When and Why do Governments Pay More? Evidence from Pharmaceuticals in São Paulo

Michael Carlos Best* Augustin Chaintreau† Joana Naritomi‡
Dimitri Szerman§

October 2024

PRELIMINARY AND INCOMPLETE. DO NOT CITE.

Abstract

This paper explores a unique market-wide view of public procurement using administrative data generated by the mandatory use of electronic invoicing in Brazil to study the São Paulo state procurement of pharmaceutical products. This data allows us to observe private sector transactions by government suppliers, as well as by firms that are not government suppliers but do transact goods purchased by the government. Thus, we can benchmark Business-to-Government (B2G) against Business-to-Business (B2B) transactions, identify the pool of potential suppliers, and determine if and when the government accesses suppliers with better prices. We begin by leveraging recent advances in Natural Language Processing to classify products based on the free-text descriptions in invoices. Then, we describe the circumstances under which governments pay more (or less) for goods in the pharmaceutical market. On average, we find that the government pays 13.6% more than the private sector for the same goods. However, when controlling for supplier fixed effects, this difference drops to -8%, indicating that while the government selects higher-priced suppliers, it successfully negotiates lower prices from those contracted. This suggests that obtaining better value-for-money depends largely on the government's access to more competitive sellers. We also find substantial heterogeneity across products and some evidence that a larger government presence in a market is associated with a smaller supplier selection issue and greater price advantages in public versus private purchases from the same supplier.

JEL CODES: H11, H57, L11, D22, C81, I18

*michael.best@columbia.edu. Columbia University, BREAD, CEPR, IFS & NBER.

†ac3318@columbia.edu. Columbia University.

‡J.Naritomi@lse.ac.uk. London School of Economics, BREAD, CEPR, & IFS.

§dimitrijoe@gmail.com. Amazon Science.

We thank the São Paulo state revenue authority (SEFAZ/SP) for collaborating in this project. No micro-data was shared with academic researchers. All data analysis was implemented by SEFAZ/SP staff. We thank Antonio Sergio Bonato and André Luis Grotti Clemente for invaliable contributions to the project. We thank Pedro D'Angelo, Anna Bottega, Xi Chen, Francesco Brandão Ferreira and Daniel Duque for excellent research assistance.

1 Introduction

Public procurement accounts for 12% of GDP worldwide and is crucial for public service delivery (Bosio *et al.*, 2022). There is a widespread perception that governments are inefficient and corrupt when spending public funds, and the literature has documented large price variations in the purchase of similar goods by governments in different countries (e.g., Bandiera *et al.* 2009, Best *et al.* 2023, Allende *et al.* 2024). However, we still know little about how government purchases compare with similar private transactions—by the same suppliers or by firms that supply the same goods to other buyers but do not participate in government procurement—and how market structure may help explain prices paid by governments.

This paper explores a unique market-wide view of public procurement to investigate whether governments pay more for goods by using the private market as a benchmark. We use administrative data generated by the mandatory electronic invoicing system in Brazil to study the procurement of pharmaceutical products in the state of São Paulo.[1] Electronic invoicing datasets provide detailed data on business-to-business (B2B) and business-to-government (B2G) transactions, as well as public procurement processes. This enables a novel comparison of B2G and B2B transactions for the same products and allows us to gather data on the pool of potential suppliers and market-wide characteristics that may influence prices.

We focus on pharmaceutical products, which represent 8% of the São Paulo state government's public procurement between 2018 and 2023. Pharmaceutical markets are particularly interesting to study. Brazil has a public health system that provides universal healthcare, so health spending is an important part of public spending, with substantive variation in the relevance of the government across product markets. Also, pharmaceutical products are regulated by a national regulatory agency Anvisa (*Agência Nacional de Vigilância Sanitária*), responsible for overseeing the safety and quality of health-related products and services, including pharmaceutical products, which leads to a systematic categorization of product definitions (i.e., active ingredient, presentation and whether the product is a generic drug).

We begin by harmonizing the transaction-level data to ensure that we are analyzing similar purchases by both government and private entities. These comparisons are complicated by the fact that product manufacturing codes are often missing from invoices and the use of non-standardized product information in invoice product descriptions. To address this, we employ artificial intelligence, specifically Natural Language Processing, to organize unstructured text in product descriptions into uniform product categories. Our analysis sample focuses on 2,194 pharmaceutical products that are supplied to the government and private firms in a total of 121,612,293 transactions in 2018-2023, and for which we have a confidence score above 80% in our classification algorithm. The sample includes 12,035 suppliers and 79,644 buyers. The average product has 301 distinct sellers, and 2,304 unique buyers.

Then, we use this data to describe differences in prices when purchased by governments

[1] The state of São Paulo is the largest state in Brazil with a population of 42 million and represents 34% of Brazilian GDP.

versus private firms while flexibly controlling for quantity, time, and product effects. We find that, on average, the government pays 13.6% more for the same good relative to the private sector. To investigate whether this price difference arises from the same supplier applying different prices to the government versus private firms, we run the same regression controlling for supplier fixed effects. Once we control for supplier fixed effects, the average price difference drops and becomes negative (-8%). Therefore, the government seems to select firms that charge higher prices, but it appears that they are able to extract rents from firms that do become government suppliers. Thus, our results suggest that the government's success in achieving better value for money primarily hinges on its ability to connect its procurement process with more competitive sellers.

To understand the mechanisms, we further investigate supplier selection and price differentials within suppliers in public versus private procurement. For off-the-shelf pharmaceutical products, unit price is the only dimension considered in the procurement selection process. Thus, the participation of lower-priced sellers should directly impact value-for-money. To better understand the selection margin, we match the transaction data to procurement processes.[2] We analyze which characteristics of the procurement process affect which firms compete for government contracts and are ultimately selected as government suppliers.

The price decomposition above also suggests that price differentiation within supplier — the same supplier charging different amounts to government versus private-sector buyers — are part of the reason that the government pays less for the same items than private-sector buyers do once we control for supplier fixed effects. However, while the regressions above control for a wide range of factors, we are not able to interpret them causally. An important potential confounder is that winning a government contract may change a supplier's costs of delivering the same product to other buyers. On the one hand, firms may have capacity constraints, meaning that satisfying a large government contract could lead to higher costs when selling to other private-sector buyers (e.g., Kroft *et al.* 2023). On the other hand, receiving large government contracts can help firms grow and benefit from economies of scale (e.g., Ferraz *et al.* 2015 and Carrillo *et al.* 2023), lowering marginal production costs. We seek to estimate the price at which the supplier who wins a given contract would have been willing to sell the same item around the same time to a private-sector buyer. To do this, we propose a Regression Discontinuity (RD) approach: we extrapolate from runner(s)-up in the auctions to estimate the private-sector price the winner would charge had they not won the government contract.

Finally, we examine how this price decomposition varies across markets. The average effects mask substantial heterogeneity across products and market characteristics. When we run the same price regressions weighted by transaction value to obtain a dollar-weighted price differential, we find that the government consistently obtains lower prices than the private sector, and

[2] As we explain in the Institutional Background section, the main procurement system we focus on is electronic procurement auctions, allowing us to observe in detail who participates in these auctions and how the bidding process unfolds. The data also include direct purchases, which are smaller procurement processes that require only a market consultation with three suppliers to justify a selection.

the selection problem disappears. This suggests that there are certain high-ticket items that the government is particularly well-positioned to procure. Next, we conduct the same decomposition analysis—comparing price differences in government versus private procurement with and without firm fixed effects—product by product to explore how the government's market presence influences prices. We find evidence that a larger government presence in a market is associated with a smaller supplier selection issue and greater price advantages in public versus private purchases from the same supplier.

This paper contributes to the literature on public procurement policy by benchmarking government purchases against private transactions, examining the relevance of supplier selection, and assessing the extent to which governments can extract rents from suppliers based on the government's market share as a buyer. This market-view perspective on public procurement is rare, as datasets containing comparable purchase data for both government and private-sector buyers are typically unavailable. Most studies on public procurement focus solely on public-sector purchases (e.g., Allende *et al.* 2024, Best *et al.* 2023, Bandiera *et al.* 2021, Bandiera *et al.* 2009). A few notable exceptions include Duggan & Scott Morton (2006), studying US pharmaceuticals procurement at the market level, Atal *et al.* (2024) studying the entry of public pharmacies in Chile, Kroft *et al.* (2023) studying the construction industry in the US, and (Carrillo *et al.*, 2023) studying small works projects in Ecuador. However, none of these studies have access to micro-level data across entire markets, nor can they observe the same firm supplying to both government and private sectors to fully identify the sources of price wedges.

2 Data & Measurement

In order to estimate the difference in the prices paid by the government and the private sector for the same goods at the same time, we need to overcome two challenges. First, we require a dataset of comparable purchases by both government and private-sector buyers. This is typically unavailable. Almost all studies of public procurement use data on public purchases only.[3] For this, we leverage unique access to the government's database of the universe of electronic invoices, which spans both public- and (formal) private-sector buyers. Second, we need to ensure that when we compare public purchases to private purchases, we are able to hold constant the exact item being purchased. In order to do this, we exploit the fact that the invoices contain textual descriptions of the precise items being purchased, and use natural language processing tools to create standardized, homogeneous products within which we can compare public and private purchases. We describe each of these in turn in more detail in the rest of this section.

[3] Among the few exceptions are Duggan & Scott Morton (2006), studying US pharmaceuticals procurement at the market level, Atal *et al.* (2024) studying the entry of public pharmacies in Chile, Kroft *et al.* (2023) studying the construction industry in the US, and (Carrillo *et al.*, 2023) studying small works projects in Ecuador.

2.1 Data Sources

We use two main sources of data for our analysis.

Electronic Invoices (Nota Fiscal Eletrônica, NF-e): In its efforts to document the economy and combat tax evasion, the government of São Paulo was a pioneer in the adoption of electronic invoicing (Naritomi, 2019). All transactions in the formal sector are required to generate an electronic invoice. We work with the universe of electronic invoices not involving consumers (i.e business to business (B2B) and business to government (B2G) transactions) from January 2018 to December 2023. The invoices contain information on the item being sold (a textual description, the commercial unit, and the Global Trade Item Number (GTIN)); transaction value and quantity sold; seller and buyer's tax IDs, legal nature, and geographic location; and the time of the purchase.

To compute the unit price of each item transacted, we subtract the discount and the sales tax (ICMS) exemption amounts from the transaction value, and then divide the remaining value by the quantity transacted.[4] This unit price is then the primary outcome in our analysis.

Table 1 summarizes the invoice data in the "Full Raw Data" column. The full dataset contains 855,517,687 invoices, of which we focus on the 756,936,875 invoices in which the buyer is either a private business (742,879,778 invoices), a São Paulo state government entity (369,533 invoices), or a NGO (13,687,564 invoices).[5] The bulk of the data are private-sector purchases, and private-sector buyers tend to make more purchases (3,438 invoices per private buyer vs 792 per government buyer). The average purchase is relatively large, involving a purchase of 81 units of a drug, and a unit price of RS. 48. The sample contains a large number of buyers of both types and so it is a useful sample to estimate the differences between government and private-sector buyers.

We see that around 18% of invoices are missing a valid GTIN product code. Moreover, this is imbalanced between government and private-sector buyers: 34% of government purchases are missing a valid product code; while only 15% of private-sector purchases are missing a valid product code. This implies that using only invoices with valid product codes risks severely biasing comparisons between government and private-sector buyers. This motivates the exercises we describe in section 2.2 below to classify invoices without valid product codes to ensure that we can make valid comparisons between government and private sector buyers.

Table 2 summarizes the buyers in our sample, and compares government and private-sector

[4]This definition of the unit price came from discussions with tax authorities in São Paulo. Although sellers most likely take into account any taxes when setting their price, we opt to not subtract these values from the transaction value in the invoices. The reason behind this decision is that it is not possible to establish a general rule for the tax regime that sellers face, especially regarding federal taxes such as payroll and import taxes. However, the sale of pharmaceuticals to the public sector is ICMS-exempt in the state of São Paulo such that we subtract the ICMS exemption value from the transaction price in order to keep the prices faced by public and private sector buyers comparable.

[5]The remaining 98,580,812 invoices are purchases by federal or municipal governments, or other special types of organizations.

buyers. Private-sector buyers tend to be bigger, making more purchases and purchasing a wider variety of products. However, government buyers tend to make bigger purchases and make purchases from a larger number of sellers.

Table 3 summarizes the sellers in our sample. There are over 30,000 sellers in our data, with the average seller making over 34,000 sales in our 6 years of data. They also sell a wide range of products to a wide range of sellers, averaging 167 distinct products sold to 88 clients. Finally, table 4 summarizes the products in the data. The full data contains over 24,000 products, sold an average of 35,000 times each. Conveniently for our analysis that seeks to compare products being bought by different types of buyers, the average product is sold by 171 sellers and bought by 2,900 different buyers.

Pharmaceutical characteristics: We use data from the Brazilian Health Regulatory Agency (Anvisa) containing the commercial name, active ingredient, presentation, and generic or branded status of each Global Trade Item Number (GTIN). GTIN codes are equivalent to the Universal Product Codes (UPCs) or "barcodes". The Anvisa data contains 25,860 unique drugs at the GTIN level.[6] These data allow us to combine barcodes (which contain brand & manufacturer information and hence are overly specific for our analysis) into homogeneous products that are identical from a clinical perspective (as described below in section 2.2.1).

2.2 Classification

To accurately measure the price difference between government and private sector purchases, we need to ensure that we are comparing the same products. The government may pay more for a particular medication because it typically buys a 30mg/ml dosage of the medication, and the private sector always buys a 15mg/ml dosage. This would make it impossible to determine whether the price difference results from the government's role as the buyer or from the variation in medication dosages. Therefore, information about pharmaceutical products, such as dosage, is essential in order to make appropriate comparisons.

Unfortunately, this information is not directly available in an invoice. However, this information is indirectly available in the product description in the invoice. To extract and structure this information, we develop a machine-learning approach, described below, that takes the unstructured text in the product descriptions and classifies them into homogeneous product categories that fully capture the clinical use of the drugs.

[6]Only 25,849 of these GTIN codes were uniquely identified in the raw data, as there were 11 instances with duplicates. In order to uniquely identify these GTIN codes, we implement the following rule: if the GTIN code is duplicated and one of the observations has a missing presentation, then we keep the one with the non-missing data; but if both observations have a missing presentation, then we randomly choose which observation to keep.

2.2.1 Definition of product

Following Bronnenberg *et al.* (2015), we define a product as a combination of three attributes: The drug's active ingredient; whether it is generic or branded; and its presentation. The main difference between our product definition and the full GTIN barcode is that we combine products that share all attributes except their brand to arrive at a definition of a product that captures all clinically-relevant differences between products but not more.[7]

Active Ingredients are chemical or natural products that provide direct biological effects in the diagnosis, cure, mitigation, treatment, or prevention of disease or affect the structure or any function of the body of humans or animals. From a clinical perspective, whether a drug is branded or generic is irrelevant Bronnenberg *et al.* (2015). Generics contain the same active ingredient, in the same dose and pharmaceutical form, with the same posology and therapeutic indication as the reference medication. As such, they are interchangeable with their branded analogs.

The presentation of the product refers to technical attributes of the drug such as dosage, pharmaceutical form, route of administration and packaging. These are reported in a standardized way dictated by the regulator Anvisa. The most common products in our data are described in two tables. Table 5 shows the top products in the full data. Panel A shows top products by the value of purchases, while panel B shows top products by the volume of purchases. Table 6 shows top products bought by the Government. Panels A and B follow the same logic of Table 5.

2.2.2 Training Data

In order to label each invoice with an active ingredient, a generic/branded status, and a format, we adopt a supervised approach in which we use a training dataset to learn the classification of the product description text. To do this, we exploit the fact that 85% of the invoices in our dataset have a valid GTIN code, though this is heavily skewed: government purchases are far less likely than private-sector purchases (66% vs 85%) to contain a valid GTIN code, making it critical to fill in the missing product data to avoid biasing our analysis.

To build our trainig data, we take the 724,073,565 invoices containing a valid GTIN and merge it with the Anvisa regulatory data by GTIN code. This gives us the 3 items we require to identify the product (active ingredient, generic status, & format) as defined in section 2.2.1 for each GTIN code. From this we keep the unique combinations of product description text, commercial unit, active ingredient, generic status, and format, yielding a final training dataset of 11,376,502 observations.

[7]That said, the inclusion of the generic status in the definition of product might be too specific, as two drugs with the same active ingredient and format should produce the same effects on the patient. Moreover, the government or the private sector might have different behaviors regarding the choice of purchasing the generic version or not. In ongoing work we explore how robust the findings are to the inclusion of the drug's generic status and how much of the variation in prices can be attributed to differential tastes for generic drugs.

2.2.3 Classification

We use the FastText classifier (Joulin *et al.*, 2017), a neural network based supervised learning model. The FastText classifier is a commonly used baseline for Natural Language Processing tasks and learns vector representations for words and then uses these in a neural network structure to make predictions about the possible labels to attach to a description.

We proceed in three steps to produce our predicted product classifications for all invoices. First, we train a model that takes a product description-commercial unit pair as input and produces a probability distribution over all the possible active ingredients. The active ingredient with the highest predicted probability is then our label for the description. Second, we train a model to predict the generic status of every product description-commercial unit pair.

Third, to predict the format of the drug, we train a separate classifier for each active ingredient. This achieves better performance by limiting the set of possible labels to be assigned to those that appear for a given active ingredient. The final layer of the classifiers' neural network is a hierarchical softmax function and so the outputs of the algorithms are probability distributions over the set of possible labels. We use these probabilities to measure the algorithm's confidence in its predictions.

With the outputs of these classification algorithms, we assign each invoice the active ingredient, generic/branded status, and format with the highest confidence score. We then compute the overall confidence score the algorithms have as the product of the confidence scores of the three sub-components of the product definition. In our baseline analysis, we treat invoices with an overall confidence score of at least 0.8 as reliable enough to be used in our analysis.[8]

2.2.4 Performance

The classifiers described above perform extremely well. Table 7 summarizes the classifiers' performance in our training data. Panel A shows that the classifier correctly predicts the active ingredient in 97.8% of cases; correctly predicts the generic/branded status for 93.9% of cases; and correctly predicts the presentation of the drug for 98.8% of cases. All three components of the product are correctly predicted for 92.8% of cases.

Panel B shows that the confidence scores are well calibrated: The average confidence score is significantly higher for correct predictions than for incorrect ones. For example, the overall confidence score averages 0.93 among correct predictions, but only 0.56 among incorrect ones. Finally, panel C shows the correlation coefficient between an indicator for the classification being correct and the confidence score, again showing that all three classifiers are well calibrated.

Figure 1 shows the distributions of the confidence scores for each component and for the product of the three confidence scores, separately by type of buyer. Panels A–C show the classifiers for each component of the product, while panel D shows the overall scores: the product of the three component scores. We see that the scores for all components tend to be very high.

[8]In ongoing work we test for robustness of our results to the threshold we use for reliability.

Moreover, we see that while the performance is slightly lower for purchases by the state government, the difference with the private sector is not very large. For example, around 85% of private sector purchases have an overall confidence score above 80% (the threshold we use for inclusion in our analysis), while for the government this drops only to 82%.

2.3 Analysis Sample

To build our sample for analysis, we start with the invoice data and apply the algorithms described in section 2.2 to attach a predicted active ingredient, generic status, format and accompanying confidence scores to every invoice[9]. To form our main analysis sample we keep only transactions that have a valid GTIN or an overall confidence score of at least 80%. In addition, we drop invoices with non-positive quantities and transaction values, as well as those with a seller that was under the SIMPLES tax regime but the transaction value was above R$4,800,000 (which is the upper limit of the annual revenue for a SIMPLES firm). The resulting data set contains 649,897,645 transactions of pharmaceutical products, which involved 14,789 products, 18,505 sellers, and 225,292 buyers.

We impose five additional restrictions to define our analysis sample. We restrict the buyers we consider: First, we drop transactions in which the buyer was part of the federal government or was located outside of the state of São Paulo, as we only observe a fraction of their purchases due to the fact that the NF-e data is restricted to the state of São Paulo. Second, we drop buyers that were part of the MEI or SIMPLES tax programs, as these firms are small by definition and as such, not comparable to government buyers.

Next, we restrict the products we study to be able to make reliable comparisons within products. Our third restriction is to keep only products that were transacted at least 100 times. Fourth, to ensure that we can reliably estimate both government and private-sector prices for the same product, we keep only products purchased at least 20 times by both the government (the state government, municipal government, or other government layers) and the private sector (for this exercise we consider NGOs and other charitable social organizations to be part of the private sector). Finally, our fifth restriction deals with outlier prices. We winsorize the prices for each product at the 2.5th and 97.5th percentiles.

The analysis sample is significantly smaller, though still boasts 121,612,293 invoices for 2,194 products, which involved 12,035 sellers and 79,644 buyers. Tables 1–4 compare the analysis sample to the full sample. The transactions in the analysis sample are, on average, for slightly higher unit prices and slightly smaller quantities. The buyers also tend to be smaller in the analysis sample than the full sample, as do the sellers, though their average transaction sizes are slightly larger. Finally, the products in our analysis sample are purchased more often, and sold by a larger number of sellers, but purchased by a slightly smaller number of buyers.

[9] Notice that invoices with a valid GTIN have both a product definition that comes from Anvisa and a predicted product definition. In this analysis, we use the variables provided by Anvisa for the invoices with a valid GTIN.

3 How Much More/Less Does the Government Pay?

To measure how much the government pays for a given good compared to the private sector, we run two main sets of regressions. First, we examine overall price wedges in government purchases relative to the private sector, flexibly controlling for relevant characteristics of the purchase. Then, we run the same regression with seller fixed effects to determine how much of the price wedge is due to the same firms applying different mark-ups to the government versus the private sector, and how much is driven by supplier selection.

3.1 Overall price wedges

To measure the overall price wedge, we run the following specification:

$$log(price_{ijgt}) = \beta \text{Gov}_i + \alpha_t + \gamma_{gt} + \log(Quantity_{ijgt}) \cdot \theta_g + \mu_g + X_{ij} + \varepsilon_{ijgt}, \tag{1}$$

Where $price_{ijgt}$ is the price observed when buyer i purchases good g from supplier j at time t. Our coefficient of interest is β as it captures the effect of dummy Gov_i that takes on value 1 when a buyer i is government and zero when it is a private firm. We control for product-by-time fixed effects (γ_{gt}) to allow each product to have a distinct time effect. We also include quantity interacted with product fixed effects ($\log(Quantity_{ijgt}) \cdot \theta_g$) to control for quantity effects, allowing these to vary by product. μ_g are product fixed effects, X_{ij} are additional buyer and buyer-seller controls, and ε_{ijgt} is the error term that is robust to heteroskedasticity.

Table 8 shows the results of specification (1). Column (1) shows that the government pays, on average, 13% more than the private sector for the same good. This difference is basically unchanged (13.6%) when adding more controls in column (3), namely, distance between buyer and seller and buyer size as measured by the log of total purchase value by firm in the sample period.

Since our sample period includes the COVID-19 pandemic, we examine the same price regression for the pre-COVID period (2018 and 2019), during COVID (2020 and 2021), and post-COVID (2022 and 2023). Columns (1), (3), and (5) in Table 9 show the results for specification (1). The results suggest there is little change in $\hat{\beta}$ across these periods, though there is a decrease in the average price wedge during the pandemic years.

3.2 Supplier selection

The average price difference estimated using specification (1) could be driven by a selection of more expensive suppliers, or due to a higher price that a given supplier may charge government purchases versus private firms' purchases. In order to she light on relevance of these two effects, we add supplier fixed effects to specification (1) interacted with product fixed effects (δ_{jg}) to allow for supplier effects to vary across products.

$$log(price_{ijgt}) = \beta \text{Gov}_i + \alpha_t + \gamma_{gt} + \log(Quantity_{ijgt}) \cdot \theta_g + \mu_g + \delta_{jg} + X_{ij} + \upsilon_{ijgt}, \tag{2}$$

With this specification $\hat{\beta}$ measures the price wedge within supplier and good, which captures a potential differential prices charged to government buyers. This within supplier difference could be positive if suppliers can apply a higher mark-up to government purchases, or if it is more costly to supply to the government relatively to the private sector. This price difference can also be negative if the competitive process of procurement auctions allow the government to extract rents from suppliers. The average effect, therefore, is an empirical question.

Table 8 shows the results of specification (2). Column (2) shows that the government pays, on average, 8.3% *less* than the private sector for the same good once we exploit variation within supplier-good. This difference is basically unchanged (-8%) when adding more controls in column (4), namely, distance between buyer and seller and buyer size as measured by the total purchase value by firm in the sample period. Therefore, the overall price wedge estimated with specification (1) seems to be entirely driven by the selection of suppliers that have, on average, higher prices. However, the government seems to be able to extract rents from suppliers that win government contracts.

Columns (2), (4), and (6) in Table 9 show the results for specification (2) before, during and after the COVID-19 pandemic. The results suggest a larger change $\hat{\beta}$ across these periods than with the overall price wedge once we exploit variation within supplier-good. Before the pandemic, the price wedge was -4% increasing to -12% during the pandemic and then slightly dropping to -10.5% after the pandemic. The results indicate that the government seems to have been able to reduce the overall price wedge during the pandemic, and also extract more rent from suppliers that secure contracts.

3.3 Mechanisms

Our findings indicate that supplier selection primarily drives the higher average prices that the government pays relative to the private sector. This remains true even after accounting for buyer size and the distance between buyers and suppliers. To explore why the government may not attract lower-priced suppliers, we match transaction-level data with public procurement data. The objective is to determine what portion of the potential supplier pool for a specific good participates in public procurement and how different procurement processes might attract a broader and more competitive subset of this supplier pool.

Figure 1 shows the distribution of how relevant the government is across suppliers as measured buy the share of their sales that goes to the government during our sample period. The picture restricts attention to shares between (0,1) excluding the two extremes, but also plots the share of firms with exactly 0 and 1. Most firms sell little or nothing to the government (above 90%) while some suppliers only sell to the government (1.5%).

[Placeholder: impact of procurement process]

Figure 2 shows that there is a wide distribution for the relevance of the government as a buyer across product markets. Possibly, the government might be able to attract a larger share of the pool of potential suppliers in markets where they are a larger buyer. Thus, there could be substantive heterogeneity behind the average price wedge effects that we document in Table 8. Table 9 shows the results of a dollar-weighted version of specification (1) and (2), i.e., weighting each observation by the transaction value. The overall average price wedge becomes negative (-12.5%) and the coefficient is essentially unchanged once we control for seller-product fixed effect. This pattern suggests that the government may perform particularly well in larger ticket purchases.

To further examine the heterogeneity of this price wedge between government and private purchases, we estimate specification (1) and (2) for each product separately. Figure 3 shows the distribution of the coefficients from regressions with and without seller-product fixed effects. The coefficients span from negative to positive values in both specifications, indicating that the average effect masks substantial heterogeneity. The coefficients with seller fixed effects show more mass around, and exactly at, zero, suggesting that the price wedge diminishes in absolute value within seller-product groups.

Figure 4 shows the correlation between price wedges across products and the share of the market purchased by the government. We observe a negative correlation between the government's market power as a buyer and the price wedge for each product. The slope is particularly steep for estimates controlling for seller-product fixed effects, suggesting that governments may be able to extract more rents from suppliers when it holds a larger share of the market.

4 Causal Analysis of Markups Charged to Government Buyers

The previous section presented descriptive evidence suggesting that the government often pays more than the private sector for the same drug at the same time. Moreover, a descriptive decomposition suggests that suppliers charge *lower* markups to government buyers, implying that the positive overall price wedge is driven by the government buying from different suppliers than the private sector. In this section we aim to identify the causal effect of being a government buyer on the markup that sellers obtain. In future work described briefly in the conclusion we aim to also identify the causal effect of being a government buyer on the selection of suppliers.

The descriptive evidence we provide in section 3.2 above on the markups charged by suppliers suggests that markups are, on average, 8% lower when the buyer is a (state) government entity than when the buyer is a private-sector business. Despite the rich set of controls in the regression, this estimate should not be interpreted causally. Conceptually, the OLS estimate could be biased in either direction. On one hand, winning a government contract might lead sellers to charge higher prices in the private sector as they divert production capacity to fulfilling the government contract (as found by Kroft *et al.* 2023 and Krasnokutskaya & Seim 2011). On the other hand, winning government contracts may allow producers to expand and benefit from economies of scale, lowering marginal costs and prices charged in the private sector (Ferraz *et al.*, 2015; Carrillo *et al.*, 2023).

4.1 Empirical Design

To address this, we develop a new strategy to estimate the counterfactual price that government suppliers would have charged for the same item around the same time had they not won the government contract. To do this, we adapt the regression-discontinuity approach in Kroft *et al.* (2023) to extrapolate from the private-sector prices charged by runners-up in close auctions to the private-sector prices that the winners of procurement auctions would have charged. This "identification at infinity" approach is analogous to the approach commonly used in the literature on discrimination to purge selection from judges decisions, for example in Arnold *et al.* (2022); Angelova *et al.* (2023).

Formally, we aim to estimate

$$\mathbb{E}_j\left[b_j\left(1\right)-\mathbb{E}_{i\in\mathcal{B}_j}\left[p_{ij}|r_{ij}=1,\text{gov}_{ij}=0\right]\right] \tag{3}$$

where $b_j\left(r\right)$ is the final bid submitted by the bidder whose bid is ranked r in auction j; $\mathcal{B}_j$ is the set of bidders in auction j; and p_{ij} is the price charged by bidder i when they sell the item being purchased in auction j in the private sector; and gov_{ij} is an indicator for winning a government contract.

The identification challenge is that the observed price at which the auction winner sells the same item in the private sector, $\mathbb{E}\left[p_{ij}|r_{ij}=1,\text{gov}_{ij}=1\right]$ need not be the same as the price at which they would have sold the item had they not won the auction, $\mathbb{E}\left[p_{ij}|r_{ij}=1,\text{gov}_{ij}=0\right]$.

To overcome this challenge, we seek to extrapolate from runners-up, who do not win the government contract, to the winner to estimate the counterfactual price the winner would charge in the private sector. In order to pool across procurement auctions, we normalize all bids and prices by the auction's winning bid, and so (3) becomes

$$\tilde{p}\left(R\right)=\mathbb{E}\left[\left(\frac{p_{ij}-b_j\left(1\right)}{b_j\left(1\right)}\right)|r_{ij}=R,\text{gov}_{ij}=0\right] \tag{4}$$

and we seek to estimate $\tilde{p}^{CF}=\lim_{r\downarrow 1}\mathbb{E}\left[\tilde{p}\left(r\right)\right]$. This can be identified from an extrapolation to the winners from the runners-up under familiar assumptions about the smoothness of the conditional expectation function $\tilde{P}\left(R\right)$ analogous to those undergirding "donut"-RD designs or, more generally, "identification at infinity" strategies (Chamberlain, 1986).

4.2 Estimation Sample

Implementing the strategy described above requires us to link our invoice data to data on public procurement purchases, since we require information on the private-sector prices (which are in the invoice data) charged by bidders in each procurement purchase, as well as their bids (which are in public procurement data).

This is not straightforward for two reasons. First, there is no direct way to identify the invoice(s) that are associated with a particular government contract. Second, products are not classified in the same way in the public procurement system as in the invoice data.

To overcome these challenges, we build a linking dataset for a subset of public procurement purchases in which we are indeed able to identify the invoices that correspond to a procurement purchase. Data on public procurement is available on the *Bolsa Eletrônica de Compras* (BEC), the official website that gathers information on all purchases made by the São Paulo government.

The process of linking procurement data to invoices begins with the procurement data itself, which includes information on bids, items, and purchases. For each item in the procurement data, identifiers such as the numbers of the purchase order and the invoice are used to connect it to the "Items Committed" to the actual Commitments (*empenhos*) dataset. Next, we match the government branch codes to specific Commitments entries. These are then connected to invoice data through a code identifying that the purchase was indeed paid. This sequence enables a comprehensive link from procurement records to the finalized invoices. Thus, we now have a dataset of invoices that are certainly linked to procurement processes.

With this linking dataset we can overcome the first challenge for the subset of procurement purchases in the linking dataset. But more broadly, we are able to overcome the second challenge about product definition. In the procurement system, products are identified by platform-specific product codes (BEC codes), while in the invoice data we identify products as described in section 2.2. From the linking dataset, however, we are able to build a dictionary of which products (as defined by us) correspond to each BEC code.

Having overcome these two challenges, we are able to produce a dataset with which to estimate markups. We test this strategy using one month of procurement data, March 2019, which has 19.657 items and 10.278 unique BEC codes. Of these, 1.589 are pharmaceuticals, 8.1%, and we have 839 unique BEC codes related to pharmaceuticals, 8.16% of the total.

In ongoing work we are scaling up our strategy. we repeat the steps to classify the products in the data described in the previous section. The invoice dataset linked to the classified procurement data gives us the homogenized description of every pharmaceutical item purchased that month. That is, for procurements where BEC codes match directly with our invoice data, we assign corresponding invoice product descriptions and commercial units. As before, we focus on joint high-accuracy predictions (higher than 80%), leaving us with 87% of our initial pharma procurement sample.

Finally, each procurement with at least one runner-up bid within a reasonable range qualifies for analysis in our main sample. This filtered dataset will form the basis of our subsequent analyses, giving us a base on which to find credible counterfactuals to perform the RD analysis.

5 Conclusion

In this paper we have studied the size and determinants of the gap between prices paid by governments, adn the prices paid by private-sector buyers for the same item at the same time. To do this we have leveraged a unique dataset spanning the universe of formal sector transactions in the state of São Paulo, Brazil, and focused on a set of standardized products — pharmaceuticals.

We show that on average, the government pays 13% more than the private sector. A descriptive decomposition shows that this overall price gap is driven entirely by the government buying from different suppliers than the private sector. Indeed, when looking at the prices the same seller charges the government versus the private sector, sellers charge the government 8% *less* than the private sector.

In ongoing work, we have developed an extrapolation-based method analogous to a "donut" regression discontinuity design to provide causal estimates of the wedge between what sellers charge the government and the private sector. We are also exploring a range of approaches to provide causal evidence on the magnitude of the gains from inducing entry into public procurement by a wider range of sellers.

Finally, we show descriptively that the wedge between public and private prices varies greatly across drugs. Most interestingly, the government performs significantly better in markets where it is a large buyer, especially in terms of the rents it is able to extract from sellers who do participate in public procurement.

These findings have wide-ranging implications for procurement policy. They suggest that the key driver of gains in procurement performance are to be found in inducing participation by low-cost sellers much more than in fine-tuning the auction format, consonant with the classic results in Bulow & Klemperer (1996).

References

ALLENDE, CLAUDIA, ATAL, JUAN PABLO, CARRIL, RODRIGO, CUESTA, JOSÉ IGNACIO, & GONZÁLEZ-LIRA, ANDRÉS. 2024. Drivers of public procurement prices: Evidence from pharmaceutical markets. *International Journal of Industrial Organization*, **96**, 103086.

ANGELOVA, VICTORIA, DOBBIE, WILL S, & YANG, CRYSTAL. 2023. *Algorithmic recommendations and human discretion*. Tech. rept. National Bureau of Economic Research.

ARNOLD, DAVID, DOBBIE, WILL, & HULL, PETER. 2022. Measuring Racial Discrimination in Bail Decisions. *American Economic Review*, **112**(9).

ATAL, JUAN PABLO, CUESTA, JOSÉ IGNACIO, GONZÁLEZ, FELIPE, & OTERO, CRISTÓBAL. 2024. The Economics of the Public Option: Evidence from Local Pharmaceutical Markets. *American Economic Review*, **114**(3), 615–44.

BANDIERA, ORIANA, PRAT, ANDREA, & VALLETTI, TOMMASO. 2009. Active and Passive Waste in Government Spending: Evidence from a Policy Experiment. *American Economic Review*, **99**, 1278–1308.

BANDIERA, ORIANA, BEST, MICHAEL CARLOS, KHAN, ADNAN QADIR, & PRAT, ANDREA. 2021. The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats. *Quarterly Journal of Economics*, **136**, 2195–2242.

BEST, MICHAEL CARLOS, HJORT, JONAS, & SZAKONYI, DAVID. 2023. Individuals and Organizations as Sources of State Capacity. *American Economic Review*, **113**, 2121–2167.

BOSIO, ERICA, DJANKOV, SIMEON, GLAESER, EDWARD, & SHLEIFER, ANDREI. 2022. Public Procurement in Law and Practice. *American Economic Review*, **112**, 1091–1117.

BRONNENBERG, BART J., DUBÉ, JEAN-PIERRE, GENTZKOW, MATTHEW, & SHAPIRO, JESSE M. 2015. Do Pharmacists Buy Bayer? Informed Shoppers and the Brand Premium. *Quarterly Journal of Economics*, **130**, 1669–1726.

BULOW, JEREMY, & KLEMPERER, PAUL. 1996. Auctions Versus Negotiations. *The American Economic Review*, **86**(1), 180–194.

CARRILLO, PAUL, DONALDSON, DAVE, POMERANZ, DINA, & SINGHAL, MONICA. 2023. *Misallocation in Firm Production: A Nonparametric Analysis Using Procurement Lotteries*. Tech. rept. National Bureau of Economic Research.

CATTANEO, MATIAS D., CRUMP, RICHARD K., FARRELL, MAX H., & FENG, YINGJIE. 2024. On Binscatter. *American Economic Review*, **114**(5), 1488–1514.

CHAMBERLAIN, GARY. 1986. Asymptotic Efficiency in Semiparametric Models with Censoring. *Journal of Econometrics*, **32**, 189–218.

DUGGAN, MARK, & SCOTT MORTON, FIONA M. 2006. The Distortionary Effects of Government Procurement: Evidence from Medicaid Prescription Drug Purchasing. *Quarterly Journal of Economics*, **121**, 1–30.

FERRAZ, CLAUDIO, FINAN, FREDERICO, & SZERMAN, DIMITRI. 2015. *Procuring firm growth: the effects of government purchases on firm dynamics*. Tech. rept. National Bureau of Economic Research.

JOULIN, ARMAND, GRAVE, EDOUARD, BOJANOWSKI, PIOTR, & MIKOLOV, TOMAS. 2017. Bag of Tricks for Efficient Text Classification. *Pages 427–431 of: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*. Valencia, Spain: Association for Computational Linguistics.

KRASNOKUTSKAYA, ELENA, & SEIM, KATJA. 2011. Bid Preference Programs and Participation in Highway Procurement Auctions. *American Economic Review*, **101**, 2653–2686.

KROFT, KORY, LUO, YAO, MOGSTAD, MAGNE, & SETZLER, BRADLEY. 2023. *Imperfect Competition and Rents in Labor and Product Markets: The Case of the Construction Industry*. mimeo: University of Chicago.

NARITOMI, JOANA. 2019. Consumers as Tax Auditors. *American Economic Review*, **109**(9), 3031–72.

Figures & Tables

TABLE 1: PURCHASE-LEVEL SUMMARY STATISTICS

Variable	Full Raw Data	Analysis Data
Total Invoices	855,517,687	121,612,293
Invoices With Valid GTIN	724,073,565	114,532,155
Invoices Without Valid GTIN	131,444,122	7,080,138
Mean Mult Confidence Score	0.852 (0.247)	0.982 (0.036)
Mean Unit Price	48.32 (14,352.07)	64.14 (628.43)
Mean Quantity Comercial	81.31 (9,292.16)	75.12 (8,777.96)
Number of Distinct Sellers	31,271	12,035
Number of Distinct Buyers	464,432	79,644
Number of Distinct Products	24,686	2,194
Number of Distinct Private Buyers	216,039	73,699
Number of Distinct SP State Buyers	466	434
Invoices Private Buyers	742,879,778	115,699,900
Invoices SP State Buyers	369,533	258,249

Notes: The table shows summary statistics of the purchases in our data. The first column shows statistics derived from the full database, while the second column shows the same statistics from our analysis sample. The table shows means with their standard deviations in parentheses.

TABLE 2: BUYER-LEVEL SUMMARY STATISTICS

Panel A: Private Sector		
Variable	**Full Raw Data**	**Analysis Data**
Number of Distinct Buyers	216,039	73,699
Pharma Sales	3,927.18 (409,513.22)	1,561.31 (108,429.44)
Pharma Purchases	3,838.04 (20,295.57)	1,569.90 (8,143.62)
Average Purchase Size	64.75 (2,197.26)	39.70 (891.16)
Average Number Sellers Purchased From	16.18 (308.05)	14.21 (61.74)
Average Number Products Purchased	308.05 (899.07)	61.74 (176.57)
Panel B: State Government		
Variable	**Full Raw Data**	**Analysis Data**
Unique Number of Buyers	466	434
Average Number of Sales	21.91 (473.02)	0
Average Number of Purchases	1,112.70 (4,826.70)	595.04 (2,430.15)
Average Purchase Size	755.06 (4,445.13)	648.28 (3,631.63)
Average Number of Sellers Purchased From	35.11 (63.64)	26.37 (49.50)
Average Number of Products Purchased	216.82 (435.25)	101.13 (212.10)

Notes: The table shows summary statistics of the buyers in our data. The first column shows statistics derived from the full database, while the second column shows the same statistics from our analysis sample. The table shows means with their standard deviations in parentheses.

Table 3: Seller-level Summary Statistics

Variable	Full Raw Data	Analysis Data
Number of Distinct Sellers	31,271	12,035
Average Number of Sales	34,360.78 (1,209,531.52)	10,104.89 (268,347.54)
Average Number Purchases	34,291.20 (1,209,531.38)	9,561.38 (268,187.94)
Average Sale Value	2,520.44 (29,734.24)	2,781.74 (28,875.96)
Average Number Buyers	88.25 (626.76)	40.57 (252.60)
Average Number Products Sold	167.32 (494.10)	54.90 (126.35)

Notes: The table shows summary statistics of the sellers in our data. The first column shows statistics derived from the full database, while the second column shows the same statistics from our analysis sample. The table shows means with their standard deviations in parentheses.

TABLE 4: PRODUCT-LEVEL SUMMARY STATISTICSS

Variable	**Full Raw Data**	**Analysis Data**
Distinct Number Products	24,686	2,194
Average Number Purchases	34,656 (108,408)	55,429 (102,790)
Average Number Purchases With Valid Gtin	29,331 (100,925)	52,202 (98,471)
Average Number Purchases Without Valid Gtin	5,325 (27,484)	3,227 (19,091)
Average Number Sellers	171.62 (292.06)	301.15 (333.12)
Average Number Buyers	2,894.91 (4,884.02)	2,304.63 (2,311.71)

Notes: The table shows summary statistics of the products in our data. The first column shows statistics derived from the full database, while the second column shows the same statistics from our analysis sample. The table shows means with their standard deviations in parentheses.

TABLE 5: TOP PHARMACEUTICAL PRODUCTS OVERALL

	Active Ingredient	Generic	Presentation	Volume of Purchases	Value of Purchases
Panel A: Top Products by Value of Purchases					
1	Semaglutida	✗	1 34 Mg Ml Sol Inj Ct X 1 Car Vd Trans X 3 Ml 1 Sist Aplic Plas Doses 1 Mg 4 Agulhas Novofine	23,261	409,965,715
2	Fumarato De Formoterol Di Hidratado Budesonida	✗	12 Mcg 400 Mcg Cap Dura Po Inal Ct Fr Plas Opc X 60 Inal	26,104	376,501,000
3	Dapagliflozina	✗	10 Mg Com Rev Ct Bl Al Al X 30	79,240	272,054,031
4	Liraglutida	✗	6 Mg Ml Sol Inj Ct X 3 Car Vd Trans X 3 Ml X 3 Sist Aplic Plas	29,267	237,758,491
5	Fumarato De Formoterol Di Hidratado Budesonida	✗	12 Mcg 400 Mcg Cap Dura Po Inal Ct Fr Plas Opc X 60	190,880	224,776,016
Panel B: Top Products by Volume of Purchases					
	Active Ingredient	*Generic*	*Presentation*	*Volume of Purchases*	*Value of Purchases*
1	Fumarato De Formoterol Di Hidratado Budesonida	✗	12 Mcg 400 Mcg Cap Dura Po Inal Ct Fr Plas Opc X 60	190,880	224,776,016
2	Cloridrato De Metformina	✗	500 Mg Com Lib Prol Ct Bl Al Plas Trans X 30	157,183	191,686,813
3	Rosuvastatina Cálcica	✗	10 Mg Com Rev Ct Bl Al Al X 30	147,604	121,000,312
4	Furoato De Fluticasona	✗	0 5 Mg G Sus Spr Nas Ct Fr Vd Amb X 120 Doses	138,101	94,714,311
5	Dextrana Hipromelose	✗	1 0 Mg Ml 3 0 Mg Ml Sol Oft Ct Fr Plas Trans Got X 15 Ml	138,077	27,649,781

Notes: The table shows the top products purchased in our data. In panel A, we display the top products by value, while panel B displays the top products by number of purchases. The *Volume of Purchases* and *Value of Purchases* are measured in yearly 2019 R$ (1USD = 4R$).

TABLE 6: TOP PHARMACEUTICAL PRODUCTS BY GOVERNMENT

Panel A: Top Products by Value of Purchases

	Active Ingredient	*Generic*	*Presentation*	*Volume of Purchases*	*Value of Purchases*
1	Somatropina	✗	12 Ui Po Liof Ct Fa Vd Inc Dil Bacteriostatico X 2ml	288	60,427,143
2	Acetato De Gosserrelina Goserrelina	✗	10 8 Mg Depot Ser Preenc Plas Trans Bs Ct Env Al Poliet X 1	6,026	163,299,638
3	Eltrombopague Olamina	✗	50 Mg Com Rev Ct Bl Al Al X 14	1,209	75,646,765
4	Esilato De Nintedanibe	✗	150 Mg Cap Mole Ct Bl Al Al X 60	878	78,832,761
5	Brometo De Tiotrapio	✗	2 5 Mcg Dose Sol Inal Ct Fr Plas 4ml 60 Doses Respimat	68	30,658,351

Panel B: Top Products by Volume of Purchases

	Active Ingredient	*Generic*	*Presentation*	*Volume of Purchases*	*Value of Purchases*
1	Cloreto De Sódio	✗	9 Mg Ml Sol Inj Iv Cx 20 Fa Plas Trans Sist Fech X 500 Ml	3,450	6,723,112
2	Hidroxicarbamida	✗	500 Mg Cap Dura Ct Bl Al Al X 100	2,553	12,291,618
3	Cloreto De Sódio	✗	9 Mg Ml Sol Inj Iv Cx 50 Fa Plas Trans Sist Fech X 100 Ml	3,173	11,381,676
4	Gabapentina	✓	300 Mg Cap Dura Ct Bl Al Plas Trans X 300	672	14,744,015
5	Fumarato De Formoterol Di Hidratado Budesonida	✗	12 Mcg 400 Mcg Cap Dura Po Inal Ct Fr Plas Opc X 60 Inal	26,104	376,501,000

Notes: The table shows the top products purchased in our data. In panel A, we display the top products by value, while panel B displays the top products by number of purchases. The *Volume of Purchases* and *Value of Purchases* are measured in yearly 2019 R$ (1USD = 4R$).

Table 7: FastText Performance Metrics

Panel A: Accuracies				
	Active Ingredient	Presentation	Generic/Branded	All Three
	0.978	0.939	0.988	0.928
Panel B: Average Confidence Scores, by Right or Wrong Predictions				
	Active Ingredient	Presentation	Generic/Branded	All Three
Right Predictions	0.983	0.991	0.949	0.927
Wrong Predictions	0.756	0.780	0.624	0.559
Panel C: Correlations between Right or Wrong Predictions and Confidence Scores				
	Active Ingredient	Presentation	Generic/Branded	All Three
	0.46	0.45	0.54	0.57

Notes: The table shows diagnostic statistics for the performance of our classifiers. Panel A shows the fraction of cases in the training data for which the classifier assigns the correct label. Panel B shows that the confidence scores are well calibrated: It shows the average confidence score among the correct and incorrect perdictions, demonstrating a large gap between them. Finally, panel C shows the correlation coefficient between an indicator for the classification being correct and the confidence score, again showing that all three classifiers are well calibrated.

TABLE 8: GOVERNMENT BUYERS PAY HIGHER PRICES

Dependent Variable:	Log(Price)			
Model:	(1)	(2)	(3)	(4)
Variables				
State Government	0.1292***	-0.0827***	0.1359***	-0.0805***
	(0.0019)	(0.0013)	(0.0019)	(0.0013)
Controls				
Distance			✓	✓
Buyer Size			✓	✓
Fixed-effects				
Product × Time	✓	✓	✓	✓
Product	✓	✓	✓	✓
Seller × Product		✓		✓
Varying Slopes				
Log(Quantity) × Product	✓	✓	✓	✓

Notes: The table shows the results of estimation of equation (1): $log(price_{ijgt}) = \beta \text{Gov}_i + \alpha_t + \gamma_{gt} + \log(Quantity_{ijgt}) \cdot \theta_g + \mu_g + X_{ij} + \varepsilon_{ijgt}$ in columns (1) and (3) and of estimation of equation (2): $log(price_{ijgt}) = \beta \text{Gov}_i + \alpha_t + \gamma_{gt} + \log(Quantity_{ijgt}) \cdot \theta_g + \mu_g + \delta_{jg} + X_{ij} + \upsilon_{ijgt}$ in columns (2) and (4). The number of observations is 121,612,293 for all specifications. Heteroskedasticity-robust standard-errors in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1.

TABLE 9: PRICE REGRESSIONS WEIGHTED BY TRANSACTION VALUE

Dependent Variable:	Log(Price)			
Model:	(1)	(2)	(3)	(4)
Variables				
State Government	-0.1250***	-0.1155***	-0.1045***	-0.1087***
	(0.0023)	(0.0021)	(0.0022)	(0.0020)
Controls				
Distance			✓	✓
Buyer Size			✓	✓
Fixed-effects				
Product × Time	✓	✓	✓	✓
Product	✓	✓	✓	✓
Seller × Product		✓		✓
Varying Slopes				
Log(Quantity) × Product	✓	✓	✓	✓

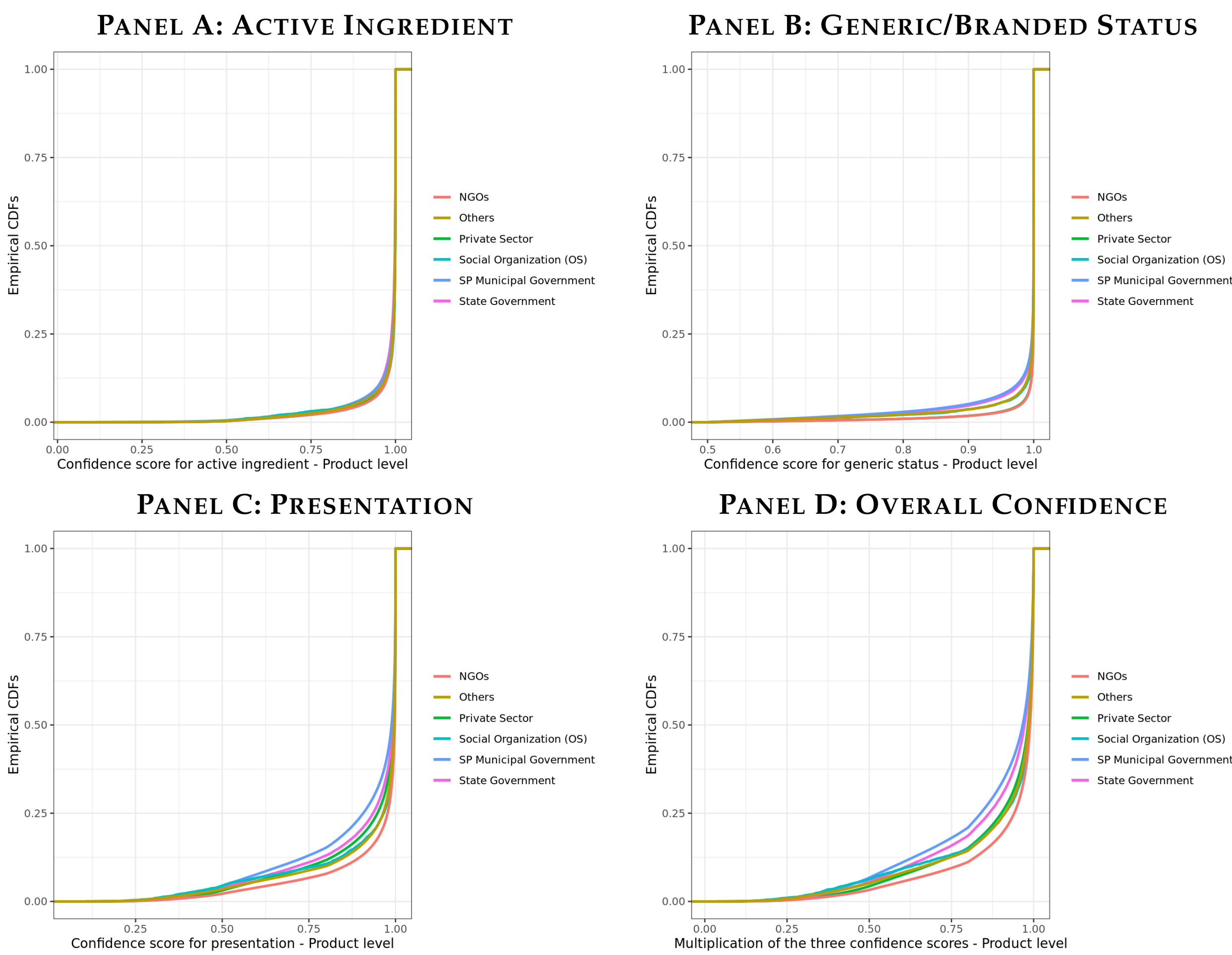
Notes: The table shows the results of estimation of equation (1): $log(price_{ijgt}) = \beta\text{Gov}_i + \alpha_t + \gamma_{gt} + \log(Quantity_{ijgt}) \cdot \theta_g + \mu_g + X_{ij} + \varepsilon_{ijgt}$ in columns (1) and (3) and of estimation of equation (2): $log(price_{ijgt}) = \beta\text{Gov}_i + \alpha_t + \gamma_{gt} + \log(Quantity_{ijgt}) \cdot \theta_g + \mu_g + \delta_{jg} + X_{ij} + \upsilon_{ijgt}$ in columns (2) and (4). In contrast to table 8, the regressions here are weighted by the value of each transaction. The number of observations is 121,612,293 for all specifications. Heteroskedasticity-robust standard-errors in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1.

TABLE 10: REGRESSIONS, BY COVID PERIODS

Dependent Variable:	Log(Price)					
COVID Period:	Pre-COVID		During COVID		Post-COVID	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
State Goverment	0.1351***	-0.0406***	0.0853***	-0.1238***	0.1201***	-0.1047***
	(0.0029)	(0.0020)	(0.0036)	(0.0024)	(0.0032)	(0.0022)
Fixed-effects						
Product × Time	✓	✓	✓	✓	✓	✓
Product	✓	✓	✓	✓	✓	✓
Seller × Product		✓		✓		✓
Varying Slopes						
Log(Quantity) × Product	✓	✓	✓	✓	✓	✓
Observations	47,707,002	47,707,002	37,104,212	37,104,212	36,801,079	36,801,079

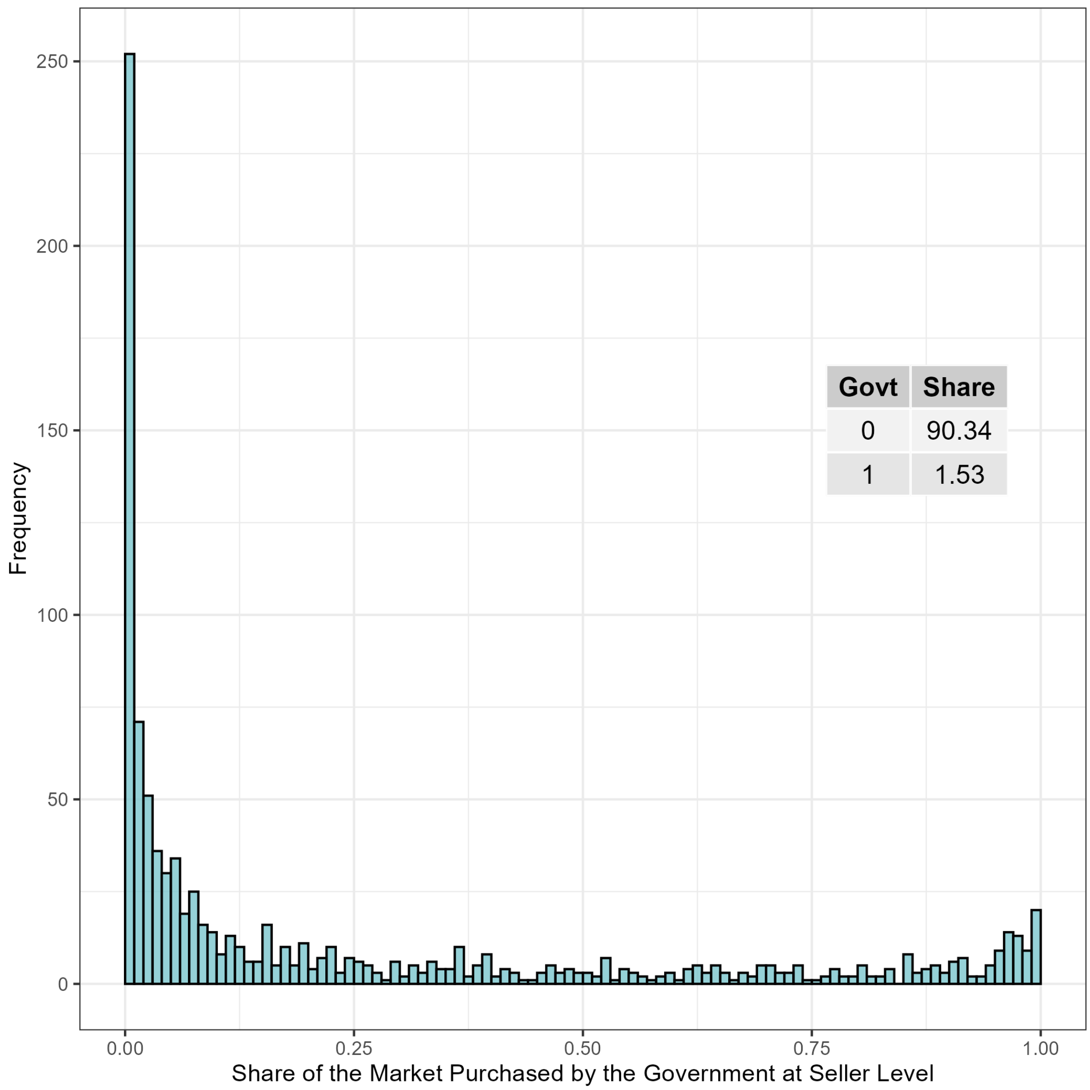
Notes: The table shows the results of estimation of equation (1): $log(price_{ijgt}) = \beta \text{Gov}_i + \alpha_t + \gamma_{gt} + \log(Quantity_{ijgt}) \cdot \theta_g + \mu_g + X_{ij} + \varepsilon_{ijgt}$ in columns (1), (3) and (5) and of estimation of equation (2): $log(price_{ijgt}) = \beta \text{Gov}_i + \alpha_t + \gamma_{gt} + \log(Quantity_{ijgt}) \cdot \theta_g + \mu_g + \delta_{jg} + X_{ij} + \upsilon_{ijgt}$ in columns (2), (4) and (6). The number of observations is 121,612,293 for all specifications. Heteroskedasticity-robust standard-errors in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1. All specifications include NGO as a control.

Figure 1: Distributions of Classifier Confidence Scores



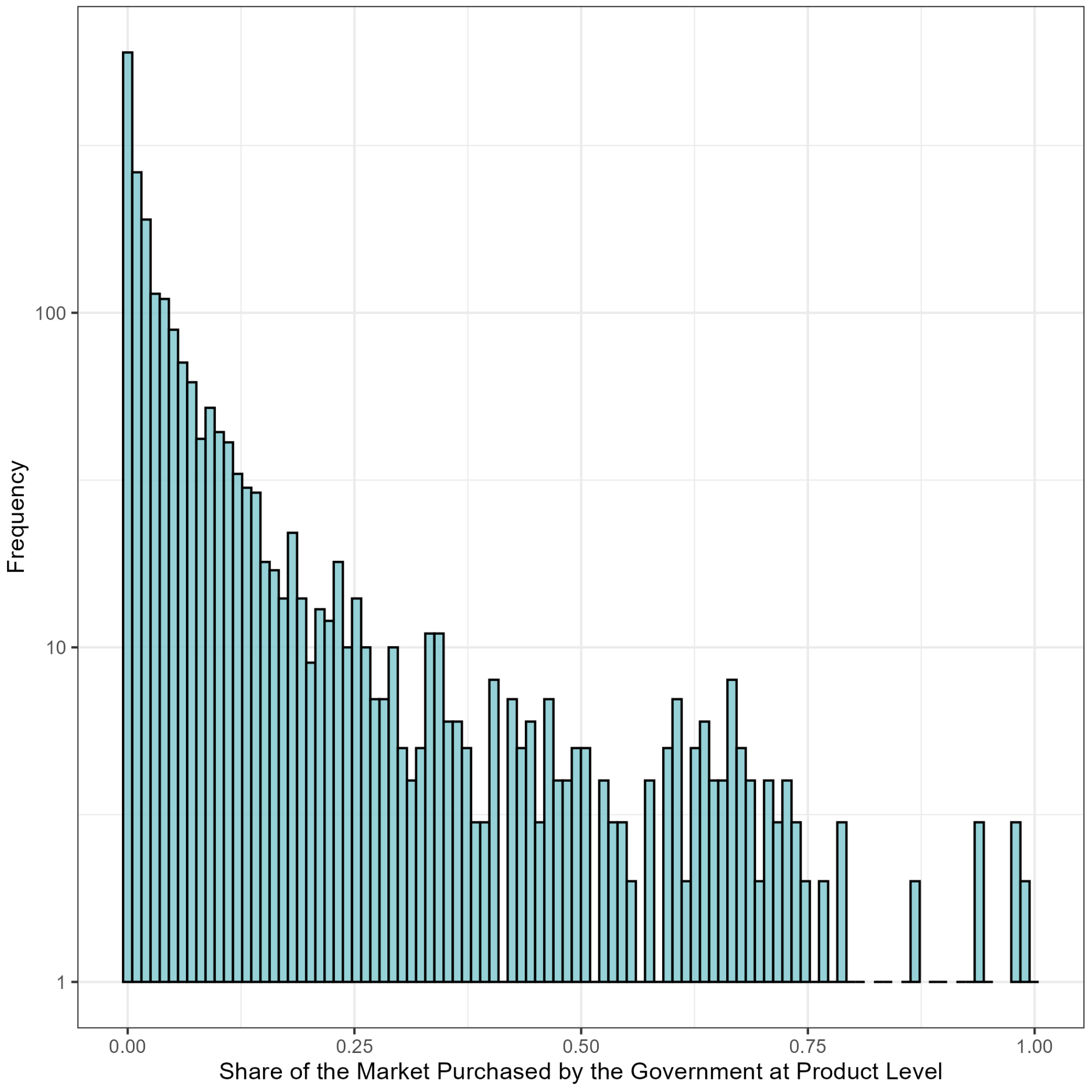
Notes: The figure shows the distributions of the Fasttext classifiers' confidence scores (the maximum score of the outputs of the last soft-max layer of the neural network, interpretable as the probability of the most preferred label being correct). Panel A shows the scores for the classifier predicting the product's active ingredient. Panel B shows the scores for the classifier predicting the product's generic/branded status. Panel C shows the scores for the classifiers predicting the product's presentation. Panel D shows the product of the three scores for each product.

FIGURE 2: DISTRIBUTION OF SHARE SOLD TO GOVERNMENT BY SUPPLIER



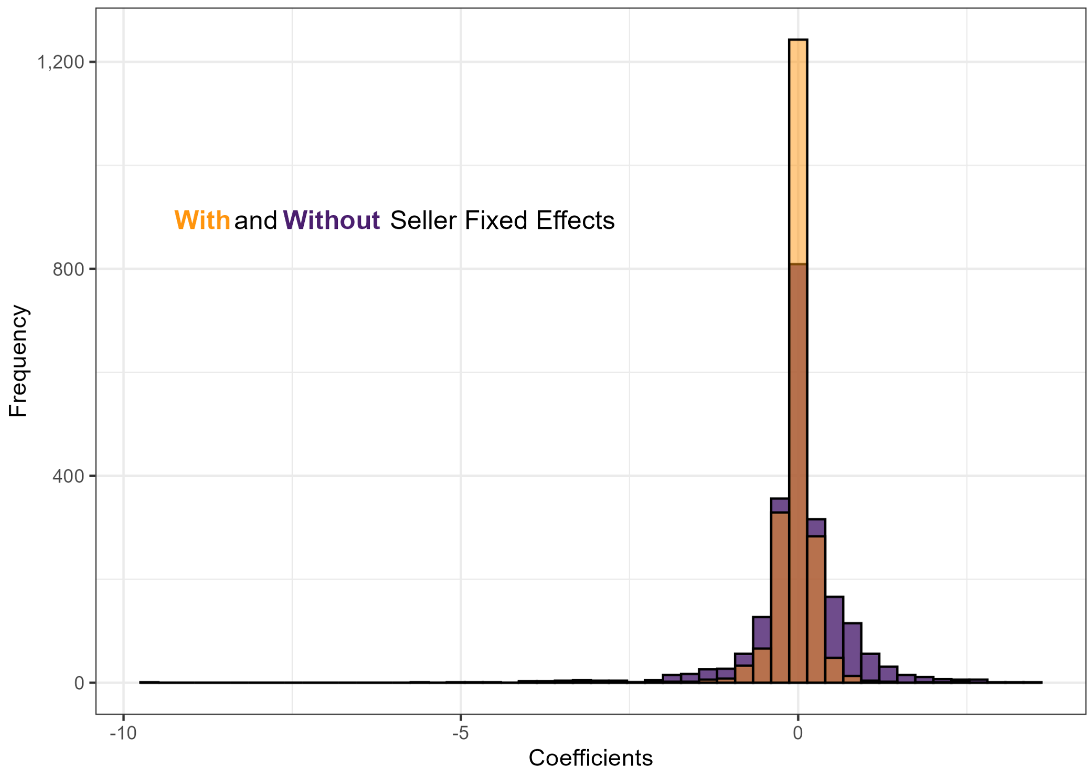
Notes: The figure plots the share of supplier sales that are purchased by the government ($Govt$) during our sample period (2018-2023). The histogram restricts attention to shares between (0,1) excluding the two extremes, but also plots the share of firms with $Govt$ that is exactly 0 and 1.

FIGURE 3: DISTRIBUTION OF SHARE SOLD TO GOVERNMENT BY PRODUCT



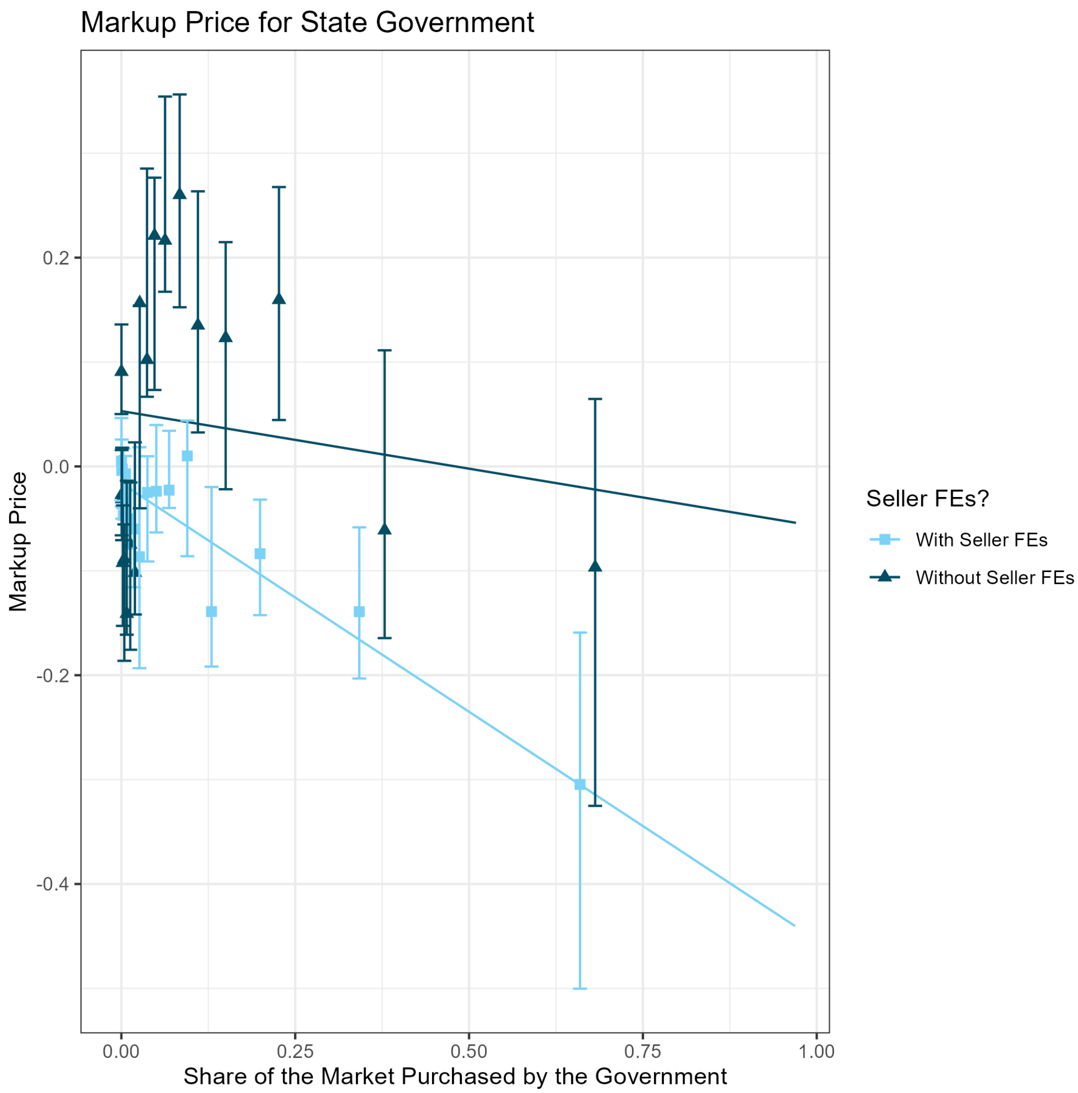
Notes: The figure plots the share of government purchases among the total sales of a given product (*Govt*) during our sample period (2018-2023).

FIGURE 4: DISTRIBUTION OF ESTIMATED COEFFICIENTS FOR STATE GOVERNMENT DUMMY ACROSS PRODUCTS, WITH AND WITHOUT SELLER FIXED EFFECTS



Notes: This displays the distribution of state government dummy coefficients obtained from two regression models. For each product, two separate regressions were estimated: one excluding seller fixed effects (Equation 1) and another including them (Equation 2). The coefficients on the state government dummies from both models were extracted and plotted as histograms. The color of each histogram indicates the model specification: purple for the model without seller fixed effects and orange for the model with seller fixed effects.

FIGURE 5: BINSCATTER REGRESSION ANALYSIS OF PRICE DIFFERENCE BETWEEN GOVERNMENT AND PRIVATE SECTOR, BY GOVERNMENT MARKET SHARE



Notes: The figure plots binscatter estimates of the relationship between Share of the Market Purchased by the Government and the government price wedge. The dark blue line represents estimates from a specification without seller fixed effects, while the light blue line shows estimates including seller fixed effects. The errorbars represent 95% confidence intervals. The binscatter estimation uses linear polynomial fit (degree = 1) and optimal number of bins following Cattaneo *et al.* (2024) methodology. Market shares are calculated using total transaction values at product level.