Invoices rather than surveys: Using ML to monitor the economy*

Pablo Acevedo
Central Bank of Chile

Emiliano Luttini
World Bank

Matías Pizarro
Central Bank of Chile

Dagoberto Quevedo
Central Bank of Chile

Marco Rojas
Central Bank of Chile

March 26, 2026

Abstract

Economic research has largely relied on survey data and national accounts for empirical analysis. Over the last decade, large databases and text data have become increasingly prevalent as new sources. In this paper, we bring both together. We classify the universe of firm-to-firm transactions in Chile into standardized product categories using a supervised multi-class text classification algorithm. We develop three applications. First, we construct real-time aggregates for machinery and equipment outlays and goods consumption that closely track national accounts counterparts without ex-post revision. Second, we study consumption patterns during the COVID-19 pandemic, documenting a reallocation toward durable goods during the recovery phase, and examine the role of stay-at-home orders and government policies allowing households to withdraw funds from their pension savings in driving these shifts. Third, we construct a flexible-weight consumer price index and find that deviations from fixed-weight measures are not a first-order concern over the pandemic period.

Keywords: National accounts, Naturally occurring data, Nowcasting, Natural Language Processing, Machine Learning

JEL Codes: C55, C82, E01, E31

*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, and "3rd Iriving Fisher Committee Workshop on Data Science in Central Banking" at the Bank of Italy, for thoughtful comments and discussions. We also thank to Michele Lenza for his detail comments. All errors are our own. Disclaimers are found at the end of this document. Email: pacevedo@bcentral.cl, 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 statistics 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, the application of machine learning 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 the universe of products that firms exchange daily and develop three applications. 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 standardized product categories useful for economic analysis. Our approach yields two aggregates that closely track their national accounts counterparts: machinery and equipment outlays and goods consumption, both available within days of the end of each month and free of ex-post revisions. Beyond nowcasting, the granularity of our classification allows us to study how consumption patterns shifted during the COVID-19 pandemic, documenting a reallocation of expenditure across durable and non-durable goods in response to stay-at-home orders and income support programs. Finally, combining our classification with price data from the invoices, we construct a flexible-weight

consumer price index and show that, while fixed and flexible weight measures diverge during specific episodes—most 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 learn about 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 an score that balances revenue, frequency, and monthly occurrence across economic sectors. We train FASTTEXT (Joulin et al., 2016) using this corpus, mapping descriptions into three complementary classification schemes—the Central Product Classification (CPC), the Unique Product Classification (UCP), and the Classification of Individual Consumption by Purpose (COICOP)—each suited to different economic applications. We validate the model using cross-validated precision, recall, and F1 scores, obtaining weighted F1 scores between 0.915 and 0.954 across classifiers. 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.

We construct aggregate series that match the evolution of machinery and equipment outlays and goods consumption series from the national accounts using the products' information that results from applying our classification algorithm. We leverage our classification algorithm using the mapping between the items the firm buys and its associated product category identified as gross fixed capital formation or consumption by the Broad Economic Categories classification used in the national accounts. We approximate final goods consumption through the retail sector purchases, assuming that once products reach retailers, they have a fast turnover and reach final consumers. When comparing the official series to our constructed ones, our series captures the official statistics' trend, seasonal pattern, and cyclical upswing. We also show that because of legal time constraints to report the electronic tax document, the production of these indices can be available approximately ten days after the month ends with virtually no posterior revisions.

A second application we study is documenting households' consumption patterns using electronic tax documents; we zero in on what happened during the COVID-19 pandemic. Statistics agencies can inform about households' consumption patterns by

conducting surveys. Our classification can provide information regarding consumption patterns as they closely correlate with the official statistics. Using data from 2018 to 2022, we document a marked reallocation of expenditure beginning in the second half of 2020: households shifted spending toward durable goods at the expense of non-durables, a pattern also documented in other countries. This shift was driven primarily by gains in Entertainment and Clothing and General Merchandise durables, while non-durable Transport expenditure contracted sharply, emerging as the single largest negative contributor. The idiosyncratic characteristics of the period can rationalize this behavior: Agents' interest in protecting themselves from the virus by limiting their exposure to social activities increases their spending on recreational goods to use in their households. In addition, we zoom in on whether stay-at-home orders or the income support programs that happened over the period can account for the observed shifts in consumption.

To cope with the spread of the virus, the Chilean government established stay-at-home orders during 2020 and 2021. We implement an event-study research design to study how these policies affected consumption. We find that stay-at-home orders compressed spending broadly across consumption categories, though the net effect between entry and exit of lockdowns points to a reallocation toward durable Entertainment goods and non-durable Clothing and general merchandise, at the expense of other categories. 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.

The Chilean government enacted three temporary pension withdrawal policies during 2020 and 2021, allowing households to access up-to ten percent of their accumulated pension savings, to complement this analysis, we study the effectiveness of this policy to stimulate household's consumption. Due to possible anticipation in the case of the second and third pension withdrawals, our focus is on the first pension withdrawal. We implement an event-study research design to estimate whether these liquidity injections affected retail purchases. We find that retail purchases increase following the initiation of the first withdrawal, with effects that build over time, consistent with retailers initially drawing down inventories before placing new orders. The positive response is driven primarily by Food and Clothing and general merchandise, while among durable goods, Entertainment exhibits the largest response. This evidence suggests that income support programs generated broad-based consumption increases, complementing the reallocation pattern observed during stay-at-home orders. Taken together, the evidence seems to

suggest a limited role of stay-at-home orders and pensions withdrawal to rationalize the change in consumption patterns that happened over the period and points to the period's characteristic the main determinant of the aggregate shifts in consumption patterns.

Despite the changes in consumption patterns, we find no significant impact on the household's cost of living measures when considering them. In the context of the COVID-19 pandemic, there were several calls to update the consumption basket in the wake of the shock. We implement a weight-varying consumer price index using price data from electronic tax documents. Using a fixed-weight approach, we first provide estimates that with the product information contained in the tax documents, we can replicate their official counterparts, especially during regular times. Differences between the official fixed-weight and the tax document measures manifest at the onset of the COVID-19 pandemic, which reflects the hurdles to collecting prices during the early stage of the episode and the possibility that the electronic tax documents keep a better track of headline inflation. Finally, we expand the analysis to compare the fixed-weight and weight-varying measures of the cost of living from the electronic tax invoice, finding that although the indices have discrepancies of up to 1.3 percent, they are not a first-order concern over the pandemic period.

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. Hoberg and Phillips (2016) is the closest to our work regarding the classification problem they face, as they classify US firms and their competitors using the forms filed by public companies that describe their business and products. Instead, we categorize the products firms exchange into standardized categories.

Our article also contributes to the literature that provides strategies 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. (2020) 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. (2020) 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 (ETD) that firms in Chile generate every time they have transactions with each other, and which must be reported to the Chilean Internal Revenue Service (SII). The documents include regular electronic invoices (EI)—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.

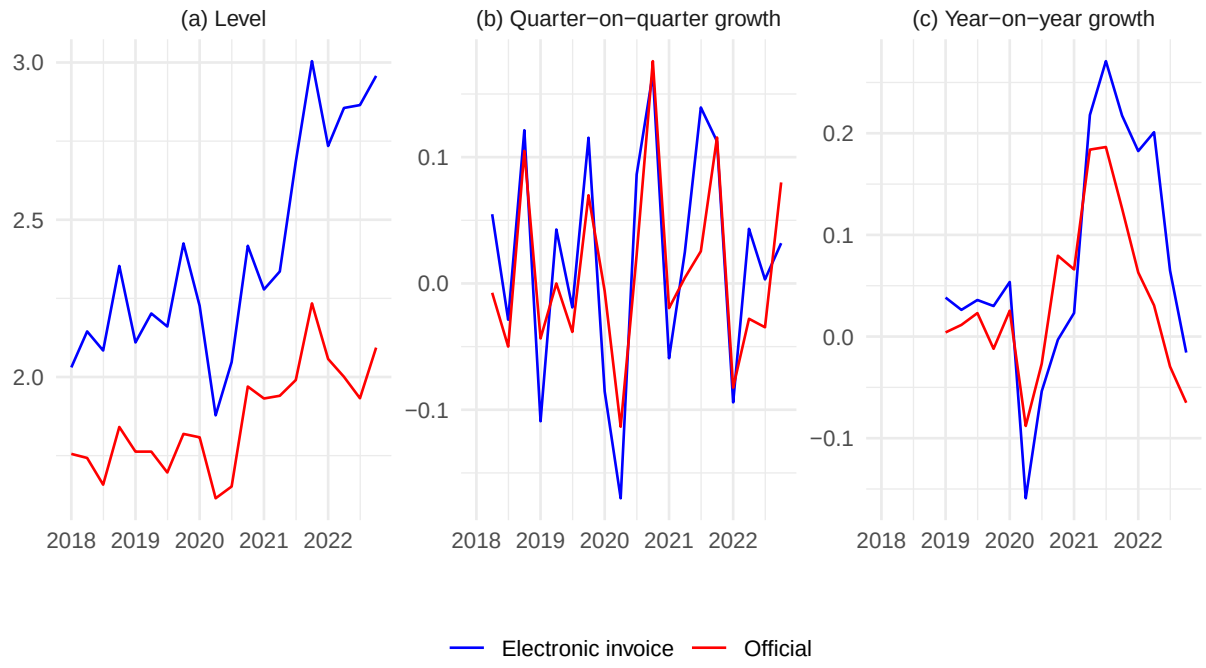
Table 1: Summary statistics of ETD and EI (in millions)

	Documents	Sellers	Buyers	Items	Descriptions
			ETD		
Monthly average	44.19	0.54	5.10	185.95	19.47
Annual average	530.28	6.47	61.16	2,231.35	233.63
Total	4,242.22	51.78	489.25	17,850.84	1,869.05
			EI		
Monthly average	26.95	0.30	1.67	111.14	10.45
Annual average	323.43	3.56	20.04	1,333.67	125.39
Total	2,587.47	28.46	160.30	10,669.35	1,003.14

Notes: The data cover 2015–2022 and report monthly and yearly averages and totals for documents, sellers, buyers, items, and unique descriptions (in millions). Both ETD and EI exclude canceled documents. EI includes only electronic invoices documents.

[1] Illustrative examples of electronic invoices can be found on the Chilean Internal Revenue Service (SII) website.

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 (EI, right axis). Panels (b) and (c) report quarter-on-quarter and year-on-year log differences, respectively

2.2 Constructing the labeled 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]

2.2.1 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 define

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

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 coverage, intensity, and economic relevance. 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

$$C = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}, \tag{2}$$

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

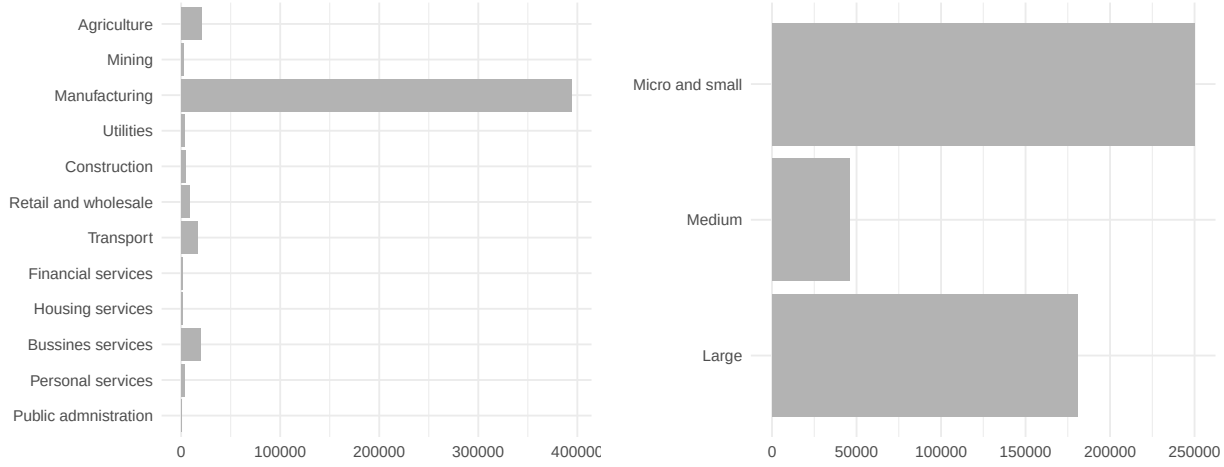
Figure 2 shows the sectoral and firm size composition of the training sample. By sector, manufacturing accounts for the largest share of labeled descriptions, reflecting the greater product variety and specificity of invoicing language in that sector. By firm size, the sample is skewed toward large firms and micro and small enterprises, with medium-sized firms representing a lower proportion.

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

[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 F1 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 2: Training sample composition



(a) UCP Division

(b) Firm sale size

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

2.2.2 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 $[a–z]$ (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

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

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 Methodology

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 must scale to millions of unlabeled entries at inference time.

3.1 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

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

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

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

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 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:

$$\min_\theta \sum_{i=1}^N \mathcal{L}(\mathbf{x}_i, y_i), \qquad \mathcal{L}(\mathbf{x}, y) = -\log p(y \mid \mathbf{x}),$$

where $y_i \in \{1, \ldots, K\}$ is the true label of observation i and $\theta = \{\mathbf{W}, \mathbf{b}, \mathbf{u}_1, \ldots, \mathbf{u}_B\}$ collects all parameters. Optimization proceeds via stochastic gradient descent (SGD) 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

$$\hat{y}(\mathbf{x}) = \arg \max_{k \in \{1,\dots,K\}} p(k \mid \mathbf{x}).$$

Where needed, we return the top-t predictions with their associated probabilities to support human-in-the-loop validation of ambiguous entries. Appendix A describe 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

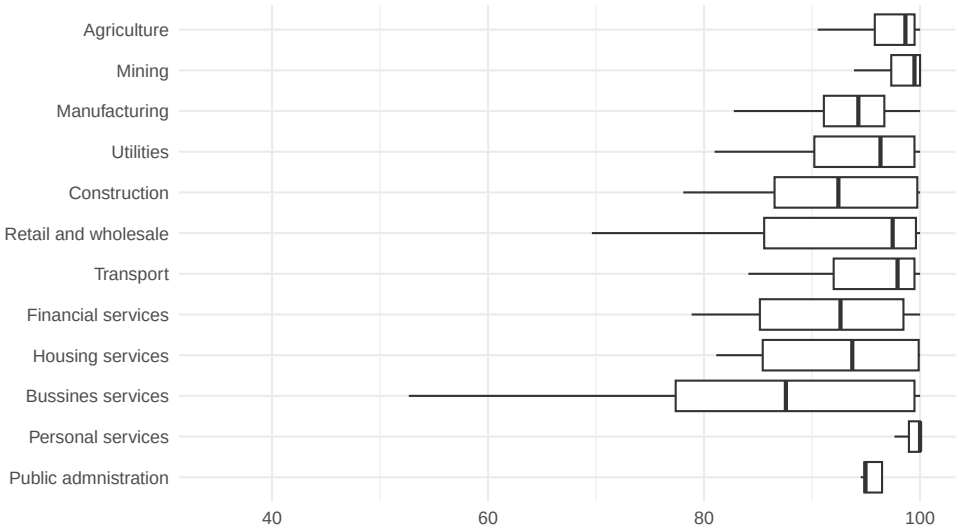
1 (perfect classification). Table 2 reports aggregate, class-weighted averages of Precision, Recall, and F1-score for the three classifiers, where weights correspond to the number of observations in class k in the training sample. Figure 3 shows the distribution of classification performance across UCP divisions.

Table 2: Quality metrics by classifier and feature model

Classifier	Precision	Recall	F1
COICOP	0.955	0.954	0.954
CPC	0.918	0.916	0.915
UCP	0.937	0.935	0.935

Notes: Metrics are computed using k-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 3: Distribution of classification performance (F1 score) across UCP divisions



Boxplots show the distribution of F1 scores for each UCP-12 division. UCP divisions correspond to economic activity groupings consistent with National Accounts classifications.

4 Applications

We develop three applications using the product classifications constructed in the previous section. First, we construct Machinery and Equipment and Goods Consumption series that closely track their national accounts counterparts, and we show that these indices are available within days of the end of each month and are not subject to ex-post revisions in official statistics. Second, we exploit the granularity of our classification to study how consumption patterns evolved during the COVID-19 pandemic, documenting a reallocation of expenditure across product categories in response to lockdowns and income support policies. Third, we use price data from the electronic invoices together with time-varying expenditure shares to construct a flexible-weight consumer price index and compare it against its fixed-weight counterpart.

4.1 Machinery and Equipment and Consumption tracking

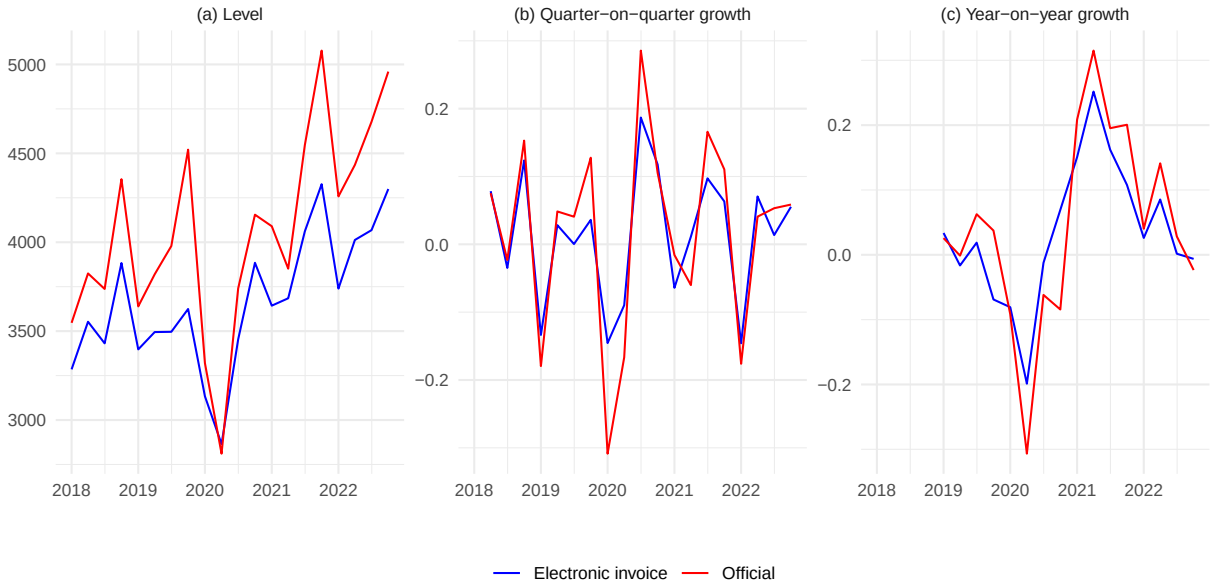
Machinery and Equipment series. We leverage our classification algorithm to construct an aggregate series that mimics the official Machinery and Equipment Outlays (M&E) series from National Accounts. We build the series using the mapping between firms' purchases and their associated CPC and UCP categories, 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 M&E outlays at quarterly frequency.[6]

Each panel in Figure 4 presents our constructed M&E series together with the corresponding national account series for M&E. 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 account 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

[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.

of the constructed M&E and national account series, is 0.93 and 0.90, respectively. Interestingly, we also computed the correlation coefficient between the first national account release of machinery and 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 4: Machinery and Equipment outlays: Official statistics and electronic invoice aggregate



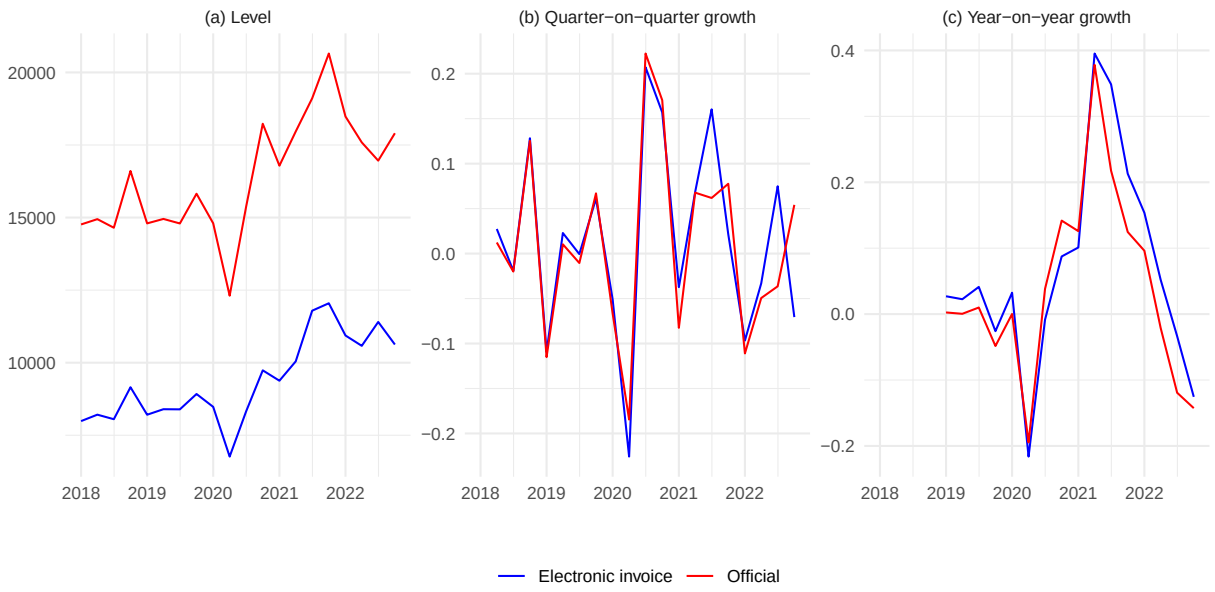
Notes: Panel (a) reports chained-volume indices for machinery and equipment outlays from official national accounts data and electronic invoices (EI, right axis). Panels (b) and (c) report quarter-on-quarter and year-on-year log differences, respectively. The EI series excludes direct imports and retail-sector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

Goods consumption. Similar to M&E, we can also construct a series for the consumption of goods. Considering that we use electronic invoices that originate in business-to-business transactions, we must 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 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 BEC 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 5 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 underestimates the official counterpart, which can be explained by the markup retailers charge consumers as well as purchases from informal retailers that are not captured in our dataset. 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 account 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 M&E indicator is explained by the greater revision M&E national account data is subject to.

Figure 5: 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 excludes services.

4.2 Real-time performance

A natural question about the production of these indices is to what extent they become available in real-time. Because of misreporting or tax design, taxpayers may file tax forms with significant delays relative to the period in which economic activity occurs. The lag

can render administrative data ill-suited for nowcasting exercises. However, this is different for electronic invoices. According to the IRS rules, firms have up to ten working days to report these documents. As a result, virtually all business-to-business transactions should be observable within a few days after the month-end.

Next, we assess the practical binding of this reporting requirement. To do so, we exploit two dates recorded in each invoice: the transaction date and the 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:

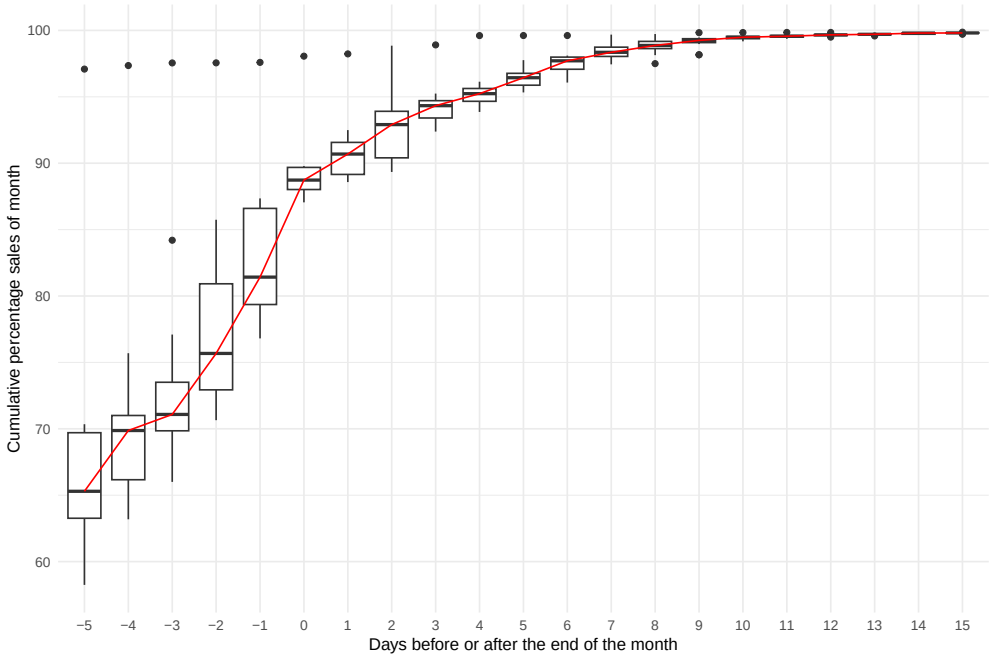
$$s_{td} = \frac{r_{td}}{r_{t\infty}},$$

where r_{td} is sales of the month t reported in d days before or after the end of month t, and $r_{t\infty}$ is sales corresponding to month t calculated with the full information for that month.

Two weeks after the month ends, nearly all the information required to produce our indices is available. Figure 6 shows how the cumulative percentage of total sales the Bank has received increases with the number of days before or after the end of the month. Each box represents the distribution of that share between 2021m8 and 2023m7. For instance, five days before the end of the month, the Bank has received 65 percent of the total. By the end of the month, this rises to almost 90 percent, and after 12 days, the Bank has virtually received the totality.

The ability to construct timely indicators that closely track official statistics highlights the usefulness of electronic invoices for nowcasting applications. Unlike official national accounts data, which are subject to ex-post revisions as additional information becomes available, our measures can be produced with minimal delay and are not subsequently revised. The results in this section support the view that combining administrative data with natural language processing techniques can complement the production of official statistics.

Figure 6: EI declared transaction date versus actual data availability



Notes: This figure plots the percentage of total sales that the Bank has received as a function of the days before and after the end of the month ($d = 0$). Each box shows the median, the interquartile and minimum-maximum values for each day. The black dots represent outliers in the serie.

4.3 Consumption patterns over time

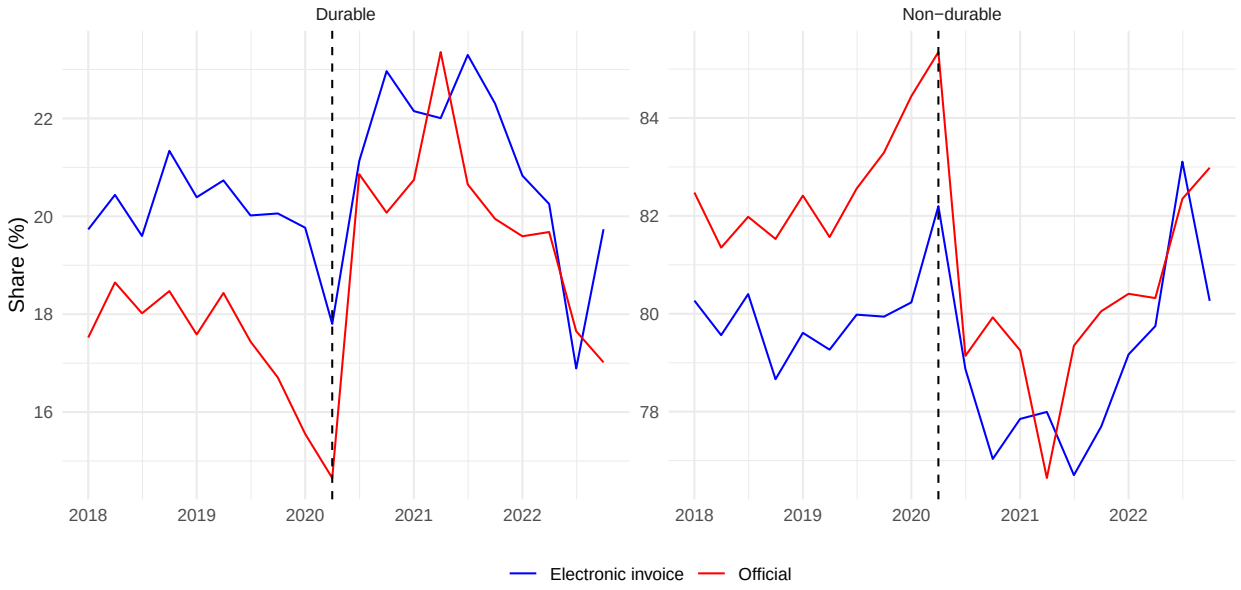
4.3.1 Durable and Non-durable consumption

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.[7] Figure 7 displays aggregate quarterly consumption shares for durable and non-durable goods between 2018 and 2022, using both our constructed series—based on retail-sector purchases recorded in electronic invoices—and the official data from GDP expenditure in personal consumption. 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

[7]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: Consumption shares: Durable and Non-Durable



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:

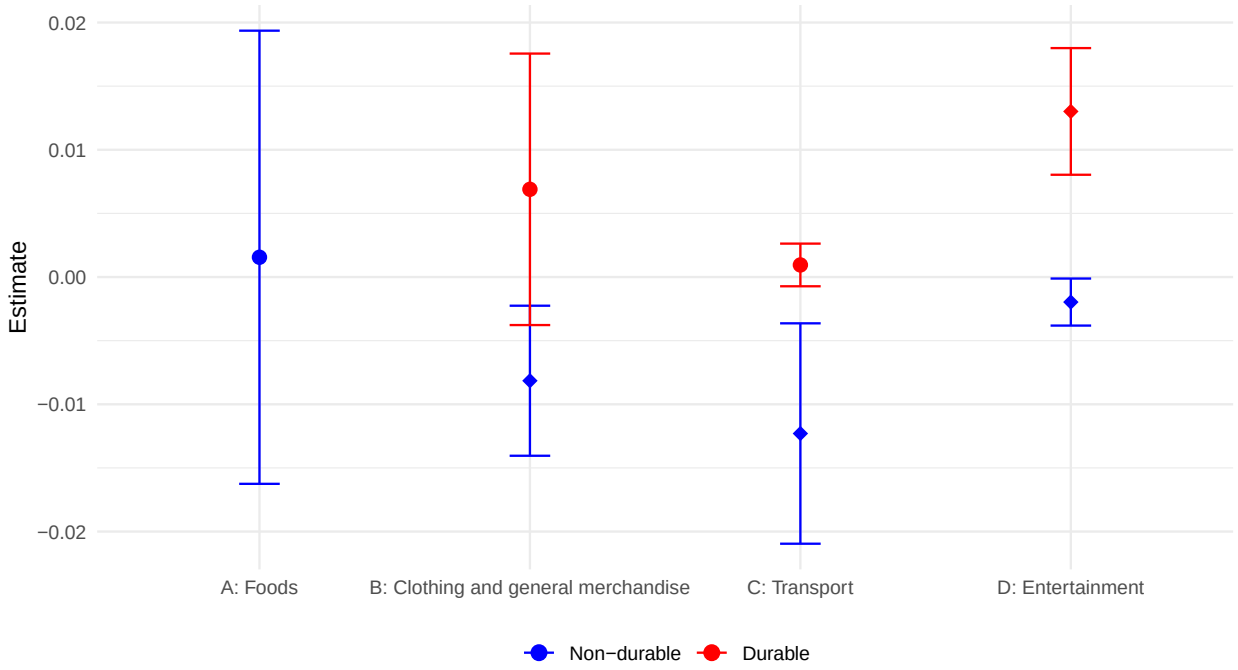
$$\omega_{jt} = \lambda_j + \beta_j D_t + \varepsilon_{jt}, \tag{3}$$

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 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. In the 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.

Figure 8: Expenditure changes during COVID



Notes: This figure shows changes in expenditure during COVID-19. The circles report the β_j coefficients from the regression defined in Equation (3) for durable and non-durable consumption categories. The range bar denote the confidence interval 5-95%.

4.3.2 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 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 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:

$$\log X_{c,t} = \lambda_c + \lambda_t + \sum_{s=4}^{4} \alpha_s \mathbb{1}_{c,t+s}^{Event} + \epsilon_{c,t}, \tag{4}$$

where $X_{c,t}$ is consumption in municipality c during week t, $Event$ can be *Enter* or *Exit*, and $\mathbb{1}_{c,t+s}^{Event}$ 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 municipal purchases around the introduction of stay-at-home orders. The error term $\epsilon_{c,t}$ captures idiosyncratic shocks to municipal purchases, 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, α_s. In the weeks before the lockdown imposition,

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

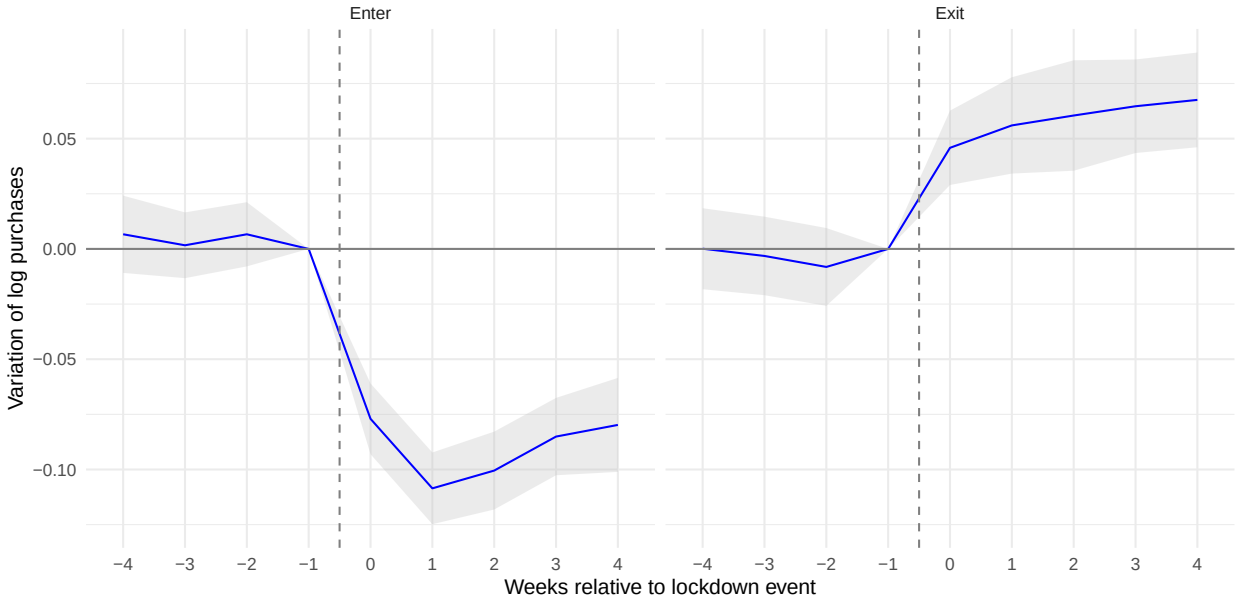
[9] Data on lockdowns, the Rotating Lockdown policy, and the Step by Step We Take Care of Ourselves plan comes from the Chilean Ministry of Science.

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, 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.

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 purchases before or during lockdown for retail sector. This figure shows $\sum_{s \in S} \alpha_s$ from regression defined in Equation 4. Gray area denote the confidence interval 5-95%.

We can further understand the changes consumption goods experience when we zero in on more disaggregate categories. Figure 10 presents the estimates for i) foods, ii) Enter-

tainment, iii) Clothing and general merchandise, iv) Transportation. Appendix Table B.1 summarizes the corresponding magnitudes.

At the onset of lockdowns, food consumption decline 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 goods 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.

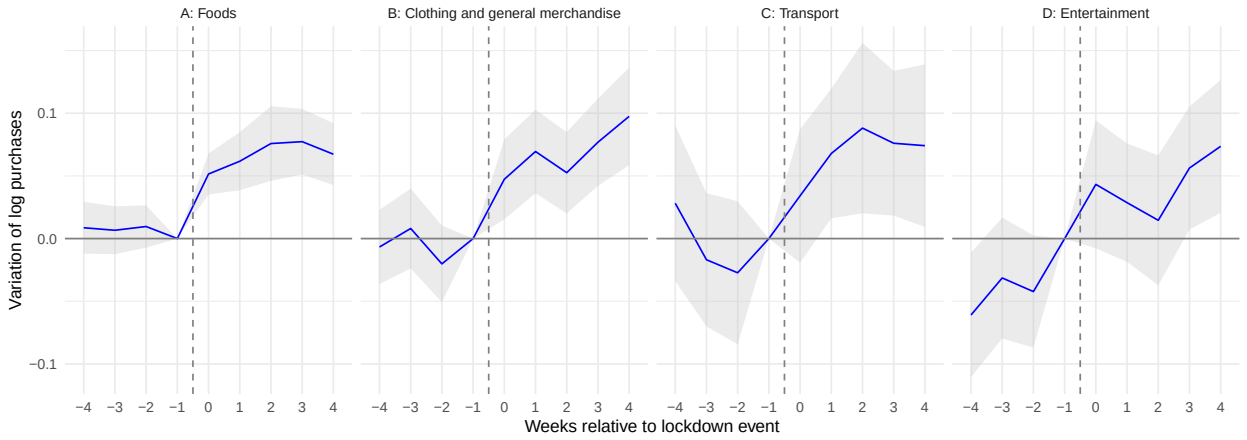
Finally, Table 3 shows further disaggregation for durable and non-durable categories and estimated effects four weeks after entering or exiting the lockdown. Durable Transport and Entertainment and non-durable Clothing and general merchandise goods are the only categories in which the net effect between 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 reports the full set of estimates.

Figure 10: Lockdown and divisions of consumption goods.

(a) Enter



(b) Exit



Notes: Variation of log purchases before or during lockdown for retail sector by product division. This figure shows $\sum_{s \in S} \alpha_s$ from regression defined in Equation 4. Gray area denote the confidence interval 5-95%. 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 retail purchases 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 form $\log X_{t,c} = \lambda_c + \lambda_t + \sum_{s \in S} \alpha_s \mathbb{1}\{s(tc) = s\} + \epsilon_{t,c}$, where $X_{t,c}$ denotes total purchases in municipality c and week t. 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 . $p < 0.15$.

4.3.3 The effects of pension withdrawals over consumption

Chile enacted three temporary pension withdrawal policies during the COVID-19 pandemic that allowed affiliates of the private pension system to withdraw up to ten percent of their accumulated mandatory contributions. 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 amount of 35 Unidades de Fomento (UF) and a maximum of 150 UF. If ten percent of an affiliate's pension balance was below the minimum threshold, the individual was allowed to withdraw the full available balance. Withdrawals were generally tax-exempt, with the exception of the second episode, and affiliates had up to 365 days to request their funds.

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. 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, sug-

gesting that withdrawals 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 retail input purchases 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 pronounced for the second and third withdrawals. Consistent with this interpretation, in appendix XXX, 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 is particular important for our application, if anticipation happens, the policy can be associated with the wrong sign with retail purchases as they create inventories to respond to increasing future consumption.

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

$$\log X_{c,t} = \lambda_c + \lambda_t + \sum_{s=-2,0,1,2} \alpha_s D_{t+s} \phi_c + \epsilon_{c,t} \tag{5}$$

The dependent variable $Y_{c,t}^k$ denotes total monthly consumption in municipality c at time t during withdrawal episode k, expressed in logarithms. Event time is defined as the number of months since the initiation of withdrawal k, denoted $s(k,t,c)$. We focus on post-withdrawal dynamics and consider event-time months $s \in \{0,1,2\}$, where $s = 0$ corresponds to the month in which the withdrawal is initiated, relative to three months before the withdrawal enactment.

Treatment intensity varies across municipalities and withdrawal episodes and is captured by $\phi_{ck} = \log(1 + \text{withdrawal}_{ck}/\text{income}_c)$, where withdrawal_{ck} is the cumulative amount withdrawn in municipality c during episode k, measured relative to the initiation month, and income_c is average municipal labor income in 2019. The specification includes municipality fixed effects, λ_c, which absorb all time-invariant differences across municipalities, and time fixed effects, ρ_t, which control for aggregate shocks common to all municipalities, including macroeconomic conditions and seasonality. The error term, $\epsilon_{t,c}$, follows the same statistical properties as the previous application.

The coefficients of interest, α_s, trace out the dynamic response of municipal purchases

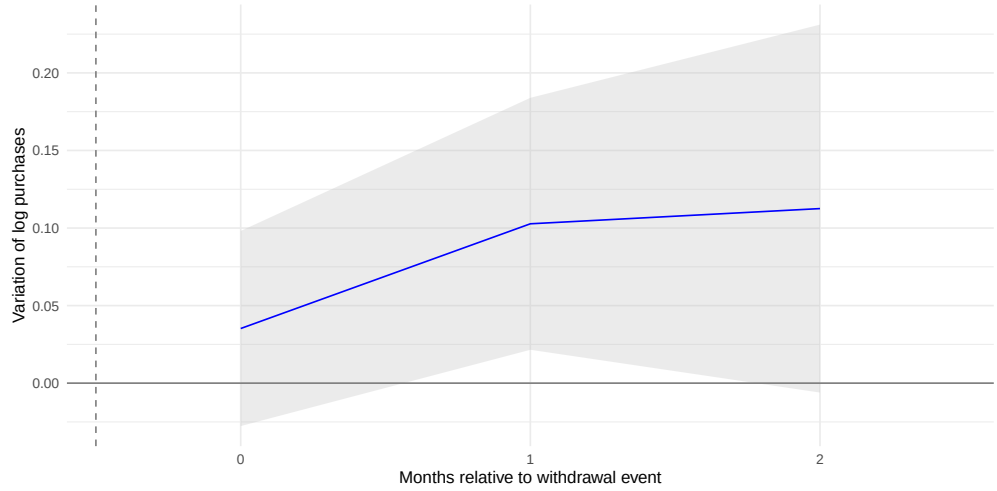
following a withdrawal. Each coefficient measures the percent change in month s associated with a one-percent increase in pension withdrawals relative to monthly labor income.

Figure 11 reports the estimated dynamic effects of the first withdrawal. Purchases of consumption goods by retailers increase following the initiation 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., 2020; 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.

We next examine heterogeneity in the response to pension fund withdrawals across broad categories of consumption goods. Figures 12 reports estimates by product-category for the first withdrawal episode using the same specification as in Equation 5.

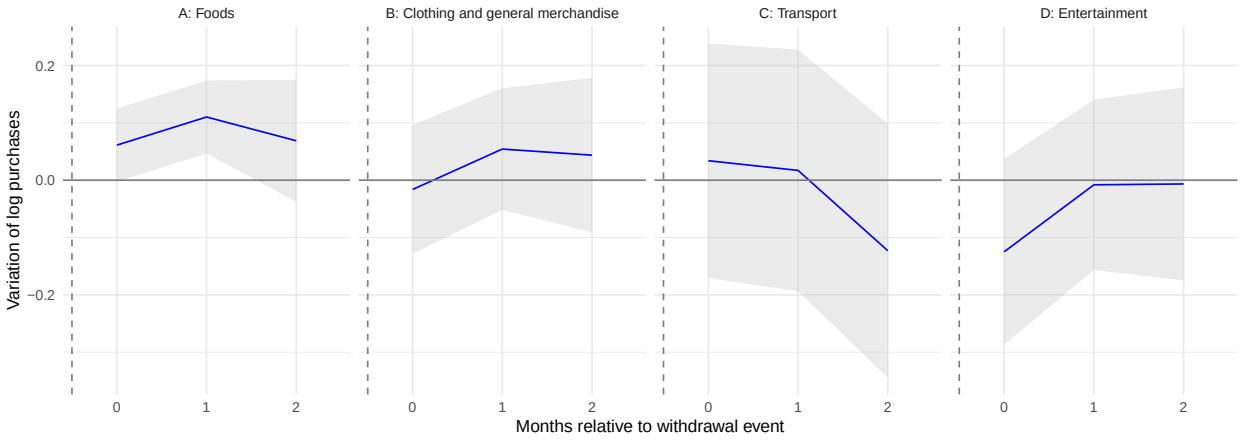
The positive response in retailers' purchases appears to be driven primarily by Food and Clothing and general merchandise, for which point estimates increase following the withdrawal and remain elevated in subsequent months. In contrast, estimates for transportation and entertainment goods are imprecisely estimated 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 non-durable 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.

Figure 11: First withdrawals effect on consumption.



Notes: Variation of log purchases after first withdrawal for retail sector. This figure shows $\sum_{s \in S} \alpha_s$ from regression defined in Equation 5. Gray area denote the confidence interval 5-95%. Standard errors are clustered at the municipality level.

Figure 12: First withdrawal and divisions of consumption goods.



Notes: Variation of log purchases after first withdrawal for retail sector by product division. This figure shows $\sum_{s \in S} \alpha_s$ from regression defined in Equation 5. Gray area denote the confidence interval 5-95%. 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 purchases after the first withdrawal for retail sector 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 5. Standard errors are clustered at the municipality level. Statistical significance is denoted by *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$, and . $p < 0.15$.

4.4 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. Using price data from electronic invoices together 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.

4.4.1 Methodology

We construct a price index analogous to the official CPI, updating consumption weights to reflect the changes in expenditure shares documented in Section 4.3.

The construction of the Consumer Price Index follows the methodology employed by the National Statistics Institute, or 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 rep-

resentative 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, EIs 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, EIs 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,

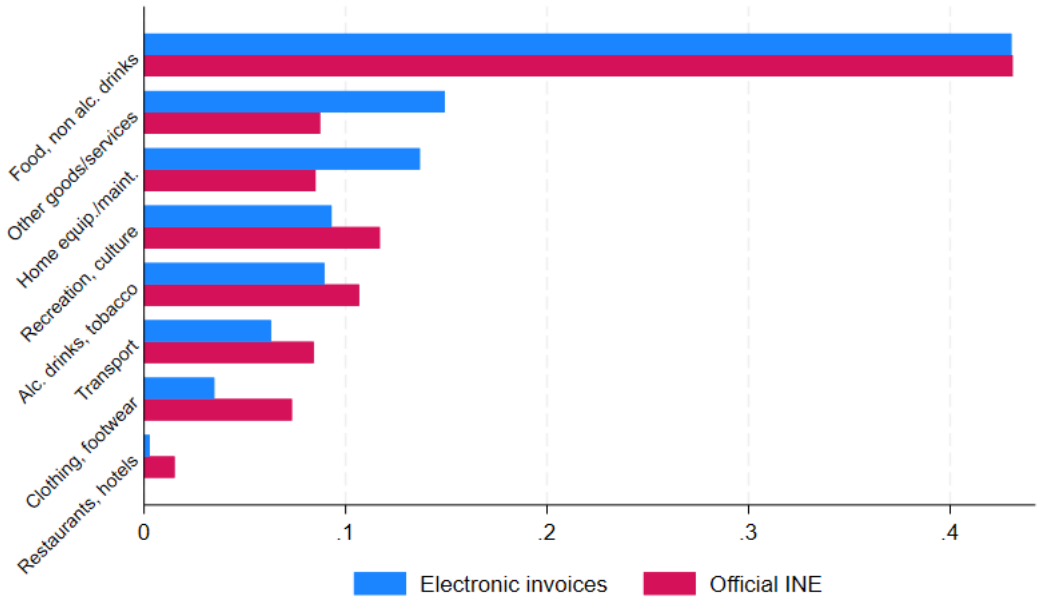
$$\Delta P_{i,t} = \left(\prod_{v \in N_{it}} \Delta P_{v(i),t} \right)^{\frac{1}{|N_{it}|}},$$

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 indexes 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 4.3. Figure 13 compares the 2018 division expenditure shares derived from this approach, based on electronic in-

voices, 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 "Restaurant & Hotel" 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 consumer-to-business transactions, particularly within service-oriented divisions

Figure 13: Consumption weights by division



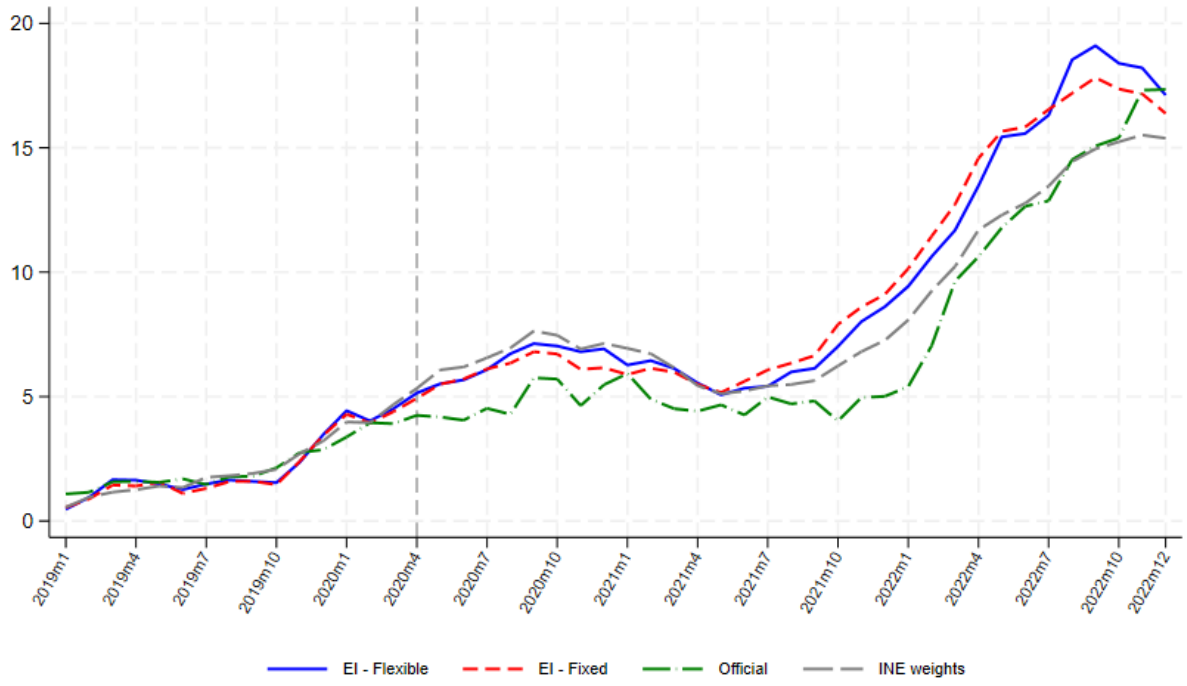
4.4.2 Results

Figure 14 presents four annual inflation series considering the 157 products we are able to classify and track. The blue line corresponds to EI price indexes with fixed weights computed from EI expenditure shares. The yellow line uses the same prices, but allows weights to vary over time. The black line replaces these with the official INE weights, while the red 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 approximately 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 simi-

lar 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 4.3.2), likely affected the collection of prices collected 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



How time-varying weight price index differ from 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.[10]

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, 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. Relatedly, Redding and Weinstein (2019) develop a framework for constructing consistent price aggregates that account for changes in preferences, which plausibly occurred during the COVID-19 pandemic and with the associated health restrictions. For that, the authors propose a price index for preferences featuring constant elasticity of substitution that treat taste shocks as equivalent to price shocks.

5 Conclusions

We showed that applying natural language processing to naturally occurring data describing the products that firms trade can produce machinery and equipment outlays and consumption of goods aggregates that closely track their national accounts counterparts without ex-post revision. Using the granularity of our classification, we documented how consumption patterns shifted during the COVID-19 pandemic: households reallocated expenditure toward durable goods during the recovery phase, while stay-at-home orders compressed spending broadly across non-durable categories. We also showed that unanticipated pension withdrawals generated sizable and persistent in-

[10] Appendix Figure B.4 shows how during these first months, Food prices were rising, while other like Transport even had negative inflation rates.

creases in retail purchases, with heterogeneous responses across consumption categories. Finally, although the flexible-weight and fixed-weight consumer price indices diverge during specific episodes—most notably the onset of the pandemic and the subsequent high-inflation period—these differences are not a first-order concern over the pandemic period as a whole.

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.

Our results carry a broader message for the production of official statistics. Naturally occurring data, like electronic invoice systems, contains 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, Nathaniel Hendren, Michael Stepner, and The Opportunity Insights Team, "The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data," NBER Working Papers 27431, National Bureau of Economic Research, Inc June 2020.

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, "Reengineering Key National Economic Indicators," Working Paper 26116, National Bureau of Economic Research July 2019.

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 Implementation

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, 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

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

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

$$\max_{h \in \mathcal{H}} f(h) \equiv J(M(h), \mathcal{D}_{\text{val}}) \quad \text{s.t.} \quad \mathsf{T} \leq Q, \tag{6}$$

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

$$h_t^\star \in \arg\max_{h \in \{h_1, \ldots, h_t\}} f(h) \tag{7}$$

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 5 shows the optimal hyperparameter configurations by classifier, which were used to train the final model and generate predictions for all electronic invoice documents.

Table 5: Set of hyperparameters configurations by classifier

Parameter	CPC	UCP	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

B Additional Figures and Tables

Figure B.1: Consumption Variations: Durable and Non-Durable

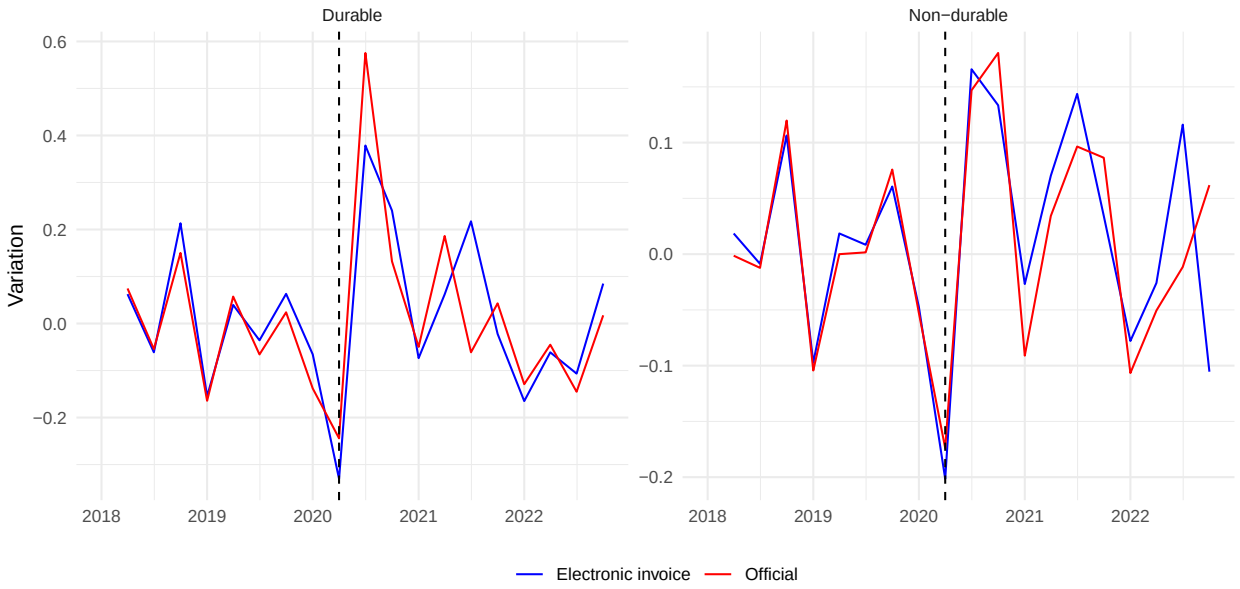


Table B.1: Lockdown and divisions of consumption goods

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 retail purchases by product division. Each column corresponds to a separate regression of the form $\log X_{t,c} = \lambda_c + \lambda_t + \sum_{s \in S} \alpha_s \mathbb{1}\{s(tc) = s\} + \epsilon_{t,c}$, where $X_{t,c}$ denotes total purchases in municipality c and week t of each product division. 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 retail purchases by product division. Each column corresponds to a separate regression of the form $\log X_{t,c} = \lambda_c + \lambda_t + \sum_{s \in S} \alpha_s \mathbb{1}\{s(tc) = s\} + \epsilon_{t,c}$, where $X_{t,c}$ denotes total purchases in municipality c and week t of each product division with durable goods. 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 retail purchases by product division. Each column corresponds to a separate regression of the form $\log X_{t,c} = \lambda_c + \lambda_t + \sum_{s \in S} \alpha_s \mathbb{1}\{s(tc) = s\} + \epsilon_{t,c}$, where $X_{t,c}$ denotes total purchases in municipality c and week t of each product division with durable goods. 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 * $p < 0.10$.

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

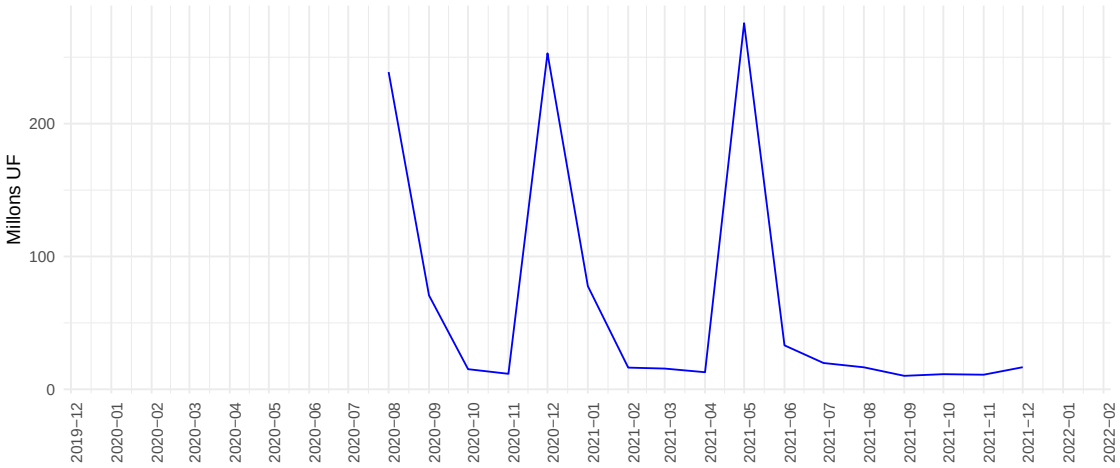


Table B.4: Withdrawals and divisions of consumption goods.

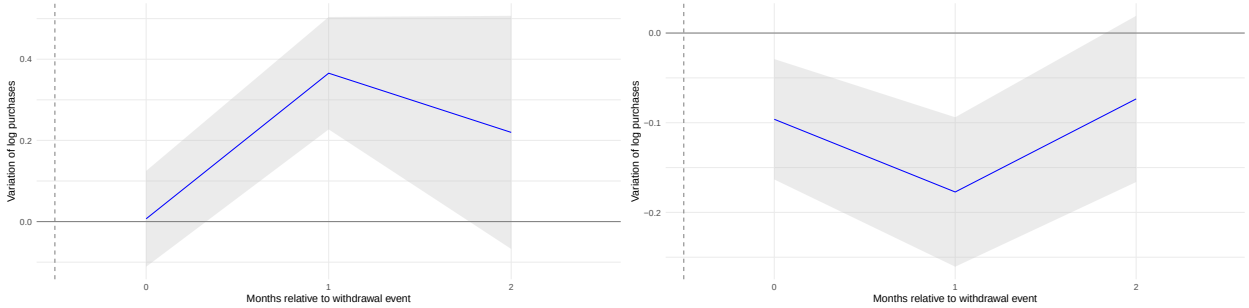
s	All	Foods	Clothing and general merchandise	Transport	Entertainment
	First withdrawal				
0	0.035	0.061.	-0.016	0.034	-0.125
	(0.032)	(0.033)	(0.057)	(0.104)	(0.082)
1	0.103*	0.110***	0.054	0.017	-0.008
	(0.041)	(0.032)	(0.054)	(0.107)	(0.075)
2	0.113.	0.069	0.044	-0.123	-0.007
	(0.060)	(0.054)	(0.069)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

Notes: Variation difference of log purchases after withdrawals for retail sector by division product. This table shows $\sum_{s \in S} \alpha_s$ from regression defined in Equation 5. Standard errors are clustered at the municipality level.

Figure B.3 presents the corresponding estimates for the second withdrawal episode. Unlike the first, 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, with effects that are larger and more precisely estimated. Purchases increase sharply in the month following policy implementation and remain elevated thereafter.

The third withdrawal episode, where anticipation was even more pronounced, tells a different story. Our interpretation is that, 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.3: Subsequent pension withdrawal and Retail and Restaurant and Hotels purchases

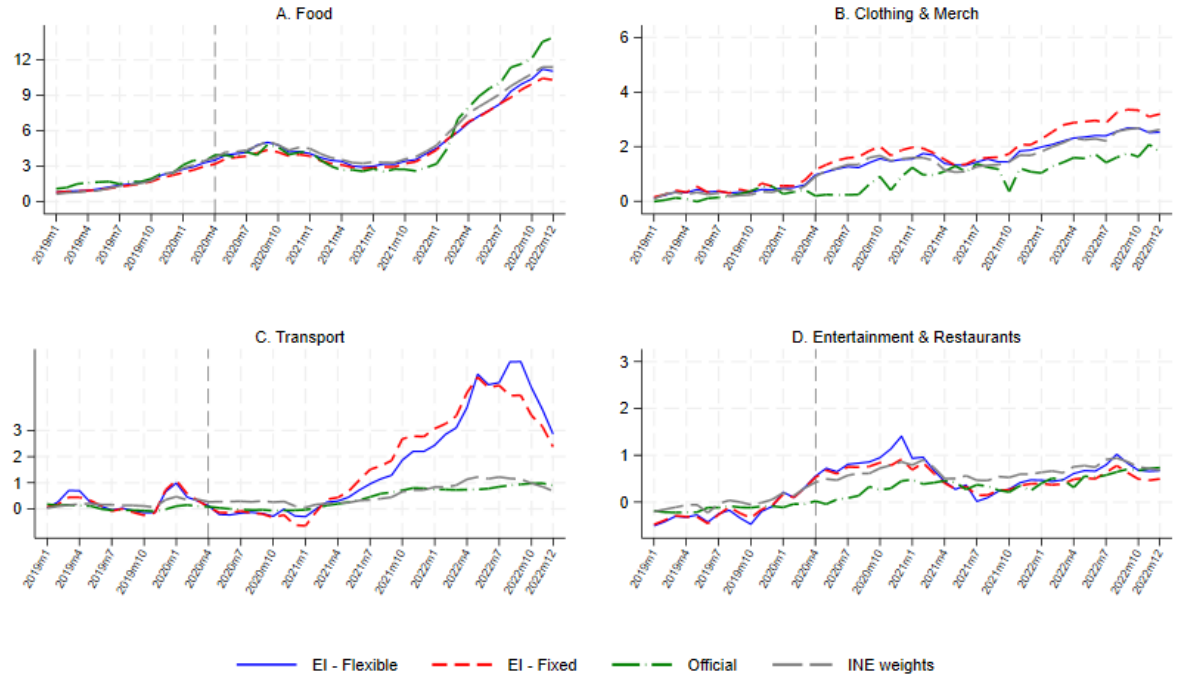


(a) Second pension withdrawal

(b) Third pension withdrawal

Notes: Variation of log purchases after second withdrawal for retail sector. This figure shows $\sum_{s \in S} \alpha_s$ from regression defined in Equation 5. Gray area denote the confidence interval 5-95%. Standard errors are clustered at the municipality level.

Figure B.4: Annual Inflation by Macro division



C Data details

C.1 Price indices

For price monitoring, we seek to have a data set 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, that is $Pmin_i * 0.9$, $Pmax_i * 1.2$. In addition, our data set is filtered to consider exclusively the economic sectors of the selling firms where we are interested in following prices (mainly retail trade). Either we consider transactions between the same firm. We require that the glosses be 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.