Department of Economics- FEA/USP

Public Sector Procurements and Reference Prices Estimation with Small Samples in Brazil

GABRIEL LYRIO DE OLIVEIRA
ANDRÉ LUIS SQUARIZE CHAGAS
DENISE LEYI LI

WORKING PAPER SERIES Nº 2022-02

DEPARTMENT OF ECONOMICS, FEA-USP
WORKING PAPER Nº 2022-02

Public Sector Procurements and Reference Prices Estimation with Small Samples in Brazil

Gabriel Lyrio de Oliveira (gabriel.oliveira@fipe.org.br)
André Luis Squarize Chagas (achagas@usp.br)
Denise Leyi Li (denise.leyi@fipe.org.br)

Research Group: NEREUS

Abstract:

This paper presents a new and simple estimator for the reference prices of products or services. These reference prices aim to work as boundaries and support the public stakeholders' decisions at the beginning of their procurements processes. We discuss the advantages of the method compared to the estimator traditionally used all over Brazil, especially in the Federal, State, and Municipal governments, which relies on the computation of averages of a few prices observed in each period. The proposed estimator decomposes the price of a product into a component of its economic sector, sector-time, supplier component representing differences in quality and productivity, and a product-specific component representing the average production-technology level of that product. With this procedure, it is unnecessary to collect the price of all known suppliers of a given product in the same month. This spawns more freedom to plan the interviews over the months, reducing data acquisition costs. To validate the new procedure, we implemented comparison tests between the adjustment of the proposed estimator (Economic) and the traditional (Statistical) to the data of real prices of inputs of the public construction sector in the country. Besides, we developed other tests based on a simulated environment where we know the actual market price and compare it with the two estimates given by the two estimators, calibrating the parameters with the ones observed in the data of real prices. Both tests showed that the proposed method performed better for every period in which not all existent suppliers of a product informed their prices. The Economic Estimator exhibited smaller variance and smaller absolute deviation to the market price.

Keywords: Public Administration; Procurements; Market Price; Reference Prices; Simulation.

JEL Codes: C53, E31, H57, H83, P35

Public Sector Procurements and Reference Prices Estimation with Small Samples in Brazil

Gabriel Lyrio de Oliveira[1]

André Luis Squarize Chagas[2]

Denise Leyi Li[3]

Abstract

This paper presents a new and simple estimator for the reference prices of products or services. These reference prices aim to work as boundaries and support the public stakeholders' decisions at the beginning of their procurements processes. We discuss the advantages of the method compared to the estimator traditionally used all over Brazil, especially in the Federal, State, and Municipal governments, which relies on the computation of averages of a few prices observed in each period. The proposed estimator decomposes the price of a product into a component of its economic sector, sector-time, supplier component representing differences in quality and productivity, and a product-specific component representing the average production-technology level of that product. With this procedure, it is unnecessary to collect the price of all known suppliers of a given product in the same month. This spawns more freedom to plan the interviews over the months, reducing data acquisition costs. To validate the new procedure, we implemented comparison tests between the adjustment of the proposed estimator (Economic) and the traditional (Statistical) to the data of real prices of inputs of the public construction sector in the country. Besides, we developed other tests based on a simulated environment where we know the actual market price and compare it with the two estimates given by the two estimators, calibrating the parameters with the ones observed in the data of real prices. Both tests showed that the proposed method performed better for every period in which not all existent suppliers of a product informed their prices. The Economic Estimator exhibited smaller variance and smaller absolute deviation to the market price.

Keywords: Public Administration; Procurements; Market Price; Reference Prices; Simulation

JEL Classification: C53, E31, H57, H83, P35

Acknowledgments: The authors are thankful for the comments of Eduardo Zylberstajn, Bruno Oliva, Maria Helena Zockun, and Jackson Rosalino to previous versions of this paper. Nonetheless, the authors are responsible for any remaining errors. The authors are also thankful to FIPE financial support granted to the development of this study.

[1] Consultant at FIPE and The World Bank / Researcher at NEREUS-USP - gabriel.oliveira@fipe.org.br
[2] Professor of Economics at the University of São Paulo (FEA/USP) - achagas@usp.br
[3] Consultant at FIPE / Researcher at NEREUS-USP - denise.leyi@fipe.org.br

1. Introduction

Purchases by public agents have specific regulations in Brazil defined by Law 8666/1993, which establishes general rules on bidding and administrative contracts relevant to works, services, including advertising, purchases, disposals, and leases within the scope of the Union, States, Federal District, and Municipalities. The procurements must be carried through a system of price registration (art. 14, item II), preceded by a broad market prices survey (art. 14 item V, par. 1). However, not always the buyer has the required expertise in market prices research. Although in common sense it might seem trivial, to develop an accurate price research demands a lot of knowledge on the details of the product or service specification, which, among other factors, is beyond the public stakeholder awareness.

Over the years, when designing the procurement process, the local custom established gathering three price queries among potential caterers achieving a minimal price registration system. Of course, the practice has no statistical confidence considering that a tiny sample from an unknown-size universe does not deliver any safety. The process of collecting three different estimates does not present any theoretical support, neither statistical nor economical, and follows only formal recommendations. However, the practice could be improved by applying scientific methods to form a reference prices list.

The price query generates crucial information for the planning of any agent, either a big one, as usually is the case of the Public Sector organizations or an individual consumer. For the Public Administration, the reference prices act as boundaries in the procurement designs, signaling an acceptable price and preventing the public resources waste (Brasil, 2015).

For important public contracts, as government construction projects, a more systematic follow-up occurs due to the money involved. In these cases, price guides exist, with a detailed list of products and services that make up the project. The inflationary process requires that the prices in these guides are constantly updated. Indeed, Brazilian legislation regulates the use of these price tables, e.g. *Sistema Nacional de Pesquisa de Custos e Índices da Construção Civil* (SINAPI), managed by the public bank *Caixa Econômica Federal* (CEF). Infrastructure and transport works use the *Sistema de Custos Referenciais de Obras*, managed by the *Departamento Nacional de Infraestrutura e Transportes* (DNIT). In the state of São Paulo, the *Índice de Preços de Obras Públicas* is computed by *Fundação Instituto de Pesquisas Econômicas* (FIPE) for over 40 years. From these different inquiries, many reference price tables were elaborated, matching the necessity of specific organizations.

Traditionally, the prices of these tables are obtained by taking the arithmetic mean of price surveys conducted periodically among potential suppliers of each product or service. Sometimes simple criteria can be applied to deal with outliers or misspecification of the product/service. We named this estimator the Market Price Statistical Estimator (MPSE) since its foundation is supposedly the statistical theory.

In real life, it is very difficult, if not impossible, for one to know and observe the prices of all suppliers of a given product in the market. For the hypothetical case in which all suppliers are known, it would still be unreal to assume that the price research is possible and financially feasible for every period for every relevant product and service. This last argument alone is enough to require developing a price estimation procedure to update the reference prices. Institutions that publish reference price tables deal with this necessity by adopting methods to estimate the unobserved prices, as shown in the SINAPI methodology (2017).

The reference price computed with a method applied to price survey data differs from the effective final price of the procurement for several reasons, as shown for Spain by García Rodriguez et al. (2019). One of them is named *query effect* (De Oliveira et al. 2010). The query procedure cannot mimic the ideally competitive bid process, resulting in higher prices than those observed in the real process. Therefore, the reference prices are higher than expected for public sector procurement. The competition in the bidding stage is required for the traded price to better reflect the market price.

A more severe problem related to the price survey estimator derives from the variation due to the different suppliers sampled in different periods. When some product or service suppliers are unknown, or even when at least one is not enquired for any reason, the Statistical Estimator will vary due to the sample change. The problem is named the *sample floating* problem, and it is more severe the smaller the number of known suppliers of a product or service.

The estimator for market prices presented in this paper is called Market Price Economic Estimator (MPEE). It relies on the economic theory of price determination and uses all available information to compute the Reference Prices. Using price observations of known suppliers and a product sector-similarity classification, we can retrieve what product price would likely have been informed by each supplier on each economic activity sector in a given period of interest, usually the current month. In other words, we obtain updated Reference Prices for all products and services regardless of which of them were observed in the current period or not.

For the MPEE existence, a given product must be classified in some activity sectors with similar products that share cost and demand components. We can retrieve an average period effect applicable by observing the prices of other similar products in the same period. Of course, we need at least one sample observation of the product in the time window considered to estimate its average specific component. Likewise, each potential supplier must have at least one price observed in the same time window to estimate each supplier's average quality and productivity in the same sector.

By considering the average product components, supplier, sector, and period in each price, we embed each Price of Reference with massive informational content through a simple structure based on economic sense. This fact is aligned with recent developments in the Prices and Price Index estimation (García Rodríguez et al. 2019, Hassim et al., 2018, Cavallo and Rigobon, 2016, Zhang, 2015) and the computational improvements that enable handling more and more data in less time.

Unlike price tables, the indexes are weighted aggregations of the reference prices estimated representing the market prices variation to a group of products, usually applied by the public or private sector to update contract terms. Therefore, it naturally comprises other potential applications of the method presented here.

In 2016, the public purchases represented 12.8% of the OECD countries GDP, and 16.4% of Brazilian GDP, according to the last Government at a Glance report (OECD, 2019). Besides this economic importance, public purchases are used as a public policy tool to foster innovation and regional development. During the recent COVID pandemic, efficient and effective public procurement becomes critical.

Even with the recent technological advances in the area, which resulted in systematization and the growing availability of data, methodological advances, especially regarding reference prices estimation, are still scarce. Papers on public procurements have focused on different matters, as assessing the aftermath performance of the process by modeling contract conditions for win-win outcomes for all parties (Brown et al., 2015) or understanding if competition leads to improved performance among contractors (Brunjes, 2019). Indeed, Trammell et al. (2020) point out low participation of Cost and Pricing modeling matters in the procurement literature and low participation of cases from the Latin America region and other developing countries. Therefore, the present paper adds a new method to estimate reference prices to this scarce literature. We apply it to the real Brazilian case.

In summary, by considering the significant role and potential impacts of public procurements, the paper presents a relevant contribution by proposing a method based on statistics and economic foundations that bring efficiency to an essential step of the procurement process. The paper is divided into four more sections. The following section presents the theoretical motivation for applying the chosen econometric model. Then, we present the econometric estimation methodology and the main characteristics of the proposed method. Next, we show the specific procedure and results of each comparison test developed, contrasting the performance of the Statistical Estimator with the Economic Estimator regarding variance and absolute deviation to the hypothetical Market Prices. Finally, we finish the paper by summarizing the main findings and usefulness of the proposed method for estimating reference prices.

2. Market Prices

Market prices can be considered random variables which successive observations can approach mathematical expected values. This principle substantiates the statistical approach for the problem of prices observation. On the other hand, the market price represents the economic equilibrium of good or service demand and supply from an economic perspective, balancing the buying and selling agent wishes in a given market. From the firm perspective, the prices need to be enough to cover the production costs, including the capital opportunity cost.

In the following subsections, we present the details of these two approaches to obtain the Reference Price of a product or service, aiming to represent its Market Price.

2.1. Market Prices: Statistical approach

Let p_i be the price of a given product for the i-th firm, in a market composed of N firms. Then, we define the average market price as

$$\overline{p} = \frac{1}{N}\sum_{\mathrm{i}=1}^{\mathrm{N}} \phi_i p_i \tag{1}$$

Where ϕ_i is the firm *i* market share. This formula states that the average market price is composed of the price of each firm of this market weighted by the frequency in which consumers face them. If consumers buy more from firm A than firm B, for any reason, the price of the first firm must have a higher weight in the average market price computation.

The survey institutes do not observe the market share of each firm on each product in practice. This is a non-observable variable. Obtaining it would mean informational gain disproportionally lower than the cost. Then, what is done is to assign equal weights to each firm, and the market price is redefined as

$$\overline{p} = \frac{1}{N}\sum_{\mathrm{i}=1}^{\mathrm{N}} p_i \tag{2}$$

Therefore, the market price is the arithmetic average of the known suppliers.

To collect the prices of all firms to compute (1) is impractical and represents another operational limitation. Once more, the cost of such an effort would not compensate for the informational gain. Then, the price is estimated by applying statistical techniques, as the arithmetic average of the sample prices.

By the Law of Large Numbers, the sample average converges to the populational one (Van der Vaart, 2000). Thus, the average market price is estimated through the following estimator

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} p_i \tag{3}$$

Where n is the number of knowns and interviewed suppliers, always smaller or at most equal to *N*. Regarding this estimator, we have that:

1. The computation of the average market price applying the observed sample data is also an estimate.
2. The estimation quality depends on the population distribution, sample size, and collection procedure of the sample data, among other factors.
3. We called this *Market Price Statistical Estimator* (MPSE) because it is based on the Statistical Theory.

2.2. Market Prices: Economic approach

The market price represents supply and demand conditions for a good or service. It is the fulfillment of a double coincidence of forces, representing the price that a set of consumers is willing to pay for a given amount of a good and the price by which the set of suppliers are willing to sell the same amount of the same good.

The whole modern microeconomic approach, heir of the classical theory of value, aim to understand price determination in different market settings: perfect competition, monopoly, oligopoly, auctions, contracts of short and long-term, among many others. In general terms, the market price depends on several factors.

On the demand side, the main factors include but are not limited to preferences structure, prices of other related goods, income, expectations, among others. On the supply side, important factors are prices of inputs like machines, wages, intermediate inputs, technology shocks, expectations, etc.

From the supply-side perspective, the market price of a good i is influenced by the cost of production in the following way

$$p_i = \mu_i c_i \tag{4}$$

where c_i is the marginal cost of production, i.e., the cost to produce an additional unit, and $\mu_i \geq 1$ is the firm mark-up, i.e., the marginal cost surplus.

2.2.1. Single-Input Firm pricing behavior

Considering a single input firm producing[4] a given good, i, its total production cost (C_i) is given by labor hours hired, l_i, multiplied by wage rate, w_i. This relation is summarized in the following equation

$$C_i = w_i l_i \tag{5}$$

Let h_i be the number of working hours needed to produce one unit of the good i. Assuming a straightforward production function that relates worked hours to the output y_i, then

$$y_i = \frac{l_i}{h_i} \tag{6}$$

That is, the number of produced goods (y_i) is fulled determined by the number of hours (l_i) divided by the technological coefficient (h_i). Replacing (6) in the total cost function (5)

$$C_i = w_i h_i y_i \tag{7}$$

The Marginal Cost, given form (7) is

$$\frac{\partial C_i}{\partial y_i} = w_i h_i = c_i \tag{8}$$

[4] In Appendix 1, we generalize the result to the multi-input firm.

Consequently, the price will be given by an equivalent to equation (4), as the following

$$p_i = \mu_i w_i h_i \tag{9}$$

Under the following assumptions on the market economic settings:

$\mu_i = \mu_f$, i.e., the mark-up is a firm decision and is similar to each product of the same firm;

$w_i = w_s$, i.e., the wage level is defined by sector; and

$h_i = h_i$, i.e., the number of hours of labor needed to produce one unit of output is determined by technical coefficients of the good (or service) and depends on the specific production-technology level.

Then, we have the following equation for the price already in logarithmic terms

$$\dot{p}_{fsi} = \dot{\mu}_f + \dot{w}_s + \dot{h}_i \tag{10}$$

Where $\dot{x} = ln\, x$. In other words, an observed price can be decomposed into components associated with the firm, sector, and good characteristics. Because the base of this model is the Economic Theory, we called it Market Price Economic Estimator (MPEE).

2.2.2. Economic Estimator Empirical Model

Starting with the equations derived in the previous sections, an econometric model based on fixed effects can be applied to decompose the observed prices and retrieve all components used to compute the reference price of a product in a given period.

Considering a sample with *N* different goods with at least one price observed along several periods and potential suppliers, we estimate the following regression with multiple fixed effects.

$$\dot{p}_{fsit} = \alpha + \hat{F}_f + \hat{S}_s + \hat{I}_i + \hat{T}_t + \epsilon_{fsit} \tag{11}$$

The left-hand side represents the logarithm of the price of the good *i*, supplied by firm *f*, in the sector *s*, and period *t*, which is decomposed in the respective fixed effects $\hat{F}_f$, $\hat{S}_s$, $\hat{I}_i$ *e* $\hat{T}_t$ and an independent part not observed by the econometrician, ϵ_{fsit}.

Because of the high number of parameters, the procedure to estimate them relies on the proposal of Guimarães and Portugal (2010), Carneiro et al. (2010), and Correia (2016)[5]. These papers propose efficient methods in computational terms that allow one to handle a vast set of fixed effects and exhibit solutions equivalent to the traditional Ordinary Least Squares.

The MPEE of good *i* in the period *t* is composed of the predicted logarithm prices for the N_i potential suppliers of this good, exponentiated and multiplied by the bias correction

[5] Appendix 2 presents additional details about the procedure.

factor[6] that is generated in the logarithmic transformation, $\hat{\gamma}$. With the predicted prices of each supplier, we compute the average that results in the Economic Estimator of the good i, $\hat{p}_{sit}$, exactly as stated in the following equation.

$$\hat{p}_{sit} = \frac{1}{N_i} \sum_{f=1}^{N_i} \hat{\gamma} \; e^{\hat{\alpha} + \hat{F}_f + \hat{S}_s + \hat{I}_i + \hat{T}_t} \quad (12)$$

3. Data and Methodology

3.1. Data

The data source used for this study was the Public Construction Prices (IPOP) from the Institute of Economic Research Foundation (FIPE). This database contains information on individual prices by product code, supplier, and reference month. In addition, depending on the item, prices are collected at monthly, quarterly, semiannual, or annual intervals, or even without defined periodicity. We use a 30-month length data, covering the period from February 2017 to July 2019.

The complete database features 12 thousand different products, around 3 thousand suppliers, and 147 sectors. Given the nature of the statistical estimator, whenever a product is selected to have its prices surveyed, the interviewers contact all registered and known suppliers of that product. This requirement of contacting all suppliers to form a reference price for the product can be relaxed with the application of the economic estimator[7], which enables a more efficient data collection strategy.

Table 1 shows some basic descriptive statistics of the 147 sectors present in the database.

Table 1 Descriptive Statistics of the 147 Sectors

Sector Characteristics	Mean	S.D.	Minimum	Maximum
Distinct products	81.96	36.16	4	234
Distinct firms	67.18	34.79	1	207
Distinct Product-Firm	296.29	149.83	9	716
Products with Observed Price on 07/2019	21.26	15.81	4	82
Product-Firm with Price on 07/2019	64.17	56.21	8	263

The table reports descriptive statistics (mean, standard deviation, minimum and maximum) of the sectors present in the database. The statistics were based on observations from the 30 months considered and refer to products, firms, and product-firm combinations by sectors, as not all firms supply all products of a sector.

[6] Under the assumption of error normality, the statistical expectation of the correction factor of this bias is given by $e^{\frac{\sigma^2}{2}}$, which is always bigger than one. For a small variance like 10%, as is the case of the FIPE dataset we have, the expectation of this factor would then be 1.0050. In other words, the reference price computed by the Economic Estimator not corrected by this factor would be, on average, underestimating the price in 0,5%. However, for a given sample, by applying the correction procedure suggested by Wooldridge (2015) that does not require the normality assumption, there is no guarantee that the factor will be bigger than 1.

[7] The method allows each supplier of a particular product to be surveyed in a different month. Still, this allows an average reference price, as we have the estimated price for all months for each supplier that has at least one observation in the 30-month interval considered.

In addition, statistics are reported for products and combinations of product-firm with at least one price observed in the last month of the period considered, in this case, July 2019.
Source: FIPE, Author's elaboration.

The table shows that, on average, sectors have 81.96 different products, 67.18 distinct firms, which do not necessarily supply all products of a sector, and 296.29 known product-firm combinations. In July 2019, an average of 21.26 products were observed in each sector. In the same month, an average of 64.17 product-firm pairs had their price collected.

From this database and using observation selection strategies, it is possible to test the performance of Statistical and Economic Estimators. The following subsection describes the strategies used to carry out comparisons between them.

3.2. Comparing both estimator MPSE and MPEE

The MPSE is calculated for the sample collected in a specific month based on the data observed in the same month. On the other hand, the MPEE uses current and past information to calculate the reference price for the month. Thus, these estimators vary depending on how samples are collected and how transitory or permanent price shocks are incorporated. We compared the performance of the two estimators in two alternative experiments: i) using simulated data and ii) using real data of prices from the FIPE's price database.

For the first experiment based on simulation, data are generated assuming parameters and distributions for the variables of interest. The product, supplier, and period effects are randomized, as well as the random term not explained by the model. With all components randomized, it is possible to generate the simulated price of each product-firm-period. An auxiliary variable is also drawn, indicating which prices should be "observed," representing the impossibility of collecting prices from all product suppliers. Then, the Economic and Statistical Estimators are calculated to assess which one is closest to the Market Price.

Two different procedures are applied to compare the two estimators for the second experiment, which considers the real data from the FIPE database. Before discussing each of these procedures, it is worth noting that an important difference between the economic estimator and the statistical estimator refers to the way each one of them deals with unobserved data. The MPSE gives weight only to prices observed, while the MPEE considers the historical prices available, even if not observed in a specific interest month.

The first procedure of the second experiment assesses the distance between the Economic and Statistical Estimators. The first estimator is calculated with a truncated sample, while the second is computed with all the available information. With this procedure, it is possible to evaluate the distance between the estimators outside the sample, in other words, considering periods for which the prices observed were not used by the MPEE.

The second procedure randomly draws a variable that represents the observation or not of each product-supplier pair. It is assumed that the complete sample, including those drawn to be ignored, represents the Market Price. Each estimator is calculated according to its method and then compared with the supposed Market Prices.

The number of observations per sector and month is controllable by the institution carrying out the price survey to a certain extent. For spurious price shocks for the products surveyed in a given month to not be fully passed on to the reference price, it is necessary to have a sufficiently large sample. In the tests performed, without prejudice to their validity, it is possible to direct the analysis only to the sectors with an adequate number of observations drawn each month.

4. Results

This section presents the different experiments carried out to compare the MPSE and MPEE in terms of variability and consistency, intending to get closer to the market prices, the supposed desired parameters.

The experiments presented in sections 4.1 and 4.3 share a common objective and can be directly compared. Both exercises define the true market price, hide some observations and compute the Economic and Statistical Estimators. Furthermore, both estimators are compared to check which one, on average, comes closest to the assumed true market price. The fundamental difference between the sections is the data used. In the first, the data are simulated, and the market prices are known by construction. In the latter, the data are real, and it is assumed that all existing suppliers in FIPE's price database constitute all firms in its market for each product.

Finally, section 4.2 evaluates the distance between the MPSE and MPEE. This section brings an important result, demonstrating that an eventual transition from one estimator to other, on average, results in a change of relatively small magnitude at the level of Reference Prices.

4.1. Experiment with Simulated Prices

We calibrated the present simulation[8] exercise parameters based on what we observed from the FIPE's Public Construction Prices Index (IPOP-FIPE) database. Data were generated 100 times for a hypothetical sector of 100 products, in 30 periods and with 100 different suppliers. As with the real database, not all firms supply all products, and not all suppliers for each product are necessarily known to the institution that collects the prices.

The procedure adopted to simulate the Market Price and test the estimators is described as follows:

1- A database with 100 products, 30 periods, and 100 suppliers is created, with an average sector effect assumed to be equal to number 3.
2- The probability of each firm supplying each product was set at 10%; in other words, each firm supplies 10 products on average in the sector, defined by drawing a binomial indicator variable.

$$Supplies_{if} \sim Binomial(1, 0.1)$$

3- The probability defined for a supplier to be never found by the institution that collects prices is 20%, defined by the indicator variable drawn by the following distribution

$$Supplier_f \sim Binomial(1, 0.2)$$

4- The product, firm, and period average effects were drawn from normal distributions with zero mean given, respectively, by

$$I_i \sim N(0, 2)$$

$$F_f \sim N(0, 0.1)$$

$$T_t \sim N(0.02\, t, 0.1)$$

An average increasing trend of 2% per month was assumed[9].

5- For each simulation, we drew a variable that indicates the observation or not of each price, as follows

$$Missing_{ift} \sim Binomial(1, 3/12)$$

Each price is expected to be observed 3 times every 12 months.

6- For each simulation, we drew a variable normally distributed to represent a random shock for each product-firm-period observation, as follows

$$\varepsilon_{ift} \sim N(0, 0.12)$$

The standard deviation of 12% was chosen as it approximates the average difference between the estimated price for the product-supplier-period and that observed in the real prices database used in this study.

7- Then, log-prices were generated by the following equation:

$$log(p_{ift}) = 3 + I_i + F_f + T_t + \varepsilon_{ift} \quad (13)$$

Where intercept value (3) matches the average data parameter. Moreover, the Market Price of each product in a given period was defined by the following formula:

$$p_{it} = \frac{1}{N_i} \sum_f e^{log(p_{ift})} \quad (14)$$

In other words, it is the average of the exponential of the logarithms of the prices of the N_i firms that supply product *i* in period *t*.

For descriptive purposes, Table 2 presents the statistics of the 100 simulations performed for the first 10 products. The simulated Market Price, the Statistical Estimator, and the Economic Estimator for the prices are reported, both calculated without the observations drawn to be hidden.

[9] Not considering such a trend proved to be indifferent to the results obtained.

Table 2 Descriptive Statistics of the 100 Simulations

Product	**Market Price** (Mean)	**S.D.**	**MPSE** (Mean)	**S.D.**	**MPEE** (Mean)	**S.D.**
1	4.03	0.12	4.02	0.37	4.03	0.06
2	1.81	0.06	1.78	0.15	1.81	0.03
3	7.10	0.41	6.96	0.61	7.17	0.18
4	4.05	0.14	4.15	0.39	4.04	0.08
5	39.13	1.57	39.03	4.01	39.27	0.68
6	158.98	5.28	157.87	16.36	158.88	2.72
7	3.26	0.11	3.16	0.27	3.27	0.05
8	81.86	3.36	82.62	9.26	82.21	1.45
9	3.57	0.17	3.61	0.48	3.57	0.09
10	79.23	2.76	80.91	8.32	78.74	1.51

MPSE is the Market Price Statistical Estimator; MPEE is the Market Price Economic Estimator.
For 10 selected products, the table shows the simulated market price's mean and the respective standard deviation (S.D.) for the last simulated period (t = 30). The following columns show the mean of the Statistical Estimator and Economic Estimator, in addition to the respective standard deviations. Source: FIPE, Author's elaboration.

The results for the calculated fit measures for the 100 simulations combined are shown in Table 3.

Table 3 Economic and Statistical Estimators Absolute Deviations from the Simulated Market Price

Product Group	**Obs.**	**Mean**	**S.D.**	**Min**	**Max**
1. All products					
Economic Estimator	10,000	3.5%	2.7%	0.0%	23.4%
Statistical Estimator	8,622	7.9%	6.7%	0.0%	57.2%
2. Products with some observation in the month (n > 1) but not the entire market (n < N)					
Economic Estimator	8,620	3.3%	2.6%	0.0%	20.8%
Statistical Estimator	8,620	7.9%	6.7%	0.0%	57.2%
3. Products with no observation in the month (n = 0)					
Economic Estimator	1,378	4.2%	3.2%	0.0%	23.4%
Statistical Estimator	ND	ND	ND	ND	ND
4. Products with many suppliers on the market (N > 5)					
Economic Estimator	9,000	3.3%	2.6%	0.0%	17.8%
Statistical Estimator	7,937	7.8%	6.7%	0.0%	57.2%
5. Products with many observations (n/N > 70%) but not the entire market (n < N)					
Economic Estimator	9	2.4%	2.5%	0.1%	7.6%
Statistical Estimator	9	2.9%	2.2%	0.6%	7.6%
6. Products with few observations (n/N < 30%)					
Economic Estimator	7,794	3.5%	2.7%	0.0%	23.4%
Statistical Estimator	6,416	8.8%	7.2%	0.0%	57.2%

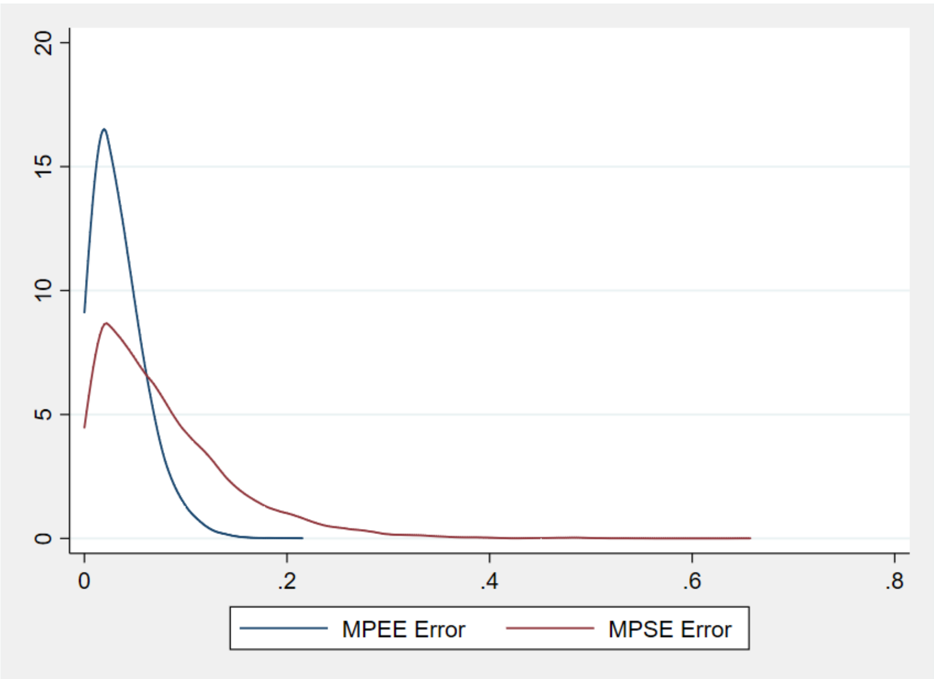
Each scenario in this table considers a specific group of products. The first scenario includes all products. In this case, there are fewer observations in the Statistical Estimator, as products in which no supplier (n) was observed do not present a Reference Price for this estimator. The second scenario selects all products

where at least one supplier has had its price chosen to be observed, but not all suppliers in the market (N). The third scenario only considers products for which no supplier has observed its price. The fourth scenario selects products whose market comprises many suppliers (N > 5), regardless of how many of these were observed. The fifth scenario selects products for which the research institute has reached most existing suppliers (n/N > 70%), but at least one remained unknown (n < N). The last scenario considers products the research institute knows less than 30% of market suppliers.
Source: Author's elaboration.

A mean absolute deviation from the true market price of 3.5% is observed for the MPEE, while the MPSE presents a mean deviation of 7.9%. Even for products in which the majority of the suppliers in the market were observed (70% < n <100%), the MPSE, though still higher (2.9% against 2.4%), approximates its mean error to the Economic estimator's. In the 100 simulations, only 9 products fit into this scenario.

Next, Figure 1 presents both estimators' absolute relative error distributions relating to the market price. This figure shows products with at least one, but not all, of its market suppliers surveyed in the month of interest. Note that the error distribution of the MPEE is more shifted to the left, therefore relatively more concentrated around zero.

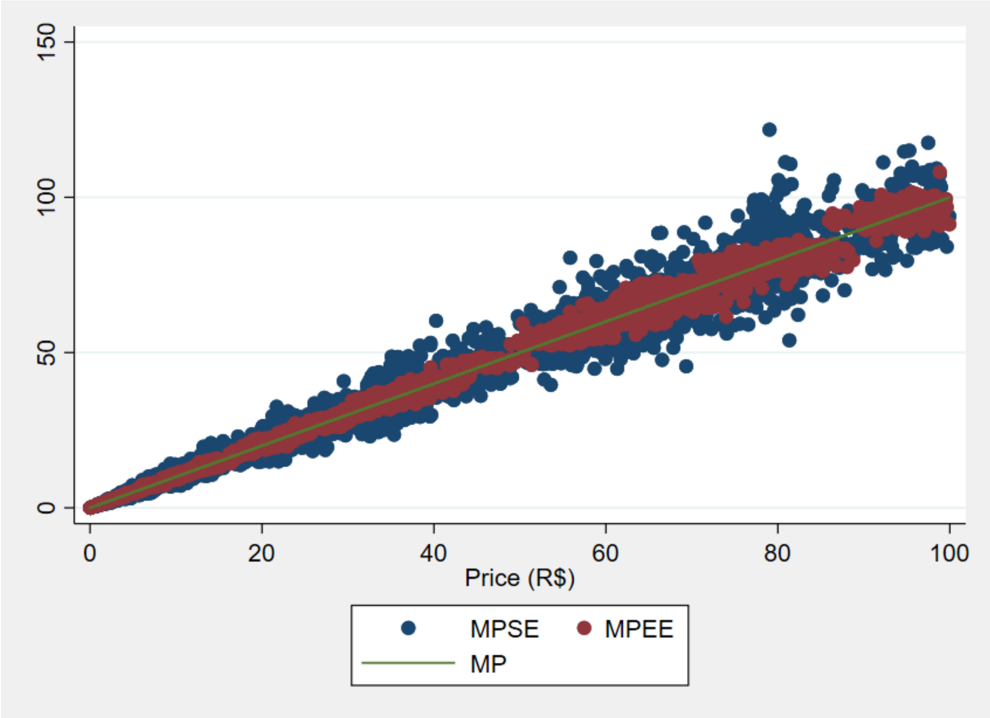
Figure 1 Error Distribution – MPEE vs. MPSE



MPSE is the Market Price Statistical Estimator; MPEE is the Market Price Economic Estimator.
Source: Author's elaboration.

The same result can be seen in Figure 2, where each red point represents a pair (Market Price, Economic Estimator) and each blue point a pair (Market Price, Statistical Estimator). The green line represents the location where the estimator and the true price coincide.

Figure 2 Reference Prices – MPEE, MPSE and true market price



MPSE is the Market Price Statistical Estimator; MPEE is the Market Price Economic Estimator; MP is the true market price.
Source: Author's elaboration.

Again, we observe a greater dispersion of the Statistical Estimator than the Economic one.

4.2. Distance between MPEE and MPSE

This experiment investigates the distance between the MPEE MPSE. For the scenario that analyzes the distance between both estimators completely out of the sample, we consider the products in which all observations for the month of interest were hidden. Through hidden observations, we were able to calculate what would be the MPSE. On the other hand, even disregarding these observations, we calculated the MPEE for these hidden periods using historical data. The steps in this procedure are outlined below.

1- An interval was determined to observe the prices of all products. We adopted a minimum interval of 12 months from one observation to another in the present case.

2- For each product, one of the months in which it has observed prices was selected, and we assume that counting from this month, only prices observed in months multiples of the defined interval remain considered. There would be some different products for each month with prices observed in each sector.

3- The observations of the months coinciding with the selected interval permitted the calculation of the Economic Estimator for all periods.

4- The logarithmic transformation bias was corrected using the method suggested by Wooldridge (2015), allowing each sector to have a different bias parameter.

5- Being the last and most recent month, July 2019 is the most interesting month in the process and was the only one kept in the database to report the desired distance measurements. To compute the distance measurements of the estimators, in one of the scenarios, only the products drawn were considered so that the July observations were hidden. Thereby, the distance measure represents an out-of-sample fit test since these are the products whose July observations were discarded in the calculation of the Economic Estimator.

After performing the steps described above, the measure of deviation between both estimators was calculated. The Absolute Deviation is defined as the module of the difference between the Reference Price generated by the Economic Estimator and that generated by the Statistical Estimator, assuming the same sample of suppliers. The formula applied is given by the following equation.

$$AD_{econ} = \left| \frac{\widehat{P_{econ}} - \widehat{P_{estat}}}{\widehat{P_{estat}}} \right|$$

Where $\widehat{P_{econ}}$ represents the Economic Estimator for a particular product in July 2019, and $\widehat{P_{estat}}$ represents the Statistical Estimator. As a result of the above procedure, the values reported in Table 4 were obtained. We hid some of the observed products to have an outside the sample comparison group. Therefore, the number of products for each measure is considerably smaller than the 12 thousand products present in the database.

Table 4 Distance between Economic and Statistical Estimators

Group Products	**# of Products**	**Mean**	**S.D.**
All the products			
A.D. (MPEE - MPSE)	3,022	7.3%	8.5%
Products not observed in the month (n = 0)			
A.D. (MPEE - MPSE)	2,137	8.9%	9.2%
Products from Sectors with at least 20 observations in the month			
A.D. (MPEE - MPSE)	1,454	6.4%	7.3%
Unobserved products from Sectors with at least 20 observations in the month			
A.D. (MPEE - MPSE)	946	7.8%	7.9%

MPSE is the Market Price Statistical Estimator; MPEE is the Market Price Economic Estimator.
The table reports the number of products in each tested scenario and the mean distance and standard deviation between the Economic and Statistical Estimator. The first scenario selects all products for which FIPE observed prices in July 2019, regardless of whether or not they were hidden to calculate the Economic Estimator. The second scenario consists of only those products where FIPE observed prices in July 2019 but were supposed to be absent from calculating the Economic Estimator. The third scenario considers the group of products belonging to sectors with at least 20 observations left to calculate the Economic Estimator after the prices of some of its products were hidden. The fourth scenario constitutes a subgroup of products from the third scenario, in which we select only products that were drawn to be hidden in the period.
Source: FIPE, Author's elaboration.

From Table 4, we note that for the out-of-sample case, that is, when the month's prices were used to calculate the Statistical Estimator but were not used to calculate the Economic Estimator for the same month, the distance between the Economic Estimator and the Statistical Estimator is on average 8.9%. This deviation is relatively low if we consider that the standard deviation of observed supplier prices for each product in July 2019 was, on average, 20.4%, that is, more than twice the average distance measure presented.

In addition, it is interesting to note that when selecting products belonging to sectors that had the highest number of observations in July 2019, the deviation was slightly smaller, around 7.8%.

The following subsection presents the last experiment developed, in which the performance of both estimators in approaching the supposed true Market Prices is compared.

4.3. Experiment with Real Prices

In this test, we use the IPOP-FIPE database again. It is assumed that each month's average price of all observed suppliers constitutes the true market price for a product. The MPSE and MPEE are calculated and compared to the supposed true parameter by randomly drawing some observations to be disregarded. The procedure is detailed below:

1- From the price history from February 2017 to July 2019, a random variable was drawn from a Binomial(1, 0.3) distribution. It means a 30% chance of success, i.e., the price Product-Firm-Period will be hidden in the estimation.
2- Computed the MPSE and MPEE with the non-hidden observations.
3- With the hypothesis that the Average Price considering all prices of each period (including the 30% hidden) is the Market Price, comparisons were made between the two estimators for the month of interest. The formulas used are similar to those presented in the previous section, as reported below:

$$AD_{econ} = |\frac{\widehat{P_{econ}} - P}{\mathrm{P}}|$$

$$AD_{stat} = |\frac{\widehat{P_{stat}} - P}{\mathrm{P}}|$$

Where P is the supposed true market price.

Next, Table 5 reports the results obtained for different groups of products. In general terms, the Economic Estimator's error is smaller than that of the Statistical Estimator, being greater only for the scenario that includes all products in which all registered suppliers in the month are observed. However, the statistic error is zero by construction in the latter case. Furthermore, this is expected because when the sample observed in the month increases, the Statistical Estimator converges to the definition supposed for the Market Price.

Table 5 Absolute Deviation from Market Price: Economic Estimator vs. Statistical

Estimator	Obs.	Mean	Min	Max
1. All products				
Economic	3,111	5.2%	0.0%	114%
Statistical	2,775	4.1%	0.0%	58.2%
2. Products with at least one observation per month (n > 1) but not all (n < N)				
Economic	1,457	4.5%	0.0%	57.3%
Statistical	1,457	7.9%	0.0%	58.2%
3. Products with no observation in the month (n = 0)				
Economic	336	8.4%	0.0%	114%
Statistical	ND	ND	ND	ND
4. Products with many suppliers on the market (N > 5)				
Economic	380	3.2%	0.0%	19.9%
Statistical	380	4.1%	0.0%	25.8%
5. Products with many observations (n/N > 70%) but not the entire market (n < N)				
Economic	410	3.9%	0.0%	25.1%
Statistical	410	4.3%	0.0%	20.6%
6. Products with few observations (n/N < 30%)				
Economic	394	7.9%	0.0%	114%
Statistical	58	13.7%	0.0%	47.5%

MPSE is the Market Price Statistical Estimator; MPEE is the Market Price Economic Estimator.
Each scenario in this table considers a specific group of products. The first scenario includes all products. In this case, there are fewer observations in the Statistical Estimator, as products in which no supplier (n) was observed do not present a Reference Price for this estimator. The second scenario selects all products where at least one supplier is chosen to be observed, but not all suppliers in the market (N). The third scenario only considers products for which no supplier has observed its price. The fourth scenario selects products whose market comprises many suppliers (N > 5), regardless of how many of these were observed. The fifth scenario selects products for which the research institute has reached most existing suppliers (n/N > 70%), but at least one remained unknown (n < N). The last scenario considers products the research institute knows less than 30% of market suppliers.
Source: Author's elaboration.

When the adjustment out-of-sample is considered, the Statistical Estimator for Market Price is not defined. Nonetheless, the Economic Estimator can be calculated and has an average deviation of 8.4%, which, as mentioned in the previous section, is relatively low compared to the average standard deviation of 20.4% of supplier prices for the same product in the same month.

When selecting products for which more than 70% of the market's suppliers are observed, the Economic Estimator still presents a better adjustment to the Market Price than the Statistical Estimator. Interestingly, when we look at less than 30% of the market, both estimators worsen their fit, which is expected because much information is lost. However, the Economic Estimator shows a much smaller distance from the Market Price (7.9% on average) than the Statistical Estimator (13.7%).

In short, the only case in which the Statistical Estimator adjusted better on average to the "Market Price" occurs when we select all products in which we observe all of its suppliers in the market in the month of interest (July 2019). This scenario is very unlikely, as even if one wants to obtain prices from all suppliers, sample attrition will prevent it. Moreover, it would be highly costly, as all suppliers should be interviewed every month for all products.

Table 6 presents an overview of the main results obtained from the different experiments applied in this paper. In summary, for small samples, when it is not possible to reach all suppliers in the market of a product, the MPEE proved to be a better and more robust alternative to the average of observed prices (MPSE).

Table 6 Results Summary

Experiment	Measure	MPEE	MPSE
Distance between Estimators	Average A.D. from mean observed price	7.3%	0
Experiment with Real Prices	Average A.D. from Market Price	4.5%	7.9%
Experiment with Simulated Prices	Average A.D. from Market Price	3.3%	7.9%

Source: FIPE, Author's elaboration.

The main results of the three types of experiments applied can be summarized as follows:

- Adjustment to the Statistical Estimator (average observed price): Reference prices obtained by the Economic Estimator are close to those obtained by the Statistical Estimator, even when comparisons are made out-of-sample.

- Experiment with Real Prices: Even supposing that the institute responsible for the price collection reached the entire market, by hiding part of it, the Economic Estimator showed to be closer to the Market Price than the Statistical Estimator.

- Experiment with Simulated Prices: In a simulated environment where we know all market suppliers and their prices, when we do not observe part of it, the Economic Estimator showed to be closer to the true Market Price than the Statistical Estimator.

5. Conclusion

We have demonstrated that the Market Price Economic Estimator for the true market price applied to surveyed prices among potential suppliers reduces variations that are not economically justifiable. Thus, it is a valid alternative as an estimator and a more robust method for constructing Reference Prices for the Public Sector.

In general terms, we show that adopting an econometric model that takes advantage of the informational content of a big price database allows for a reduction in the number of visits to suppliers throughout the year. It keeps reference prices more stable and updated than the Marke Price Statistical Estimator, and therefore it can be considered more robust for Public Sector users. Furthermore, in the tests carried out in which the true Market

Price is assumed, the presented Economic Estimator performed better by getting more closer to it, on average, than the Statistical Estimator.

A characteristic of field data collection that can easily be observed in our database is its extreme variability, with successive price rises and falls. Some of these variations may reflect the very short-term economic condition of the supplier (e.g., considering discounts). However, such measures may not be adequate to reflect the market price that will be practiced in public construction or used to support price adjustments in public contracts. Other price variations consist of pure measurement errors that can occur with changes of the suppliers, the price collector, or during the information registry process (e.g., errors in the questionnaire, typing errors). It is not uncommon for these errors in the data collection process to be corrected months later. However, the corrections introduce a spurious variation into the price series: an increase followed by a sharp fall, or vice versa when in reality, the "market price" had no variation.

The MPEE is an alternative approach, founded on statistics and Economic Theory, to determine the market price from a sample of observed prices. The difference is the use of full information available, current and past observations, and sector and supplier characteristics, and not only a few observations in a given period.

The Economic Estimator method - which should be seen as a potential family of statistical models - is more robust as it uses more extensive information to obtain the market price. This method is less sensitive to contemporary price variations, one of its main advantages. Depending on the situation, admitting a price variation when it has not occurred can be more harmful than smoothing it when it occurs. In this regard, analysts of price institutes deal arbitrarily with sharp variations by asking for price confirmation in the following months. The presented estimator makes this procedure statistically justifiable without relying on human judgment, which can be flawed.

Some of the extensions and future research for the proposed method are straightforward but interesting paths. For instance, the prediction of future tender prices (Wong, 2010) or adjustments to help transform the suppliers' informed estimates of their tender price into the procurements' winner-prices (García Rodríguez et al., 2019). As argued, the public procurement's economic importance is currently not represented in the number of methodological studies in the literature, especially regarding the generation of the most important parameter of the process, the reference price. Hopefully, many studies may come in the next years presenting extensions and improvements in the topic.

6. References

ABOWD, J. M., CREECY, R. H., & KRAMATZ, F., (2002). Computing person and firm effects using linked longitudinal employer-employee data. Longitudinal Employer-Household Dynamics Technical Papers, No. 2002-06. ***Center for Economic Studies***, U.S. Census Bureau.

BRASIL, F. (2015) Preço de Referência em Compras Públicas (Ênfase em Medicamentos) – ***Cartilhas, Manuais e Tutoriais - TRIBUNAL DE CONTAS DA UNIÃO (TCU)*** – available in: <https://portal.tcu.gov.br/biblioteca-digital/preco-de-referencia-em-compras-publicas-enfase-em-medicamentos.htm>

BROWN, T. L., POTOSKI, M.; SLYKE, D. Van, (2016). Managing complex contracts: A theoretical approach. ***Journal of Public Administration Research and Theory***, v. 26, n. 2, p. 294-308, 2016.

BRUNJES, B. M., (2020) Competition and federal contractor performance. ***Journal of Public Administration Research and Theory***, v. 30, n. 2, p. 202-219, 2020.

CARNEIRO, A., GUIMARÃES, P., & PORTUGAL, P. (2012). Real wages and the business cycle: Accounting for worker, firm, and job title heterogeneity. ***American Economic Journal: Macroeconomics***, *4*(2), 133-52.

CAVALLO, A., RIGOBON, R. (2016). The billion prices project: Using online prices for measurement and research. ***Journal of Economic Perspectives*** 30, no. 2 (2016): 151-78.

CORREIA, S., (2016). REGHDFE: Estimating linear models with multi-way fixed effects. ***Stata Users Group***. Available in <https://ideas.repec.org/s/boc/scon16.html>

DE OLIVEIRA, L., LIMA, M. C., & MACIEL, R. G., (2010). Efeito barganha e cotação: fenômenos que permitem a ocorrência de superfaturamento com preços inferiores às referências oficiais. ***Revista do TCU***, (119), 29-36. Available in <https://revista.tcu.gov.br/ojs/index.php/RTCU/article/view/188>

GARCÍA RODRÍGUEZ, M.J., RODRÍGUEZ MONTEQUÍN, V., ORTEGA FERNÁNDEZ, F. and VILLANUEVA BALSERA, J. M., (2019). Public procurement announcements in Spain: regulations, data analysis, and award price estimator using Machine Learning. ***Complexity***, 2019.

GUIMARAES, P., PORTUGAL, P. (2010) A simple feasible procedure to fit models with high-dimensional fixed effects. ***The Stata Journal***, v. 10, n. 4, p. 628-649.

HASSIM, S., MUNIANDY, R., ALIAS, A.H. and ABDULLAH, P., (2018). Construction tender price estimation standardization (TPES) in Malaysia: Modeling using fuzzy neural network. ***Engineering, Construction and Architectural Management***.

OECD (2019). Government at a Glance 2019. Paris: ***OECD Publishing***. Available in <https://www.oecd-ilibrary.org/governance/government-at-a-glance-2019_8ccf5c38-en>

SINAPI (2017) Sistema nacional de pesquisa de custos e índices da construção civil: Métodos de Cálculo / IBGE, ***Coordenação de Índices de Preços***. – Rio de Janeiro: IBGE, 2017.

TRAMMELL, E., ABUTABENJEH, S., DIMAND, A., (2020) A review of Public Administration research: Where does public procurement fit in? ***International Journal of Public Administration***, v. 43, n. 8, p. 655-667, 2020.

WOOLDRIDGE, J. M., (2010). ***Econometric analysis of cross section and panel data*****.** MIT press.

WOOLDRIDGE, J. M., (2015). ***Introductory econometrics: A modern approach***. Cengage learning.

VAN DER VAART, A. W., (2000). ***Asymptotic statistics*****.** Vol. 3. Cambridge University Press.

ZHANG, Y., LUO, H., & HE, Y., (2015). A system for tender price evaluation of construction project based on big data. ***Procedia Engineering***, *123*, 606-614.

Appendix 1: Multi-input Firm pricing behavior

In this section, we extend the theoretical structure presented in the text to the case in which firms may use more than one production input. Despite being more complicated, the result is still similar to the single-input firm case.

The total cost of production is given by the market value of the quantity of input used per product unit. Let w_{ij} is the cost of input *j* (*j* = 1,…, *J*) per unit of product *i*, and let h_{ij} be the quantity of input *j* needed to produce a unit of product, then the total cost of production (C_i) will be

$$C_i = \sum_{j=1}^{J} w_{ij} h_{ij} y_i \tag{A1.1}$$

The marginal cost is the derivative of C_i concerning the product y_i, that is

$$\frac{\partial C_i}{\partial y_i} = \sum_{j=1}^{J} w_{ij} h_{ij} \tag{A1.2}$$

With, substituting in the price equation (5), is

$$p_i = \mu_i \sum_{j=1}^{J} w_{ij} h_{ij} \tag{A1.3}$$

In logarithm, equation (A1.3) becomes

$$\ln p_i = \dot{p} = \ln \mu_i + \sum_{j=1}^{J} \ln w_{ij} + \sum_{j=1}^{J} \ln h_{ij} + \ln\left(\frac{\sum_{j=1}^{J} w_{ij} h_{ij}}{\prod_{j=1}^{J} w_{ij} h_{ij}}\right)$$

$$= \dot{\mu}_i + \sum_{j=1}^{J} \dot{w}_{ij} + \sum_{j=1}^{J} \dot{h}_{ij} + c_i \tag{A1.4}$$

where

$$c_i = \ln\left(\frac{\sum_{j=1}^{J} w_{ij} h_{ij}}{\prod_{j=1}^{J} w_{ij} h_{ij}}\right) \tag{A1.5}$$

As before, the mark-up is a fixed effect firm-specific, the input prices are defined in the market, and it is reasonable that they are sector-specific, the input fixed technical coefficients are good-specific, and, for *J* sufficiently big (but not infinite), c_i is negligible.

Appendix 2: Econometric Model

Each price is decomposed into an average product effect, an average supplier effect, sector, and an average month effect. For this purpose, we adapted Abowd et al. (2002) methodology, which is still evolving (Correia, 2016) and deals with large groups of fixed effects. In our case, we use a database with more than 10 thousand products and with 4 effect groups, namely Sector, Product, Firm, and Time.

We specify an economic model for reference prices considering the following hypotheses:

- Similar products are subject to common temporal components;

- Suppliers differ in terms of quality and price level in each market sector;

- Each product has a medium-term equilibrium price, which is only shifted in the event of technological or permanent demand shocks;

The method applied to generate the reference prices can be summarized as follows: the items are divided into groups of products with some degree of similarity, called Inference Groups; with some observations for each product, each supplier, and each month, the following equation is estimated:

$$\log(p_{fsit}) = \alpha + S_s + I_i + F_f + T_t + \varepsilon_{ift} \qquad \text{(A2.1)}$$

Where *s* indicates the sector, *i* indicates product, *f* indicates the firm, and *t* indicates the reference month of the observation. This model decomposes each price into average effects of the sector (*S*), the product (*I*), the firm/supplier (*F*), and the time (*T*), and an unobserved random shock (ε).

Using econometric language, this model is called the Fixed Effects Model[10] and is estimated by incorporating indicator variables (dummies) for each sector, product, supplier, and period. Due to computational issues[11], this procedure can be circumvented by more efficient algorithms, called High Dimensional Fixed Effects (Abowd et al., 2002, Guimarães and Portugal, 2010, and Correia, 2016), which are numerically equivalent to the results of the method of Least Squares incorporating indicator variables.

Once this equation is estimated, it is possible to select which sample of suppliers will be considered for each product and each month[12] and then take the average of the estimated prices of the suppliers considered in the reference month.

For each new period, this model is estimated using the recent 30 months' time window. Older information will no longer directly influence the estimated product, supplier, period effects, and consequently the reference price. This way, if the product has a trend different

[10] Although the Fixed Effects are updated every month, we chose to keep the name traditionally used by the Panel Econometrics literature (Wooldridge, 2010).

[11] The inclusion of a variable for each product, each supplier, and each period can lead to problems involving a lack of computing capacity. Therefore, the alternative method of the equivalent numerical solution is relevant to the procedure.

[12] Usually, one is interested in the most recent month of the 30-month time window. In the following month, new observations will enter, and the oldest month's observations will be discarded.

from its sector, as new observations of this product enter the database over time, its fixed effect will gradually absorb this trend. The same is true for the firm-specific-effect trends.

Since the model is estimated in logarithm, we consider the exponential of each estimated price before performing the average to reconstruct the average price. This operation generates an estimation bias (Wooldridge, 2015 pg. 190-192), which can be resolved by regressing the price against the expected price without correction and an intercept. That is, with the price estimated by (A2.1), the following regression is performed:

$$p_{ift} = \gamma \, \hat{p}_{ift} + \xi_{ift} \tag{A2.2}$$

Where:

$$\hat{p}_{ift} = \exp\left(\hat{\alpha} + \widehat{S_s} + \widehat{I_\iota} + \widehat{F_f} + \widehat{T_t}\right) \tag{A2.3}$$

Estimating γ by (A2.2) leads to the corrected estimated price:

$$\widehat{\widehat{p_{ift}}} = \hat{\gamma} \, \hat{p}_{ift} \tag{A2.4}$$

The sections that present the tests always report the adjustment measures for the prices corrected by this procedure for both simulated and real prices.