from EI, and the dashed gray line is the official CPI inflation for the set of products tracked in that macro division.

C Data details

C.1 Other datasets

Unemployment Insurance Administration. In order to compute the average income at the municipality level and pension withdrawal amounts, we rely on the information firms provide to the Unemployment Insurance Administration. This is because the amount deposited in a worker's account depends on her salary and other social contributions. In addition, we have the worker's municipality of residence together with a firm identifier.

Ministry of Science. We use municipality-level quarantine information from Chile's COVID-19 Data Repository. We construct a quarantine-events panel that records the timing (start and end dates) and the corresponding quarantine status/phase for each municipality. The public release corresponds to the repository's state as of August 31, 2023, and it has not been updated since the end of the health emergency.

C.2 Price indices

For price monitoring, we seek to have a dataset that is similar to the varieties monitored by the INE. For this, we keep information on products whose price is within the INE price range, which is between $0.9 \times P_i^{min}$ and $1.2 \times P_i^{max}$. In addition, our dataset is filtered to consider exclusively the economic sectors of the selling firms where we are interested in following prices (mainly retail trade). We remove transaction within the same firm, require that the text descriptions are in the sample for at least 30 consecutive periods, and that the annual variation of the index for each product have a correlation greater than 0.2 with the official INE series.

For the calculation of the weights, the most relevant filter is that we exclusively consider purchases from the Retail Trade and Hotels and Restaurants sector. Additionally, we removed rentals, cars and motorcycles because they ostensibly unbalanced the weights, and we removed gasoline. Finally of the 12 official product divisions, we consider 8. We drop Health, Housing Services, Education and Communications divisions.

Disclaimers:

The views expressed are those of the authors and do not necessarily represent the views of the Central Bank of Chile (CBC) or its board members.

This study was developed within the scope of the research agenda conducted by the CBC in economic and financial affairs of its competence. The CBC has access to anonymized information from various public and private entities, by virtue of collaboration agreements signed with these institutions.

To secure the privacy of workers and firms, the CBC mandates that the development, extraction and publication of the results should not allow the identification, directly or indirectly, of natural or legal persons. Officials of the Central Bank of Chile processed the disaggregated data. All the analysis was implemented by the authors and did not involve nor compromise the SII, Aduanas and AFC.

The information contained in the databases of the SII is of a tax nature originating in self-declarations of taxpayers presented to the Service; therefore, the veracity of the data is not the responsibility of the Service.

Documentos de Trabajo **Banco Central de Chile**	**Working Papers** **Central Bank of Chile**
NÚMEROS ANTERIORES	PAST ISSUES
La serie de Documentos de Trabajo en versión PDF puede obtenerse gratis en la dirección electrónica: www.bcentral.cl/esp/estpub/estudios/dtbc. Existe la posibilidad de solicitar una copia impresa con un costo de Ch$500 si es dentro de Chile y US$12 si es fuera de Chile. Las solicitudes se pueden hacer por fax: +56 2 26702231 o a través del correo electrónico: bcch@bcentral.cl.	Working Papers in PDF format can be downloaded free of charge from: www.bcentral.cl/eng/stdpub/studies/workingpaper. Printed versions can be ordered individually for US$12 per copy (for order inside Chile the charge is Ch$500.) Orders can be placed by fax: +56 2 26702231 or by email: bcch@bcentral.cl.

DTBC - 1088
From Invoices to Economic Monitoring Using Machine Learning
Emiliano Luttini, Matías Pizarro, Dagoberto Quevedo, Marco Rojas

DTBC - 1087
Inflation Illiteracy: What Does Microdata Reveal About Chilean Consumers?
Carlos Medel

DTBC - 1086
Structural Synchronicity: A Granular Analysis of China Shocks and Global Inflation
Benjamín Álvarez, Gabriel Zelpo

DTBC – 857 (Updated)
Immigration in Emerging Countries: A Macroeconomic Perspective Agustín H. Arias, Juan Guerra-Salas

DTBC – 1085
Anatomy of the investment network in a commodity-dependent economy
Roberto Gillmore, Rodrigo Heresi, David Kohn, Dagoberto Quevedo, Nicolás Rivera

DTBC – 1084
Inversión Agregada en Chile: Factores Globales, Domésticos, y Mineros
Arturo Claro, Camilo Levenier, Carlos A. Medel

DTBC – 1083
Regulating Vertical Markets through Delegation
Andrea Canales, Nicolás Figueroa, Hugo E. Silva

DTBC – 1082
Sovereign Wealth Funds and Optimal Foreign Reserves
Miguel Acosta-Henao, Humberto Martínez, Carlos Rondón-Moreno

DTBC – 1081
Conditional Bayesian Quantile Regressions for Forecasting the GDP Growth Distribution in a Small Open Economy
Jorge Fornero, Carlos Molina

DTBC – 1080
Fundamental Drivers of Financial Conditions
Elías Albagli, Guillermo Carlomagno, Javier Ledezma, María Teresa Reszczynski

DTBC – 1079
Hospital Choice, C-sections, and long-term maternal health
Ramiro de Elejalde, Eugenio Giolito

DTBC – 1078
Zero Energy Day: How Nationwide Blackouts Affect the Economy
Luis Gonzales, Koichiro Ito, Mar Reguant

DTBC – 1015 (Updated)
Fiscal Consolidations in Commodity-Exporting Countries: A DSGE Perspective
Manuel González-Astudillo , Juan Guerra-Salas , Avi Lipton

DTBC – 1077
Un sistema de proyección de demanda por efectivo en Chile: Actualización y propuesta
Nicolás Leiva, Carlos A. Medel

DTBC – 1076
Inflation Heterogeneity and Differential Effects of Monetary and Oil Price Shocks
Felipe Martínez

DTBC – 1075
Consumption Insurance over the Life Cycle
Enzo Cerletti, Tomás Cortés

DTBC – 1074
Precios de viviendas en Chile: Herramientas para Evaluar Desalineamientos y sus Efectos sobre la banca
Serio Díaz V., Mauricio Salas G., Francisco Vásquez L.

banco central
Chile

DOCUMENTOS DE TRABAJO Agosto 2026