NBER WORKING PAPER SERIES

HOW MANAGERS CAN USE PURCHASER PERFORMANCE INFORMATION TO IMPROVE PROCUREMENT EFFICIENCY

Pablo A. Celhay
Paul Gertler
Marcelo Olivares
Raimundo Undurraga

Working Paper 32141
http://www.nber.org/papers/w32141

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2024

We thank Piero Zanoco for excellent research assistance and acknowledge funding from Fondo Evaluación de Impacto DIPRES 2020, FONDEF Project ID19I10303, Instituto Sistemas Complejos de Ingenieria and the Institute of Market Imperfections and Public Policy, and UC Berkeley's Lab for Inclusive Financial Technology. Undurraga acknowledges the funding and support provided by the Millenium Nucleus MIGRA, ANID - MILENIO - NCS2022_051 as well as Fondecyt Regular Project No. 1230510. Olivares acknowledge financial support from ANID PIA/APOYO AFB220003. Undurraga and Olivares further acknowledge funding from Fondecyt Regular Project No. 1230416. The authors declare that they have no material or financial interests in the results of this paper. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

How Managers Can Use Purchaser Performance Information to Improve Procurement Efficiency
Pablo A. Celhay, Paul Gertler, Marcelo Olivares, and Raimundo Undurraga
NBER Working Paper No. 32141
February 2024
JEL No. D23

ABSTRACT

We examine the effect of performance monitoring in public procurement through the lens of organizational culture in a principal-agent model where the manager (principal) and buyers (agents) may have different beliefs about how much the government values efficiency. We show that the effect of performance information not only increases efficiency but is greater when the buyer's belief is stronger than the manager's belief. We leverage a new e-procurement system in Chile to test these ideas by randomizing monthly reports on the purchasing performance of buyers and further whether the individual performance reports were disclosed to managers. We find that the reports generated sizable reductions in overspending— with savings reaching a 15% reduction or 0.1% of GDP — but only when individual performance was observable to managers. This is consistent with extrinsic motivation rather than intrinsic motivation driving buyer behavior. Consistent with the theoretical model, we also find that the gain in efficiency is concentrated in procurement units where buyer belief that the government cares about efficiency is stronger than manager belief. Our results highlight the key role played by organizational culture in mediating the impact of purchasing performance information on preventing the misuse of public resources.

Pablo A. Celhay
School of Government
and Economics Institute
Pontificia Universidad Católica de Chile
Avda. Vicuña Mackenna 4860
Macul, Santiago
Chile
pacelhay@uc.cl

Paul Gertler
Haas School of Business
University of California, Berkeley
Berkeley, CA 94720
and NBER
gertler@haas.berkeley.edu

Marcelo Olivares
Universidad de Chile
Industrial Engineering Department
molivares@uchile.cl

Raimundo Undurraga
Department of Industrial Engineering
and Center for Applied Economics
University of Chile
Beauchef 851, Santiago
Chile
undurraga.raimundo@gmail.com

1. Introduction

In public procurement, efficiency is rarely a primary concern. Procurement officers (buyers) spend other people's money and procurement itself is characterized by incomplete contracts and significant transaction costs, all of which give rise to serious moral hazard concerns (Laffont and Tirole, 1993; Bajari and Tadelis, 2001). Buyers may put little effort into finding the best prices, as the absence of proper organizational incentives discourages them from achieving value for money (Liebman and Mahoney, 2017, Bandiera et al., 2021b). Wishful thinking and willful blindness that cause denial of any inefficiency, can easily pervade organizations, leading to the belief that efficiency is inconsequential (Bénabou, 2013). However, pursuing efficiency can prove to be a means of large budget savings that is especially valuable for organizations facing tight resource constraints. Indeed, there exists significant potential for such savings due to substantial price variation in public purchases (Best et al., 2022), much of which might be attributed to passive waste (Bandiera et al., 2009b)[1].

One popular approach to improving efficiency is through monitoring the performance of individual buyers.[2] Managers can use performance information to identify, retrain and better motivate under-performing buyers, and to make informed decisions regarding promotions and salaries. The information can also be used to provide feedback directly to buyers so that those who are intrinsically motivated might use that information to correct mistakes and increase their effort to improve efficiency.

Nonetheless, there are challenges associated with performance monitoring. In particular, the costs of implementing monitoring systems can be high, potentially leading to increased bureaucracy and regulatory burdens that might exacerbate inefficiency rather than reduce it (Kelman, 1990, 2005). For instance, establishing Public Oversight Boards for conducting external and internal audits can incur significant expenses (OECD, 2013). Auditing may deter wasteful spending (Engel et al., 2021; Shi, 2023), but it might also backfire by discouraging the use of complex administrative rules for auditors to check (Gerardino et al., 2022). Finally, monitoring systems must adhere to civil service laws and procedures, which could limit their scope due to concerns related to workplace environment, discrimination, privacy, and whistleblower protection (Barnard, 2011).

Many of the problems that make performance monitoring difficult are mitigated in e-procurement systems. Chile recently implemented a new e-procurement system based on an online electronic platform that resembles Amazon, in which pre-qualified suppliers and products are chosen to

[1] Bandiera et al. (2009b) draws a critical differentiation between active and passive waste. Active waste involves either direct or indirect gains for the procurement officer, often manifesting as corruption. For example, the agent might pay a higher price for products from a specific seller in return for a kickback. Conversely, passive waste offers no advantage to the agent and usually stems from agents either lacking the necessary skills or not having the incentives to put in the effort to achieve efficiency.

[2] Another popular approach is to grant more autonomy in the exercise of public spending and introducing financial incentives to reward efficient behavior (Bandiera et al., 2021a). Yet introducing more autonomy and pecuniary incentives may require modifying public sector labor laws. Also, there is a risk that monetary incentives distort intrinsic motivation (Bénabou and Tirole, 2003).

be listed in the system through competitive bidding following Framework Agreements (FAs).[3] The new system fosters increased competition among suppliers, thereby enhancing the bidding process for products of comparable quality. While this reduces the chances for active waste, such as corruption, passive waste may remain high. Importantly, the transparency of posted prices enables a low-cost monitoring technology that can be leveraged to tackle passive waste through efficient tracking of individual buyer purchasing behavior.

In collaboration with Chile's Public Budget Office (DIPRES) and Public Procurement Office (*ChileCompra*), we leverage the transaction-level information on the platform (i.e., price paid, attributes of purchased items including product type, brand, and quantity, among others) to develop measures of individual buyer overspending on a selection of items widely acquired by procurement units. We utilize Natural Language Processing for quality-adjusted standardization to facilitate price comparisons. We then aggregate the transaction-level overspending information to the buyer level and create automated monthly reports about individual buyers' purchasing performance. Note that the primary information is embedded in the e-procurement platform so that automating the production and delivery of individual performance reports is at virtually zero marginal cost.

We then assess the impact of these performance reports on overspending using a cluster randomized field experiment with 184 public service purchasing units and over 3,500 procurement officers (buyers). We randomly assigned each of the purchasing units into one of three arms: (1) a "Public" information arm where both the buyer and her manager receive the buyer's individual performance report; (2) a "Private" information arm where only the buyer is given their individual performance report and the manager receives aggregate information on the overall performance of their purchasing unit but not the buyer-level reports; and (3) a pure control group where no performance reports are delivered at any level.

Having both Public and Private information arms allows us to examine the role of *intrinsic* versus *extrinsic* motivation as mediators of the effects of performance information on efficiency (Bénabou and Tirole, 2003). Since the Private treatment does not allow the manager to monitor individual performance, the reports on individual performance can only improve efficiency if buyers are intrinsically motivated.

Our results show that performance monitoring greatly boosts public spending efficiency. Specifically, the performance reports generated large and statistically significant reductions in overspending in the Public information arm but not in the Private information arm. This suggests that efficiency gains are driven by *extrinsic* as opposed to *intrinsic* motivation. On average, the reductions in overspending are on the order of 15% of the control group mean or about US $4.5

[3] Over the past two decades, many governments have begun implementing e-procurement tools to enhance the efficiency and transparency of public procurement processes. The realm of e-procurement encompasses a broad array of technological solutions, ranging from the online publication of all procurement bids through a centralized web platform to electronic tendering, comprehensive e-procurement encompassing contract and payment administration, and advanced functionalities such as e-catalogs and e-marketplaces, among others. The World Bank Procurement Framework Initiative offers an e-procurement world map showing the status of e-procurement around the world; see https://wbnpf.procurementinet.org/featured/e-procurement-world-map.

MM dollars over a span of five months for the items under study.[4] If we extrapolate the annualized treatment effect to all transactions made in the public procurement system, the potential savings are on the order of US $0.15 billion, i.e., 1.2% of Central Government procurement expenditure or about 0.1% of Chilean GDP in 2019.[5]

Turning to mechanisms, we examine how organizational culture, in this case the extent to which the beliefs about the government's value of efficiency are *shared* between managers and buyers, mediates the effectiveness of the performance reports. Theoretically, when beliefs are not shared, communication is less informative (Crawford and Sobel, 1982) and managers are less likely to delegate (Aghion and Tirole, 1997). Divergent beliefs create organizational distrust (Rotemberg and Saloner, 1995; Dewatripont and Tirole, 1999). As a result, in public organizations where beliefs about the value that government's authority places on efficiency are misaligned, the manager has to exert more persuasion effort to convince the buyer to use cost saving technologies and practices.

We formalize this discussion in the context of a simple principal-agent model based on Van den Steen (2010) that examines misalignment between the manager's and buyer's belief about how much the government values efficiency. In the model, performance information increases the marginal productivity of manager's effort to get buyers to adopt more efficient purchasing practices. This seems reasonable, for example, by facilitating the manager's ability to identify, retrain and better motivate under-performing buyers, and to make informed decisions regarding promotions and salaries.

The model predicts that the effect of additional buyer performance information on efficiency varies with the manager's belief about the government's prioritization of efficiency relative to the buyer's belief. Intuitively, if the buyer's belief is greater than that of the manager, then any information revealing buyer's inefficiency may trigger *motivated reasoning*, i.e., the manager surprisingly realizes that the buyer is not performing up to her potential. In this case, since the buyer has stronger beliefs in government's prioritization of efficiency, the manager believes that the marginal cost of persuading the buyer to improve her efficiency performance is low, which encourages the manager to invest in persuasion effort. Conversely, when the manager's belief is greater than the buyer's, information on buyer's inefficiency induces *confirmation bias*: i.e.,the manager believes that the buyer's inefficiency is due to her low belief that the government values efficiency. In this scenario, the manager perceives the marginal cost of persuading the buyer is high, which discourages investment in persuasion effort.

We test the model's predictions by quantifying organizational culture through a baseline survey

[4]We compared our research findings to the views of the scientific community, government officials, and technocrats by eliciting ex ante forecasts of the potential effects of our intervention (Dellavigna et al., 2019). We find the order of magnitude of the predictions closely aligns with the experimental estimates. Importantly, experts' and non-experts' predictions are not statistically different, suggesting a consensus between the scientific and non-scientific communities for a potential program's scale-up.

[5]The Central Government of Chile purchases approximately $13 billion worth of goods and services per year, i.e., around 5% of 2019 GDP. This account for public expenditures purchased through *ChileCompra* only, yet it does not account for Regional Government purchases and other expenditures such as procurement by State Firms.

index that measures discrepancies between managers and buyers beliefs about the government's efficiency priority. Consistent with the model, we find that the impact of making individual performance information public to both officers and managers is significantly larger in more misaligned organizations; and this occurs when buyers have stronger beliefs about the government's emphasis on efficiency compared to managers, but not the other way around. Overall, our findings suggest that individual buyer performance information can not only significantly enhance efficiency but also help mitigate the adverse effects resulting from organizational culture misalignment.

We then extend the model to allow the manager to collect performance information herself through monitoring. In this scenario, the (positive) effect of manager's access to performance information on efficiency is smaller, as performance information becomes a substitute of monitoring, offsetting the marginal value of information. Essentially, then, in the presence of strong monitoring systems, additional performance information has diminishing marginal returns. We find our experimental data empirically validates this prediction.

Related Literature. Our study contributes to a burgeoning literature that focuses on *addressing inefficiency in public procurement*.[6] This literature has primarily centered on the optimal architecture of public procurement units, specifically around the debate of rules *versus* discretion in the administration of public resources and how to strike the right balance between the two. On one hand, OECD (2009) advocate for enhancing monitoring through bureaucratic organizations that implement strict rules and external controls with explicit costs for those engaging in inefficient resource utilization. Conversely, Kelman (1990) argues for increased discretion, highlighting that governments face multiple constraints in implementing effective monitoring plans without incurring additional bureaucracy costs and regulatory burdens, which could lead to more inefficiencies in the form of passive waste in government spending.[7] Our paper contribute to this literature by investigating how innovative management practices meant to monitor and motivate procurement officers affect efficiency, while holding architecture (i.e., rules and discretion) fixed.[8]

Our work relates to the findings of Bandiera et al. (2021a), who find that providing financial incentives to procurement officers in Pakistan does not significantly improve performance, except when officers face a monitor who promptly approves purchases. This suggests the impact of incentives on performance depends on the allocation of authority between agents. Our paper complements these findings by revealing that low-cost monitoring technologies, such as the

[6]see Dimitri et al. (2006) and Bandiera et al. (2021b) for a comprehensive review.

[7]Several recent studies have demonstrated the potential benefits of discretion in public procurement units (Coviello et al., 2018; Carril, 2019; Decarolis et al., 2020; Bandiera et al., 2021a; Szucs, 2023) Less optimistic, Bosio et al. (2021) use cross-country comparisons to show that laws that relax procurement rules are beneficial in most low income countries but detrimental in richer countries. This could in part be due to limited capacity of bureaucracies to monitor procurement contracts (Spagnolo, 2012; Palguta and Pertold, 2017; Liscow et al., 2023).

[8]We do so through the lens of a single-layer principal-agent framework, similar to other papers studying the determinants of productivity in organizations, e.g., the role of pecuniary and non-pecuniary rewards (Bandiera and Rasul, 2011, 2013; Khan et al., 2016, 2019), management practices (Bloom and Van Reenen, 2007; Bloom et al., 2019; Rasul and Rogers, 2018), decentralization (Dal Bó et al., 2021; Balán et al., 2022) or monitoring technologies (Hubbard, 2000, 2003; Baker and Hubbard, 2003, 2004; Halac and Prat, 2016; de Rochambeau, 2021; Kelley et al., 2021; Mattsson, 2022).

automatic provision of efficiency performance reports, can substantially enhance efficiency in public procurement, provided that managers have knowledge of the reports. A significant advantage of this technology is that it conveys efficiency effects without the need to introduce financial incentives or make changes to the organizational structure and contracts. This approach helps avoid potential political economy frictions associated with hierarchical changes, thereby expanding its potential scalability.

With a specific focus on e-procurement platforms like *ChileCompra*, we address common concerns associated with manual procurement practices, such as limited access to bid information, collusion among bidders, and corruption. For instance, Lewis-Faupel et al. (2016) investigate the impact of implementing electronic procurement systems on procurement outcomes related to public works projects in India and Indonesia. Although they find no evidence of price reductions, they did find that e-procurement led to notable improvements in quality, as indicated by the quality of roads and the reduction in delays. Such improvements can be attributed to e-procurement systems facilitating the entry of higher quality contractors in the bidding process. Our research demonstrates that e-procurement systems can also bring about efficiency gains when the transactional data embedded within them is leveraged to boost performance monitoring, thereby broadening our understanding of the impact of electronic platforms in public procurement.

Our paper also contributes to the literature exploring the influence of *organizational culture* on performance[9]. Gibbons and Henderson (2013) argue that management styles and intangible factors within an organization act as complements in explaining the success or failure of various managerial practices. Specifically, intangible factors like relational contracts (Macleod and Malcomson, 1989; Baker et al., 2002; Levin, 2003; Helper and Henderson, 2014; Blader et al., 2020), ideology (Spenkuch et al., 2023), or trust (Ichniowski et al., 1997) can be critical in determining an organization's adoption and effective use of new management practices. Our paper shows that the effectiveness of a given management practice may vary across organizational cultures with more or less belief alignment among its members.

Our focus is on the incidence of top-down relationship on the performance of organizations. Previous works in this line include Bloom et al. (2012) who show that firms headquartered in high-trust regions are significantly more likely to decentralize, increasing aggregate productivity by affecting the organization of firms. Guiso et al. (2015b) show that when employees perceive top managers as trustworthy and ethical, firm's performance is stronger. Using personnel data from a large high-tech firm, Hoffman and Tadelis (2021) show that survey-measured people management skills have a strong negative relation to employee turnover. Bandiera et al. (2009a) study how social connections between workers and managers (an indirect attribute of organizational culture) affects productivity in firms, and find that while social connections increase the performance of connected workers, favoring connected workers is detrimental for the firm's overall performance. Our paper adds to this literature by examining how top-down belief misalignment interacts with

[9] For a comprehensive review, refer to Guiso et al. (2015a) and Gibbons and Henderson (2013), along with interesting avenues for future research discussed in Martinez et al. (2015).

monitoring efforts oriented to improve efficiency performance in public organizations.

Closer to our work, an insightful study by Spenkuch et al. (2023) documents that at any point in time in the U.S., a considerable share of bureaucrats is ideologically misaligned with their political leaders, and examine the performance implications of this misalignment specifically for procurement officers. The authors find that procurement contracts overseen by misaligned officers tend to exhibit greater cost overruns and delays, a form of passive waste. This phenomenon is attributed to a "morale effect" where misaligned officers are less motivated to pursue the organizational mission, thereby affecting performance outcomes. Our evidence complements this work by showing that top-down belief misalignment in non-political but related dimensions like the government's valuation of efficiency can also be detrimental for the performance of procurement officers, but inexpensive monitoring tools, such as the automatic generation of efficiency performance reports, can mitigate part of the negative consequences of organizational misalignment.

In a similar vein, Beer et al. (2021) implement a computerized laboratory experiment consisting of a procurement game designed to test the impact of increased transparency on delegated purchasing behavior. The authors show that buyers who are observed by their peers reduce overspending, and that this is mostly driven by buyers willing to comply with social norms that favor efficiency. Also related, Blader et al. (2020) find that providing truck drivers with information on their driving performance as well as their relative performance with respect to peer drivers is more effective than solely providing individual-level performance information. Interestingly, this effect is positive only in workplace environments that favor competition-based managerial practices. In contrast, the effect is negative in settings with cooperative-based relational contracts, a backfire effect that is possibly explained because the employees perceived the pro-competition intervention was inconsistent with the organizational culture promoted by their leaders. Their results, like ours, highlight that the organizational culture may play a key role in mediating the impact of performance monitoring on individual performance.

Our paper differs from these in at least three ways. First, we examine how performance information affects individual performance when this is private versus publicly shared with the manager (not the peers), i.e., we focus on the principal-agent axis of information flow while keeping the agent-to-agent, horizontal flow unchanged. Second, we focus on the top-down misalignment in the beliefs about the government's valuation of efficiency and its interaction with monitoring. Finally, our setting is the public sector, where the incentive scheme faced by procurement officers may well differ from that faced by fictitious lab-buyers or truck drivers in a private organization.

The paper is organized as follows. Section 2 details the institutional setting governing the procurement system in Chile. Section 3 provides a theoretical framework on the mechanisms operating in the relationship between purchasing performance information and procurement efficiency. Section 4 introduces the performance information experiment, while Section 5 present the data and method used to measure efficiency. Section 6 test the model predictions, Section 7 contrasts model predictions with predictions from experts, and Section 8 concludes.

2. Public Procurement in Chile

Like most public procurement systems, Chile's combines some degree of centralization combined with input from local purchasing units.(OECD, 2019; Carpineti et al., 2006). In 2003, Chile adopted new regulations that mandated the creation of a Central Procurement Body (CPB), *Chilecompra*, located in the Ministry of Finance. *Chilecompra* creates the rules and mechanisms for purchases, but procurement itself is decentralized to procurement units located in the ministries. CPBs like *Chilecompra* have a number of advantages: (i) tighter control of expenditures; (ii) economies of scale with fewer suppliers; (iii) more bargaining power; (iv) lower administrative costs; and (v) higher productivity. In contrast, local bodies can be more effective when they have an information advantage in selecting local suppliers and there is heterogeneity in local purchaser preferences.

Each local procurement unit is managed by a Director and a Finance Executive, who lead a team of procurement officers (Buyers). In 2019, there were 224 public procurement units with 9,300 registered buyers in the central government.[10]

Chile's Public Budget Office (DIPRES) is responsible for allocating resources annually to the local procurement units. The procurement units have little incentive to be be efficient. Each unit is given an annual budget and next year's budget is driven by the extent to which they fully spent this year's allocation. The largest proportion of purchases is in December at the end of the fiscal year, suggesting wasteful end of year spending (Engel et al., 2021).[11] This could be due to the prevailing belief that DIPRES does not care about passive waste. In our baseline survey, 87% of managers agreed with the statement "*there is pressure from DIPRES to fully spend Utilizing this year's budget*," and 44% recognized that "*sometimes purchases need to made at high prices to comply with budget expecution by the end of the year*".

The 2003 reform created three purchasing mechanisms:

(i) Auctions. These are open calls where suppliers submit bids based on publicly announced rules that describe the characteristics and volumes of the goods/services to be purchased, the bid format, and the criteria used to determine the winners. Auction enhance competition through objective and transparent selection criteria, but involve significant time and effort to define the rules and implement the auction (Bajari et al., 2009). Auctions are required for purchases above USD $62,500, and in 2019 were used for more than 50% of government expenditures and roughly 40% of the transactions.

(ii) Direct Purchases. Direct purchases from a supplier chosen by the buyer can be used for small expenditures (< USD $625). During 2019, direct purchases accounts for 17% of the transactions. This mechanism provides the buyer more flexibility to select among a wider variety of products/services and suppliers, but is discouraged due to the lack of transparency are subject to

---

[10] This excludes municipalities, public universities, and public enterprises, which collectively contribute an additional 6,000 buyers to the procurement system. However, these public entities operate independently of the Ministry of Finance's direct oversight.

[11] Similar findings have been documented in other contexts. For instance, Liebman and Mahoney (2017) show that spending in by the U.S. Federal Government in the last week of the year is 4.9 times higher than the rest-of-the-year.

scrutiny by the General Comptroller.

(iii) Framework Agreements (FAs). FAs use competitive bidding to select a set of products and services to be listed on an electronic platform from which local buyers make purchases. It seeks a balance between using a CBP competitive bidding process to lower costs and maintain quality while allowing local purchaser flexibility in the choice of specific products and suppliers (Albano and Nicholas (2016)). During 2019 FAs accounted for about 20% of government expenditures and more than 40% of the transactions, and it has been growing steadily.

Framework Agreements (FAs) are implemented in two stages. In the *First Stage*, *Chilecompra* uses competitive auctions to select a group of suppliers and products qualified to be listed in a digital online platform from which local buyers can make their purchases. The auctions establish product quality standards, delivery times and costs, and ceiling prices of selected products. The system generates a large number of pre-qualified suppliers, thereby providing a broad variety of products and suppliers in the online marketplace. Awarded contracts last between one and four years.

In the *Second Stage*, *Chilecompra* operates a digital online marketplace platform (www.mercadopublico.cl), similar to e-commerce retailing like Amazon, on which the winners of the first stage list their products and from which local purchasing unit buyers make purchases. Purchasing is fully decentralized, i.e., the local buyer has complete flexibility to choose among the posted products and suppliers with no restrictions.

The objective of FAs is to ensure competitive prices and reduce buyers' transaction costs both in term of administrative costs and the search effort needed to find products and services that satisfy the preferences of their local governments. FAs use competition both to *enter* the market (first stage) and *within* the market (second stage) (Saban and Weintraub, 2021).[12] ChileCompra awards multiple suppliers for similar products in the first stage to provide more flexibility to buyers to access larger product variety and create competition in the second stage.[13] Following OECD (2019), *Chilecompra* developed a standardized product classification of products offered on the FA online marketplace based on attributes that describe both vertical and horizontal product differentiation. The product standardization is intended to enhance competitive bidding among suppliers trying to enter the market in the first stage as well as reduce search costs for buyers in the second stage (Dinerstein et al., 2018).

By increasing competition, the implementation of FAs reduces the chances for active waste, such as corruption. However, passive waste may still remain high. We investigate how buyer performance information can be used to lower such waste. We leverage the transaction-level information from the FA online market platform to develop measures of individual buyers' purchasing performance (overspending) that are given to the buyer and her manager.

[12]While competitive bidding can lower procurement prices (Bulow and Klemperer, 1996), there is a risk that under-qualified contractors win bids Carril et al. (2022).

[13]Allende et al. (2023) find that the introduction of an extra vendor reduces average procurement prices by 12%, highlighting the significant role of supply-side factors in determining public procurement prices.

3. Theoretical Framework

We explore how providing buyers' performance information to managers could reduce wasteful spending in a simple principal-agent model based on Van den Steen (2010). An important feature of the model is organizational culture – in this instance, the extent to which the manager (principal) and buyer (agent) share beliefs about the value that the government (DIPRES) places on efficiency.[14] Theoretically, when beliefs are not shared, communication is less informative (Crawford and Sobel, 1982) and managers are less likely to delegate (Aghion and Tirole, 1997). Divergent beliefs create organizational distrust (Rotemberg and Saloner, 1995; Dewatripont and Tirole, 1999). As a result, in public organizations where beliefs about the value that government places on efficiency are misaligned, the manager has to exert more effort to persuade the buyer to use cost saving technologies and practices.

In the model, performance information increases the marginal productivity of manager's effort to get buyers to adopt more efficient purchasing practices. This seems reasonable, for example, by facilitating the manager's ability to identify, retrain and better motivate under-performing buyers, and to make informed decisions regarding promotions and salaries.

However, the effect of additional buyer performance information on efficiency varies based on the manager's belief in how much the government values efficiency relative to the buyer's belief. Intuitively, when the manager believes the government cares more about efficiency than does the buyer, information on buyer inefficiency likely induces *confirmation bias*, i.e., the manager believes that buyer's inefficiency is due to her low belief about the government's prioritization of efficiency. If true, this could increase the manager's perceived cost of persuasion, which in turn discourages effort to persuade the buyer. On the other hand, if the buyer believes that the government cares more about efficiency than does the manager, then revealing the buyer's inefficiency triggers *motivated reasoning*, i.e., the manager surprisingly realizes that the buyer is not performing up to her potential. Consequently, because the buyer believes that the government cares about efficiency, the manager thinks that the buyer is persuadable (low cost of persuasion effort) and thus invests in effort to persuade the buyer to be more efficient, resulting in improved efficiency.

3.1. Setup

Consider a procurement unit composed of a Manager (principal) J and a single (agent) Buyer I. The Buyer must choose between two possible actions, A and B. Option A represents the *status quo* action performed by the unit and is known by both individuals. In our case, action A might be fully spending the assigned budget to avoid the risk of not receiving the full budget next year due

[14] Others like Gorton et al. (2021) define organizational culture as "...elements like norms, values, knowledge, and customs...[based in] unwritten codes, implicit rules, and regularities in interactions." Schein (2004) defines it as the combination of *artefacts* (e.g., organizational structure or management practices), *espoused values* (e.g., values that the organization attempts to pursue), and *basic underlying assumptions* (e.g., deep values and beliefs associated to the individuals origins), and predicts that the effect of changes of *artefacts* on organizational performance depends on how the *espoused values* and the *basic assumptions* adapt to those changes.

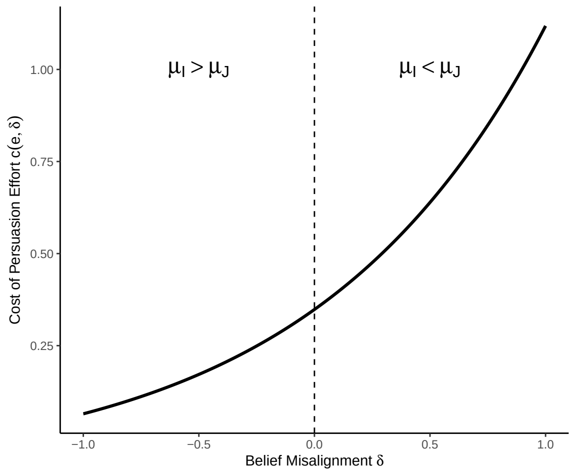
to under-spending. Action B is an alternative such as improving efficiency by searching for low prices while allowing for under-spending of the full budget.

Both Manager J and Buyer I are extrinsically motivated and risk neutral. Their actions are guided by their beliefs about how much the government (DIPRES) values action A versus action B. Both have the same belief about how much DIPRES values action A's payoff. However, the Manager and Buyer may have different individual beliefs, μ_J and μ_I, about DIPRES's valuation of action B's payoffs. This is reflected in $\delta = \mu_J - \mu_I$, the misalignment in the beliefs about how much DIPRES values efficiency. Note that the direction of δ could be either neutral ($\mu_J = \mu_I$), positive ($\mu_J > \mu_I$) or negative ($\mu_J < \mu_I$), and that will play a role on J's and I's optimal decisions (defined below).

The Manager has access to a given level of information about buyer performance, $s > 0$. The Manager exerts effort e to try to convince the buyer to choose the option that she believes is more important for DIPRES with probability $P(e, s)$, and at a cost of effort $c(e, \delta)$. We assume functions $P(e, s)$ and $c(e, \delta)$ are smooth; that $P(\cdot, e)$ and $P(s, \cdot)$ are concave functions for any (s, e); and that $c(0, \delta) = P(0, 0) = 0$.

Importantly, the effectiveness of persuasion effort depends positively on s and e, i.e., $\frac{\partial P(e,s)}{\partial e}$ and $\frac{\partial P(e,s)}{\partial s}$ are both strictly positive. Second, persuasion effort and information are complements in $P(e, s)$, meaning $\frac{\partial^2}{\partial e \partial s} P(e, s) > 0$. This means Buyer's performance information increases the productivity of persuasion effort. Third, the persuasion cost is increasing and convex in persuasion effort, i.e., $\frac{\partial c(e,\delta)}{\partial e} > 0$ and $\frac{\partial^2 c(e,\delta)}{\partial e^2} > 0$.

Figure 1: Persuasion Cost and Belief Misalignment



Note: Relationship between Persuasion Cost, $c(e, \delta)$, and Belief Misalignment, $\delta = \mu_J - \mu_I$.

Finally, persuasion cost increases with δ (i.e., $\frac{\partial c(e,\delta)}{\partial \delta} > 0$), and $\frac{\partial^2 c(e,\delta)}{\partial \delta \partial e} > 0$ (see Figure 1). That is, as Manager's belief that DIPRES cares about efficiency (μ_J) gets too large relative to that of the Buyer (μ_I), then it becomes increasingly costly for the Manager to convince the Buyer to be efficient. In contrast, when μ_I becomes greater than μ_J, then δ becomes more negative and it is increasingly easier for the Manager to persuade the Buyer to buy at lower prices.[15] Moreover, we assume the Manager's persuasion cost decreases at increasing rates with the distance of μ_I relative to μ_J.

3.2. Payoffs

We characterize the Manager's optimal level of persuasion effort in terms of the levels of misalignment δ and performance information s. Let action A's payoff be a random variable $x \sim U(0,1)$, which is publicly drawn before decisions are made. The Manager and Buyer have the same belief about how much DIPRES values action A's payoff. In contrast, action B's payoff is an unknown random variable, and the Manager and Buyer have potentially distinct beliefs about DIPRES's valuation of action B's expected payoff, denoted by μ_J and μ_I, respectively. The Manager would prefer A if x is larger than μ_J, otherwise she prefers B. Likewise, the Buyer would prefer A if x is larger than μ_I, otherwise she prefers B.

From the Manager's perspective, if she were choosing the action, then her expected payoff would be:

$$\int_0^{\mu_J} \mu_J dx + \int_{\mu_J}^1 x dx = \frac{1+\mu_J^2}{2}, \tag{1}$$

where $\int_0^{\mu_J} \mu_J dx$ reflects the expected utility when the belief that DIPRES's expected payoff of choosing B is greater than that of A, and $\int_{\mu_J}^1 x dx$ is the expected utility when the belief that DIPRES's expected payoff of choosing A is greater than B's one.

On the other hand, if Buyer chooses the action, then Manager's payoff would be:

$$\int_0^{\mu_I} \mu_J dx + \int_{\mu_I}^1 x dx = \frac{1+\mu_J^2-\delta^2}{2} \tag{2}$$

Hence, if Manager decides to exert persuasion effort e to convince Buyer to pursuing the action she believes DIPRES values the most, then her expected payoff is given by:

$$\pi(e,\delta,s) = P(e,s)\left(\frac{1+\mu_J^2}{2}\right) + (1-P(e,s))\left(\frac{1+\mu_J^2-\delta^2}{2}\right) - c(e,\delta). \tag{3}$$

That is, Manager's expected payoff is the weighted probability that the Buyer executes the action that the Manager believes DIPRES values the most (first term in the right side of equation 3), and that the Buyer executes the action that he believes DIPRES values the most less the cost of

[15]For instance, suppose that a Buyer believes DIPRES's goal is to reach an overspending below 10% of total expenses. If the Manager believes that DIPRES's goal is, say, 20% or lower, then convincing the buyer to reach an overspending below 20% is easy since the Buyer expects herself to reach a 10% goal.

persuasion effort (last two terms).[16] It follows then that as $P(e, s)$ increases, managers that believe action B is better than action A run more efficient units.

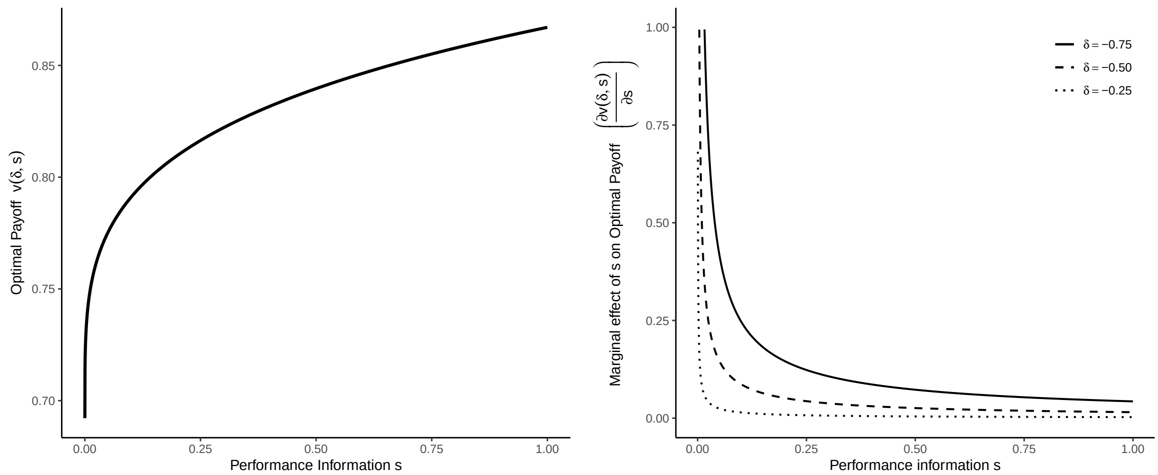
3.3. Optimality

Let $e^* = e^*(\delta, s)$ be the optimal persuasion effort that maximizes Manager's payoff (equation 3), which we assume is an interior solution. We show that if the Buyer has a weaker belief than the Manager, then the optimal payoff is decreasing in δ because decreasing the Buyer's belief μ_I (relative to μ_J) increases the cost of exerting persuasion effort. As a result, decreasing μ_I would lead to a lower optimal payoff $v(\delta, s) = \pi(e^*, \delta, s)$. In contrast, when $\mu_I > \mu_J$, the higher the Buyer's belief μ_I, the lower is the associated cost of effort. Indeed, if Manager's cost reduction is large enough, then increasing μ_I would lead to a larger optimal payoff $v(\delta, s) = \pi(e^*, \delta, s)$. See Appendix C for proofs.

On the other hand, increasing information s makes the Manager's effort more productive in convincing buyers to make more efficient purchases. It follows then that the optimal effort e^* is increasing in s. Moreover, the optimal payoff is increasing in s, but at a decreasing rate (i.e., $\frac{\partial}{\partial s}v(\delta, s) > 0$, and $\frac{\partial^2}{\partial s^2}v(\delta, s) < 0$). This leads to the following proposition:

Proposition 3.1. *Increasing Manager's performance information leads to increases in efficiency (Figure 2, left panel).*

Figure 2: Performance Information and Efficiency



Note: Left panel shows a simulation of the effect of degree of Manager's access to performance information s on optimal payoff. We assume s and $e \in (0, 1)$, and $\delta \in (-1, 1)$. Also, $P(e, s) = e^{0.2}s^{0.2}$; and $c(e, \delta) = 0.1(e^{1+x+\delta} - 1)$. Right panel shows the marginal effect of degree of access to performance information on efficiency when $\mu_I > \mu_J$.

[16]Note that the disutility generated by belief misalignment enters twice in the expected payoff: first as the inherent disutility of having to tolerate that the Buyer is not implementing the Manager's preferred action, and secondly, as the effort cost associated to persuade the Buyer to implement the Manager's preferred action. Indeed, if $\delta = 0$, it follows the payoffs associated to these scenarios are equal, and the monitoring cost is zero.

The effect of performance information s depends on the marginal impact on the optimal payoff $v(\delta, s)$, which is affected by the degree of belief misalignment δ. In particular, we show that if $\mu_I > \mu_J$, then the larger the absolute difference $|\delta|$, the larger is the payoff derived from improving manager's access to performance information. This is because on one side increases in $|\delta|$ reduces the persuasion costs, and on the other side more s makes the Manager's persuasion effort more productive, all of which leads to larger optimal effort.

Intuitively, performance information is likely to reveal some level of inefficiency on the part of the Buyer. Therefore, when the Buyer's belief about government's prioritization of efficiency is greater than that of the Manager, revealing the Buyer's inefficiency triggers *motivated reasoning*, i.e., the manager surprisingly realizes that the buyer is not performing up to her potential. Consequently, because the buyer believes that the government cares about efficiency, the manager thinks that the buyer is persuadable (low cost of persuasion effort) and thus invests in effort to persuade the buyer to be more efficient, resulting in improved efficiency.

In contrast, when $\mu_I < \mu_J$, increasing the distance $|\delta|$ leads to a higher marginal cost of effort. This is because information on buyer's inefficiency induces *confirmation bias*, i.e., the manager believes that buyer's inefficiency is due to her low belief about the government's prioritization of efficiency, which in turn increases the manager's perceived cost of persuasion effort, lowering both optimal persuasion and payoff.

Proposition 3.2. *When $\mu_I > \mu_J$, the larger the belief misalignment, the bigger is the effect of performance information on efficiency (Figure 2, right panel). When $\mu_I < \mu_J$, the effect is smaller than when $\mu_I > \mu_J$.*

3.3.1. Monitoring

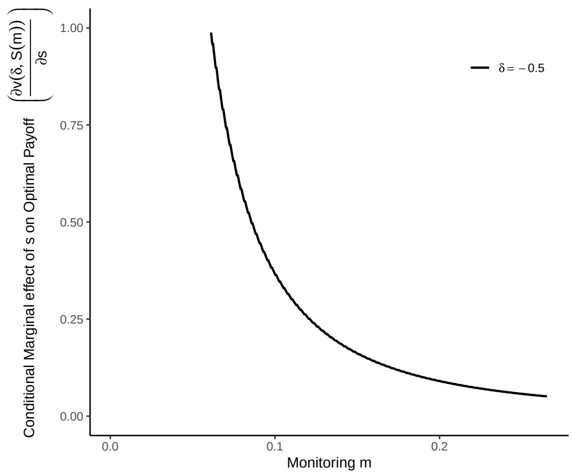
Finally, we extend the model to allow the manager to collect performance information herself through internal monitoring and ask what is the effect of more information from an external monitoring system.

Let the amount of internal monitoring by the Manager be a smooth function $m(e, s)$, which we assume it is increasing in e and s so that $\frac{\partial m(e,s)}{\partial e} > 0$ and $\frac{\partial m(e,s)}{\partial s} > 0$. Then, for any given degree of access to existing performance information s, the Manager will implement an optimal level of monitoring $M(\delta, s) = m(e^*, s)$, which is again, increasing in s. Reciprocally, in order to implement a level of monitoring m, the Manager requires access to performance information $S(m)$ (for simplicity, we omit the dependency of S on δ). It follows then that $S(m)$ is an increasing function of m. Therefore, a higher level of monitoring will be reflected in the optimal utility through two mechanisms: (i) a higher degree of access to performance information $S(m)$; and (ii) a higher optimal persuasion effort e^*.

We prove that $\frac{\partial}{\partial m} v(\delta, S(m)) > 0$, i.e., a larger level of monitoring will imply a higher optimal utility for the Manager. Moreover, given the complementarity between persuasion effort and performance information, it follows that $\frac{\partial}{\partial m}(\frac{\partial v(\delta, S(m))}{\partial s}) < 0$. That is, the (positive) effect of Manager's additional performance information from an external monitoring system on optimal utility will be smaller when the Manager implements larger levels of monitoring, as performance information becomes a substitute for internal monitoring, offsetting the marginal value of information. In other words, if the manager invests substantial effort to monitor performance herself, then additional performance information from an external monitoring system has diminishing marginal returns (see Appendix C for proofs). This leads to the following proposition:

Proposition 3.3. *Internal monitoring purchases has a positive impact on efficiency. However, as the level of internal monitoring conducted by the Manager increases, the marginal effect of external performance information on efficiency decreases (Figure 3).*

Figure 3: Performance Information, Monitoring, and Efficiency



Note: Marginal effect of increasing access to performance information on efficiency as a function of monitoring.

4. Performance Information Intervention

In collaboration with *ChileCompra* and DIPRES, we constructed monthly measures of overspending for each buyer (see Section 5 for details). We then gave each buyer a monthly buyer performance report of their overspending compared to peer buyers in the same procurement unit. The reports were signed by DIPRES and included explanations of how *ChileCompra* calculates overspending.[17] Before the intervention, buyers were also given a 10-minute training video of concrete, step-by-step examples on how to manage the marketplace platform to make efficient purchases.[18]

Our model predicts that performance information leads to increases in efficiency conditional on buyers being extrinsically motivated and managers being able to access buyers' performance

[17] See Appendix Figure B.1 for an example of the buyer-level report, and Appendix Figure B.3 for explanations. The overpricing data are confidential not available to the public or government institutions.

[18] See YouTube link.

information (Proposition 3.1). However, it is possible that performance information could influence performance even without managers having access to the information if buyers are intrinsically motivated. In order to test these prediction we created two versions of the reports: *Private versus Public Reports*.

In the public report intervention, managers were given a performance report for their purchasing unit aggregated across all buyers in the unit in which individual buyer performance could not be separately identified. The manager was also given a list of each individual buyer's overspending (ranked by largest to lowest) and her relative contribution to unit's total overspending (see Appendix Figure B.4). In addition, Managers had access to the individual buyer reports so that they can check details of individual buyer overspending behavior. Finally, the buyer's report includes a message informing them their manager has access to their performance report ("Your individual report has been sent to the Unit's Manager."). From the buyer's perspective, these reports are "public" since both the buyer and the manager receive information on buyer performance and buyers are made aware that the manager has access to their individual performance reports.

In the private report intervention, managers were only given a performance report for their purchasing unit aggregated across all buyers in the unit in which individual buyer performance could not be separately identified. This basic version of the manager's report is considered "private" from the buyer's perspective in that it does not include information on his performance. This is highlighted in the buyer's report with a message that "Your individual report has not been sent to the Unit's Manager.", i.e., buyers were aware that their individual performance reports are private. In this case, we would expect the private report to improve buyer efficiency only if buyers were *intrinsically* motivated. In contrast, we would expect the public report to improve buyer efficiency if buyers were *extrinsically* motivated.

4.1. Experimental Design

We use a cluster randomized field experiment to assess the impact of the monthly performance reports on buyer overspending in 184 procurement units in 22 ministries. The procurement unit is the unit of randomization and randomization was stratified within Ministry and by whether the level of overspending of the unit was above or below the median in the previous fiscal year. We randomly assigned each of the purchasing units into one of three groups:

1. *"Public" information group*: both buyers and manager are first exposed to the training video and then receive the "Public Performance Reports" with information about individual buyer performance and the overall performance of the unit. The buyer's report includes a message highlighting that "Your individual report has been sent to the Unit's Manager.", so that buyers are made aware that the manager knows their purchasing efficiency.
2. *"Private" information group*: both buyers and managers are exposed to the training video and then buyers are given their individual performance report but the manager only receives information on the overall performance of the unit and not the buyer-level reports. The

buyer's report includes a message highlighting that "Your individual report has not been sent to the Unit's Manager.", so that buyers are made aware that the manager does not knows their purchasing efficiency.

3. *Control group*: no training and no performance reports are delivered at any level.

Note the intervention may have impacted overspending through the reports calling attention to overspending (Hawthorne effect) as opposed to the information about overspending included in the reports. To test for this, within Public and Private groups, we randomly assign one fourth of buyers to receive only a "Placebo treatment". The placebo consists in the buyer receiving the video training and, on a monthly basis, a simple message indicating that her overspending is being monitored, but do not provide any type of information about individual- or unit-level performance.

Sample Sizes. Table 1 summarizes the experimental design and sample sizes. There are 7,285 buyers in the 184 purchasing units in the study. However, the intervention is only appropriate for those buyers who use the FA purchasing platform, of which there are 5,648. We further restricted the sample to those buyers who specialize in purchasing products offered in the Food, Office Supplies, or Computers categories as those are the ones for which we can measure overspending (more in Section 5), yielding a sample of 3,323 buyers for the study. We randomly chose 2,743 to participate in the main study, reserving the remaining 580 for the placebo test.

Intervention and Data Collection Timeline. Figure 4 displays the timeline of the experiment. Between February and April 2020, a baseline survey was administered to managers and buyers to measure their perspectives, attitudes, and beliefs about the government's prioritization of efficiency. Overall, we were able to collect baseline surveys of managers in 89% of procurement units, while buyers in all 184 procurement units responded the survey, for a total of 2,661 responses. In June 2020, we randomized the procurement units into the experimental groups, and sent email invitations to managers in the treatment groups for the 10-minute training video and with instructions to share the video link with buyers within their unit. The training video take-up rate, measured by the manager clicking the video link, was 85% in the Public group and 93% in the Private one.

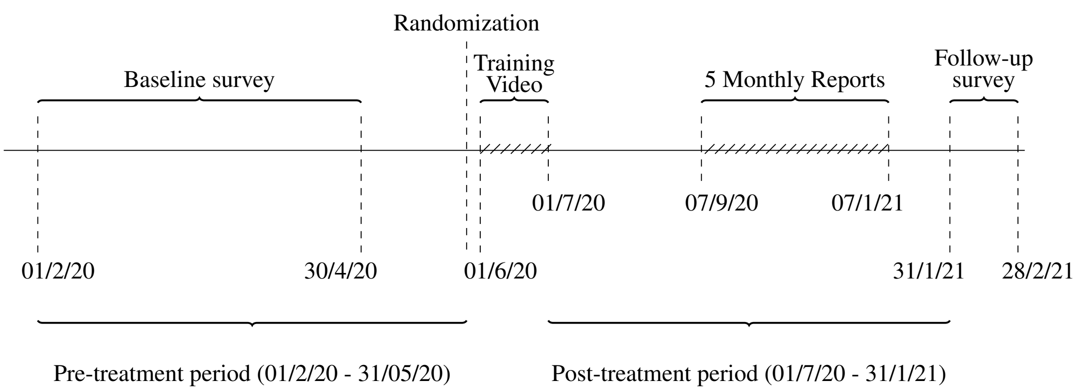
Starting in September 2020 through January 2021, managers and buyers received the monthly overspending performance reports via email. The Reports were sent during the first week of each month covering the previous month's performance. The emails contained login credentials to a centralized platform (www.gastoeficiente.cl) where each buyer and manager had an individual homepage from where they could download the monthly reports. The reports remained available so that that users go back to them at any point in time. We track whether the user logged in the reports' platform, which helps as a proxy measure for the treatment's take-up rate. The across-months average take-up rate was 99% in the case of managers, and 36% among buyers. Both manager and buyer take-up rates are well balanced across experimental groups. Finally, after sending the last report (January 2021), we collected a follow up survey to understand the mechanisms underlying the potential impact of the reports.

Table 1: Experimental Design

	Public Treatment	Private Treatment	Control	Total
Procurement Units	61	62	61	184
Managers	61	62	61	184
Receive 10-minutes Training Video?	Yes	Yes	No	
Receive Performance Reports at Unit-level?	Yes	Yes	No	
Receive Performance Reports at Buyer-level?	Yes	No	No	
Buyers	2,468	2,547	2,270	7,285
Buyers who specialize in FA channel	1,894	1,931	1,823	5,648
Analysis Sample: Buyers who specialize in Food, Office Supplies, or Computers FAs, and for which we can calculate overspending (and thus produce performance reports)	1,069	1,145	1,109	3,323
A. Experimental Buyers	790	844	1,109	2,743
Receive 10-minutes Training Video?	Yes	Yes	No	
Receive Performance Reports at Unit-level?	Yes	Yes	No	
Receive Performance Reports at Buyer-level?	Yes	Yes	No	
Receive Public Information Message? (“Your individual report has also been sent to the Unit’s Manager")	Yes	No	No	
B. Placebo Buyers	279	301	0	580
Receive 10-minutes Training Video?	No	No		
Receive Performance Reports at Unit-level?	No	No		
Receive Performance Reports at Buyer-level?	No	No		
Receive Public Information Message? (“Your individual report has also been sent to the Unit’s Manager")	No	No		
Receive Placebo Message? (“Your overspending is being monitored")	Yes	Yes		

Note: Distribution of Procurement units, managers, and buyers, by experimental group, for the analysis period (Feb. 2020 - Jan. 2021).

Figure 4: Timeline of Experiment



5. Measuring Efficiency in the Online Marketplace Platform

We measure efficiency by comparing procurement prices paid to to the lowest prices listed within groups of products of comparable quality. We worked with *Chilecompra* and the Budget Office (DIPRES) to create *reference groups* of products that are close substitutes in terms of quality. We use a Natural Language Processing algorithm adapted from Sun et al. (2014) to characterize *tagged data* of observable product attributes extracted from free-text product descriptions. We focus on products in the *Food*, *Office Supplies* and *Computers* categories, which typically include items that are widely acquired by most purchasing units (e.g., instant coffee, brackets, laptops, etc.).

We use the tagged data to create reference groups comprising products that exhibit identical attribute values. In order to minimize processing errors, we used crowd-sourced manual verification of the automated classifications to assess the adequacy of the identified attributes within reference groups in distinguishing quality variation among products. The verification ensures that the established reference groups accurately represent likely close substitutes. While FA contracts regulate the service conditions such as stock availability, delivery times, shipping rates, among others, they are not adequate for price comparisons. Products in reference groups, however, are not only similar in terms of service conditions but also in terms of quality attributes and therefore can be used for price comparisons. The details of this procedure and examples are provided in Appendix A.

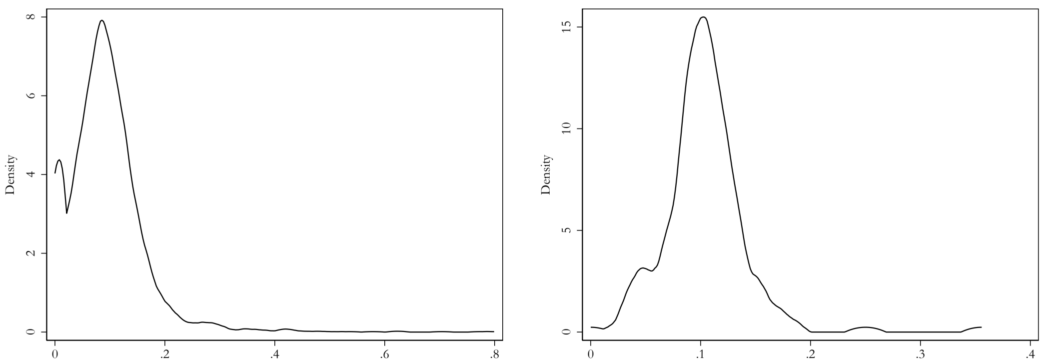
The open-text descriptions of products are very detailed, allowing to us to identify 3 to 10 attributes per product with a median of 5. However, the text can in some cases include more attributes than our algorithm is able to classify, meaning the buyer could choose the product based on attributes that reflect unobserved quality. In order to preserve the accuracy of efficiency metrics, *Chilecompra* opted to exclude from the monitoring system those products for which the information provided in the open-text description is too limited to assess product quality (see Appendix A.1 for concrete examples). Specifically, *Chilecompra* considered 37,885 products in the Computers, Office Supplies, and Food categories; of which they excluded 18,611 products due to classification limitations, leaving a total of 19,274 products for the efficiency analysis.

We calculate the amount overpaid —termed 'overprice'— for each purchase by comparing the per-unit price paid to the lowest-priced alternative within the reference group, all at the daily level (i.e., considering the posted prices within the date of the purchase). The 'overprice' value serves to highlight the extent of passive waste in current purchasing choices and represents the potential savings that could be realized by opting for the cheapest alternative. More formally, define p_{irt} as the per-unit transaction price of purchased product i from reference group r in date t, and p_{rt}^{min} as the per-unit reference price, corresponding to the lowest price in the t-daily price distribution for products from reference group r. Then, the 'overprice' op_{irt} of purchased product i is calculated as:

$$op_{irt} = \frac{p_{irt} - p_{rt}^{min}}{p_{rt}^{min}} \tag{4}$$

We are able to estimate overprice for 36% of *Food* purchases; 80% of *Office Supplies* purchases; and 100% of *Computer* purchases. Over the 12-months experimental period of analysis (Feb. 2020 to Jan. 2021), we estimated overprice for 206,704 transactions made by 3,323 buyers working in 184 purchasing units in 22 ministries.[19]

Figure 5: Distribution of Average Overprice: Experiment Period (Feb 2020 - Jan 2021)



Note: Distribution of mean overprice at the buyer level (left) and procurement unit level (right) for the period February 2020 to January 2021. Sample of analysis considers 206,704 unit transactions made by 3,323 buyers working in 184 procurement units from the central government apparatus.

Table 2: Descriptive Statistics of Overprice for the period Feb. 2020 - Jan. 2021

	Purchase Level			Buyer Level (mean)			Unit Level (mean)		
	Obs.	Mean	Median	Obs.	Mean	Median	Obs.	Mean	Median
All	206,704	0.105	0.052	3,323	0.092	0.086	184	0.103	0.101
Computers FA	313	0.126	0.038	185	0.137	0.063	68	0.148	0.095
Food FA	66,120	0.090	0.046	1,141	0.084	0.082	74	0.084	0.082
Office Supplies FA	140,271	0.112	0.054	2,941	0.095	0.087	167	0.106	0.104

Note: Mean and median overprice for the period February 2020 to January 2021. Sample of analysis considers 206,704 unit transactions made by 3,323 buyers working in 184 procurement units from the central government. Following equation 4, overprice of a unit transaction is defined as the relative difference between the per-unit price paid and the lowest price in the daily price distribution for products from the reference group. Purchase level panel shows the mean and median overprice at the unit transaction level. Buyer level panel shows the mean and median of the mean overprice at the buyer level. Unit level panel shows the mean and median of the mean overprice at the procurement unit level.

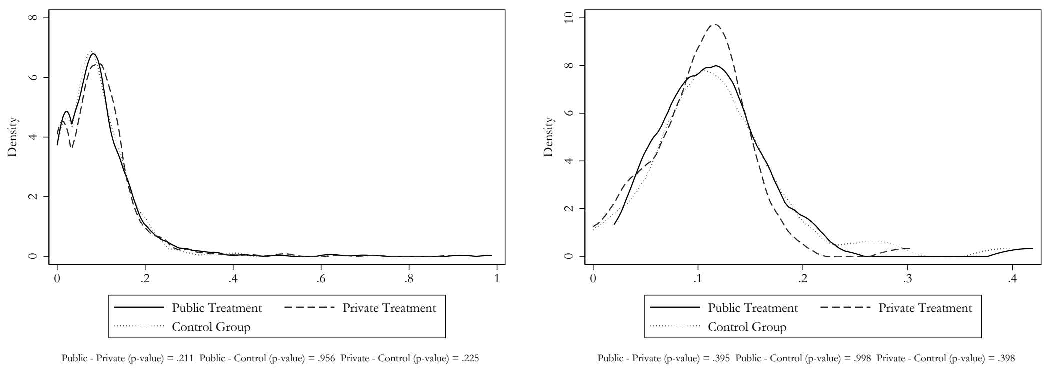
Table 2 and Figure 5 depict the distribution of overprice across buyers and purchasing units. The average overprice is around 10%, yet there is substantial variation both across buyers as well as across procurement units. The central moments of the overprice distribution seem to behave similarly across the 3 different product categories, with purchases from *Computers* FA being the

[19]The ministries included in the analysis are listed in Appendix Table B.1.

less efficient, on average. While number of Computer transactions is a bit more than 300, the *Food* and *Office Supplies* categories encompass a wider array of products and the number of transactions is roughly 65,000 and 140,000, respectively. Buyers do not specialize by product categories, but buy products from different categories at the same time. Finally, while nearly all purchasing units procure *Office Supplies*, only around half purchase *Food* and *Computers*.

Baseline Balance. Figure 6 presents the baseline distribution of average overprice per unit transaction in each experimental group at the pre-treatment period (February - May, 2020). The left panel reports the distribution of average overprice at the buyer level, while the right panel at the procurement unit level. The overprice distributions are not statistically different across groups based on Kolmogorov-Smirnov tests, and this is the case both at the transaction and the buyer levels.

Figure 6: Baseline Distribution of Average Overprice by Treatment Status



Note: Distribution of overprice per unit transaction in the pre-intervention period (February 2020 - May 2020), by treatment group. Left panel shows the average distribution at the buyer level. Right panel shows the average distribution at the unit level. Observations with overprice above the 99^{th} percentile are excluded.

Appendix Table B.2 additionally reports the means for the 3 groups for variables measured at baseline at the procurement unit group level fro the period between February and May 2020, including number of buyers per procurement unit that made purchases throughout that period, number of purchases made, average overprice per product transaction, and whether overspending is above or below the median relative to other procurement units. We also test for mean differences across groups using data from survey data, including variables associated with manager-buyer misalignment on the beliefs about the government's valuation of efficiency, and perceived level of monitoring. Overall, the table shows that experimental units are statistically balanced across these dimensions, consistent with the experiment being internally valid.

6. Treatment Effects

We estimate treatment effects on overprice by estimating the following OLS regression model at the purchase level:

$$op_{icbjt} = \alpha + \delta_1 T_j^{Public} + \delta_2 T_j^{Private} + \phi_c + \eta_t + X_j'\beta + \varepsilon_{icbjt} \quad (5)$$

where op_{icbjt} is the per-unit purchased overprice of transaction i of a product in product type c, made by buyer b who works in procurement unit j, in calendar week t, considering all weeks from July 2020 to January 2021, the post-treatment period. T_j^{Public} and $T_j^{Private}$ are equal to one if the buyer's procurement unit was assigned to either the Public or Private treatment, respectively, and zero otherwise. The ϕ_c are 332 product type fixed effects.[20] η_t are week fixed effects.[21] The vector X_j is a set of control characteristics that includes strata fixed effects and the average overprice of purchases in unit j during pre-intervention period (February to May 2020). ε_{icbjt} is the error term and the standard errors of the coefficients are clustered at the level of randomization, i.e., the procurement unit.

6.1. Effects on Efficiency

Column (1) in Table 3 presents estimates of the treatment effects on *overprice* per unit of transaction. We find that public information reports have a statistically significant impact of reducing overprice by 1.6 *pp.*, but no effects are found for private information reports. This is consistent with model's Proposition 3.1 in that buyers' purchasing behavior is predominantly influenced by *extrinsic* motivation rather than *intrinsic* motivation.

The order of magnitude of the public treatment effect is large and meaningful. It represents a 15% reduction relative to the control mean and translates into US $4.5 million over a span of five months for the FA products being examined in the study. Extending the annualized treatment effect to encompass all transactions conducted within the Chilean public procurement system would yield approximately 0.15 billion dollars in savings, equivalent to 1.2% of Central Government procurement expenditure or about 0.1% of Chile's GDP in 2019.

6.2. Alternative Explanations

Quantity. Rather than finding lower prices, buyers might attempt to cut costs by acquiring items in greater volumes to take advantage of discounts for bulk buying, thus the lower overprice effect of the public information treatment may be due to buyers purchasing larger quantities of items. We

[20] Product type classifies products in groups of products with similar functionality. For example, the Food FA includes product types like Instant Coffee, Rice, Tomato, etc. The Office Supplies FA includes product types such as Brackets, Pencils, or Staplers, among many others. Computers FA includes Desktop, Laptop, and All-in-one product types. See Appendix A for details.

[21] Our results are generally robust to using day or month fixed effects instead of week fixed effects, suggesting that time-specific shocks are well captured on a weekly basis.

examine this possibility by using the same model specification but with the number of units in the transaction as the dependent variable. The results in column (2) in Table 3 show no significant differences on the number of units purchased for both interventions.

Effort. Another concern is that buyers compensated for the extra amount of work required to find lower prices by reducing effort elsewhere and in particular reducing the number of purchase orders (P.O.) processed in a given time period. As such, any efficiency gains achieved from reducing overspending could be lost by reducing the number of purchase orders executed. Related, in any given P.O. buyers may purchase more than one product from a single seller, although all contained products in the P.O. must belong to the same FA (either *Food*, *Office Supplies*, or *Computers*). We test for these possibilities by examining the treatment effects on the number of P.O.'s executed in a month by each buyer. Here the unit of analysis is the buyer, thus we run the same treatment effects regression in equation (5) but only controlling for strata fixed effects and for the outcome at baseline (see Table 3, column 3).[22] Again, we find no evidence of either treatment on the number of purchase orders.

Extensive Margin. Performance reports are based on purchasing data from Framework Agreements (FA), and thus buyers may choose a different purchasing mechanism in order to avoid being monitored. If buyers respond to the performance information reports by shifting purchases off the FA platform, then sample selection bias could be introduced into the estimates of the treatment effects. We examine this possibility by estimating the extent to which being exposed to the performance reports had any incidence on the share of purchase orders (P.O.) buyers made through the FA channel instead of alternative procurement mechanisms like *Procurement Auctions* or *Direct Purchases*. Specifically, we regress the share of purchase orders made through the FA channel against the treatment dummies, controlling for strata fixed effects and the average outcome in the unit during the pre-intervention period (Table 3, column 4). Outcome here includes the universe of P.O. made by buyers in either of the three purchasing mechanisms. We find that receiving Public or Private performance reports had no effect on buyers' propensity to use the FA platform. This result implies that the estimated treatment effects on overspending based on FA purchasing data is not subject to sample selection bias.

This result could also be interpreted through the lens of Bandiera et al. (2009b)'s framework distinguishing active from passive waste. Active waste, such as corruption, is less likely to be observed in purchases made through the FA channels than in those made through *Procurement Auctions* or *Direct Purchases*, and this is because in the FA channel the competition, transparency, and oversight is notably higher. Therefore, the lack of observed changes in the use of the FA channel suggests that our intervention did not drive buyers to engage in purchasing mechanisms that are commonly linked to active waste.

Training Effects. In order to separately identify the effect of the training video from the

[22] This variable has a large number of zeros in some procurement units, and thus we estimate the model through a Poisson regression model. Appendix Figure B.5 shows the distribution of Purchase Orders.

Table 3: Intensive and Extensive Margin Effects

	Intensive Margin (FA only)			Extensive Margin
	Overprice	Log(Q)	#P.O.	% P.O. in FA
	OLS	OLS	Poisson	OLS
	(1)	(2)	(3)	(4)
Public Reports	-0.016**	0.112	-0.299	-0.025
	(0.007)	(0.086)	(1.048)	(0.015)
Private Reports	0.000	-0.027	0.782	-0.005
	(0.005)	(0.087)	(1.000)	(0.014)
No. Observations	93,792	93,792	2,076	416,390
No. Buyers	2,076	2,076	2,076	5,368
Control Mean	0.103	3.408	9.220	0.370
p-val. Public=Private	0.013	0.280	0.304	0.065

Notes: This table shows the intention-to-treat effects of being exposed to the Public and Private performance reports considering the full post-treatment period of analysis (July 2020 to January 2021). Intensive margin outcomes all refer to purchases made through the FA mechanism, which naturally limit the number of buyers and procurement units under consideration. Regressions in models (1) and (2) are at the unit-transaction level, and consider only purchases made through Food FA, Office Supplies FA, and Computers FA, and for which overprice can be calculated (i.e., we can find a reference group of products). Following equation 4, overprice of a unit transaction is defined as the relative difference between the per-unit price paid and the lowest price in the daily price distribution for products from the reference group. Log(Q) stands for the amount of items purchased per unit transaction (in logs). In buyer-level model (3), estimates are derived through a Poisson regression model, and #P.O. counts the number of Purchase Orders made by a buyer, which may contain more than one product (wholesales). Model regressions (1)-(3) control for strata fixed effects and the average outcome in procurement unit during the pre-intervention period. Since models (1) and (2) are at the unit-transaction level, we also control for product type fixed effects and week fixed effects. The extensive margin outcome (column 4) refers to the share of Purchase Orders made through FA (Framework Agreement) instead of alternative procurement mechanisms like *Procurement Auctions* or *Direct Purchases*. Outcome includes the universe of P.O. made by buyers in either of the three purchasing mechanisms, and the regression controls for strata fixed effects and the average outcome in the procurement unit during the pre-intervention period (Feb. - May 2020). In all, standard errors clustered at the procurement unit level are shown in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Bottom row shows the p-value corresponding to the null hypotheses of no differences between Public and Private treatment effects.

added effects associated to the performance reports, we estimate regression model 5 restricting the analysis to the period covering July-August 2020, i.e., the two-months period after the implementation of the training phase (June 2020) but before the performance reports began to be sent (September 2020). Since both Public and Private groups were exposed to the very same training video, we report the training effects through a single dummy that equals one if the purchase was made in an assigned-to-treatment purchasing unit (either Public or Private) and zero otherwise. Appendix Table B.3 shows the results. We observe no effects at all, implying that the performance reports component of the intervention is what drove the reductions in overspending among buyers in the Public treatment group and not the training component.

Hawthorne effects. The influence of the performance reports can arise either from the direct impact of the information presented in the reports or from the indirect effect of buyers realizing that they are under observation from DIPRES. In order to isolate the role of performance information from awareness of external monitoring, placebo buyers in both Public and Private group units received a monthly message indicating that her overspending is being monitored but do not receive individual performance reports, meaning they know they are being observed but do not have information about their performance. As is shown in Appendix Table B.4, the treatment effects are found to be null for placebo buyers in units assigned to either the Public or Private groups. This finding implies that merely being aware of external supervision is insufficient to influence the purchasing behavior of buyers. Instead, it highlights the crucial role played by the information presented in the Public reports in enhancing efficiency in public procurement.

Unobserved Quality. Finally, it is possible that buyers responded by purchasing products that were cheaper because they had lower quality that was unobserved, i.e., not due to one of the attributes used to construct the reference groups. Our approach for constructing reference groups relies on observable products' attributes to identify products with similar functionality that capture potential substitutes of similar quality. However, attributes may not fully capture the quality of products, thereby allowing for variation in unobserved quality across products within the same reference group. As a result, overprice dispersion may not be explained by inefficient behavior of buyers but by unobserved quality of purchased products. We examine this possibility in two ways.

First, since each product's attribute contribute to the product's quality, we expect unobserved quality to be lower among products with a larger number of attributes. In our analysis sample, the number of attributes per product goes from 3 to 10. Hence, we test for the role of unobserved quality on the treatment effects by adding interactions of the number of attributes per product with the treatment dummies in our main regression. Column (1) of Appendix Table B.5 shows that as the number of attributes for a product increases, the overprice tends to be lower. This is because more attributes make a product more specific, leading to a smaller, more similar group of comparable products, which mechanically reduces price dispersion, and that explains its negative correlation with overprice. More importantly, we find no heterogeneous treatment effects of the performance reports across number of attributes per product.

Second, in column (2), we replicate the exercise but restrict the sample to purchases of products whose reference group is composed only by identical products (same SKU), i.e., products for which we would expect the quality differences across products within the reference groups are negligible. Again, we find no heterogeneous treatment effects. Overall, these results suggest unobserved quality is not playing a role on the effectiveness of the intervention.

6.3. Manger Perceptions of Being Monitored and Buyer Attention to Efficiency

One way in which the performance reports might have changed behavior is by making agents aware of their performance and that they were being monitored. After 5 months of treatment exposure, we asked managers to report whether they were informed of the extent of overspending in their organization. Being exposed to the reports increased knowledge of overspending by a remarkable 60 *pp.* in both treatment groups, confirming they are more informed about it (see Table 4, column 1). Second, we examine whether managers changed their perception that DIPRES was monitoring their unit's performance. We asked managers to rate their perceived level of monitoring that DIPRES exerts using a 1-7 scale where 1 means "*No Monitoring at all*" and 7 "*High Level of Monitoring*". Managers in both the Public and Private information groups report significantly higher ratings, on the order of 32% and 21% relative to the control group mean, respectively.

Table 4: Treat. Effects on Awareness, Perceived Monitoring, and Buyers' Preferences for Efficiency

	Managers		Buyers		
	Manager knows Unit's level of Overspending	Perceived Monitoring from DIPRES (1-7 scale)	If DIPRES audit 10% of purchases, buyer chooses Price criteria	If DIPRES audit 5% of purchases, buyer chooses Price criteria	If DIPRES audit 1% of purchases, buyer chooses Price criteria
	OLS	OLS	OLS	OLS	OLS
	(1)	(2)	(3)	(4)	(5)
Public Reports	0.609***	1.371***	0.061	0.081**	0.135***
	(0.086)	(0.316)	(0.038)	(0.036)	(0.038)
Private Reports	0.595***	0.900**	0.015	0.007	0.065
	(0.087)	(0.351)	(0.043)	(0.040)	(0.040)
No. Observations	156	158	961	961	961
Control Mean	0.358	4.352	0.578	0.534	0.454
p-val. Public=Private	0.869	0.176	0.316	0.080	0.088

Notes: This table shows the intention-to-treat effects of being exposed to the Public and Private performance reports on post-treatment survey outcomes. Regressions use outcomes from follow-up surveys administered to managers (columns 1 and 2) and buyers (columns 3 to 5). Dep. var. in column (1) is a dummy that takes the value of 1 if the manager knows the level of overspending within the procurement unit, and zero otherwise. Dev. var. in column (2) is the manager's perceived level of monitoring that DIPRES exerts over procurement unit's overspending in a 1-7 scale (1=No Monitoring at all, and 7=High level of Monitoring). Dep. var. in column (3) is a dummy that takes the value of 1 if the buyer reports to consider price (instead of quality) as the main criterion for selecting products when DIPRES audit 10% of public purchases, and zero otherwise. Regressions in columns (4) and (5) replicate the same but for audit thresholds of 5% and 1%, respectively. All regression control for strata fixed effects and the outcome at baseline. Regression models in columns (1) and (2) are at the unit (manager) level, thus robust standard errors are shown in parentheses. Regression models in columns (3) to (5) are at the buyer level, thus standard errors are clustered at the procurement unit level (shown in parenthesis). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Bottom row shows the p-value corresponding to the null hypotheses of no differences between Public and Private treatment effects.

We also examined whether the interventions changed buyers attention to getting better prices. Specifically, we conducted multiple inquiries regarding their inclination to select a product with the lowest price. These inquiries were based on different scenarios about the audit rates implemented by DIPRES. We proposed a high audit rate scenario (10%). a medium rate (5%) and a low rate (1%). Columns (3) to (5) in Table 4 show the results. If DIPRES audited 10% of total purchases (a relatively high audit rate), neither Public nor Private group buyers were more likely to prioritize the lower prices. However, as the audit rate decreases from 10% to 5%, we find the treatment effect on preferences for efficiency increases substantially, but only for buyers in the Public treatment group. Again, this is consistent with model's Proposition 3.1 in that increases in efficiency are due to the *extrinsic motivation* of buyers as they will prioritize lower prices only when they are aware their manager has information on their purchasing performance. The effects are not negligible, for an effect size of 15% relative to the control group mean. Notably, farther reducing the audit rate margin to 1% leads to a nearly 100% increase in the effect size among Public group buyers.

Interestingly, there is a decreasing trend in the control group mean as we reduce the hypothetical audit rate. This is because when a small proportion of purchases undergo auditing, buyers are less concerned about facing penalties for overspending, and thus less inclined to base their decisions on price criteria. Yet this is only expected when the extrinsic motivation is inactive, as it is the case of control group and Private group buyers. This is not the case of Public group buyers who are aware of that their overspending is being closely monitored by the authority, hence they prioritize price over quality regardless of the audit margin, and that explain the positive treatment effects on preferences for efficiency.[23]

6.4. Organizational Culture

Our model predicts that manager's access to buyer's performance information produces larger efficiency gains in units where buyer's belief about DIPRES's valuation of efficiency is higher than that of the manager (*Proposition* 3.2). We test this prediction by using baseline survey data collected from 2,411 buyers and 161 unit managers participating in the experiment to measure manager-buyer differences in their expectations, perspectives, and attitudes toward the government's emphasis on efficiency.

*6.4.1. **Belief Misalignment***

The baseline survey consists of 10 questions designed to explore beliefs related to budget execution, savings, and budget allocation decisions from the lens of DIPRES. The questions ask managers and buyers to report their beliefs about the impact of under-spending their unit's budget allocation, the impact of efficiency savings on budget allocation, the role of path-dependence of budget allocation,

[23]Consider the opposite case where all purchases were subject to auditing. If so, then the treatment effect on preferences for efficiency is expected to be negligible since buyers would be compelled to prioritize efficiency in all purchases to avoid punishment.

the extent to which public spending is monitored, the degree of pressure to avoid under-spending, and the incidence of passive waste in procurement practices. These are framed using exactly the same wording for both managers and buyers, and the responses use ordinal scales so that we can rank the responses (see Appendix Table B.6 for a detailed description).The values of the responses are oriented across questions such that higher values are associated with beliefs that go *against* common wisdom, i.e., that DIPRES does not care about efficiency. Hence, the higher the score, the higher is the belief that DIPRES cares about efficiency.[24] The questions are not designed to gauge the intrinsic value that managers and buyers attribute to efficiency, but rather beliefs about how DIPRES values efficiency. Collectively, these work as a proxy to measure the level of misalignment in beliefs about DIPRES's prioritization of efficiency.

We measure the difference between manager and buyers beliefs within each purchasing unit by comparing the responses of the manager and the buyers for each of the ten baseline questions. Let define a manager-buyer question-specific discrepancy as δ_{kij}, where $k = 1, \ldots, 10$ indexes the question, $i = 1, \ldots, N_j$ indexes the buyer in unit j, and $j = 1, \ldots, J$ indexes the manager in unit j, with:

- $\delta_{kij} = 1$ if manager's response to question k scores more than buyer's response;
- $\delta_{kij} = -1$ if manager's response to question k scores less than buyer's response; and
- $\delta_{kij} = 0$ if manager and buyer responses to question k are equivalent.

Then, for each question k in unit j, we compute the proportion of buyers who disagree with the manager as $p_{kj} = \frac{1}{N_j}\sum_{i=1}^{N_j} |\delta_{kij}| \in (0, 1)$, and follow Anderson (2008) to compute a weighted average of the ten p_{kj} values and obtain the top-down Belief Misalignment Index, $MA_j \in (0, 1)$. The closer is MA_j to 1, the larger is the manager-buyers belief misalignment within unit j[25]. The mean level of belief misalignment is 0.73 (with a standard deviation of 0.10), suggesting manager and buyers are generally not aligned in their beliefs about DIPRES's valuation of efficiency.

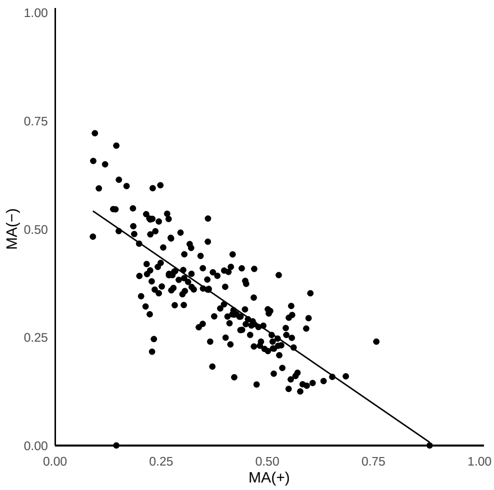
Next, we define two alternative measures that give direction to the belief misalignment index. In particular, for each unit j and question k we compute $p^-_{kj} = \frac{1}{N_j}\sum_{i=1}^{N_j} 1[\delta_{kij} = -1]$, the proportion

[24]For instance, in question 9 we ask buyers and managers to report their level of agreement with the statement "Utilizing the maximum budget is a pressure within units", for which we use a 5-points agreement scale. Since there is the general belief that units have a pressure to avoid under-execution of assigned budget (otherwise they could get budget cuts next year), then being in *disagreement* with the statement implies the respondent believes DIPRES do care about efficiency, and thus a higher value is assigned to that response. Likewise, in question 10, responses that disagree more with the statement "Public purchases are often made at high prices to comply with budget execution" have a higher value compared to responses that agree more with it, since disagreeing with that statement goes counter to the common wisdom that DIPRES do not value efficiency. Another example is the common wisdom that savings have a negative impact on next year's budget. Questions 3 and 6 are in the domain of the role of savings in the next year's budget. In particular, question 3 is about expectations of budget allocation for the following year if savings are generated despite not fully executing the current year's budget, with potential answers being "Decrease", "Same", or "Increase", in that order. As such, the value of the option "Increase" is larger than the value of the option "Decrease", since increasing the next-year budget due to demonstration of savings goes counter to the common wisdom that DIPRES do not value efficiency.

[25]This approach is similar to the one used by Van den Steen (2010) to measure the distance in beliefs between workers in the same firm.

of buyers in unit j who are in disagreement with the manager, but where the manager has a lower belief in that DIPRES cares about efficiency ($\delta_{kij} = -1$). Then, again, using Anderson (2008) we compute a weighted average of the ten p_{kj}^{-} values to obtain the negative top-down misalignment measure, $MA_j(-) \in (0,1)$. A higher value of $MA_j(-)$ suggests that the divergence in beliefs is primarily influenced by the buyers' stronger belief that DIPRES values efficiency, more so than the belief held by the manager. In contrast, $MA_j(+)$ captures the opposite movement, with $p_{kj}^{+} = \frac{1}{N_j}\sum_{i=1}^{N_j} 1[\delta_{kij} = 1]$. These two measures allow us to empirically test whether the model's Proposition 3.2 holds or not.

Figure 7: Relationship between Positive and Negative Belief Misalignment

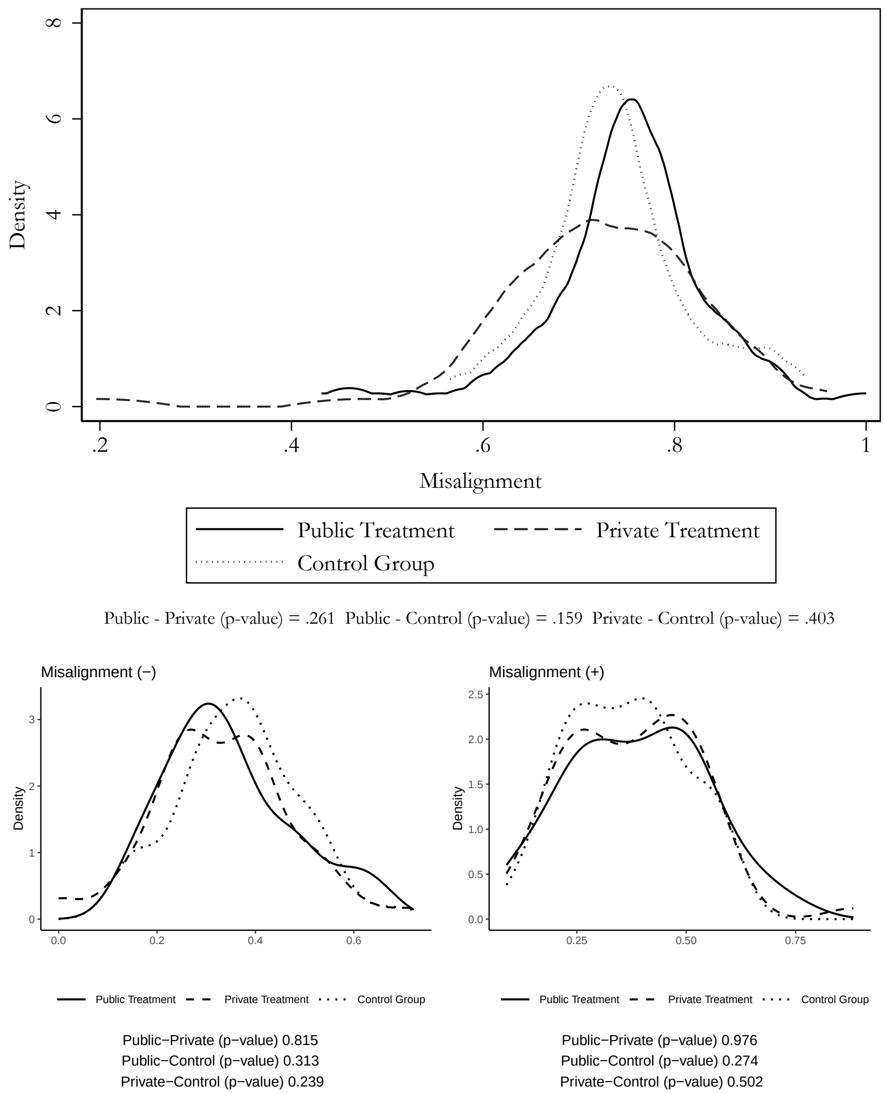


Note: Two-way scatter plot mapping $MA_j(-)$ against $MA_j(+)$. Each dot is a specific unit. The linear correlational coefficient is -0.67.

The average $MA_j(-)$ is 0.35, while the average $MA_j(+)$ is 0.38, meaning the overall misalignment is not unidirectional but it reflects units with positive-dominated misalignment (i.e., managers believe more than buyers that DIPRES values efficiency, on average) as well as units with negative-dominated misalignment (on average, managers believe less than buyers that DIPRES values efficiency). As expected, $MA_j(-)$ and $MA_j(+)$ are negatively correlated, i.e., units in which the manager tends to believe less than the buyers that DIPRES values efficiency have a high $MA_j(-)$ and a low $MA_j(+)$, and *viceversa* (see Figure 7).

Figure 8 displays the distribution of top-down $MA_j(+)$ across different units for each experimental group (top panel). The bottom panel replicates the exercise for $MA_j(-)$ and $MA_j(+)$. According to the Kolmogorov–Smirnov equality-of-distributions test, the belief misalignment index is statistically balanced across groups, and this is regardless of the direction of misalignment.

Figure 8: Distribution of Belief Misalignment by Experimental Group



Note: Baseline distribution of Belief Misalignment Index, by experimental group. Top panel shows the overall misalignment (MA_j), while bottom left and bottom right panels show the distribution of negative ($MA_j(-)$) and positive ($MA_j(+)$) misalignment, respectively. *p-values* for Kolmogorov –Smirnov equality-of-distributions test are shown at the bottom of each panel.

*6.4.2. **Estimation Results***

We estimate the main overprice regression model in equation 5 adding the interaction of the public treatment status with MA_j, controlling for main effects of MA_j, for the public intervention and control group sample excluding the private information group. We standardize MA_j to ease the interpretation of the coefficients, i.e., we subtract the baseline mean and divide by its standard deviation. Hence, the estimates can be interpreted at the mean levels of misalignment.

Table 5 presents the results. At mean levels of misalignment the Public treatment reduced overprice by 1.3 *pp.* relative to the control group (column (1)). However, for transactions made in units that are 1 *s.d.* above the mean of belief misalignment, the Public treatment effect almost doubles, generating an overprice reduction of 2.4 *pp.*

Table 5: Belief Misalignment

	Overprice OLS (1)	Overprice OLS (2)	Overprice OLS (3)
(β_1) Public Reports	-0.013**	-0.013	-0.003
	(0.005)	(0.010)	(0.007)
(β_2) Z-score MA_j	-0.004*	0.000	-0.012***
	(0.002)	(0.008)	(0.004)
(β_3) Public Reports $\times$ Z-score MA_j	-0.011**	-0.027***	0.006
	(0.004)	(0.009)	(0.007)
(β_4) Manager Has High Beliefs		-0.002	
		(0.007)	
(β_5) Z-score MA_j $\times$ Manager Has High Beliefs		-0.003	
		(0.008)	
(β_6) Public Reports $\times$ Manager Has High Beliefs		0.018	
		(0.015)	
(β_7) Public Reports $\times$ Z-score MA_j $\times$ Manager Has High Beliefs		0.030***	
		(0.010)	
(β_8) Buyer Has High Beliefs			0.009
			(0.009)
(β_9) Z-score MA_j $\times$ Buyer Have High Beliefs			0.014**
			(0.006)
(β_{10}) Public Reports $\times$ Buyer Have High Beliefs			-0.004
			(0.012)
(β_{11}) Public Reports $\times$ Z-score MA_j $\times$ Buyer Has High Beliefs			-0.030**
			(0.012)
Observations	54,842	54,842	54,842
N. Units	95	95	95
N. Buyers	1,234	1,234	1,234
Control Mean	0.103	0.103	0.103
$\beta_1 + \beta_3$ (mean+*s.d.*)=0	-0.024 [0.000]	-0.040 [0.002]	0.003 [0.796]
$\beta_1 + \beta_3 + \beta_7$ (mean+*s.d.*)=0		-0.011 [0.350]	
$\beta_1 + \beta_3 + \beta_{11}$ (mean+*s.d.*)=0			-0.027 [0.040]

Notes: Data analysis considers the full post-treatment period of analysis (July 2020 to January 2021) and includes only purchases made in procurement units assigned to either Public or Control groups, and for which the Belief Misalignment Index (MA_j) is observed at baseline (95 units). Following equation 4, overprice of a unit transaction is defined as the relative difference between the per-unit price paid and the lowest price in the daily price distribution for products from the reference group. MA_j is the z-score share of buyers per procurement unit whose beliefs about DIPRES's valuation of efficiency are different compared to the corresponding beliefs of their manager, i.e., we subtract the mean share and divide by its standard deviation. "Manager Has High Beliefs" is a dummy that equals 1 if the manager has a z-score belief index above the across-units' average. "Buyer Has High Beliefs" is a dummy that equals 1 if the buyers' average of the z-score belief index within-unit is above the across-units' average. All regressions control for baseline outcome at the procurement unit level, strata fixed effects, product type fixed effects, and week fixed effects. Standard errors clustered at the procurement unit level are shown in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Bottom rows show the estimates and p-values (in brackets) corresponding to the specified null hypotheses.

We also examine how heterogeneity in the strength of manager and buyer beliefs influences the treatment effects. We create a dummy that equals 1 if the manager has "high beliefs" – i.e., a belief score that is above the median and zero otherwise, and another dummy that equals 1 if the buyer has "low beliefs" measured analogously. We find belief misalignment amplifies the treatment effect, but only when either the manager has “low beliefs" (column 2), or buyers have “high beliefs" (column 3). In particular, the effect size is large and statistically significant when the manager has “low beliefs" ($\beta_1 + \beta_3$ in column 2), but it is null when the manager has “high beliefs" ($\beta_1 + \beta_3 + \beta_7$). In contrast, the effect size is null when the buyer has “low beliefs" ($\beta_1 + \beta_3$ in column 3), but it is large and statistically significant when the buyer has “high beliefs" ($\beta_1 + \beta_3 + \beta_{11}$).

Recall, however, that *Proposition* 3.2 establishes that MA_j facilitates efficiency gains derived from performance information only when $\mu_I > \mu_J$, and thus the treatment effects when $\mu_I > \mu_J$ are larger than when $\mu_I < \mu_J$. We directly test *Proposition* 3.2 in Table 6, where we show the extent to which the treatment effect varies with increases in the *within-unit* share of buyers with higher beliefs than the manager ($MA_j(-)$). We observe that it is the negative misalignment that intensifies the treatment effect and this result is robust to controlling for the share of buyers with lower beliefs than the manager ($MA_j(+)$) (see column 2). Moreover, the role of $MA_j(+)$ is null as both γ_6 and γ_7 are close to zero and statistically insignificant (column 3).

Finally, *Proposition* 3.3 predicts that manager's internal monitoring has a positive effect on efficiency, but lowers the effect of performance information from an external monitoring system. We test this prediction in Table 7 where add to equation 5 the interaction of the public treatment status with MA_j and a Monitoring Index MI_j, controlling for main effects of MA_j and MI_j.[26] First, the effect of MA_j on treatment remains intact after controlling for MI_j (Column 1). Second, our analysis reveals a negative and significant main effect of internal monitoring on overprice (α_4), confirming that increasing monitoring enhances efficiency. However, as monitoring increases, the effect of the public intervention on efficiency appears to diminish, as is shown by the positive and significant coefficient associated to the interactive variable (α_5). Indeed, for units that are 1 *s.d.* above the monitoring mean, the effect of accessing to Public performance reports is null ($\alpha_1 + \alpha_5$ (mean+*s.d.*)=0 not rejected).

For robustness, we replicate the analysis but using $MA_j(-)$ instead of MA_j, and indeed, as we show in regression column (2), the results are robust to this specification. Finally, in columns (3) and (4) we split the sample in purchases made at units with high and low levels of monitoring at baseline (above *vs.* below the median), respectively, and again, in line with model predictions, we find all the treatment action is concentrated in units with low monitoring. This result reinforces the hypothesis of performance reports having diminishing returns as internal monitoring increases.

[26] The Monitoring Index MI_j is built based on baseline survey data, where we asked buyers to report on the perceived level of monitoring of public spending within the unit, in a 1-7 scale (1= no monitoring; 7=high monitoring). The median score is 6, hence MI_j is the share of buyers per unit reporting that the perceived level of monitoring is 6 or more, i.e., the larger the monitoring index, the larger is the share of buyers within the unit perceiving a relatively “high" level of monitoring. As with MA_j, we standardize MI_j to ease the interpretation of the coefficients.

Table 6: Direction of Belief Misalignment

	Overprice OLS (1)	Overprice OLS (2)	Overprice OLS (3)
(γ_1) Public Reports	-0.013**	-0.001	-0.002
	(0.005)	(0.006)	(0.006)
(γ_2) Z-score MA_j	-0.004*		
	(0.002)		
(γ_3) Public Reports $\times$ Z-score MA_j	-0.011**		
	(0.004)		
(γ_4) Z-score $MA_j(-)$		-0.001	-0.001
		(0.005)	(0.005)
(γ_5) Public Reports $\times$ Z-score $MA_j(-)$		-0.018***	-0.018**
		(0.008)	(0.008)
(γ_6) Z-score $MA_j(+)$			-0.001
			(0.005)
(γ_7) Public Reports $\times$ Z-score $MA_j(+)$			0.002
			(0.009)
Observations	54,842	54,842	54,842
N. Units	95	95	95
N. Buyers	1,234	1,234	1,234
Control Mean	0.103	0.103	0.103
$\gamma_1 + \gamma_3$ (mean+*s.d.*)=0	-0.024 [0.000]		
$\gamma_1 + \gamma_5$ (mean+*s.d.*)=0		-0.020 [0.008]	-0.019 [0.061]
$\gamma_1 + \gamma_7$ (mean+*s.d.*)=0			0.000 [0.991]

Notes: Data analysis considers the full post-treatment period of analysis (July 2020 to January 2021) and includes only purchases made in procurement units assigned to either Public or Control groups, and for which the Belief Misalignment Index (MA_j) is observed at baseline (95 units). Following equation 4, overprice of a unit transaction is defined as the relative difference between the per-unit price paid and the lowest price in the daily price distribution for products from the reference group. MA_j is the z-score share of buyers per procurement unit whose beliefs about DIPRES's valuation of efficiency are different compared to the corresponding beliefs of their manager, i.e., we subtract the mean share and divide by its standard deviation. Z-score $MA_j(-)$ is the standardized share of buyers whose z-score beliefs index about DIPRES's valuation of efficiency are higher compared to the corresponding beliefs of their manager. Z-score $MA_j(+)$ is the standardized share of buyers whose z-score beliefs index about DIPRES's valuation of efficiency are lower compared to the corresponding beliefs of their manager. All regressions control for baseline outcome at the procurement unit level, strata fixed effects, product type fixed effects, and week fixed effects. Standard errors clustered at the procurement unit level are shown in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Bottom rows show the estimates and p-values (in brackets) corresponding to the specified null hypotheses.

Table 7: Belief Misalignment and Internal Monitoring

	All	All	High Baseline Monitoring	Low Baseline Monitoring
	Overprice OLS (1)	Overprice OLS (2)	Overprice OLS (3)	Overprice OLS (4)
(α_1) Public Reports	-0.010**	-0.002	0.024	0.007
	(0.004)	(0.006)	(0.026)	(0.009)
(α_2) Z-score MA_j	-0.005*			
	(0.003)			
(α_3) Public Reports × Z-score MA_j	-0.015***			
	(0.004)			
(α_4) Z-score Baseline Monitoring Index (MI_j)	-0.010***	-0.007***		
	(0.002)	(0.002)		
(α_5) Public Reports × MI_j	0.007**	0.008**		
	(0.003)	(0.003)		
(α_6) Z-score $MA_j(-)$		-0.001	0.011	0.004
		(0.003)	(0.011)	(0.009)
(α_7) Public Reports × Z-score $MA_j(-)$		-0.017***	-0.034	-0.021**
		(0.004)	(0.021)	(0.009)
Observations	54,842	54,842	16,029	38,813
N. Units	95	95	45	50
N. Buyers	1,234	1,234	462	772
Control Mean	0.103	0.103	0.101	0.103
$\alpha_1 + \alpha_3$ (mean+*s.d.*)=0	-0.025 [0.000]			
$\alpha_1 + \alpha_5$ (mean+*s.d.*)=0	-0.003 [0.564]	0.005 [0.445]		
$\alpha_1 + \alpha_7$ (mean+*s.d.*)=0		-0.020 [0.008]	-0.009 [0.479]	-0.014 [0.015]

Notes: Data analysis considers the full post-treatment period of analysis (July 2020 to January 2021) and includes only purchases made in procurement units assigned to either Public or Control groups, and for which the Belief Misalignment Index (MA_j) is observed at baseline (95 units). Following equation 4, overprice of a unit transaction is defined as the relative difference between the per-unit price paid and the lowest price in the daily price distribution for products from the reference group. MA_j is the z-score share of buyers per unit whose beliefs about DIPRES's valuation of efficiency are different compared to the corresponding beliefs of their manager, i.e., we subtract the mean share and divide by its standard deviation. Z-score $MA_j(-)$ is the standardized share of buyers whose z-score beliefs index about DIPRES's valuation of efficiency are higher compared to the corresponding beliefs of their manager. Baseline Monitoring Index (MI_j) is the standardized share of buyers per procurement unit reporting that the z-score perceived level of monitoring of public spending within the unit is above or equal to the median in a 1-7 scale (1= no monitoring; 7=high monitoring). Regressions in columns (1) and (2) consider the full sample of unit transactions. Column (3) regression considers only unit transactions made at procurement units where the mean level of MI_j is above the median (High Baseline Monitoring), while Column (4) regression considers only unit transactions made at procurement units where the mean level of MI_j is below or equal to the median (Low Baseline Monitoring). All regressions control for baseline outcome at the procurement unit level, strata fixed effects, product type fixed effects, and week fixed effects. Standard errors clustered at the procurement unit level are shown in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Bottom rows show the estimates and p-values (in brackets) corresponding to the specified null hypotheses.

7. Comparison to Forecasts from the Prediction Platform

We compare our research findings to the views of the scientific community and technocrats, which has proven useful to improve the informativeness and interpretation of research results (Dellavigna et al., 2019). Specifically, we elicit forecasts from experts and non-experts on public procurement about the potential effects of our experimental intervention. Our forecasting survey first collects basic information on the working experience of respondents, then introduces the experimental design and interventions, and finally asks the respondents to predict what would be the post-treatment difference in the level of overspending (in standard deviations) across public, private, and control groups.

Forecasts are elicited through the Social Science Prediction Platform, which enables the systematic collection and assessment of expert forecasts of the effects of untested programs.[27] The invitation to participate in the forecast was voluntary and did not include pecuniary nor non-pecuniary incentives. The survey was posted in the platform website before the intervention was implemented and remained open up to December 2021, i.e., it ended before any of the results of this study had been publicly released. 61 individuals completed the survey, half of which are classified as experts in public procurement, i.e., they report to have had experience (and thus some expertise) working in one or more of the following areas: public procurement, public budgeting, state-level efficiency, state capacity, public administration, or corruption.

More than 60% of respondents forecast that procurement officers receiving the Public performance reports will reduce overspending, for a mean predicted effect of -0.070 *s.d.*, and a 95% confidence interval of [0.001; -0.141]. This is not statistically different to the average causal effect of being exposed to the Public performance reports (-0.078 *s.d.*), suggesting respondents does not underestimate its role on enhancing efficiency in public procurement.

Importantly, expert and non-experts predict similar effect sizes on average (-0.076 *s.d.* vs. -0.065 *s.d.*, respectively). The latter reveals that having working experience in areas related with public procurement does not add too much precision to the treatment effects prediction, giving room to a broad consensus across experts and non-experts regarding the expected effectiveness of the intervention. Second, the mean predicted effect for the Private performance reports is 0.002 *s.d.* [-0.059; .064], and again, expert and non-experts show similar average predictions (0.009 *s.d.* vs. -0.003 *s.d.*, respectively). This is consistent with the null experimental effect we find for this treatment arm.

Overall, our experimental findings closely align with the predictions made by both professionals and non-professionals in the public procurement field. This consensus strengthens the viability of scaling-up the public intervention to other service units within the State.

[27] See https://socialscienceprediction.org/s/x46yya for details on the invitation to forecast the impact of our interventions.

8. Conclusion

Public procurement is often characterized by moral hazard issues, incomplete contracts, and bureaucratic leisure, which ultimately result in inefficiencies. Chile recently implemented a new procurement system based on an online electronic platform that resembles Amazon, in which pre-qualified suppliers and products are chosen through competitive bidding following Framework Agreements (FAs). The new system has fostered increased competition among suppliers, thereby enhancing the bidding process for products of comparable quality, which in turn reduces the chances for active waste, such as corruption. However, passive waste may still remain high.

Together with Chile's Public Budget Office and Public Procurement Office, we designed a cluster randomized field experiment to examine the impact of providing individual performance reports to procurement officers on passive waste. The results indicate that performance monitoring reduces overspending, but this is only when buyer performance was observable for managers, suggesting that buyer's efficiency performance is mostly driven by *extrinsic* motivation.

We also investigated how organizational culture, in this case the extent to which the beliefs about the government's valuation of efficiency are shared between managers and officers, moderated the effectiveness of performance information. We find that performance information proves to be more effective in procurement units where the manager has a lower belief that the government cares about efficiency than does the buyer. We hypothesize this is due to that managers engage in *motivated reasoning*, i.e., the manager surprisingly realizes that the buyer is not performing up to her potential. Consequently, because the buyer believes that the government cares about efficiency, the manager invests in effort to persuade the buyer to be more efficient, resulting in improved efficiency. Conversely, when the manager believes the government cares about efficiency more than does the buyer, information on buyer's inefficiency induces *confirmation bias*: the manager believes that buyer's inefficiency is due to her low beliefs on government's prioritization of efficiency. This increases the manager's perceived persuasion cost, which in turn discourages pro-efficiency persuasion efforts. Our findings demonstrate that managers' access to buyers' performance information can greatly enhance the efficiency of public procurement, especially under adverse effects resulting from cultural obstacles tied to organizational misalignment.

Overall, we find the providing managers and buyer performance reports reduces overspending by about 15%. Extending the annualized treatment effect to encompass all transactions conducted within the Chilean public procurement system would yield approximately US $0.15 billion dollars in savings, which is equivalent to 1.2% of Central Government procurement expenditure or about 0.1% of Chile's GDP. Since the primary information is embedded in the e-procurement platform, automating the production and delivery of the performance reports incurs virtually zero cost. However, scaling-up the intervention may require a broad consensus among policy makers. Our experimental findings somewhat seem to align with the predictions made by both professionals and non-professionals in the public procurement field, which indicate there might be some promise in considering the feasibility of scaling-up the intervention.

References

AGHION, P. AND J. TIROLE (1997): "Formal and real authority in organizations," *Journal of Political Economy*, 105, 1–29.

ALBANO, G. L. AND C. NICHOLAS (2016): *The Law and Economics of Framework Agreements*, Cambridge University Press.

ALLENDE, C., J. ATAL, R. CARRIL, J. CUESTA, AND A. GONZÁLEZ-LIRA (2023): "Drivers of Public Procurement Prices: Evidence from Pharmaceutical Markets," *mimeo*.

ANDERSON, M. L. (2008): "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 103, 1481–1495.

BAJARI, P., R. MCMILLAN, AND S. TADELIS (2009): "Auctions Versus Negotiations in Procurement: An Empirical Analysis," *Journal of Law, Economics, and Organizations*, 25, 372–399.

BAJARI, P. AND S. TADELIS (2001): "Incentives versus Transaction Costs: A Theory of Procurement Contracts," *RAND Journal of Economics*, 32, 387–407.

BAKER, G., R. GIBBONS, AND K. MURPHY (2002): "Relational Contracts and the Theory of the Firm," *The Quarterly Journal of Economics*, 117, 39–84.

BAKER, G. AND T. HUBBARD (2003): "Make Versus Buy in Trucking: Asset Ownership, Job Design and Information," *The American Economic Review*, 93, 551–572.

——— (2004): "Contractability and Asset Ownership: On-Board Computers and Governance in U.S. Trucking," *The Quarterly Journal of Economics*, 119, 1443–1479.

BALÁN, P., A. BERGERON, G. TOUREK, AND J. WEIGEL (2022): "Local Elites as State Capacity: How City Chiefs Use Local Information to Increase Tax Compliance in the Democratic Republic of the Congo," *American Economic Review*, 112, 762–97.

BANDIERA, O., B. I. AND I. RASUL (2011): "Field Experiments With Firms," *Journal of Economic Perspectives*, 25, 63–82.

——— (2013): "Team Incentives: Evidence From a Firm Level Experiment," *Journal of the European Economic Association*, 11, 1079–1114.

BANDIERA, O., I. BARANKAY, AND I. RASUL (2009a): "Social Connections and Incentives in the Workplace: Evidence From Personnel Data," *Econometrica*, 77, 1047–94.

BANDIERA, O., M. C. BEST, A. Q. KHAN, AND A. PRAT (2021a): "The allocation of authority in organizations: A field experiment with bureaucrats," *The Quarterly Journal of Economics*, 136, 2195–2242.

BANDIERA, O., E. BOSIO, AND G. SPAGNOLO (2021b): "Procurement in Focus: Rules, Discretion, and Emergencies," *CEPR Press*.

BANDIERA, O., A. PRAT, AND T. VALLETTI (2009b): "Active and Passive Waste in Government Spending: Evidence from a Policy Experiment," *American Economic Review*, 99, 1278–1308.

BARNARD, C. (2011): "Using Procurement Law to Enforce Labour Standards," *in The Idea of Labour Law, edited by Guy Davidov and Brian Langille. Oxford University Press*.

BEER, R., I. RIOS, AND D. SABAN (2021): "Increased Transparency in Procurement: The Role of Peer Effects," *Management Science*, 67, 7511–7534.

BÉNABOU, R. (2013): "Groupthink: Collective delusions in organizations and markets," *Review of Economic Studies*, 80, 429–462.

BÉNABOU, R. AND J. TIROLE (2003): "Intrinsic and Extrinsic Motivation," *The Review of Economic Studies*, 70, 489–520.

BEST, M., J. HJORT, AND D. SZAKONYI (2022): "Individuals and Organizations as Sources of State Effectiveness," *American Economic Review*, Forthcoming.

BLADER, S., C. GARTENBERG, AND A. PRAT (2020): "The Contingent Effect of Management Practices," *Review of Economic Studies*, 87, 721–749.

BLOOM, N., E. BRYNJOLFSSON, L. FOSTER, R. JARMIN, M. PATNAIK, I. SAPORTA-EKSTEN, AND J. VAN REENEN (2019): "What Drives Differences in Management Practices?" *American Economic Review*, 109, 1648–1683.

BLOOM, N., R. SADUN, AND J. VAN REENEN (2012): "The Organization of Firms Across Countries," *Quarterly Journal of Economics*, 127, 1663–1705.

BLOOM, N. AND J. VAN REENEN (2007): "Measuring and Explaining Management Practices Across Firms and Countries," *Quarterly Journal of Economics*, 122, 1351–1408.

BOSIO, E., S. DJANKOV, E. GLAESER, AND A. SHLEIFER (2021): "Public Procurement in Law and Practice," *American Economic Review*, Forthcoming.

BULOW, J. AND P. KLEMPERER (1996): "Auctions Versus Negotiations," *American Economic Review*, 86, 180–194.

CARPINETI, L., G. PIGA, AND M. ZANZA (2006): "The variety of procurement practice: evidence from public procurement," .

CARRIL, R. (2019): "Rules versus Discretion in Public Procurement," *Job Market Paper*.

CARRIL, R., A. GONZALEZ-LIRA, AND M. WALKER (2022): "Competition under Incomplete Contracts and the Design of Procurement Policies," *BSE Working Paper #1327*.

COVIELLO, D., A. GUGLIELMO, AND G. SPAGNOLO (2018): "The Effect of Discretion on Procurement Performance," *Management Science*, 64, 715–38.

CRAWFORD, V. P. AND J. SOBEL (1982): "Strategic Information Transmission," *Econometrica*, 50, 1431–1451.

DAL BÓ, E., F. FINAN, N. LI, AND L. SCHECHTER (2021): "Information technology and government decentralization: experimental evidence from Paraguay," *Econometrica*, 89, 677–701.

DE ROCHAMBEAU, G. (2021): "Monitoring and Intrinsic Motivation: Evidence from Liberia's Trucking Firms," *mimeo*.

DECAROLIS, F., R. FISMAN, P. PINOTTI, AND S. VANNUTELLI (2020): "Rules, Discretion, and Corruption in Procurement: Evidence from Italian Government Contracting," *CEPR Discussion Paper No. 14794*.

DELLAVIGNA, S., D. POPE, AND E. VIVALT (2019): "Predict Science to Improve Science," *Science*, 366, 428–429.

DEWATRIPONT, M. AND J. TIROLE (1999): "Advocates," *Journal of Political Economy*, 107, 1–39.

DIMITRI, N., G. PIGA, AND G. SPAGNOLO (2006): *Handbook of procurement*, Cambridge University Press.

DINERSTEIN, M., L. EINAV, J. LEVIN, AND N. SUNDARESAN (2018): "Consumer price search and platform design in internet commerce," *American Economic Review*, 108, 1820–59.

ENGEL, E., F. JORDAN, T. RAU, AND A. REPETTO (2021): "Audit Threats and Wasteful Spending by Government Agencies: Experimental Evidence from Chile," *mimeo*.

GERARDINO, M., S. LITSCHIG, AND D. POMERANZ (2022): "Can Audits Backfire?" *NBER Working Paper No 23978*.

GIBBONS, R. AND R. HENDERSON (2013): "What Do Managers Do?" *in Gibbons, R. and Roberts, J. (eds) The Handbook of Organizational Economics, (Princeton, NJ: Princeton University Press)*, 680–731.

GORTON, G. B., J. GRENNAN, AND A. K. ZENTEFIS (2021): "Corporate culture," Tech. rep., National Bureau of Economic Research.

GUISO, L., P. SAPIENZA, AND L. ZINGALES (2015a): "Corporate Culture, Societal Culture, and Institutions," *American Economic Review Papers and Proceedings*, 105, 336–39.

——— (2015b): "The Values of Corporate Culture," *Journal of Financial Economics*, 117, 60–76.

HALAC, M. AND A. PRAT (2016): "Managerial Attention and Worker Performance," *American Economic Review*, 106, 3104–32.

HELPER, S. AND R. HENDERSON (2014): "Management Practices, Relational Contracts, and the Decline of General Motors," *Journal of Economic Perspectives*, 28, 49–72.

HOFFMAN, M. AND S. TADELIS (2021): "People Management Skills, Employee Attrition, and Manager Rewards: An Empirical Analysis," *Journal of Political Economy*, 129, 243–285.

HUBBARD, T. (2000): "The Demand for Monitoring Technologies: The Case of Trucking," *The Quarterly Journal of Economics*, 115, 533–560.

——— (2003): "Information, Decisions, and Productivity: OnBoard Computers and Capacity Utilization in Trucking," *The American Economic Review*, 93, 1328–1353.

ICHNIOWSKI, C., K. SHAW, AND G. PRENNUSHI (1997): "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines," *American Economic Review*, 87, 291–313.

KELLEY, E., G. LANE, AND D. SCHÖNHOLZER (2021): "Monitoring in Small Firms: Experimental Evidence from Kenyan Public Transit," *mimeo*.

KELMAN, S. (1990): "Procurement and Public Management: The Fear of Discretion and the Quality of Government Performance," *AEI Press*.

——— (2005): "Unleashing change: A study of organizational renewal in government," *Washington, D.C.: Brookings Institution*.

KHAN, A., A. KHWAJA, AND B. OLKEN (2016): "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors," *Quarterly Journal of Economics*, 131, 219–71.

——— (2019): "Making Moves Matter: Experimental Evidence on Incentivizing Bureaucrats through Performance-Based Postings," *American Economic Review*, 109, 237–70.

LAFFONT, J.-J. AND J. TIROLE (1993): *A theory of incentives in procurement and regulation*, MIT press.

LEVIN, J. (2003): "Relational Incentive Contracts," *The American Economic Review*, 93, 835–857.

LEWIS-FAUPEL, S., Y. NEGGERS, B. OLKEN, AND R. PANDE (2016): "Can Electronic Procurement Improve Infrastructure Provision? Evidence from Public Works in India and Indonesia," *American Economic Journal: Economic Policy*, 8, 258–83.

LIEBMAN, J. AND N. MAHONEY (2017): "Do Expiring Budgets Lead to Wasteful Year-End Spending? Evidence from Federal Procurement," *American Economic Review*, 107, 3510–49.

LISCOW, Z., , W. NOBER, AND C. SLATTERY (2023): "Procurement and Infrastructure Costs," *Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract*$_id = 4522676$.

MACLEOD, W. AND J. MALCOMSON (1989): "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment," *Econometrica*, 57, 447–480.

MARTINEZ, E., N. BEAULIEU, R. GIBBONS, P. PRONOVOST, AND T. WANG (2015): "Organizational Culture and Performance," *American Economic Association: Papers and Proceedings*, 105, 331–335.

MATTSSON, M. (2022): "Information Systems, Service Delivery, and Corruption: Evidence from the Bangladesh Civil Service," *mimeo*.

MIKOLOV, T., I. SUTSKEVER, K. CHEN, G. S. CORRADO, AND J. DEAN (2013): "Distributed representations of words and phrases and their compositionality," *Advances in neural information processing systems*, 26.

OECD (2009): "Principles for Integrity in Public Procurement," *Organisation for Economic Co-operation and Development, Tech. Report*.

——— (2013): "Monitoring of Public Procurement," Tech. rep., Available at https://www.sigmaweb.org/publications/Brief27_MonitoringofPP_2013.pdf.

——— (2019): *Reforming Public Procurement: Progress in Implementing the 2015 OECD Recommendation*, OECD Publishing.

PALGUTA, J. AND F. PERTOLD (2017): "Manipulation of Procurement Contracts: Evidence from the Introduction of Discretionary Thresholds," *American Economic Journal: Economic Policy*, 9, 1–24.

RASUL, I. AND D. ROGERS (2018): "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service," *Economic Journal*, 128, 413–446.

ROTEMBERG, J. J. AND G. SALONER (1995): "Overt Interfunctional Conflict (and Its Reduction Through Business Strategy)," *The RAND Journal of Economics*, 26, 630–653.

SABAN, D. AND G. WEINTRAUB (2021): "Procurement mechanisms for assortments of differentiated products," *Operations Research*, 69, 795–820.

SCHEIN, E. (2004): "Organizational Culture and Leadership," *San Francisco, CA: Jossey-Bass*.

SHI, M. (2023): "Monitoring for Waste: Evidence from Medicare Audits," *Quarterly Journal of Economics*, Forthcoming.

SPAGNOLO, G. (2012): “Reputation, Competition and Entry in Procurement,” *International Journal of Industrial Organization*, 30, 291–296.
SPENKUCH, J., E. TESO, AND G. XU (2023): “Ideology and Performance in Public Organizations,” *Econometrica*, Forthcoming.
SUN, C., N. RAMPALLI, F. YANG, AND A. DOAN (2014): “Chimera: Large-scale classification using machine learning, rules, and crowdsourcing,” *Proceedings of the VLDB Endowment*, 7, 1529–1540.
SZUCS, F. (2023): “Discretion and Favoritism in Public Procurement,” *Journal of the European Economic Association*, 0, 1–44.
VAN DEN STEEN, E. (2010): “Culture clash: The costs and benefits of homogeneity,” *Management Science*, 56, 1718–1738.

A. Appendix: Building Reference Groups

Measuring efficiency in public purchases requires to compare adjusted-quality prices across products. This in turn implies defining groups of products of comparable quality that are potential substitutes of each other such that, in the margin, buyers do not significantly reduce utility from choosing the cheapest option. In practice, *Chilecompra* publicly announce a product catalogue including all the products that will be auctioned to suppliers in each Framework Agreement (FA). The products are classified using three levels of aggregation:

- *Framework Agreement:* it specifies the general category of products. For instance, our study includes three FAs: Food, Office Supplies, and Computers.
- *Product Type:* it classifies products in groups of products with similar functionality. For example, the Food FA includes products like Instant Coffee, Rice, Tomato, etc. The Office Supplies FA includes product types such as Brackets, Pencil, Stapler or Batteries, among many others. Computers FA includes Desktop, Laptop, and All-in-one product types.
- *Product ID:* it identifies different products within product type. It is the most granular definition of a product, equivalent to a Stock Keeping Unit (SKU). Importantly, each *Product ID* is associated to a product description (open text), which provides the specific attributes and characteristics of the product.

The catalogue classifies the products based on their functionality, yet this is not sufficient for the purpose of constructing reference groups that capture potential substitutes of similar quality. Indeed, descriptive statistics reveal significant price dispersion across products within the same *Product Type*, for a coefficient of variation of posted prices ranging from 0.5 to 1.8. This could be in part due to vertical differentiation. On the other side, price dispersion is much lower within the same *Product ID* —coefficient of variation between 0.06-0.08 (more on this in Section A.3 below). Still, we observe that several Product IDs have similar attributes and therefore could be considered substitutes for the purpose of measuring overprices.

Hence, we worked in collaboration with *Chilecompra* to built an intermediate product aggregation based on *tagged data*, i.e., observable product attributes that can be extracted from open-text product descriptions. By leveraging this tagged data, we created reference groups comprising products that exhibit identical attribute values. We first extract the attributes from the open-text descriptions of the products, and then evaluate the adequacy of the identified attributes in distinguishing quality variations among products, ensuring that the established reference groups accurately represent potential substitutes of similar quality.

A.1. Identifying Product Attributes

The first step is to partition the set of catalogued products into product types and identify which attributes are frequently included in the product description. Attributes can be separated into numerical and categorical, which will become useful to define product similarities. This process is conducted by a category manager at *Chilecompra*, assisted with automated text processing tools. Table A.1 below presents some examples.

Product Type	Category	Categorical Attributes	Numerical Attributes
Desktop	Computers FA	Brand, Processor, Operating System, Hard Drive type	RAM, Hard Drive Capacity, Screen size
Pencil	Office Supplies FA	Brand, Size, Color	Units per package, Weight, Number of packages/Box
Batteries	Office Supplies FA	Type, Size, Rechargeable	Units per Package, Voltage
Instant Coffee	Food FA	Brand, If Decaffeinated	Weight

Table A.1: Examples of product types and their attributes.

For instance, Desktop is a product type of Computers FA, whose categorial attributes are *Brand, Processor, Operating System* and *Hard Drive type*, and numerical attributes are *RAM, Hard Drive Capacity* and *Screen Size*. If the Desktop does not include a screen, it takes the value "not available" in that attribute.

A second step consists in assigning values to each identified attribute. First, non-supervised Natural Language Processing (NLP) algorithms are used to identify possible values of each attribute. Specifically, Word2Vec (Mikolov et al. (2013)) is applied to the text descriptions to identify common words (tokens). For example, for products in the Pencil category, the tokens "10 cm", "12 cm", "15cm" and "20cm" appear on each product description. Tokens are grouped in this manner and revised by a human (the category manager at *ChileCompra*), to indicate the attribute to which these tokens belong; in this example, the tokens correspond to the values of the *Height* attribute of the Pencil. Since this is a numerical attribute it is assigned a measurement unit – "centimeters" in our example. Table A.2 provides more examples of product description and the identified attribute values.

This manual assignment of tokens identified through Word2Vec is repeated several times until each attributes has at least one assigned value, thereby providing a training data set that can be used with the *supervised* NLP classification algorithm. Specifically, new products can be processed by calculating a distance measure between the product description and the attribute values already identified in the training data set. If the distance is below a specified threshold, the attribute value is assigned automatically. Otherwise, the product description is assigned for manual processing and added to the training set by incorporating the new identified attribute values. The threshold to accept/reject the automated classification can be optimized in order to achieve a desired level of classification error. The algorithm keeps learning as more and more products are classified, improving its precision to identify attributes and thereby reducing the need of manual supervision. Overall, the precision of the classification algorithm was validated using a manually classified testing dataset. For the case of Computers FA, the algorithm correctly classified the attributes of 100% of products that were transacted in 2019. However, in the case of Office Supplies, only 80% of products attributes were correctly identified, and for the case of Food, this was 36%.

Table A.2: Examples of Products Description and Identified Attribute Values

FA: Computers (Desktop)

Id	Product description	Type	Procesor	Brand	HD Cap	RAM	OS	HD Type	Screen
1537456	COMPUTADOR SP LABS INTEL CORE I5-7500, MSI B250M PRO-VH LGA 1151, DDR4-2400 M.2, HDMI, VGA, USB 3.1 MICRO-ATX, MEMORIA RAM 8GB DDR4 , HDD 1 TB , FREE DOS, GRABADOR DE DVD DESKTOP	Desktop	INTEL CORE I5	SP LABS	1TB	8GB	FREEDOS	HDD	None
1615084	LENOVO DESKTOP M720Q , PORCESADOR INTEL CORE I7-8700T,8GB DDR4 2666MHZ SODIMM, 256GB SOLID STATE DRIVE, WIFI INTEL 3165+BT 1X1AC, WINDOWS 10 PRO, MONITOR LENOVO S22E 21.5"" 1920X1080	Desktop	INTEL CORE I7	LENOVO	256GB	8GB	WINDOWS 10 PRO	SSD	21.5

FA: Office Supplies (Print Paper)

Id	Product description	Type	Size	Brand	Pages/Unit	Weight	Color	Units/Pkg
1533457	PAPEL MULTIPROPÓSITO CHAMEX ECO CARTA 75GR UNIDAD ECO CARTA 75GR UNIDAD MEDIDA 216X279MM, RESMA 500 HOJAS	Print Paper	Letter	CHAMEX	500	75	BLANCO	1
1006598	PAPEL MULTIPROPÓSITO XEROX OFICIO 75GR ALBURA 90-95% RESMA 500 HJS.	Print Paper	OFICIO	XEROX	500	75	BLANCO	1
1006553	PAPEL MULTIPROPÓSITO DIAZOL A4 SPECTRA 500HJ AMARILLO	Print Paper	A4	DIAZOL	500	NA	AMARILLO	1

FA: Food (Chicken)

Id	Product	Type	Brand	Maritated	Weight
552243	frozen chicken boneless breast not marinated caja 15 k	frozen chicken		not marinated	15000.0 gr.
552447	frozen chicken boneless breast las camelias not marinated caja 15 k	frozen chicken	las camelias	not marinated	15000.0 gr.
1471381	frozen chicken boneless breast not marinated bolsa 1k	frozen chicken		not marinated	1000.0 gr.
1505907	frozen chicken boneless breast seara not marinated bolsa 2k	frozen chicken	seara	not marinated	2000.0 gr.

FA: Food (Tomato)

Id	Product	Type	Model	Weight
1473761	fresh tomato cherry 1 kg. aprox.	fresh vegetables	cherry	1000.0 gr.
1473779	fresh tomato pomarola small size 1 kg. aprox.	fresh vegetables	pomarola small size	1000.0 gr.
1473794	fresh tomato pomarola medium size 1 kg. aprox.	fresh vegetables	pomarola medium size	1000.0 gr.
1473808	fresh tomato pomarola large size 1 kg. aprox.	fresh vegetables	pomarola large size	1000.0 gr.
1473819	fresh tomato pomarola extra large size 1 kg. aprox.	fresh vegetables	pomarola extra large size	1000.0 gr.
1473833	fresh tomato larga vida small size 1 kg. aprox.	fresh vegetables	pomarola small size	1000.0 gr.

A.2. Products Selection

As can be shown in the example from Table A.2, the open-text description for products in the Computers FA is very detailed, allowing to characterize eight attributes per desktop product. Likewise, in the case of Pencil, it was possible to identify six attributes. Still, the text might include more attributes than the ones our algorithm is able to classify, meaning the buyer could choose the product based on unidentifiable attributes reflecting unobserved quality of products. Two extreme cases are Chicken and Tomato, where the information provided in the open-text description is much more limited to assess the quality of the product; this limitation was observed for most products of fresh produce and meat. Indeed, in order to preserve the accuracy of efficiency metrics, *Chilecompra* opted to remove all fresh products and meat products from the overprice reports used in the monitoring system. More generally, out of the 37,885 products offered in the Computers, Office Supplies, and Food framework agreements in 2019, *Chilecompra* excluded 18,611 of them from the efficiency reports due to classification limitations. That is, our analysis consider about half of offered products.

A.3. Constructing Reference Groups.

In order for reference groups to effectively reflect comparable products in terms of functionality and quality, *ChileCompra* match products with same identifiable attributes. To be on the safe side, they evaluate whether the attributes used to construct reference groups are adequate to minimize differences in quality across products, for which examine price dispersion across four different reference group criteria:

1. Group products from the same Product Type (functionality), regardless of whether they share same attributes.
2. Group products from the same Product Type (functionality) and ID, which by definition share exactly the same attributes, and this is regardless of whether the attributes are identifiable or not.
3. Group products from the same Product Type (functionality), yet not necessary with the same ID, but that share exactly the same attributes.
4. Group products from the same Product Type (functionality), yet not necessary with the same ID, but that share exactly the same attributes, except brand. This is because for several product types, it is not clear whether different brands are comparable in terms of quality.

Notice that FA contracts regulate the service conditions that suppliers must deliver, such as stock availability, delivery times, shipping rates, among others. Consequently, products in a reference group are similar in terms of both product attributes and service quality.

Table A.3 shows the average coefficient of variation of posted prices across reference groups for each reference group criteria.

Table A.3: Coefficient of Variation of Posted Prices across different aggregation levels, by FA

	Match on Same Product Type regardless of Attributes		Match on Same Product Type and Same Product ID		Match on Same Product Type and Same Attributes		Match on Same Product Type and Same Attributes except Brand	
FA	CV	N	CV	N	CV	N	CV	N
Office Supplies	1.85	597.80	0.06	2.87	0.08	3.73	0.17	7.87
Computers	0.73	7,322.34	0.06	10.46	0.09	22.21	0.11	29.02
Food	0.50	175.81	0.07	3.13	0.08	4.38	0.10	6.40

Notes: FA is the Framework Agreement. CV is the average coefficient of variation of prices of grouped products based on the corresponding reference group criteria. N is the average number of available products per group.

When grouping products by Product Type regardless of whether products share the same attributes, we observe prices vary significantly, for an average coefficient of variation ranging between 0.50 and 1.85, suggesting large differences in quality across products, even though they share the same functionality. Indeed, the average number of available options (N) included within the same Product Type is quite large, ranging between 175 to 7,322. In contrast, when products are grouped by Product ID, i.e., products that share the same functionality and exactly the same attributes, price dispersion is substantially lower, with average CV ranging 0.06-0.07. However, the average number of products per reference group reduces dramatically, which generate numerous singleton products for which there is not a set of reference products to compare with.

Next, when reference groups are constructed by matching on the same Product Type and same identifiable attributes, we observe average levels of price dispersion are comparable to the ones observed when products are grouped by Product ID, yet the average number of options per reference group is larger, enriching the reference group size. This suggests that identifiable product attributes are effective in capturing most of the price variation within Product ID, capturing well differences in quality due to vertical differentiation. Finally, when excluding brand as a matching criteria, the average number of products included per group increases, yet the price dispersion also increases, with average coefficient of variations in the range of 0.10 to 0.17, increasing the risk of grouping products with similar functionality but different quality. Given this, *ChileCompra* conservatively opted for the third strategy and built reference groups based on products from the same Product Type and with same identifiable attributes, including brand, regardless of whether the products share the same Product ID. This approach ensured that products where brand is associated with quality would be adequately controlled in the price comparison used to measure overprice.

Table A.4 shows, for each FA, the fraction of the price variation within Product Type that is explained by the reference groups when these are built based on the third grouping strategy, i.e., group products from the same Product Type (not necessary same ID), but that share exactly the same attributes, including brand. Overall, for the large majority of the Product Types included in the analysis, reference groups explain more than 95% of price variation within Product Type, suggesting grouping products using the attributes identified from the open-text description was effective at neutralizing differences in quality across products of the same reference group.

Table A.4: Reference Groups' explanatory power of Price Variation within-Product Type

FA	p5	p25	p50	p75	p95
Office Supplies	0.981	0.994	0.997	0.999	1
Computers	0.966	0.969	0.974	0.981	0.986
Food	0.419	0.810	0.942	0.986	0.998

Notes: Quantiles of the distribution of R^2 when regressing posted price against reference group dummies within product type.

A.4. Calculating Overprice: Example

A buyer is interested in purchasing a jar of instant coffee, which is sold through the *Food* Framework Agreement. Appendix Figure A.1 shows the webpage after filtering "Instant Coffee". Each product listing includes a brief description and a per-unit price. Reference groups are determined by matching products across 3 attributes, including Brand, Weight, and If Decaffeinated. The red colored boxes exemplify products within the same reference group, in this case Not Decaffeinated Nescafé Instant Coffee of 170 gr. Note that the grouping involves some subjective criteria. For instance, "Traditional" Not Decaffeinated Nescafé coffee products are considered in a different reference group than "Fine Selection" Not Decaffeinated Nescafé coffee products, the latter being considered of higher quality. Secondly, the reference group is composed solely of Nescafé products, ensuring that it captures any perceived differences in quality that may be attributed to brand recognition. Note that this criteria (i.e., do not group products of different brands) applies to all reference groups of any FA, meaning that our estimates of overprice control for brand-premium.

The platform offers the buyer the possibility of comparing unit prices across products. Figure A.2 illustrates an example for the four products in the reference group "Not Decaffeinated Nescafé Instant Coffee of 170 gr". All of them have exactly the same attributes (Brand, Weight, If Decaffeinated), yet they differ in terms of physical designs and structures. Indeed, coffee jars differ in aspects such as size, shape, color, lid type, labeling, and material used. Additionally, certain products may be available for sale only when a minimum quantity is bought. Table A.5 details the standardized attributes of each product in the reference group.

Figure A.1: Example of Instant Coffee products displayed in the online catalogue. The image shows a screenshot of the Instant Coffee category in the Food FA. Attributes of products marked in red have all the same values and thus belong to the same reference group.

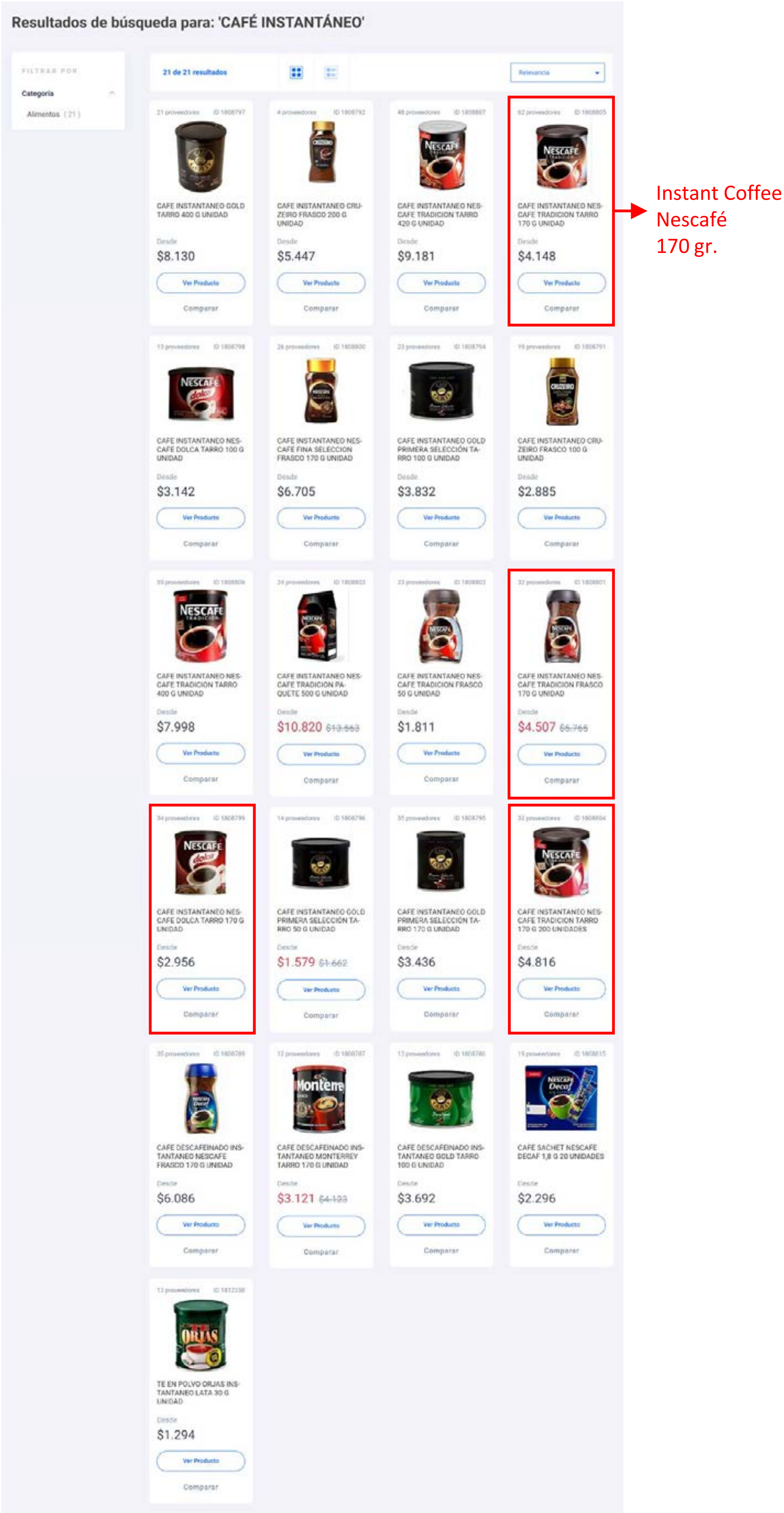


Figure A.2: Example of price comparison functionality in the online platform.

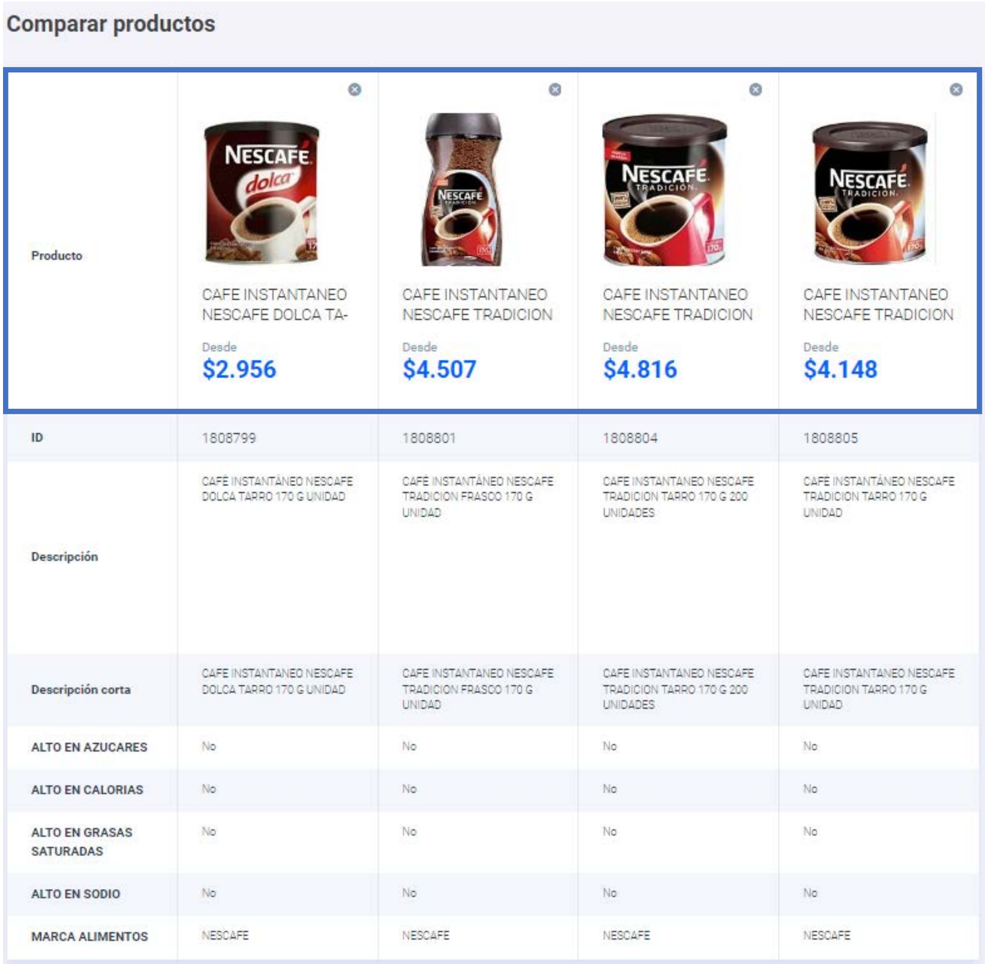


Table A.5: Example of reference group for Instant Coffee in the Food FA

id	Type	Brand	Weight	Units	Decaffeinated	Unit Price
1808799	Instant coffee	NESCAFÉ	170 gr.	1	Not decaffeinated	$2,956
1808801	Instant coffee	NESCAFÉ	170 gr.	1	Not decaffeinated	$4,507
1808804	Instant coffee	NESCAFÉ	170 gr.	200	Not decaffeinated	$4,816
1808805	Instant coffee	NESCAFÉ	170 gr.	1	Not decaffeinated	$4,148

The first item, ID 1808799, has the lowest unit price, equal to $2,956 CLP. This can be used as the *reference price* to measure the overprice of transactions involving any of these 4 products. Suppose the user buys the second item, ID 1808801, sold at $4,507 CLP. Compared to the reference price, the buyer incurred in an overprice of $1,551 CLP or 1,551/2,956=52.46% from not choosing the lowest price product in this reference set[28].

[28]We may think on an e-procurement platform that allows the buyer to choose the product group (e.g., "Not Decaffeinated Nescafé Instant Coffee of 170 gr") but prevent her from choosing the specific product to be purchased (1808799, 1808801, 1808804, or 1808805), a decision that is ultimately made by an algorithm that warrants choosing the product with lowest price. This is out of the scope of the actual technology, yet it is one of the potential innovation that *ChileCompra* is evaluating to introduce in the near future.

B. Appendix Tables and Figures

Figure B.1: Buyer-level Report

DIPRES

CSP

Monthly Overspending Report

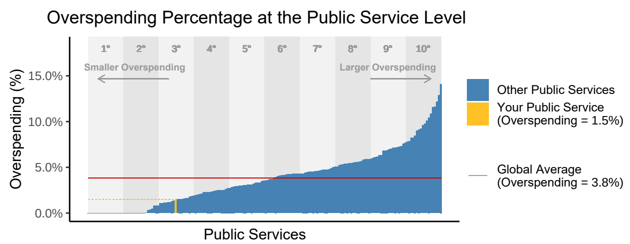
Public Service's name

October 2020

Dear **Recipient**,

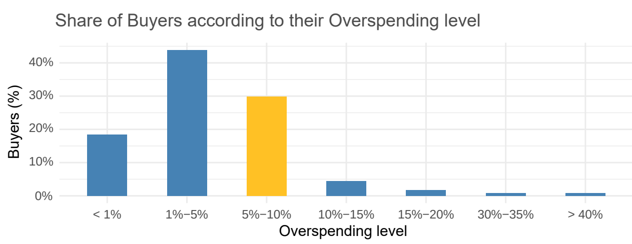
This report presents relevant information regarding the state of public spending executed by the Public Service in which you are employed, covering the period September 2019 - September 2020. This report comprises an analysis of aggregated purchases made by your Public Service, as well as a detailed breakdown of your purchases within the Public Market platform. We would like to inform you that **your average Overspending as a buyer is 7%. You are among the top 25% of buyers in your Service with the highest Overspending**. You are welcome to direct any inquiries regarding this report to gastoeficiente@dipres.gob.cl.

Remember that planning the expenses you make for your Service allows you to execute more efficient purchases, potentially leading to significant savings. For this purpose, ChileCompra's users have access to various procurement mechanisms such as Framework Agreement, Compra Ágil, or Procurement Auctions, which enable more strategic purchasing compared to other alternatives like Direct Purchases.



This chart displays the Overspending associated to each Public Service of the State, arranged in ascending order from the lowest to the highest Overspending (see definition in Appendix B). The relative position of your Public Service is indicated in **yellow**[1].

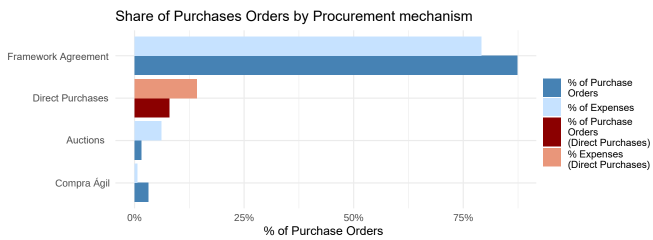
The average Overspending in your Public Service is equal to 1.5%, which positions it among the 50 Services with the lowest Overspending out of a total of 184 Services (excluding Autonomous Services and Subsecretariats). As an illustrative example, considering purchases made exclusively through Framework Agreement the Overspending in your Public Service amounts to approximately $200 million pesos during this period.



This chart displays the share of buyers in your Public Service according to their level of Overspending (%). For instance, 18% of the buyers within the Public Service have a level of Overspending below 1%.

The **yellow** bar indicates the range where you are positioned relative to the rest of buyers in your Service. Your average Overspending as a buyer is 7%.

On average, the buyers of your Service have incurred an Overspending level of 5.1%.



This chart illustrates the distribution of your purchases by procurement mechanism.

You executed the 87.3% of your purchases through Framework Agreement, which corresponds to 79.2% of the total expenditure you have incurred.

Remember to prioritize procurement procedures such as Framework Agreement, Auctions, and Compra Ágil, as they facilitate better expenditure planning and obtaining better prices.

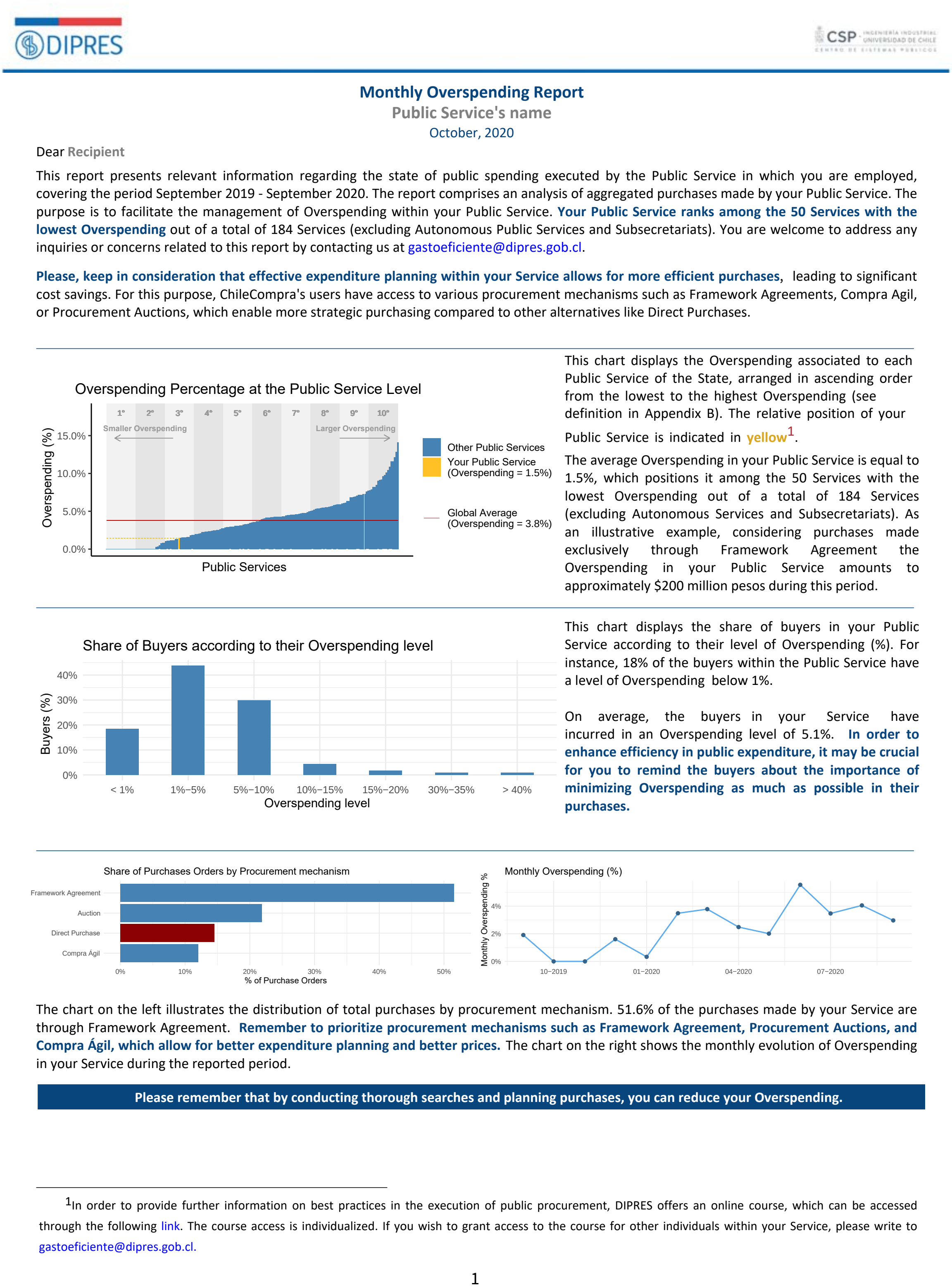
Please remember that by conducting thorough searches and planning purchases, you can reduce your Overspending.

[1] In order to provide further information on best practices in the execution of public procurements, DIPRES offers an online course, which can be accessed through the following link. The course access is individualized. If you wish to grant access to the course for other individuals within your Service, please write to gastoeficiente@dipres.gob.cl.

1

Note: This is the Buyer-level report. It shows the information displayed in the monthly reports sent to buyers within Public and Private Treatment units.

Figure B.2: Manager-level Report



DIPRES

CSP · Ingeniería Industrial Universidad de Chile · Centro de Sistemas Públicos

Monthly Overspending Report

Public Service's name

October, 2020

Dear Recipient

This report presents relevant information regarding the state of public spending executed by the Public Service in which you are employed, covering the period September 2019 - September 2020. The report comprises an analysis of aggregated purchases made by your Public Service. The purpose is to facilitate the management of Overspending within your Public Service. **Your Public Service ranks among the 50 Services with the lowest Overspending** out of a total of 184 Services (excluding Autonomous Public Services and Subsecretariats). You are welcome to address any inquiries or concerns related to this report by contacting us at gastoeficiente@dipres.gob.cl.

Please, keep in consideration that effective expenditure planning within your Service allows for more efficient purchases, leading to significant cost savings. For this purpose, ChileCompra's users have access to various procurement mechanisms such as Framework Agreements, Compra Agil, or Procurement Auctions, which enable more strategic purchasing compared to other alternatives like Direct Purchases.

This chart displays the Overspending associated to each Public Service of the State, arranged in ascending order from the lowest to the highest Overspending (see definition in Appendix B). The relative position of your Public Service is indicated in yellow[1].

The average Overspending in your Public Service is equal to 1.5%, which positions it among the 50 Services with the lowest Overspending out of a total of 184 Services (excluding Autonomous Public Services and Subsecretariats). As an illustrative example, considering purchases made exclusively through Framework Agreement the Overspending in your Public Service amounts to approximately $200 million pesos during this period.

This chart displays the share of buyers in your Public Service according to their level of Overspending (%). For instance, 18% of the buyers within the Public Service have a level of Overspending below 1%.

On average, the buyers in your Service have incurred in an Overspending level of 5.1%. **In order to enhance efficiency in public expenditure, it may be crucial for you to remind the buyers about the importance of minimizing Overspending as much as possible in their purchases.**

The chart on the left illustrates the distribution of total purchases by procurement mechanism. 51.6% of the purchases made by your Service are through Framework Agreement. **Remember to prioritize procurement mechanisms such as Framework Agreement, Procurement Auctions, and Compra Ágil, which allow for better expenditure planning and better prices.** The chart on the right shows the monthly evolution of Overspending in your Service during the reported period.

Please remember that by conducting thorough searches and planning purchases, you can reduce your Overspending.

[1] In order to provide further information on best practices in the execution of public procurement, DIPRES offers an online course, which can be accessed through the following link. The course access is individualized. If you wish to grant access to the course for other individuals within your Service, please write to gastoeficiente@dipres.gob.cl.

1

Note: This is the Manager-level report. It shows the information displayed in the monthly reports sent to managers of Public and Private Treatment units.

Figure B.3: Backpage of Buyer-level and Manager-level Reports

Appendix B

Computation of Overspending

The Overspending is an efficiency measure for purchases of standardized products, i.e. products with comparable attributes to other products in the market. The Overspending associated with the purchase of a product measures how much could have been saved if the purchase had been made at a reference price. The reference price corresponds to the lowest price among the set of comparable products that were available in the purchasing region at the time of the purchase. For each purchase executed by a buyer, there exists an associated Overspending Amount and an associated Overspending Percentage.

Consider the purchase of a product on a specific date. At the time of the purchase, there is a number of comparable products that meet the buyer's needs, whose availability will depend on the product's region of origin. As mentioned above, the reference price corresponds to the lowest price among those comparable products available in the purchasing region. The Overspending Amount for this purchase will then be calculated as follows:

$$\textit{Overspending Amount} = (\textit{Purchase Price} - \textit{Reference Price}) \times \textit{Purchase Quantity}$$

A buyer may execute multiple purchases within a specific period, and each purchase will have a different Overspending Amount. To compute the Overspending Percentage of all the purchases made by a buyer, we sum the Overspending Amount associated with each purchase and divide the total amount by the total expenditure incurred in those purchases.

Example: Let's assume that a user makes two purchases in a specific period. The first purchase is a ream of printer paper priced at 2100 CLP, with a reference price of 1800 CLP. The associated Overspending Amount for that purchase will be $300 \times 1 = 300$ CLP. The second purchase consists of 50 disposable face masks, priced at 190 CLP each, with a reference price of 130 CLP. The associated Overspending Amount for that purchase will be $60 \times 50 = 3000$ CLP. Thus, the Overspending Percentage associated to the two purchases executed by the buyer is computed as follows:

$$\textit{Overspending \%} = \left(100 \times \frac{300 + 3000}{1800 + 130 \times 50}\right)\% = 39.7\%$$

Thus, if we wish to compute the Overspending Percentage at the Public Service level, we sum the Overspending Amount associated with all purchases made by each buyer of the Public Service, and divide it by the total expenditure associated with those purchases. Please notice that not every purchase executed by the Public Service is associated to a standardized product. Therefore, the results presented in this report may not comprise all the purchases executed by the Public Service during the analyzed period.

Recommendations

- Remember that you can use Compra Ágil as an alternative to Direct Purchases. In this manner, you will be able to access a wider range of products and sellers, leading to lower prices.
- When executing a purchase through Compra Ágil or Direct Purchases, clearly specify the product you want to procure. We recommend searching for the product in other catalogs and identifying its relevant characteristics, so you can specify them in your request. This way, you will ensure receiving a broad set of products meeting your purchasing needs.
- If you want to reduce your Overspending, consider planning your purchases and grouping them into a single purchase order using the electronic catalog of products for Framework Agreements, which might provide lower unit prices.

2

Note: This is the backpage included in both buyer-level and manager-level reports. It informs buyers and managers on how *ChileCompra* calculates Overspending as well as their recommendations to make efficient purchases.

Figure B.4: Ranking of Buyers by Overspending, received only by Managers in Public Treatment

Appendix A

Please find below the list of active buyers of your service. The list indicates each buyer's Overspending level and the buyer's share of expenditure within the Public Service associated to the analyzed purchases. The buyers are organized in descending order according to their Overspending level. This arrangement aims to simplify the identification of individuals in need of more immediate adjustments regarding their purchasing behavior.

Name	Email	Overspending (%)	Exp. (%)
		> 40%	< 0.1%
		30%-35%	< 0.1%
		15%-20%	0.1%
		15%-20%	< 0.1%
		10%-15%	0.1%
		10%-15%	0.5%
		10%-15%	< 0.1%
		10%-15%	0.3%
		10%-15%	0.6%
		5%-10%	1.1%
		5%-10%	0.4%
		5%-10%	0.1%
		5%-10%	0.2%
		5%-10%	0.3%
		5%-10%	0.7%
		5%-10%	0.4%
		5%-10%	0.2%
		5%-10%	0.9%
		5%-10%	0.2%
		5%-10%	0.2%
		5%-10%	0.5%
		5%-10%	2.9%
		5%-10%	0.4%
		5%-10%	0.4%
		5%-10%	< 0.1%
		5%-10%	0.3%
		5%-10%	0.3%
		5%-10%	< 0.1%
		5%-10%	0.5%
		5%-10%	0.5%
		5%-10%	0.1%
		5%-10%	0.1%
		5%-10%	< 0.1%
		5%-10%	< 0.1%
		5%-10%	0.1%
		5%-10%	< 0.1%
		5%-10%	0.2%
		5%-10%	0.1%
		5%-10%	0.4%
		5%-10%	3.9%
		5%-10%	0.2%
		5%-10%	0.2%
		5%-10%	0.5%
		1%-5%	3.6%
		1%-5%	0.4%
		1%-5%	0.6%
		1%-5%	1.6%
		1%-5%	7.1%
		1%-5%	3.8%
		1%-5%	3.5%
		1%-5%	0.1%
		1%-5%	0.5%

2

Note: The list contains the names and emails of all buyers in the procurement units and the range of their overspending, ranked by largest to lowest, as well as the share of expenditures incurred by each buyer. This information is displayed in an additional backpage of monthly reports, but it is included only for managers in units assigned to the Public treatment group.

Table B.1: Experimental Sample

Ministry	No. of Service Units
Ministry of Agriculture	5
Ministry of Science and Technology	1
Ministry of Defense	14
Ministry of Social Development and Family	5
Ministry of Economics, Development, and Tourism	12
Ministry of Education	19
Ministry of Energy	3
Ministry of Finance	8
Ministry of Justice and Human Rights	5
Ministry of Women and Gender Equality	1
Ministry of Culture and Arts	1
Ministry of Environmental Protection	2
Ministry of Mining	2
Ministry of Public Infrastructure	14
Ministry of International Relationships	4
Ministry of Health	36
Ministry of Transport and Telecommunications	1
Ministry of Housing and Urbanism	17
Ministry of Sport	1
Ministry of Interior and Public Safety	23
Ministry of Labor and Social Security	9
Ministry of General Secretariat of Government	1
Total No. of Experimental Service Units	184

Notes: List of Ministries and number of selected Service Units per Ministry to participate in the experiment.

Table B.2: Baseline Balance Test

	Mean Control (1)	Mean Public (2)	Mean Private (3)	Control vs Public (4)	Control vs Private (5)	Public vs Private (6)
Panel A. Baseline Procurement Data						
No. Experimental Buyers per Procurement Unit	13.943	14.453	15.130	-0.091	-3.611	3.520
				(6.638)	(6.674)	(7.786)
No. Purchase Orders (P.O.) per Procurement Unit in FA	92.132	72.755	87.389	-15.876	-30.031	14.434
				(33.636)	(32.916)	(39.161)
Procurement Unit Avge. Overprice	0.101	0.098	0.091	-0.011	-0.020	0.010
				(0.015)	(0.013)	(0.014)
Panel B. Baseline Survey Data						
Misalignment Index in Procurement Unit (%)	0.740	0.747	0.721	0.012	-0.016	0.028
				(0.021)	(0.021)	(0.025)
Manager's Perceived Monitoring from DIPRES (1-7)	3.982	3.827	4.055	-0.178	0.010	-0.188
				(0.348)	(0.380)	(0.370)
Avge. Buyers' Perceived Monitoring from HP in Procurement Unit (1-7)	5.648	5.816	5.786	0.188	0.128	0.060
				(0.145)	(0.150)	(0.135)

Notes: Experimental sample of procurement units (N=184). Columns (1)-(3) report the baseline mean for each treatment arm. Procurement Data (Panel A) comprises the baseline period (February 2020 to May 2020). Baseline Survey Data (Panel B) was conducted between February and April 2020. Columns (4)-(6) report OLS estimates of the mean differences across groups and its associated standard errors (in parenthesis), controlling for strata fixed effects. We have too many zeros for the No. of Purchase Orders (P.O.), hence for that case we report marginal effects of a Poisson regression model. ***p<0.01, **p<0.05, * p<0.10.

Figure B.5: Baseline Number of Purchase Orders (P.O.) between February 2020 and May 2020

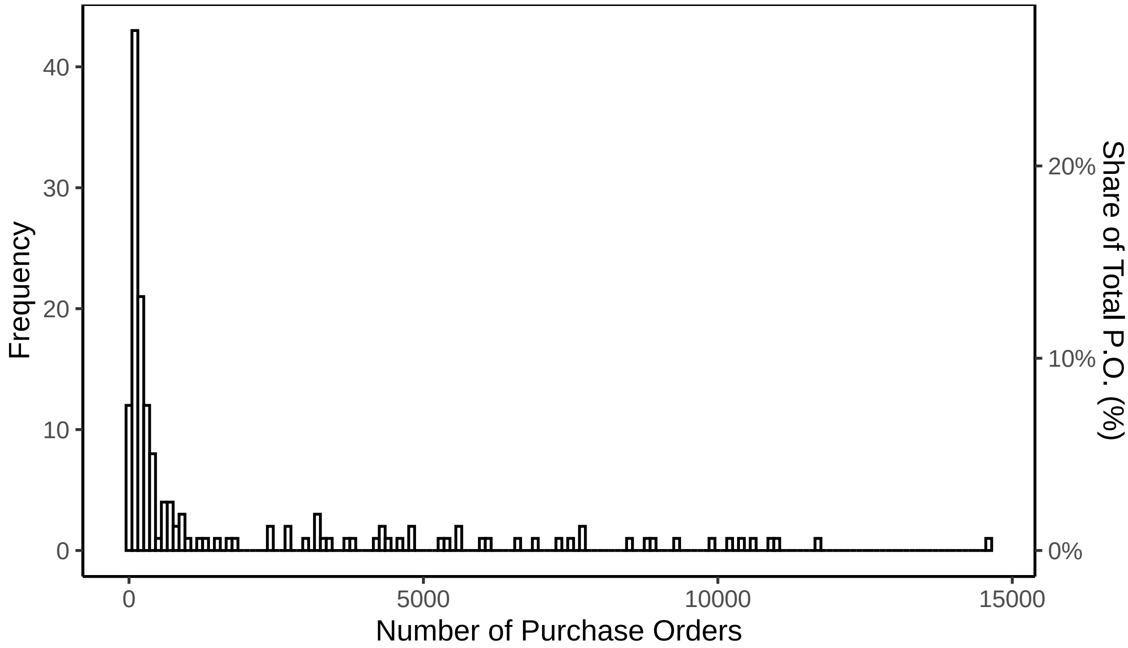


Table B.3: Extensive and Intensive Margin Effects of being exposed to Training Video.

	Intensive Margin (FA only)			Extensive Margin
	Overprice	Log(Q)	#P.O.	% P.O. in FA
	OLS	OLS	Poisson	OLS
	(1)	(2)	(3)	(4)
Training Video	-0.001	-0.053	-0.017	-0.017
	(0.004)	(0.096)	(0.500)	(0.014)
No. Observations	24,340	24,340	1,128	115,583
No. Buyers	1,128	1,128	1,128	4,699
Control Mean	0.103	3.448	4.840	0.360

Notes: This table shows the intention-to-treat effects of being exposed to the Training Video. Both Public and Private assigned-to-treatment groups were exposed to the very same Training Video, thus we report the training effects through a single dummy that equals one if the purchase was made in an assigned-to-treatment unit (either Public or Private) and zero otherwise. The Training Video effects are estimated for the two-month post-training-pre-reports period (July to August 2020). Intensive margin outcomes all refer to purchases made through the FA mechanism, which naturally limit the number of buyers and procurement units under consideration. Regressions in models (1) and (2) are at the unit-transaction level, and consider only purchases made through Food FA, Office Supplies FA, and Computers FA, and for which overprice can be calculated (i.e., we can find a reference group of products). Following equation 4, overprice of a unit transaction is defined as the relative difference between the per-unit price paid and the lowest price in the daily price distribution for products from the reference group. Log(Q) stands for the amount of items purchased per unit transaction (in logs). In buyer-level model (3), estimates are derived through a Poisson regression model, and #P.O. counts the number of Purchase Orders made by a buyer, which may contain more than one product (wholesales). Model regressions (1)-(3) control for strata fixed effects and the average outcome in procurement unit during the pre-intervention period. Since models (1) and (2) are at the unit-transaction level, we also control for product type fixed effects and week fixed effects. The extensive margin outcome (column 4) refers to the share of Purchase Orders made through FA (Framework Agreement) instead of alternative procurement mechanisms like *Procurement Auctions* or *Direct Purchases*. Outcome includes the universe of P.O. made by buyers in either of the three purchasing mechanisms, and the regression controls for strata fixed effects and the average outcome in the procurement unit during the pre-intervention period (Feb. - May 2020). In all, standard errors clustered at the procurement unit level are shown in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table B.4: Extensive and Intensive Margin Effects on Placebo Buyers

	Intensive Margin (FA only)			Extensive Margin
	Overprice	Log(Q)	#P.O.	% FA
	OLS	OLS	Poisson	OLS
	(1)	(2)	(3)	(4)
Public Reports	0.018	0.015	2.383	0.023
	(0.018)	(0.126)	(3.261)	(0.022)
Private Reports	0.003	0.156	2.694	0.014
	(0.007)	(0.123)	(2.049)	(0.023)
No. Observations	61,928	61,928	1,285	3,124
No. Buyers	1,285	1,285	1,285	3,124
Control Mean	0.103	3.408	9.220	0.419
p-val. Public=Private	0.351	0.319	0.899	0.709

Notes: This table shows the intention-to-treat effects for placebo buyers, i.e., it compares purchases made by placebo buyers in Public and Private treatment groups with purchases made by control group buyers, considering the full post-treatment period of analysis (July 2020 to January 2021). Every month, placebo buyers received a message indicating that her overspending is being monitored, but do not provide any type of information about individual performance. The extensive margin outcome refers to the share of purchases per buyer made through FA (Framework Agreement) instead of alternative procurement mechanisms like *Auctions* or *Direct Purchases*. Intensive margin outcomes all refer to purchases made through the FA mechanism, which naturally limit the number of buyers and procurement units under consideration. Regressions in models (1) and (2) are at the unit-transaction level, and consider only purchases made through Food FA, Office Supplies FA, and Computers FA, and for which overprice can be calculated (i.e., we can find a reference group of products). Following equation 4, overprice of a unit transaction is defined as the relative difference between the per-unit price paid and the lowest price in the daily price distribution for products from the reference group. Log(Q) stands for the amount of items purchased per unit transaction (in logs). In buyer-level model (3), estimates are derived through a Poisson regression model, and #P.O. counts the number of Purchase Orders made by a buyer, which may contain more than one product (wholesales). Model regressions (1)-(3) control for strata fixed effects and the average outcome in procurement unit during the pre-intervention period. Since models (1) and (2) are at the unit-transaction level, we also control for product type fixed effects and week fixed effects. The extensive margin outcome (column 4) refers to the share of Purchase Orders made through FA (Framework Agreement) instead of alternative procurement mechanisms like *Procurement Auctions* or *Direct Purchases*. Outcome includes the universe of P.O. made by buyers in either of the three purchasing mechanisms, and the regression controls for strata fixed effects and the average outcome in the procurement unit during the pre-intervention period (Feb. - May 2020). In all, standard errors clustered at the procurement unit level are shown in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Bottom row shows the p-value corresponding to the null hypotheses of no differences between Public and Private treatment effects.

Table B.5: Heterogeneous Effects by Number of Attributes per Product

	Full Sample	Identical Products in Reference Group
	Overprice	Overprice
	OLS	OLS
	(1)	(2)
Public Reports	-0.017**	-0.030***
	(0.007)	(0.008)
Private Reports	0.000	-0.006
	(0.006)	(0.007)
# Attributes	-0.012***	-0.031**
	(0.002)	(0.015)
# Attributes × (Public Reports)	-0.001	0.044
	(0.003)	(0.032)
# Attributes × (Private Reports)	-0.003	0.006
	(0.002)	(0.025)
No. Observations	93,792	25,254
No. Buyers	2,076	1,174
Control Mean	0.103	0.089

Notes: This table shows the heterogeneous effects of being exposed to the Public and Private performance reports on overprice, by number of attributes per product, considering the full post-treatment period of analysis (July 2020 to January 2021). Following equation 4, overprice of a unit transaction is defined as the relative difference between the per-unit price paid and the lowest price in the daily price distribution for products from the reference group. All regressions are at the unit-transaction level, and consider only purchases made through Food FA, Office Supplies FA, and Computers FA, and for which overprice can be calculated (i.e., we can find a reference group of products). All regressions control for strata fixed effects, the average outcome in procurement unit during the pre-intervention period, product type fixed effects and week fixed effects. Model in column 1 uses the full sample of purchases. Model 2 restricts the sample to purchases of products whose reference group is composed by identical products (same SKU). Standard errors clustered at the procurement unit level are shown in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table B.6: Baseline Survey Questions used to construct Belief Misalignment Index (MA)

Question
1. Suppose that the unit in which you work fails to execute 100% (or a percentage close to 100%) of the budget allocated for this year. Regardless of the economic projections that DIPRES manages for the following year, what would you expect to happen with the budget approved for your unit the following year? *Alternatives: 1) I would expect the unit's budget to decrease from this year's budget, 2) I would expect the unit's budget to remain the same as this year's budget, 3) I would expect the unit's budget to increase from this year's budget.*
2. This question is about what you think the head of purchases of the unit in which you work believes. What do you think he or she would answer about what would happen to next year unit's budget if 100% of the budget (or close to it) allocated for this year is not executed? *Alternatives: 1) I think he/she believes the unit's budget would decrease next year, 2) I think he/she believes the unit's budget would stay the same next year, 3) I think he/she believes the unit's budget would increase next year.*
3. Suppose that the unit in which you work does not manage to execute 100% (or a percentage close to 100%) of the budget allocated for this year, but manages to demonstrate that it generated considerable savings in the use of resources. Regardless of the economic projections that DIPRES manages for the following year, what would you expect to happen with the unit's budget the following year? *Alternatives: 1) If savings are generated, I would expect the unit's budget to decrease from this year's budget, 2) If savings are generated, I would expect the unit's budget to remain the same as this year's budget, 3) If savings are generated, I would expect the unit's budget to increase from this year's budget.*
4. This question is about what you think the head of purchases of the unit in which you work believes. What do you think he/she would answer about what would happen to next year unit's budget if the unit does not manage to execute 100% of the budget (or a close percentage) allocated for this year, but manages to demonstrate that it generated considerable savings in the use of resources? *Alternatives: 1) If savings are generated, I think he or she believes the unit's budget would decrease next year, 2) If savings are generated, I think he or she believes the unit's budget would remain the same next year, 3) If savings are generated, I think he or she believes the unit's budget would increase next year.*
5. Next year's budget is largely determined by how much of this year unit's budget is executed. *Alternatives: 1) Strongly disagree, 2) Disagree, 3) Neither agree nor disagree, 4: Agree, 5: Strongly agree*
6. If savings are generated, the unit's budget will be reduced next year. *Alternatives: 1) Strongly disagree, 2) Disagree, 3) Neither agree nor disagree, 4: Agree, 5: Strongly agree*
7. Executing the budget allocated to your unit involves making spending decisions and savings strategies. Some believe that the level of public savings is a relevant variable for DIPRES in the process of negotiating the budget for the following year. Others, on the other hand, believe that DIPRES has little interest in the level of public savings generated by each unit and does not consider this variable in the budget negotiation process. From your perspective, what is the level of relevance that DIPRES gives to the public savings generated by your unit when making budget allocation decisions? Being 1 "Not relevant" and 7 "Extremely relevant"
8. One of the main mechanisms through which the budget is executed is the Public Procurement System ("Auctions", "Framework Agreements" or "Direct Purchases"). Some believe that DIPRES monitors and supervises on a recurrent basis the purchases made by each unit. Others, on the other hand, believe that DIPRES does NOT monitor or supervise the purchases made by each unit. From 1 to 7, what do you think is the level of monitoring exercised by DIPRES on public purchases made by your unit? Being 1 "No monitoring" and 7 "High level of monitoring".
9. Utilizing the maximum unit's budget is a pressure. *Alternatives: 1) Strongly disagree, 2) Disagree, 3) Neither agree nor disagree, 4: Agree, 5: Strongly agree*
10. Sometimes purchases are made at high prices to comply with budget execution. *Alternatives: 1) Strongly disagree, 2) Disagree, 3) Neither agree nor disagree, 4: Agree, 5: Strongly agree*

C. Appendix: Theoretical Results

Manager's optimization problem can be written as follows:

$$\max_{e>0} \pi(e, \delta, s) = \frac{1 + \mu_J^2}{2} - (1 - P(e, s))\frac{\delta^2}{2} - c(e, \delta) \tag{C.1}$$

Let $e^* = e^*(\delta, s)$ be an interior solution of (C.1). From Envelope Theorem we have:

1. $\frac{\partial}{\partial \delta} v(\delta, s) = -\delta(1 - P(e^*, s)) - \frac{\partial c(e,\delta)}{\partial \delta}$. Then, if $\delta = \mu_J - \mu_I > 0$, then $\frac{\partial}{\partial \delta} v(\delta, s) < 0$. However, when $\delta = \mu_J - \mu_I < 0$, then the sign of $\frac{\partial}{\partial \delta} v(\delta, s)$ depends on the magnitude of $\frac{\partial c(e,\delta)}{\partial \delta}$.
2. $\frac{\partial}{\partial s} v(\delta, s) = \frac{\partial P(e,s)}{\delta s} \frac{\delta^2}{2} > 0$

Regarding the monotonicity of optimal efforts, it follows from Topkis's Theorem that:

3. $\frac{\partial e^*(\delta,s)}{\partial \delta} > 0$ if and only if $\frac{\partial^2 \pi(e,s)}{\partial e \partial \delta} > 0$. Then,

$$\frac{\partial^2 \pi(e, s)}{\partial e \partial \delta} = \frac{\partial P(e, s)}{\partial e}\delta - \frac{\partial^2 c(e, \delta)}{\partial e \partial \delta}$$

Thus, if $\delta < 0$, then $\frac{\partial e^*(\delta,s)}{\partial \delta} < 0$. Hence, the larger $|\delta|$, the larger is the optimal effort $e^*(\delta, s)$. In turn, if $\delta > 0$, it follows the sign of $\frac{\partial e^*(\delta,s)}{\partial \delta}$ would depend on the magnitud of $\frac{\partial^2 c(e,\delta)}{\partial e \partial \delta}$.

4. $\frac{\partial e^*(\delta,s)}{\partial s} > 0$ if and only if $\frac{\partial^2 \pi(e,s)}{\partial e \partial s} > 0$. Since

$$\frac{\partial^2 \pi(e, s)}{\partial e \partial s} = \frac{\partial^2 P(e, s)}{\partial e \partial s}\frac{\delta^2}{2} > 0,$$

it follows that $\frac{\partial e^*(\delta,s)}{\partial s} > 0$.

Further, the optimal effort e^* satisfies:

$$\frac{\partial v(\delta, s)}{\partial \delta} = -\delta(1 - P(e^*, s)) - \frac{\partial c(e, \delta)}{\partial \delta},$$

which implies that

$$\frac{\partial^2 v(\delta, s)}{\partial s \partial \delta} = \delta \left(\frac{\partial P(e^*, s)}{\partial e}\frac{\partial e^*}{\partial s} + \frac{\partial P(e^*, s)}{\partial s} \right) - \frac{\partial^2 c(e, \delta)}{\partial e \partial \delta}\frac{\partial e^*}{\partial s}. \tag{C.2}$$

Then, if $\delta < 0$, $\frac{\partial^2 v(\delta,s)}{\partial s \partial \delta} < 0$, which implies that increasing $|\delta|$ leads to larger optimal payoffs. However, if $\delta > 0$ this result does not necessarily hold. For instance, if increasing δ generates large increases in the marginal cost of persuasion effort ($\frac{\partial^2 c(e,\delta)}{\partial e \partial \delta}$ is large enough), then increasing $|\delta|$ will decrease the marginal payoff from improving s. More generally, when $\delta < 0$, the effect of improving s is smaller than when $\delta > 0$.

C.1. Monitoring

Given an effort e and information s, let $m(e, s)$ the monitoring exerted by the Manager. We understand monitoring as a systematic procedure of obtaining information, which is a combined function of effort and information. Thus, let us assume that:

1. Monitoring is increasing in s, i.e. $\frac{\partial m(e,s)}{\partial s} > 0$: the more information, the better the monitoring the manager can achieve.
2. Monitoring is increasing in e, i.e. $\frac{\partial m(e,s)}{\partial e} > 0$: The persuasion effort exerted by the manager in convincing the agent of implementing the decision the manager wants is reflected in more monitoring.

Then, for an optimal level of persuasion effort $e^*(\delta, s)$, the manager will exert an optimal level of monitoring according to information s, $M(s) = m(e^*, s)$. Thus,

$$\frac{dM(s)}{ds} = \frac{\partial m(e^*, s)}{\partial e}\frac{\partial e^*}{\partial s} + \frac{\partial m(e^*, s)}{\partial s}.$$

Let be $S(m) = M^{-1}(m)$ the information required to implement a level of monitoring m (when efforts are optimal). Thus,

$$\frac{d}{dm}S(m) = \frac{1}{M'(S(m))}.$$

Then,

$$\frac{\partial v(\delta, S(m))}{\partial m} = \frac{\partial v(\delta, S(m))}{\partial s}\frac{d}{dm}S(m).$$

Since $\frac{\partial v(\delta,S(m))}{\partial s}$ is positive, the sign of $\frac{\partial v(\delta,S(m))}{\partial m}$ will depend on the sign of $\frac{d}{dm}S(m)$.

Furthermore,

$$\frac{\partial}{\partial m}\left(\frac{\partial v(\delta, S(m))}{\partial s}\right) = \frac{\partial^2 v(\delta, S(m))}{\partial s^2}\frac{d}{dm}S(m) < 0,$$

where $\frac{\partial^2 v(\delta,S(m))}{\partial s^2} = \frac{\partial^2 P(e,S(m))}{\partial s^2}\frac{\delta^2}{2}$.