Price setting in Chile: Micro evidence from consumer on-line prices during the social outbreak and Covid-19*

J. Peña† and E. Prades ‡

March 2021

Abstract

In this paper we analyze the price setting behavior in Chile by using scraped data from public websites of the main retailers including supermarkets, a pharmacy retailer and car dealerships. The data collection started in July 2019 and the dataset covers two major recent events: (1) the social outbreak and (2) the state of emergency declaration due to Covid-19, both episodes led to disruptions in the economy. With information on product varieties that accounts for 22% of the CPI basket, we document several empirical findings as regards price setting behaviour in terms of stickiness, that is, frequency, implied duration and the size of price adjustments. We find that in spite of facing large shocks, prices adjusted very little, at a lower frequency and at a smaller size than prior to these two events. We also find that there was a reduction on product variety availability on-line, a typical feature that also has been found during natural disasters such as earthquakes. The reduction in product availability poses additional difficulties to construct CPI indexes and to properly capture price rigidities, which are relevant for monetary policy.

Keywords: online price data, CPI, prices stickiness, retail distribution.
JEL-Codes: E01,E31,L81.

*We are very grateful for useful comments to internal seminar participants, both in BoS and the CBCh, and to participants of the Conference on "Non-traditional data and statistical learning with applications to macroeconomics" organized by the Bank of Italy and the Federal Reserve Board and to Isabel Sánchez García. The views expressed herein are those of the authors and do not necessarily reflect the views of Banco de España nor the Central Bank of Chile.

†Central Bank of Chile (CBCh). E-mail: jpena@bcentral.cl

‡Bank of Spain (BoS). This paper was written while visiting the Central Bank of Chile under the agreement signed with Banco de España. The hospitality of the Central Bank of Chile is is gratefully aknowledged. E-mail: elvira.prades@bde.es

Resumen

En este trabajo analizamos la fijación de precios en Chile utilizando técnicas de *webscraping*, donde recolectamos los precios de productos de consumo a través de las páginas web de los principales establecimientos de comercio minorista del país. La base de datos se incluye precios de productos de supermercados, farmacias y comercializadoras de automóviles. La recolección de datos se inició en julio de 2019 y contiene información precios durante la ocurrencia de dos hechos relevantes: (1) el estallido social y (2) la declaración del estado de emergencia por el covid-19. Ambos episodios conllevaron a disrupciones en el funcionamiento normal de la actividad. Con información de productos que representan aproximadamente el 22% de la canasta del IPC, documentamos el comportamiento de los precios en términos de frecuencia y duración y el tamaño de estos ajustes, indicadores que informan sobre la flexibilidad o rigidez del ajuste ante perturbaciones de esta magnitud y comparamos con respecto a tiempos normales. Encontramos que durante estos episodios la rigidez en el ajuste aumentó, ya que los precios se ajustaron con menor frecuencia y las variaciones, en promedio, fueron de menor tamaño. También encontramos que se produjo una importante reducción en la variedad de productos disponibles durante la pandemia, un hecho que suele producirse durante eventos disruptivos como los desastres naturales. Esta reducción en las variedades supone un reto adicional para la elaboración del IPC así como una mayor complejidad para capturar las rigideces nominales, que son relevantes para la política monetaria.

Palabras clave: Precios on-line, índice de precios al consumidor, rigideces de precios, distribución minorista.

Códigos JEL: E01,E31,L81.

1 Introduction

The ability of monetary policy to influence in the short-run real economic activity it much depends on nominal rigidities. This paper intends to contribute to understand from a micro perspective the behavior of price setting in Chile during the 2019 social outbreak and the Covid-19 pandemic.[1] The price setting behavior is even more relevant in a context with heightened uncertainty as regards how inflation will behave over time.[2]

To this end we will make use of a dataset collected over internet sites of main retailers in Chile. The use of Big Data by policy makers and researchers became more popular and accepted in recent years. The main source is Internet data sets. Since online shopping substitutes, or at least supplements, offline shopping, online prices can also be used as a substitute, or supplement, of offline prices. One concern may lie on the differences between on-line and off-line prices, the discrepancy between these two channels is expected to be small as in Chile third-party delivery services are widely used and have a higher penetration compared to other countries.[3]

Online data have many advantages in the context of the Covid-19 Pandemic where there are lockdowns, mobility restrictions and social-distancing rules. Data collection over the internet is called web scraping. This technique provides flexibility and extreme automation. There is abundant anecdotal evidence that scraped prices is a potentially useful tool for nowcasting and short-term forecasting of CPI inflation as well as retail sales variables. Online CPI provides policy-makers with a reasonably good "pulse" as to the direction being taken by inflation in real time and at a higher frequency. The Billion Prices Project (BPP) at MIT is a solid evidence that has been functioning for more than a decade.[4]

The objective of this paper is to provide evidence on the behavior of prices/price setting in Chile based on the dataset collected by the "Online Prices Project" by the Central Bank of Chile. The dataset covers two episodes of particular interest that led to disruptions in the economy, particularly in terms of the degree of price rigidities, by using online prices. We want to answer questions such as: How nominal rigidities have changed during the social outbreak and the pandemia episode in Chile? Do we

[1] There is a renewed interest this front as there are recent initiatives to study pricing behaviour from a micro perspective such as the Price-setting Microdata Analysis Network (PRISMA) and the Inflation Persistence Network (IPN) by the European Central Bank. PRISMA is a new ESCB research network established in 2018 with a mandate from the Governing Council, this network will study price-setting behaviour of individual firms and in the retail sector using micro price datasets.

[2] See Blanchard (2020) https://voxeu.org/article/there-deflation-or-inflation-our-future.

[3] For a discussion on the representativeness of on-line versus off-line data and implications see Cavallo (2017).

[4] For more details see http://www.thebillionpricesproject.com/.

observe product/sectoral differences? What is the role of seasonal/sales discounts in a context where a substantial share of products have been dropped out from on-line retailers?

In June 2019 the "Online Prices Project" was initiated at the Central Bank of Chile. The purpose was to follow-up price developments focused on goods typically available at supermarkets using web scraping techniques; this provides flash estimates of CPI available on a daily basis, that is, with higher frequency and more timely that the one provided by the national statistical office. This article describes the behavior of online prices in Chile, taking a deeper look into the micro-data that we have collected on a daily basis between July 2019 till November 2020, from the two of the largest retailers of fast moving consumer goods (groceries, personal care and household equipment), a health retailer and auto dealers in Chile. We provide evidence on the key features on price-setting behaviour.

The dataset collected includes products corresponding to expenditure categories that cover around 22% of the official Chilean CPI. In spite the short-time span, due to the recent start of this project, our sample covers two major events that took place in the Chilean economy. Along this period two major shocks occurred which are unambiguous examples of exogenous and unanticipated aggregate shocks, which makes them very interesting to analyze how prices are being set and to anticipate what could be the implications for CPI developments. Firstly, the 18 of October (18-O) the social outbreak started, during a week a severe curfew was set and from this date on regular demonstrations took place with different intensities. All these actions led to some disruptions in economic activity, mainly in the retail sector. Strikes and demonstrations dramatically reallocate store traffic and opening hours which led to an increase of on-line sales and delivery services.

Then, a second shock took place, this time of a global nature, the Covid-19 pandemia. The first Covid-19 case was registered in Chile in early March and by mid-March, in order curb the spread of the virus and to avoid the collapse of the health system, the government introduced some measures in order to maintain social distance. To this end, the government announced dynamic and selective quarantines, curfews, the closure of restaurants, malls, schools, universities and, when possible, stay at home work was encouraged. Few activities such as supermarkets and pharmacies where allowed to function normally. And, again, it was registered a heightened demand for online shopping. Both episodes implied the closure of some retail stores and an increased demand for delivery services. The pandemia also put some pressures from the supply side. Under this episode we can also expect some supply disruptions as some manufacturing activities require high social interaction.

We end up with a dataset following prices on a daily basis for approximately 47 product categories out of the 305 included by the National Statistical Office, and start following around $4,755$ product-varieties that account for 22% of the CPI Chilean basket. The main stylized facts on the price setting was calculated out for four differentiated periods: (1) "pre-shocks", that goes from June 2019 to 18 of October, (2) "social outbreak" from the end of October till mid-March and (3) "pandemia" from 18 March and (4) from the partial uplift of mobility restrictions from July till November 2020 (502 days at this point). We start by constructing common statistics of price setting based on averages across products and time. These include measures focused on three dimensions: frequency of price changes, implied duration and the magnitude of price changes.

The main conclusions are that both product availability and the price setting behavior changed during Covid-19 and the social outbreak episodes. In particular, the frequency of price changes fell during both episodes, suggesting that firms were delaying their price adjustments. In sum, after the two shocks prices became moderately more sticky. In terms of availability we found that there was a significant reduction in online availability of products during pandemia and that they are recovering at a slow pace. This reduction occurred after March 18, the date in which the state of alarm was declared and confinement measures where put in place in Chile due to Covid-19. The goods that disappear quickly from the stores are mostly perishable and some nonperishable items that people are likely to hoard in fear of future supply disruptions, that is to say, a focus on essential goods. However, during the social outbreak the reduction in online availability was not significant.

Along these episodes the CLP also registered a depreciation against the USD. Therefore, if passed through, we may expect imported goods to become more expensive. To explore this, we have identified the brand for each product and therefore we can infer the country of origin. This allows us to explore additional features such as the role of exchange rate pass-through or whether there have been supply-chain disruptions.[5]

This paper is organized as follows. After this introduction, in **section 1**, we review the literature on microdata on prices and webscraping in **section 2**. In **section 3** we describe the data collection method and the characteristics of the scraped data. In **section 4** we report the main results on product availability, frequency and size of adjustments comparing all 4 periods. Finally, in **section 5**

[5]We are aware that this assumption may be simplistic in a context where some domestic firms might be using imported goods and at the same time we are identifying the country of origin of final goods, that at the same time if involved in global value chains the true origin of the product can be from other or from multiple origins.

we conclude with the preliminary results drawn so far.

2 Literature Review

Price setting analysis based on official micro data.– The existence of price rigidities is a determinant of the effectiveness of monetary policy. In general, the literature refers to two ways to measure the degree of price flexibility: one constitutes a macro approach through modelling as in Gali and Gertler (2000), and another is the micro approach. The latter is based on four types of data sources: on surveys directed at companies, on microdata compiled in each reporting establishment and for each variety of article by the agency in charge of generating the consumer price indicator, on scanner data and more recently web scraped data. This paper focuses on the micro approach using scraped online data.

Three studies have analyzed the patterns of price setting of Chile, evidence from 1999-2005 has been reported by Medina et al. (2007) using micro data provided by the national statistical office, and has been recently updated by Canales and López (2021) with information up to March 2020. Both studies find substantial heterogeneity in frequency of adjustment across product divisions. Chaumont et al. (2011) provide evidence with scanner data 2007-2009 in the supermarket industry, coverage of 7.23% of the CPI basket.

Price setting analysis based on webscraping techniques.– This paper contributes to up-date the facts for Chile but with online or web scraped data. One of the first one to use online data to construct a price index in Chile was Cavallo (2013) and due to its good performance and it has proven to be a representative source of information. Online data has proved to provide many advantages in particular during the social distancing episode. The practice of social distancing means staying at home and away from others as much as possible to help prevent of Covid-19. Online data also has advantages in the context of natural disasters, for more details see Cavallo et al. (2014), where they analyze the price setting behavior after the earthquake in Chile in 2010 and in Japan 2011. Notwithstanding, online data have also limitations. Even when retail sales are carried online, their services could stop operating during crisis, so there is no way to scrape data.

The applications in empirical literature of online or web scraped data are in general three-folded and include: (1) inflation measurement, (2) inflation forecasting (including nowcasting) and (3) the analysis of micro price setting mechanism. In 2008, the Billion Prices Project was created at MIT. It has remained the largest project focused on web scraping and online prices analysis, for more details see

Cavallo and Rigobon (2016). The usage of big data in Central Banks also is becoming a common practice. In addition to the above mentioned PRISMA led by ECB. Some central banks have used big data tools to track prices online. Macias and Stelmasiak (2019), they evaluate the ability of web scraped data to improve nowcasts of Polish food inflation. They also explore product selection and classification problems, their importance in constructing web price indices and other limitations of online datasets. The Sveriges Riksbank, Hull et al. (2017), they developed an automatic internet data collection process to collect sales prices daily for selected fruits and vegetables from a number of Swedish online retailers. Their results indicate that the information from the daily data could increase the precision in short-term inflation forecasts in Sweden.

Once we have mentioned the penetration that web scraping has had in central banking as a great ally of big data, we want to use its benefits to provide evidence on price setting and inflation during Covid-19 pandemic lockdowns. A lockdown implies that many of typical products that people buy are no longer available. During this period there are changes both on the supply and on the demand side that could be reflected in prices, generating either inflationary pressures or deflation. Cavallo (2020) has found that consumers spend relatively more on food and other categories with rising inflation, and relatively less on transportation and other categories experiencing significant deflation. The coronavirus pandemic has radically changed what households are buying. People are spending a relatively more money on food at grocery stores and less on transportation, healthcare and clothes. Also, since overall spending is down, housing costs now represent a larger share of most people's expenditures; see Diewert and Fox (2020), Carvalho et al. (2020), Dunn et al. (2020) and Cavallo (2020) for evidence of dramatically changing consumer expenditure patterns arising from the pandemic.

Similarly, by using scanner data from UK supermarkets covering the Great Lockdown episode Jaravel and O'Connell (2020) for fast-moving consumer goods also report a sharp decline in product variety and a reduction in promotions.

3 Data description

The "Online Prices Project" was initiated by in June 2019 by the Central Bank of Chile, with the objective of collecting public prices posted online by main retailers in Chile. At the initial stages of the project the aim was to construct an index based on fast-arriving online prices of food and alcoholic beverages and other fast moving consumer goods usually sold in supermarkets. By December 2019, the scope of products was expanded and started to scrap the prices of other product such as clothing, footwear, medicines as well as vehicles. In this paper we will focus on goods typically available in two supermarkets, a pharmacy and ten car dealers where we have collected for more than a year of daily data.

Although online transactions are still a small fraction of all retail sales in most countries, we are confident that the data collected online also provides information about offline prices and product availability. In Chile delivery services are widely used and moreover, due to the health crises derived from Covid-19 the switch towards on-line shopping was more widespread among consumers than usual in order to avoid social contact.[6]

3.0.1 Products and retailers selection

To select which products to track, we comply the criteria suggested by **?**: (1) the product should be from a retailer with a high market share and (2) this retailer should exploit different channels to sell their products (multi-channel). We only track official pages of retailers and not of intermediaries.

With this in mind, for grocery products we track prices from the main largest retailers in Chile. These are the two main players in the retail for food and supermarket supplies with operations along the country. Both account for more than 70% market share. For health products we obtain prices from the main pharmaceutical retailer in Chile for 25% of the market share in the pharma market. For car dealers we have collected ten models of passenger cars and sport utility vehicle (suv).

Another element to consider when selecting the goods to track is the definition of variety and product. The National Statistical Office (Instituto Nacional de Estadística - INE) defines a variety as "a good or service that forms the basic or elemental unit of the IPC basket and is defined according to a set of attributes or pre-established specifications, such as the brand, description, size, content, packaging and provenance, among other specific characteristics", while a product is the "generic term to refer to a good or service that it has a defined purpose and for which it has an expense weight. In terms of the CPI, it is the representation of an elementary aggregate". A product is made up of varieties, as an example, we have the Pisco product that has Pisco 35 grades or 40 grades; from 700 to 1000 cc varieties.

[6]For a discussion on the representativeness of on-line versus off-line data and implications see Cavallo (2017).

3.0.2 Web scraping

We have collected information through web scraping techniques conducted in Python using Selenium, Beautiful Soup as well as auxiliary libraries to fetch and process data. We collect online prices every day and we perform data acquisition while minimizing the burden on web stores owner. We build an automated procedure that on a daily basis scans the retailers web and searches for the web code associated with each product variety and collects the price, and other characteristics and stores this information. As illustrated in **figure 1** we proceed in three steps:

- **First step.** Each day, between 10:00 p.m and midnight, we download from each selected url the information on the varieties and prices.

- **Second step.** The code finds the relevant information. An example of a product-variety in our data is "Brand Name, Pisco, 1 liter". The scraping software automatically collects data on every single product on display on the retailer's website each day. We only extract goods are listed on the website if they are in stock, are available for on-line sale and that are not labelled as "agotado", that is out of stock. Some out of stock items are labelled as such or they immediately disappear from the website, and only reappear on the day that they are offered for sale again (if ever). We can therefore build statistics to study how the set of goods available for purchase changes over time.

- **Third step.** The code stores the following information for each product: its price, the unit of measure (e.g. kilograms, millimeters,..), the product description, whether it is under a promotion, the *Stock Keeping Unit* (SKU) and the date. The price includes taxes and sale discounts, to account for the final price paid by the consumer. We do not include shipping/delivery costs, which may vary according to the location of the purchaser and are an exclusive feature of the online purchasing experience. In addition, missing prices in the data were often caused by scraping errors on days when the software fails to automatically start.

In sum, the collection of data follows the best practices for statistical collection delaying the access to time frames where little traffic is expected and we use Python given the large amount of information we handle and its flexibility. In **table 1** we summarize the coverage by division compared with the official methodology.

3.0.3 Data cleaning and filters

When analyzing prices at high frequencies, it is observed that the *actual* prices they often temporarily move away from a slow-moving trend line called the *regular* price, but after a temporary price change, the nominal price often returns exactly to its pre-existing level. These distinctive features imply that even though an individual price series has a great deal of high-frequency price flexibility (the *actual* price changes frequently), the series also has a great deal of low-frequency price stickiness (the *regular* price changes infrequently).

Before we can address how to treat sales in constructing a measure of price stickiness, we must first be able to accurately identify sales. Sales or promotions are short-lived price changes that are reverted. Constructing a dependable algorithm for identifying sales has proven challenging for researchers. Researchers often turn to other methods of detecting temporary price changes, such as asserting that sale prices have a particular shape. A common identifier of sales is V-shaped temporary discounts, similar to those employed by Nakamura and Steinsson (2008). We also apply the running mode filter proposed by Kehoe and Midrigan (2015)'s algorithm categorizes each price shift as either temporary or regular, based on each price's relative position to the mode price over a given window of time within a particular time series; notably, the algorithm differentiates between price increases and price decreases.

To account for missing information we take the price information from previous periods. For the price index calculations, if a good disappears and later reappears in our sample, we fill missing prices with the previously available values until a new price is observed within 150 days.[7]

4 Empirical findings

The data sample was broken into four periods labelled as follows: (1) "pre-shocks", that goes from June 2019 up to 18 of October, (2) "social outbreak" from mid October till mid-March, (3) "pandemia" from 18 March till end of June and (4) from the "gradual uplift" of mobility restrictions in July till November 2020. The dataset includes price information of food, drinks, toiletries, cleaning products and household equipment.

We end up with three price series for each product-variety, that is, (1) the raw dataset, (2) the price filtered following Nakamura and Steinsson (2008) (NS) and (3) the price filtered series using Kehoe and Midrigan (2015) (KM). Our collected data allows us to observe that one of the retailers under study follows a more dynamic pricing strategy. We identify retailers as R_i, $i = 1, 2, 3$ and 4. To calculate the different measures of frequency of price changes each product is weighted by

[7]During the pandemia a substantial amount of goods were not being offered or taking longer to reappear, this parameter has been revised at each update.

the number of observations, for more details on the calculations see the appendix.

Product availability.– One of the most remarkable features along the sample is the decline in on-line product availability across division and groups, see **figure 2**. In the sample of our two supermarkets we observe a decline of 55.7% on average with respect to the average available before March 18.[8] In spite of the different nature of the shock, we can compare this drop with the evidence on natural disasters provided by Cavallo et al. (2014) for Chile and Japan. They show that the earthquake that took place in Chile in 2010 had an immediate impact on product availability. A large share of goods went out of stock within days. The fall was gradual but larger in Chile (than in Japan), where the number of products available fell by 32% in the first two months after the earthquake, recovering slowly after that. Supply shocks originate in the destruction of production capacity and the disruption of supply chains. Demand shocks are linked to the panic that consumers may experience after natural disasters, with people rushing to the stores to purchase basic necessities and hoarding goods that they fear could become unavailable in the days to come.[9]

The lowest availability of on-line products was registered during April and by the end of the month it began to recover, but without reaching the levels shown at the beginning of March. The recovery in availability occurred in food and beverages, while technology and white line products continued at lower levels, with respect to those registered at the beginning of data collection. This period of less availability coincides with the months of greatest imputation by the INE.[10]

We also have information as regards the brands supplied in each supermarket as well as the number of varieties supplied by each brand. We observe that there is a reduction in the number of varieties offered by each supplier but we do not observe such a decline in the number of suppliers. This points to a decline in the varieties possibly with the aim of focusing their production on their basic products, see **figure 3**.

Analytical classification of goods.– In order to capture differentiated behaviour we classify product-varieties according to their nature: emergency, whether it is perishable, nonperishable or durable. A first group include

[8]This reduction is computed by comparing the average number of items available before March 18, the date in which the state of alarm was declared and confinement measures where introduced in Chile.

[9]During the first days after the state of alarm and given the increased demand in fear of shortages or the derived uncertainty some retailers focused on essential products, in a way of simplifying the delivery process.

[10]During the pandemia the National Statistical Office faced several problems to collect the information of prices and used imputation methods.

"emergency goods": pasta, water, rice and flour. A second group of "perishable goods": eggs, meat, milk, cheese, bread, juice and yogurt. Nonperishable goods: such as powdered milk, mashed potatoes, sodas. Finally durable goods: beds, washing machine, refrigerators, stoves, water heaters, clothes iron, TVs; a similar classification is used in Cavallo et al. (2014).

Notwithstanding, there are category level differences and with different underlying explanations. Some (basic) products may see a higher demand than usual as consumer gets scared and suppliers have little time to react by increasing supply and changing prices. We found goods that during Covid-19 pandemic disappear quickly from the stores are mostly perishable and nonperishable items that people are likely to hoard in fear of future supply disruptions, that is to say, a focus essential goods. Additionally, we classify whether goods are imported as we have information of the brand. We also identify high and low priced items in narrowly defined products.

Frequency of prices adjustments.– The pricing behaviour is different along four periods: "pre-shocks", "social outbreak", "pandemia" and "gradual uplift". While we observe a lot of action in pricing during the "pre-shocks" period, there is a substantial decline in the frequency of price changes (*extensive margin*), across types of products, possibly explained by a reduced activity in promotions. In all periods, price increases are more frequent than price decreases. According to our results, see **table 2**, prices change on average every 2.2 months. Although there is substantial heterogeneity across divisions under study. The frequency of price changes in food and beverages is 30.8 percent (i.e. on average, in a given month, 30.8 percent of prices are changed) and the average (implied) duration of a price spell is 3.0 months. These statistics show a marked differences during the four sample periods and across divisions. With differentiated intensities they show a common pattern, during the shock periods the frequency of price changes declined and the implied duration of prices increased. This are calculations based on the filtered series and given the important amount of missing products we have to be somewhat careful in drawing conclusions. These results, in terms of division heterogeneity are broadly in line with the recent up-date with ONS micro prices by Canales and López (2021) with information up to March 2020, but direct comparison cannot be made as there are substantial differences in terms of product-variety coverage. An additional observation is that prices adjust more frequently than in advanced economies.

Size of price changes.– **Figure 5** shows the distribution of price changes according to their magnitude (*intensive margin*) considering four periods. In all of them, there is a considerable preponderance of variations of low magnitude. As regards the size of the changes, the increases are on average of bigger size than the reduction in prices. The distribution of size changes, in general, shows a large mass of small price increases being of smaller size during the shocks. Notwithstanding, this is not the case for the health division where we can observe a higher dispersion in price changes during the pandemia.

Price Indices.– We check the ability of the *webscraped* data to follow the aggregate dynamics of inflation. To construct price indexes for each division group pair, as well as an overall index, we use the methodology employed by the National Statistical Office (NSO) in terms of weighting. The NSO uses fixed weights and the current base year is 2018. This methodology requires the observation of all product varieties in all periods, and if not observed, missing prices are imputed.

The joint behaviour in the frequency of price changes and the size of these changes explains CPI developments. For the product categories for which we have collected information we compare the inflation rates with the official information and our sample matches quite well in some divisions, see **figure 6**.

Given the exceptional circumstances along this year, we may consider to deliver an alternative measure by changing the weights to take into account the actual consumption as proposed by Cavallo (2020). He proposes to modify the basket weights based on credit card expenditure, under this circumstances it may provide a more realistic view based on those products which are demanded and can be purchased.

Online vs. offline prices.– To get a sense on the differences between offline and online prices we compare the prices we obtained from large "multi-channel" retailers (that sell both online and offline) to the micro data collected by National Statistical Office (NSO) used to compute the official CPI.[11,12]

The NSO relies on a large number of data observations per division-group, however, retailer and product details remain confidential and is available with certain lag. The webscraped information, while restricted to fewer observations is available more promptly. As the ONS dataset provides anonymized information of the products we cannot compare whether there are discrepancies between online and offline prices. But to check the degree of similarity, we compare the price distributions and the price evolution of selected division-groups **figure 7**. We find that usually webscraped data covers a wider range of prices than the one collected by the ONS price collectors. This is because that the data collector has also a sense

[11] The granular information is available at https://www.ine.cl/estadisticas/economia/indices-de-precio-e-inflacion/indice-de-precios-al-consumidor.

[12] Cavallo (2017) explores in a more systematic way this issue by making use of crowdsourcing platforms and mobile phone app to compile offline prices and compare levels of online prices in a wide range of countries. He finds that price levels are identical 72% of the times.

on which goods are usually bought and disregards from the sample those products that she is aware that they are not usually sold. This is usually biased towards low and medium priced goods within a variety, leaving aside highly priced products and probably not usually bought by consumers.[13]

5 Conclusion

This paper characterizes the dynamics of price adjustments in Chile using micro-level data. We make use of web scraping techniques to collect prices from public websites of the main retailers including supermarkets, pharmacy retailer and car dealerships. The internet provides an additional source of information to closely monitor price developments within the country and the use of Big data tools allows it to be obtained. Although the time span is still short, the preliminary results answer questions about pricing decisions under two major exogenous and unanticipated shocks faced by the Chilean economy, the social outbreak and Covid-19.

We compute the frequency of price adjustments for 22% goods included in the CPI basket, as well as measures of the frequency and size of price changes. We found that there was a significant reduction in product availability and that they are recovering at a slow pace. This reduction occurred after March 18, the date in which the state of alarm was declared and confinement measures in Chile due covid-19. The goods that disappear quickly from the stores are mostly perishable and nonperishable items that people are likely to hoard in fear of future supply disruptions, that is to say, a focus on essential goods. During the social outbreak the reduction in online availability was not significant.

After the two shocks prices became moderately more sticky. According to our pre-shock filtered sample the weighted average frequency of price changes within a month was 39%, dropping to 30% during the social outbreak and at 36% during the pandemia. Alternatively, the direct computation of the duration of price spells indicates that average implied duration is slightly above 2 months, and reaching 3 months afterward. In sum, prices adjusted at a lower frequency, there were fewer sales and discounts and the size of changes became smaller, possibly due to the "fear of anger" by retailers.

Online data have many advantages in the context of natural disasters or under exceptional circumstances such as the pandemia. It was possible to collect the data remotely, in real time, and without requiring any resources from the retailers

[13] One of the limitations of on-line data is that there is no information on expenditure, leading to potential measurement errors and bias in some calculations, based on the price retail distribution Antoniades et al. (2019) propose a methodology to impute expenditure shares.

involved. But there are also important limitations that need to be kept in mind. We have been able to collect information for relatively few retailers and still few categories of goods are currently available online compared to off-line. The underlying behaviour of retailers as simply by stopping to update prices and stock information online, voluntarily suspending their online service during some time may add some concerns on its representativeness. The reliability of online data sources under extreme circumstances needs to be determined, and will likely depend on the country and context.

As regards the way ahead, the data collection under the "CPI Project" will continue and we will broaden categories, such as clothing and footwear. It would be interesting for future research to tackle issues such as the challenges derived from changes in consumption patterns during and after Covid-19. This has led to a vivid discussion whether weights should be modified at a higher frequency as in Cavallo (2020). An additional avenue could be to explore the degree of exchange rate pass-through at retail level, as along this period, the Chilean Peso depreciated against the USD. As we track the prices of individual goods we could also explore the role of exchange rate pass-through further downstream in the supply chain. We can study the extent to which the price increases faced by importers passed through into higher retail prices or were instead absorbed by lower retailer profit margins.

References

Alvarez, L. J. and I. Hernando (2006): "Price setting behaviour in Spain: Evidence from consumer price micro-data," *Economic Modelling*, 23, 699–716.

Antoniades, A., R. C. Feenstra, and M. Xu (2019): "Using the Retail Distribution to Impute Expenditure Shares," *Under Revision AER.*

Canales, M. and B. López (2021): "Dinámica de precios en Chile: Stylized Facts from Individual Consumer Prices," *Documento de Trabajo, Banco Central de Chile*, XXX.

Carvalho, V., S. Hansen, Ortiz, J. Ramón García, T. Rodrigo, S. Rodriguez Mora, and J. Ruiz (2020): "Tracking the COVID-19 Crisis with High-Resolution Transaction Data," *Cambridge-INET Working Paper Series, University of Cambridge.*

Cavallo, A. (2013): "Online and official price indexes: Measuring Argentina's inflation," *Journal of Monetary Economics*, 60, 152 – 165.

——— (2017): "Are Online and Offline Prices Similar? Evidence from Large Multi-channel Retailers," *American Economic Review*, 107, 283–303.

——— (2020): "Inflation with Covid Consumption Baskets," Working Paper 27352, National Bureau of Economic Research.

Cavallo, A., E. Cavallo, and R. Rigobon (2014): "Prices and Supply Disruptions during Natural Disasters," *Review of Income and Wealth*, 60, 449–471.

Cavallo, A. and R. Rigobon (2016): "The Billion Prices Project: Using Online Prices for Measurement and Research," *The Journal of Economic Perspectives*, 30, 151–178.

Chaumont, G., M. Fuentes, F. Labbé, and A. Naudón (2011): "Dinámica de precios en Chile: evidencia con datos de supermercado," *Documento de Trabajo, Banco Central de Chile*, 642.

Diewert, W. E. and K. J. Fox (2020): "Measuring Real Consumption and CPI Bias under Lockdown Conditions," Working Paper 27144, National Bureau of Economic Research.

Dunn, A., K. Hood, and A. Driessen (2020): "Measuring the Effects of the COVID-19 Pandemic on Consumer Spending using Card Transaction Data," *U.S. Bureau of Economic Analysis Working Paper WP2020-5.*

Gali, J. and M. Gertler (2000): “Inflation Dynamics: A Structural Econometric Analysis,” Working Paper 7551, National Bureau of Economic Research.

Hull, I., M. Löf, and M. Tibblin (2017): “Price information collected online and short-term inflation forecasts,” *IFC-Bank Indonesia Satellite Seminar on “Big Data” at the ISI Regional Statistics Conference 2017*, Bank for international settlements.

Jaravel, X. and M. O’Connell (2020): “Real-time price indices: Inflation spike and falling product variety during the Great Lockdown,” *Journal of Public Economics*, 191, 104270.

Kehoe, P. and V. Midrigan (2015): “Prices are sticky after all,” *Journal of Monetary Economics*, 75, 35–53.

Macias, P. and D. Stelmasiak (2019): “Food inflation nowcasting with web scraped data,” *Narodowy Bank Polski*, 302.

Medina, J., D. Rappoport, and C. Soto (2007): “Dinámica de ajuste de precios: evidencia de datos microeconómicos para Chile,” *Economía chilena, Vol. 10, Nº. 2, 2007, pags. 5-26*, 10.

Nakamura, E. and J. Steinsson (2008): “Five Facts about Prices: A Reevaluation of Menu Cost Models,” *The Quarterly Journal of Economics*, 123, 1415–1464.

——— (2013): “Price Rigidity: Microeconomic evidence and macroeconomic implications,” *Annual Review of Economics*, 5, 133–163.

6 Figures

Figure 1:
STYLIZED DATA COLLECTION

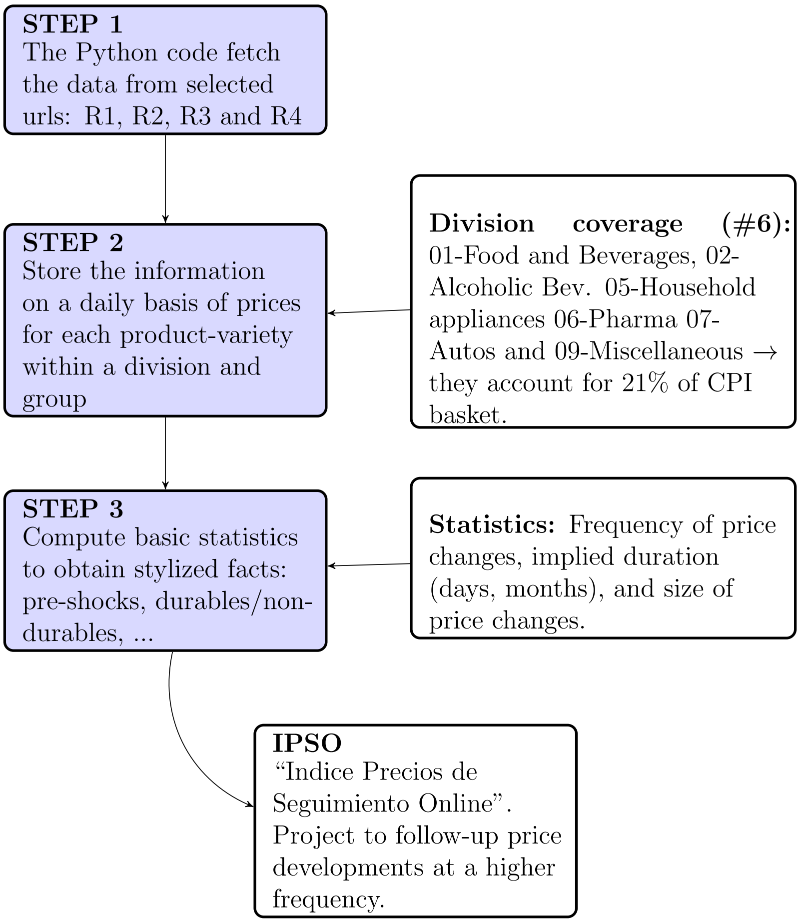
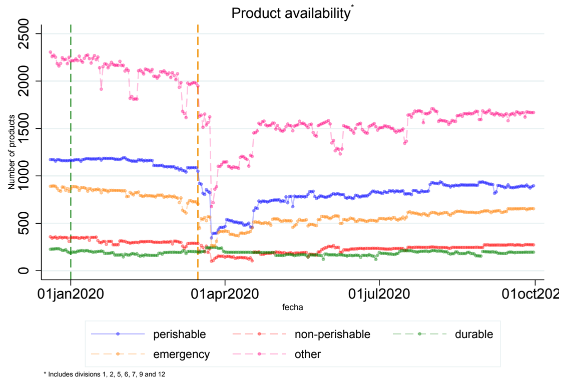


Figure 2:
PRODUCT AVAILABILITY

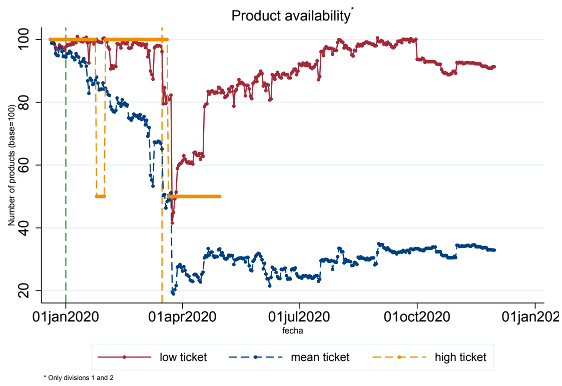
(a) Number of prices quoted

(b) By product category



(c) Number of price quoted
(Dec_2019=100)

(d) By product category
(Dec_2019=100)

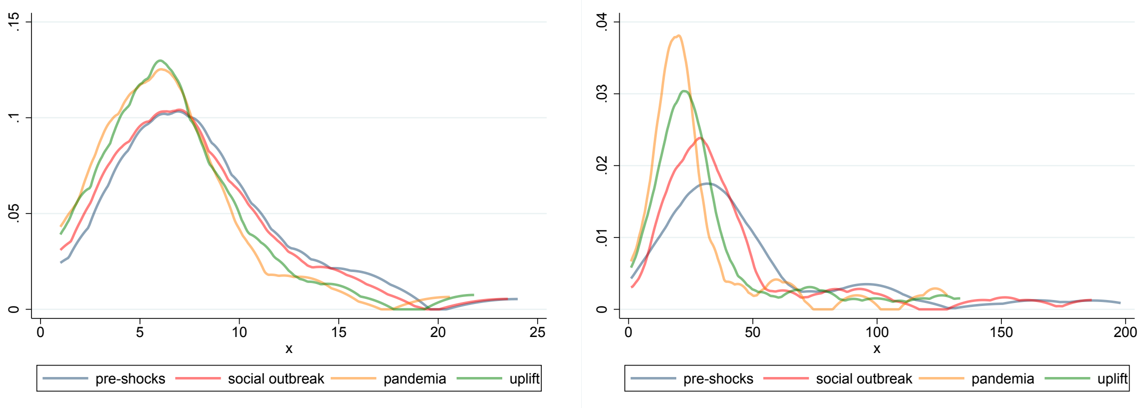


Note: In **panel (a)** we plot the evolution of the total number of varieties quoted on-line. While during the social outbreak product availability remained almost unchanged, since the start of COVID-19 the number of varieties decreased substantially and they are recovering slowly, in spite the uplift of mobility restrictions, but still behind pre-covid levels. In **panel (b)** we decompose the number of quoted varieties based on their analytical classification. In **panel (c)** we re-base the number of prices quoted to 100 in December 2019 and we can observe that the quotes remain below pre-crisis levels. In **panel (d)** we breakdown into high ticket, average ticket and low ticket goods within each group.

Source: "Online Prices Project" Central Bank of Chile.

Figure 3:
Number of varieties

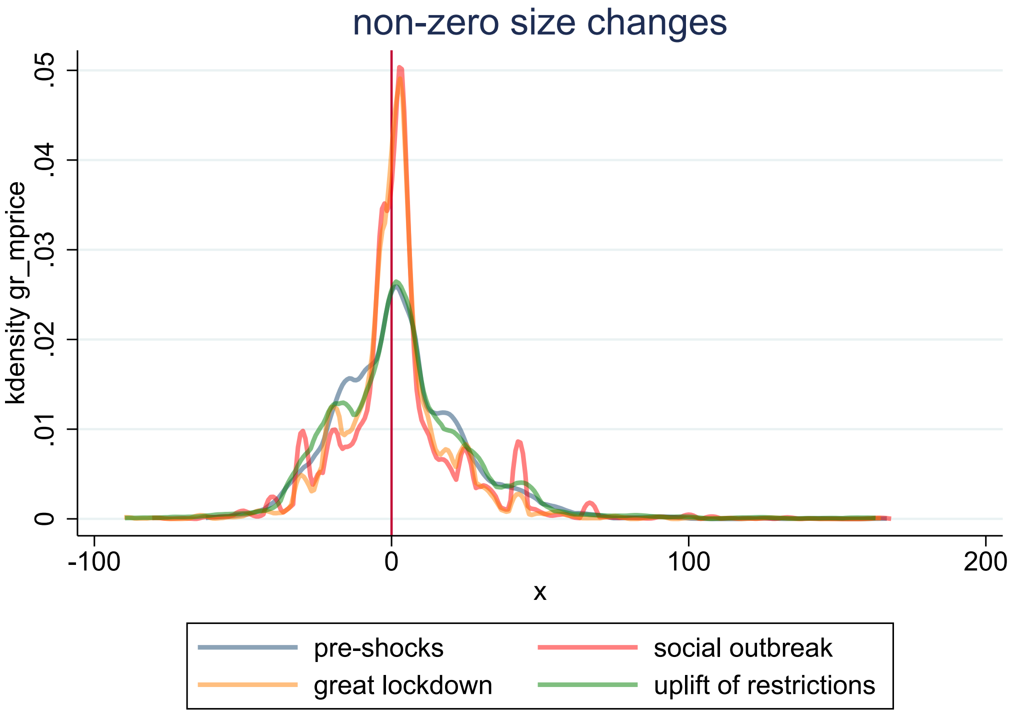
(a) Number of brands (b) Number of product varieties



Notes: Figure in **panel (a)** plots the distribution of the number of brands that supply each product. In **panel (b)** the number of varieties being offered by each product. There is a reduction on the number of brands and even more acute in the number of varieties.

Source: "Online Prices Project" Central Bank of Chile.

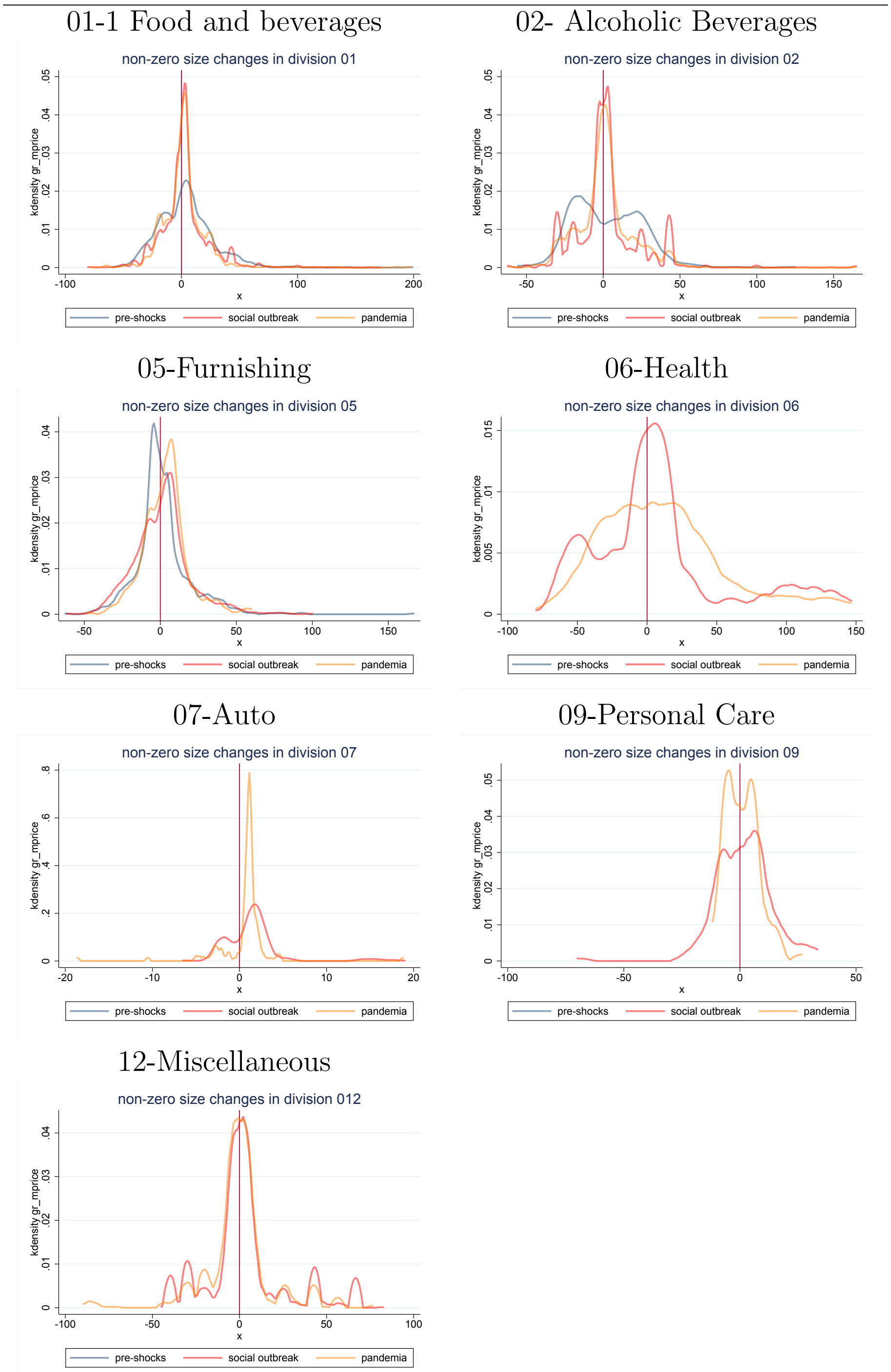
Figure 4:
SIZE OF PRICE CHANGES



Notes: The size of price changes are tilted towards small positive changes. During the pandemia episode the distribution is even more skewed towards small changes.

Source: "Online Prices Project" Central Bank of Chile.

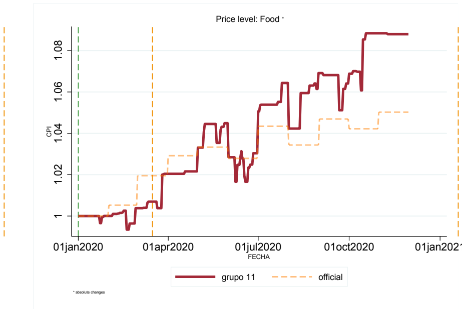
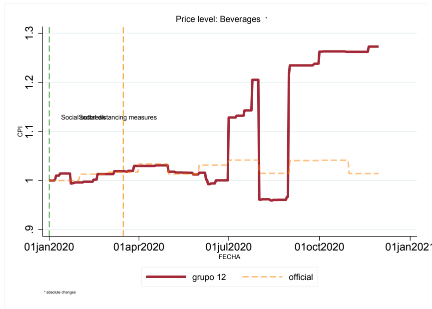
Figure 5:
Price changes by COICOP 1 digit

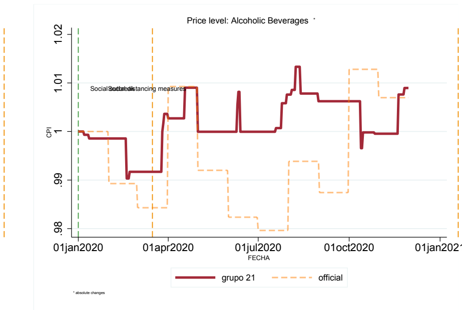
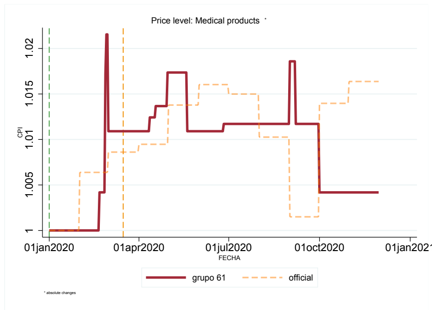


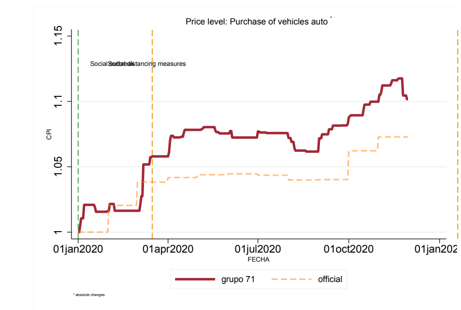
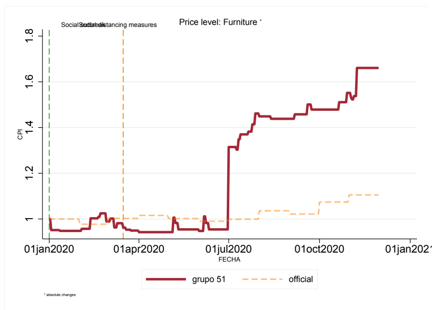
Notes: The distribution (non-filtered series) as regards size and sign of price changes has changed. In general, after the pandemia price changes tend to be smaller and of a positive sign. It seems that part of the activity due to sales and promotions.

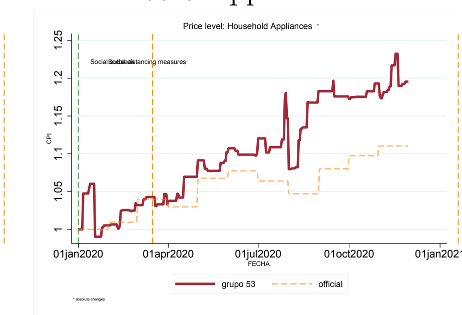
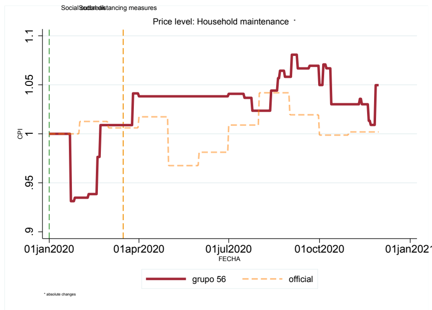
Source: "Online Prices Project" Central Bank of Chile.

Figure 6:
CPI BY COICOP 2 DIGIT

01-1 Food	01-2 Beverages
	

02-1 Alcoholic B.	06-1 Medical
	

07-1 Vehicles	05-1 Furniture
	

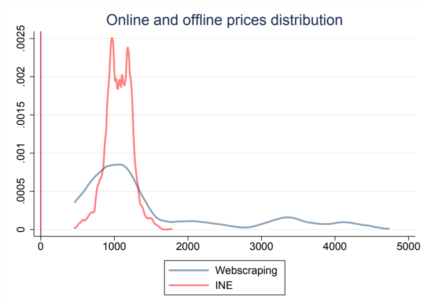
05-3 Appliances	05-6 Routine maintenance
	

Notes: Comparison between INE official index and the index built by using web scraped data re-based to 1 in January 2020.

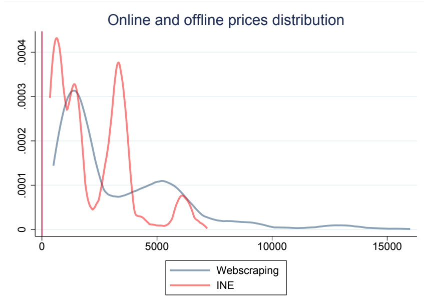
Source: "Online Prices Project" Central Bank of Chile and INE Chile.

Figure 7: Online vs. offline prices at Product Level

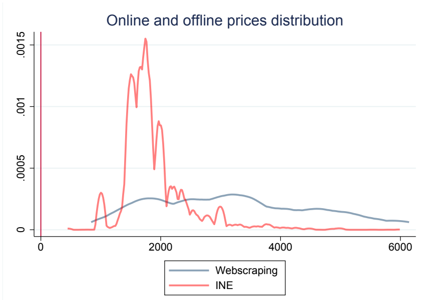
01-1 Food
RICE



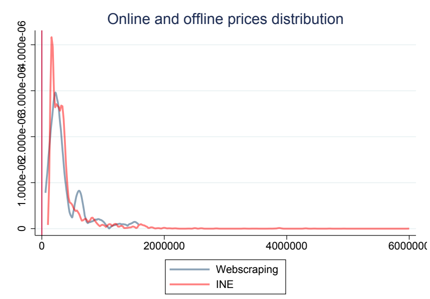
02-1 Alcoholic B.
BEER



12-1 Personal Care
TOOTHPASTE



09-1 Recreation and culture
TELEVISION



Notes: In these graphs we explore the representativeness of the data collected by the NSO and our webscraped dataset. Webscraped data collects data from all the products quoted in the websites while INE data collectors drop from the data collection those products which are not so common in a consumer basket.

Source: "Online Prices Project" Central Bank of Chile and INE Chile.

7 Tables

Table 1:
Analytical classification and CPI weights

coicop 1-digit	Type: not class.	perishable	emergency	durable	service	energy	Total	Cov.
01. Food and beverages (n=76)	12.6	6.0	0.7	0.0	0.0	0.0	19.3	12.1
02. Alcoholic Beverages (n=8)	4.8	0.0	0.0	0.0	0.0	0.0	4.8	2.88
03. Clothing and Footwear (n=28)	3.4	0.0	0.0	0.0	0.1	0.0	3.5	-
04. Housing (n=16)	7.7	0.0	0.0	0.0	7.1	0.0	14.8	-
05. Furnishings (n=36)	1.5	0.0	0.0	2.3	0.0	2.6	6.5	0.87
06. Health (n=22)	7.8	0.0	0.0	0.0	0.0	0.0	7.8	2.43
07. Transport (n=24)	10.3	0.0	0.0	2.9	0.0	0.0	13.1	2.86
08. Communication (n=6)	5.5	0.0	0.0	0.0	0.0	0.0	5.5	-
09. Recreation (n=37)	6.6	0.0	0.0	0.0	0.0	0.0	6.6	-
10. Education (n=11)	6.6	0.0	0.0	0.0	0.0	0.0	6.6	-
11. Restaurants (n=7)	6.4	0.0	0.0	0.0	0.0	0.0	6.4	-
12. Miscellaneous (n=32)	5.2	0.0	0.0	0.0	0.0	0.0	5.2	xx
Total (n=303)	78.3	6.0	0.7	5.1	7.2	2.6	100.0	21.1

Note: Analytical classification of CPI groups/varieties are provided by INE. In brackets the number of groups within each division. Column indicates the coverage of the webscraped sample.

Source: INE Chile (National Statistical Office).

Table 2:
FREQUENCY OF PRICE CHANGES

	1. Food and Beverages – 12.2% (19.5%)					
	raw		***NS***		***KM***	
	mean	Median	mean	Median	mean	Median
Daily frequency	1.8	1.5	1.2	1.1	1.2	1.1
Monthly frequency	40.0	36.8	29.6	28.9	28.4	27.6
Implied duration (days)	69.3	65.4	95.3	88.1	100.0	92.8
Implied duration (months)	2.3	2.2	3.2	2.9	3.3	3.1
	2. Alcoholic Beverages – 2.8% (4.8%)					
	raw		***NS***		***KM***	
	mean	Median	mean	Median	mean	Median
Daily frequency	2.4	2.4	1.4	1.1	1.2	1.1
Monthly frequency	49.2	51.6	32.1	27.8	29.6	27.0
Implied duration (days)	56.3	41.4	94.7	92.0	103.1	95.2
Implied duration (months)	1.9	1.4	3.2	3.1	3.4	3.2
	5. Furnishing – 0.8% (6.5%)					
	raw		***NS***		***KM***	
	mean	Median	mean	Median	mean	Median
Daily frequency	2.7	2.6	2.2	2.1	1.9	1.9
Monthly frequency	53.6	56.1	47.0	48.4	43.3	44.2
Implied duration (days)	53.7	36.5	65.1	45.3	74.5	51.5
Implied duration (months)	1.8	1.2	2.2	1.5	2.5	1.7
	6. Health					
	raw		***NS***		***KM***	
	mean	Median	mean	Median	mean	Median
Daily frequency	0.4	0.2	0.3	0.2	0.3	0.2
Monthly frequency	14.0	8.3	9.9	5.8	9.4	5.9
Implied duration (days)	524.3	348.8	489.7	497.9	504.6	496.5
Implied duration (months)	17.5	11.6	16.3	16.6	16.8	16.6
	7. Transport– Auto					
	raw		***NS***		***KM***	
	mean	Median	mean	Median	mean	Median
Daily frequency	2.8	2.2	2.7	2.3	2.6	2.2
Monthly frequency	53.7	48.4	54.1	50.1	53.7	49.1
Implied duration (days)	42.5	45.4	39.6	43.1	40.1	44.5
Implied duration (months)	1.4	1.5	1.3	1.4	1.3	1.5
	9. Recreation and Culture– TVs					
	raw		***NS***		***KM***	
	mean	Median	mean	Median	mean	Median
Daily frequency	3.5	2.9	1.3	1.3	1.1	1.1
Monthly frequency	61.2	58.4	31.9	31.5	27.9	27.7
Implied duration (days)	35.5	34.2	92.5	79.2	156.4	92.5
Implied duration (months)	1.2	1.1	3.1	2.6	5.2	3.1

Note: Calculation of frequency of price changes and implied duration on *raw* prices and filtered series using Nakamura and Steinsson (2013) methodology (NS) and Kehoe and Midrigan (2015) methodology (KM).

Source: Own calculations.

Table 3:
FREQUENCY OF PRICE CHANGES BY PERIOD

	1. Food and Beverages – 12.2% (19.5%)				
	full sample mean	***pre-shocks*** mean	***social outbreak*** mean	***pandemia*** mean	***uplift*** mean
Daily frequency	1.2	1.4	1.2	1.1	1.2
Monthly frequency	29.6	32.7	30.2	28.2	27.4
Implied duration (days)	95.3	81.1	89.8	102.9	107.3
Implied duration (months)	3.2	2.7	3.0	3.4	3.6

	2. Alcoholic Beverages – 2.8% (4.8%)				
	full sample mean	***pre-shocks*** mean	***social outbreak*** mean	***pandemia*** mean	***uplift*** mean
Daily frequency	1.4	1.9	1.1	1.0	1.4
Monthly frequency	32.1	41.4	28.3	25.3	33.4
Implied duration (days)	94.7	75.8	99.5	119.4	84.2
Implied duration (months)	3.2	2.5	3.3	4.0	2.8

	5. Furnishing– 0.8% (6.5%)				
	full sample mean	***pre-shocks*** mean	***social outbreak*** mean	***pandemia*** mean	***uplift*** mean
Daily frequency	2.2	3.3	2.2	1.2	2.0
Monthly frequency	47.0	58.5	47.7	32.6	47.2
Implied duration (days)	65.1	39.3	50.7	113.7	63.6
Implied duration (months)	2.2	1.3	1.7	3.8	2.1

	6. Health–			
	full sample mean	***social outbreak*** mean	***pandemia*** mean	***uplift*** mean
Daily frequency	0.3	0.5	0.2	0.1
Monthly frequency	9.9	18.3	6.7	5.2
Implied duration (days)	489.7	164.2	497.2	806.4
Implied duration (months)	16.3	5.5	16.6	26.9

	7. Transport –New autos			
	full sample mean	***social outbreak*** mean	***pandemia*** mean	***uplift*** mean
Daily frequency	2.7	3.5	2.1	2.3
Monthly frequency	54.1	65.4	46.9	50.1
Implied duration (days)	39.6	28.3	47.4	43.1
Implied duration (months)	1.3	0.9	1.6	1.4

	9. Recreation and Culture –TVs			
	full sample mean	***social outbreak*** mean	***pandemia*** mean	***uplift*** mean
Daily frequency	2.6	2.5	1.8	3.4
Monthly frequency	53.0	52.4	42.5	64.1
Implied duration (days)	41.3	40.4	54.2	29.3
Implied duration (months)	1.4	1.3	1.8	1.0

Note: Calculation of frequency of price changes and implied duration on filtered series using Nakamura and Steinsson (2013) methodology (NS). The daily frequency reads as follows, each day x% of products change price, and the implied duration is x days. We turn this info into monthly implied duration, and the monthly frequency is the % of goods that change prices within a month.

Source: Own calculations.

Table 4:
Dataset description

	Retailer 1 (R1)	Retailer 2 (R2)	Retailer 3 (R3)	Retailer 4 (R4)
Initial date	18/Jul/2019	12/Dec/2019	1/Dec/2019	4/Dec/2019
Final date	30/Nov/2020	30/Nov/2020	30/Nov/2020	30/Nov/2020
# of Days	502	293	304	301
Product description	✓	✓	✓	✓

Note: Basic information as regards each of the retailers.

Source: "Online Prices Project" Central Bank of Chile.

A Appendix

For each product group/category we proceed as follows:

- We collect the price of product variety i. We collect two type of prices real and actual, the latter takes into account the final price payed by the consumer after sales and discounts are applied.
- We collect additional information on quantity, brand and retailer. We categorize each product-variety to a division and group according to the National Statistical Office classification.

A.1 Variables

In this section we describe the statistics utilized to characterize the behavior of prices at establishment level. In general, the statistics are similar to those used by Alvarez and Hernando (2006). We define the following binary variables for each price quote ijt where i, j and t, denote the establishment, variety (CPI) and time, respectively:

$$DEN_{ijt} = \begin{cases} 1 & \text{if } P_{ijt} \text{ and } P_{ijt-1} \\ 0 & \text{if } P_{ijt} \text{ exists but not } P_{ijt-1} \end{cases} \tag{1}$$

$$NUM_{ijt} = \begin{cases} 1 & \text{if } P_{ijt} \neq P_{ijt-1} \\ 0 & \text{if otherwise} \end{cases} \tag{2}$$

$$NUM\_UP_{ijt} = \begin{cases} 1 & \text{if } P_{ijt} > P_{ijt-1} \\ 0 & \text{if otherwise} \end{cases} \tag{3}$$

$$NUM\_DW_{ijt} = \begin{cases} 1 & \text{if } P_{ijt} < P_{ijt-1} \\ 0 & \text{if otherwise} \end{cases} \tag{4}$$

DEN is a dummy variable indicating that the product i observed in t was also observed in $t-1$. NUM is a dummy variable indicating that the price of product i has changed in t.

A.2 Calculation of basic statistics

On the basis of the above variables we define (at the variety level) the main variables analysed:

- Frequency of price changes:

$$F_j = \frac{\sum_{i=1}^{n_j} \sum_{t=2}^{\tau} NUM_{ijt}}{\sum_{i=1}^{n_j} \sum_{t=2}^{\tau} DEN_{ijt}} \tag{5}$$

- Implied median price duration:

$$T_j^{50} = \frac{ln(0.5)}{ln(1 - F_j)} \tag{6}$$

- Implied mean price duration:

$$T_j = -\frac{1}{ln(1 - F_j)} \tag{7}$$

- Frequency of price increases:

$$F_j^{+} = \frac{\sum_{i=1}^{n_j} \sum_{t=2}^{\tau} NUM\_UP_{ijt}}{\sum_{i=1}^{n_j} \sum_{t=2}^{\tau} DEN_{ijt}} \tag{8}$$

- Frequency of price decreases:

$$F_j^{-} = \frac{\sum_{i=1}^{n_j} \sum_{t=2}^{\tau} NUM\_DW_{ijt}}{\sum_{i=1}^{n_j} \sum_{t=2}^{\tau} DEN_{ijt}} \tag{9}$$

- Duration of price spells:

$$d_j = \frac{\sum_{i=1}^{N_j} \sum_{s=1}^{NS_i} \frac{d_{ijs}}{NS_i}}{N_j} \tag{10}$$

Where N_j is the number of establishment selling product j and NS_i refers to the number of spells observed for establishment i.

B The dataset

Variable description All the .dta files contain the following variables:

- **date:** date is stata %td format.
- **actual price:** product price in CLP after discounts.
- **regular price NS:** product price in CLP after applying filter to account for temporary sales and missing information, based on Nakamura and Steinsson (2008).
- **regular price KM:** product price in CLP after applying filter to account for temporary sales, temporary price increases and missing information based on Kehoe and Midrigan (2015).
- **coicop5:** products are classified according to COICOP classification https://www.ilo.org/public/english/bureau/stat/download/cpi/coicop.pdf by (1) division, (2) group, (3) class, (4) subclass and (5) product.
- **retailer:** R1, R2, R3 and R4.
- **weights:** Each group is weighted according to the basket weights from the National Statistical Office https://www.ine.cl/docs/default-source/%C3%ADndice-de-precios-al-consumidor/metodologias/base-anual-2018-100/metodolog%C3%ADa.pdf
- **unit:** ml, kg, pair, etc.
- **brand:** The brand is used to compute the product varieties supplied by each brand-supplier. Additional, for technological goods such as cars, TVs and other household appliances, the brand serves to identify the country of origin of each product, therefore whether these are imported.
- **classification:**
  - *Emergency goods:* pasta, water, rice and flour.
  - *Perishable goods:* eggs, meat, milk, cheese, bread, juice and yogurt.
  - *Nonperishable goods:* powdered milk, mashed potatoes, sodas...
  - *Durable goods:* beds, washing machine, refrigerators, stoves, water heaters, clothes iron, TVs.
- **ht:** within a group we breakdown the sample into high ticket, and low ticket based on price percentiles (p25,p75) during the pre-shock sample period.