In the Shadow of Sunshine Regulation: explaining disclosure biases*†

Thomas Bolognesi,‡ Géraldine Pflieger§

Abstract

Performance reporting in sunshine regulation is subject to disclosure biases, because agents may game the regulation or encounter difficulties in complying. These biases limit the appraisal of the impact of sunshine regulation on performance. We investigate the behavioral causes of such disclosure biases by focusing on transaction costs economizing and opportunism. We provide an original methodology to take into account information asymmetries in principal-agent relationships. We focus on water utilities management in France. Our dataset includes 795 observations covering water utilities and performance indicators characteristics. It allows for comparisons of revealed and observed performance and identifies different types of disclosure biases. Findings indicate that opportunism is a significant motivation for disclosure biases while, unexpectedly, transaction costs are not a direct trigger of disclosure biases.

Keywords: Regulation, Disclosure, Performance management, Public management, Organisation

JEL Codes: L51, L95, D23, D78, K42

*This paper is part of the project InfraGouv funded by the SNSF (Project 169521)

†We gratefully acknowledge comments and suggestions from Christian Bréthaut, Marion Chabrost, Louise Gallagher, Tanya Heikkila, Simon Hug, Zoé Le Squeren, Jérémy Lucchetti, Simon Porcher, Stéphane Saussier, Pascal Sciarini, Luc Tonka and Frédéric Varone. We also thank audiences of the 6th Conference of the Florence School of regulation and of the Political Science departement seminar (Univ. of Geneva). Any errors remain ours.

‡Corresponding author, University of Geneva thomas.bolognesi@unige.ch

§University of Geneva, geraldine.pflieger@unige.ch

1 Introduction

Since the 1970s, the principal-agent problem is a central topic for regulation and public administration scholars and practitioners (Laffont and Martimort, 2009; Laffont and Tirole, 1993; Ross, 1973). It consists of the strategic use of information asymmetries by the agent (the regulatee), knowing that the principal (the regulator) accesses with difficulty information on real performance. To prevent this, regulators benchmark operators through yardstick competition (Chong and Huet, 2009; Shleifer, 1985) or sunshine regulation (Baldwin and Black, 2008; Decker, 2015). Sunshine regulation requires transparency on specific tasks, mostly through the reporting of performance indicators. The regulator aims at pointing out bad practices to motivate operators to deliver high public service quality; hence it is sometimes called "naming and shaming" regulation because agents fear a negative reputational effect. It is a less systematic benchmark than yardstick competition.[1] Consequently, the use of performance indicators expands since the nineties.

Several scholars have examined the link between sunshine regulation increases and performance of the service delivery. As sunshine regulation serves ex-post enforcement of the whole regulation, Baldwin and Black (2008) claim that it provides with dynamic beneficial effects in contributing to set a "smart" and "responsive" regulation. Diverse empirical studies confirm a positive impact of sunshine regulation in the health sector in UK (Bevan and Hamblin, 2009; Bevan and Hood, 2006; Bevan and Wilson, 2013) or in education and health in US and UK (Propper, 2003). A broader picture draws ambiguous impact. For instance, in the Dutch public sector, incentive-oriented performance measurement is negatively correlated to organizational performance while the exploratory use of performance indicators positively impacts on performance (Speklé and Verbeeten, 2014). Heckman et al. (1997) in the US job training system or Witte and Saal (2010) in the Dutch water sectors confirm these findings. To synthesize all empirical evidence, Gerrish (2016) carried-out a meta-analysis of 2188 effects from 49 original studies, which conclude on small average effects of performance management in public organizations.

A mismatch arises between the recent empirical conclusions on one side and the more frequent use of performance indicators in public utilities on the other side. Some methodological limitations in empirical assessment from the results of research on the effects of performance management could explain this mismatch, opening an avenue for new research. Empirical

[1] Yardstick competition is a systematic comparison of services characteristics to predict their costs, ceteris paribus. Then regulator sets prices. Yardstick competition has already been widely studied (Chong and Huet, 2009; Lannier and Porcher, 2014; Romano and Guerrini, 2011).

evidence on sunshine regulation and practices diverge. If sunshine regulation is inefficient why do we observe a multiplication of performance indicators? As sunshine regulation has a lot to do with disclosure processes, attention needs to be paid to methodological limitations in the evaluation of disclosure practice. Previous studies or meta-analysis are not properly designed to identified causation (Leuz and Wysocki, 2016). Besides, data collected reflects revealed performance and not the observed performance of agents. Therefore, studies face the same problem of asymmetric information that regulators face. Disclosure biases, i.e. poor reporting, may cause measurement errors in the dependent variable, i.e. performance of agents. This has already been envisaged (Heckman et al., 1997; Pollitt, 2013; Van Thiel and Leeuw, 2002) and recent evidence found that disclosure biases exist in the French water sector (Brochet et al., 2016; Tsanga-Tabi and Verdon, 2014).

The understanding of disclosure biases is therefore important for assessing the impact of sunshine regulation on performance. To determine if disclosure biases reduce the efficiency of sunshine regulation, we need to know if they come from opportunistic behaviors (strategic use of information asymmetry) or are rooted in the design of sunshine regulation (functionalist view of the coordination). Our research question is: which behavioral mechanisms could explain disclosure biases in sunshine regulation? We use a unique dataset of disclosure quality in the French water sector. The dataset gathers 795 observations that allow us to compare observed and revealed performance of agents in the context of the implementation of sunshine regulation in the French water sector. We observed the performance of agents by accessing data of each water service, and compared it with revealed performance as declared by the agents who completed official surveys. Results indicate that opportunistic behaviors are a trigger of disclosure biases. Regarding transaction costs, we come to unexpected conclusions. The complexity of reporting is not significantly positively correlated with disclosure biases. This counter-intuitive empirical result may mean that agents use transaction costs as rooms for maneuver rather than reducing them.

The paper is organized as follows. Section 2 presents theoretical explanations of the emergence of disclosure biases in sunshine regulation and hypotheses to be tested. Section 3 presents the empirical strategy and the dataset. Section 4.2 describes the empirical model. Section 5 provides results and section 6 discusses them and concludes.

2 Theory and hypotheses: two stories about disclosure biases in sunshine regulation

2.1 Sunshine regulation and performance: empirical evidences limitations

Leuz and Wysocki (2016) point out significant limitations in research focusing on disclosure, and, noticeably, sunshine regulation is a form of disclosure. The main limitations identified comes from the absence of a quasi-random assignment of treatment and unaffected groups, which prevents inferring causality.[2]. Such design, i.e., non-(quasi)experimental, reduces the ability to infer causality from data. With such research design, the causal relationship should be theoretically proved. In that respect, the study of Pawson (2002) proposes a complementary method to tackle this issue. He suggests that the comparative analysis should be dedicated to theoretical purpose, then some results could be extrapolated. This "realist synthesis" is a complementary approach to traditional meta-analysis and narrative review. The second main limitation is impacts of sunshine regulation and disclosure mechanisms are inter-dependent with the institutional environment, which is hard to control. Third main limitation, it is noteworthy that sunshine regulation is used in the context of agency problems but data often come from agents themselves (Heckman et al., 1997; Pollitt, 2013). The collection by academics is likely to suffer from the same information asymmetry as regulators (Brochet et al., 2016). It is therefore difficult to ensure that data reflects the real performance of services.

2.2 Sunshine regulation and the performance paradox: the scope for disclosure biases

Meyer and Gupta (1994); Pollitt (2013); Van Thiel and Leeuw (2002) emphasize the existence of a performance paradox that performance indicators do not reflect actual performance, causing a failure in discriminating bad practices from good ones over the time.

We define disclosure biases as a difference in revealed performance by agents and their observed performance. They could take two forms of behaviours: *(i)* distorting the information, *(ii)* hiding it. The analysis of performance paradox sheds light on three plausible

[2]Witte and Saal (2010) try to manage this issue by comparing utilities before and after the implementation of sunshine regulation

behaviours regarding disclosure biases (Brochet et al., 2016; Pollitt, 2013; Van Thiel and Leeuw, 2002).

First, actors could game regulation to affect performance measurements according to their private interest. Second, they could also decide to cheat (Bevan and Hood, 2006). Gaming ("bending the rule") differs from cheating ("breaking the rule"), the former is mostly an ex-ante strategy while the latter occurs ex-post.

Three, definitional drifts have been identified as a source of performance mismeasurement. Reports may be not consistent with each other and actors could consider definitional drifts as opportunities for reporting in ways that are favorable to them (Bevan and Hamblin, 2009; Christensen et al., 2006). The complexity of reporting may explain biases. The first behaviour, actors gaming, does not enter the scope of our study because such behavior occurs before implementation of regulatory measures. We thus investigate *actors cheating* (2.3) and *definitional drifts* (2.4).[3]

In the two following sub-sections, we present the rationale behind these two behaviors. For each behavior, we develop on the causal mechanisms at work to then formulate hypotheses linking the behavior to disclosure biases.

2.3 The role of opportunism in disclosure biases

Cheating is a classic case of opportunism and may be frequent in sunshine regulation. Performance indicators are tools of regulation enforcement (Baldwin and Black, 2008). The state of nature is non-verifiable, and bargaining power is limited, especially due to public procurement liberalization[4]. Regarding the Laffont and Martimort (2009)'s definition of regulation games, it corresponds to a limited enforcement case . Such configuration favors opportunism in two ways: moral hazard and free-riding.

[3]Symbolic use of information is also plausible cause. The disclosure may be used to show how are functioning services to discriminate modern ones from the others. It contributes to legitimate a particular type of management. As an illustration, the use of performance indicators is a crucial tool of mindset changes in the modernization of the urban water systems in Europe (Bolognesi, 2018; Canneva and Guérin-Schneider, 2011; Renou, 2017). Consequently, actors could choose to resist to the use of performance measurement anticipating future symbolic uses of the disclosed information, e.g., reforming the sector in favor of significant actors and changing managerial practices of public services. In this paper, we do not focus on this issue as it is firstly a question of political capture and game rather than a consequence of the principal-agent dilemma.

[4]Laffont and Martimort (2009) consider that this configuration leads to a hold-up game where the agents have high incentives to under-investigate. The case of the privatization of water utilities in UK provides proofs of this hold-up mechanism (Bakker, 2010; Bolognesi, 2018)

Laffont and Martimort (2009) and Armstrong and Sappington (2006) emphasise the significance of moral hazard in principal-agent relation. Agents might be tempted to adopt an opportunistic behavior to enhance their private welfare because they consider unlikely the Principal realize that their cheating. There is low incentive to comply, as the institutional environment is considered weak. In our case, it could lead to two different disclosure biases, cream-skimming, and over-estimation. Heckman et al. (1997) already underlined the risk of cream-skimming with regulation based on disclosure but found no links between cream-skimming and increase of the performance over time. Agents disclose performance indicators that are good and do not reveal the other indicators. Over-estimation proceeds of the same willingness of inflating performance. Agents communicate results that are higher than reality. In the over-estimation case, we consider only the trade-off between gains from cheating and risks of sanctions.

Free-riding is slightly different from over-estimation albeit both cases result in the disclosure of erroneous information. Free-riding is grounded in institutional inconsistencies that could create incentives to adopt an opportunistic behavior. These inconsistencies make the gains from cheating higher than the benefits and costs of complying with the sunshine regulation.

Assumption 1

Opportunistic behavior leads to disclosure biases. If hiding or reporting false values of a given indicators benefits to the service, disclosure biases are more frequent.

2.4 The impact of transaction costs on disclosure biases

Definitional drifts remind us of the core role of bounded rationality and transaction costs on actors' behavior. They fuel implementation issues and reduce coordination efficiency (Coase, 1960; Simon, 1982; Williamson, 2005). Actors economize transaction costs. They do so by establishing a credible governance structure and institutional arrangement to frame the transaction they are involved in. Then, theory predicts that the sub-transaction is well done (Künneke et al., 2010). However, if they can not create this credible environment, economizing transaction costs may lead to not to fulfill their duty. Disclosure is a non-critical sub-transaction of the core transaction (service delivery), which stresses this risk.

Analysis of policy measures and transaction costs highlight the significant cost of information treatment for private actors (McCann et al., 2005). Authors approximate these private transaction costs related to policy requirements by the work time of employees and the cost of outsourcing the policy compliance tasks (Mettepenningen et al., 2009).[5] Brown and Potoski (2003) demonstrate strong empirical evidence on the impact of transaction costs on government service production decision in the US. In the context of environmental protection, they could represent up to 15% of the total cost of goods and services production (Mettepenningen et al., 2009; Phan et al., 2017). Sunshine regulation, especially in its disclosure phase, is about information treatment. Informational transaction costs may appear and lead actors to report biased data.

The rationale for this behavior links transaction costs with rationality in two ways because of the nature of uncertainty (Dosi and Egidi, 1991). Complex calculation, e.g. indicator with a long and technical formula, produces procedural and substantive uncertainty. Procedural uncertainty is the inability of agents to use information. They struggle to manage with the variables required to assess the performance indicators. As a result, they could disclose erroneous performance. Weber and Mayer (2014) recently highlight that interpretive uncertainty exists and poses coordination problems among firms. Substantive uncertainty ensues from difficulty in gathering information. The more numerous and complex required data are, the more they are costly to assemble. Transaction costs economics states actors tend to minimize transaction costs (Williamson, 2005). In consequence, agents may bias their disclosure by deciding not to report or using proxies in order to minimize the impact of transaction costs on their activity. Complex and unclear calculation methods should favor disclosure biases.

Hypothesis 2

Transaction costs create incentives to report not consolidated data or make reporting infeasible. The more complex calculation of indicators is, the more likely disclosure biases are.

Figure 1 sorts explanations of the two leading causes of disclosure biases considered in the paper, and links them to the expected disclosure bias. These two leading causes correspond to five behaviors. It is noteworthy that the "No information" outcome is different in the case of nonfeasibility and uncertainty. In the former, the service provider could not access the data because they are too costly. He can not disclose his performance. In the latter, the complexity of the calculation trouble the service provider and he finally does not report.

[5]For a general discussion on measuring transaction costs see Sykuta (2010); Wang (2003).

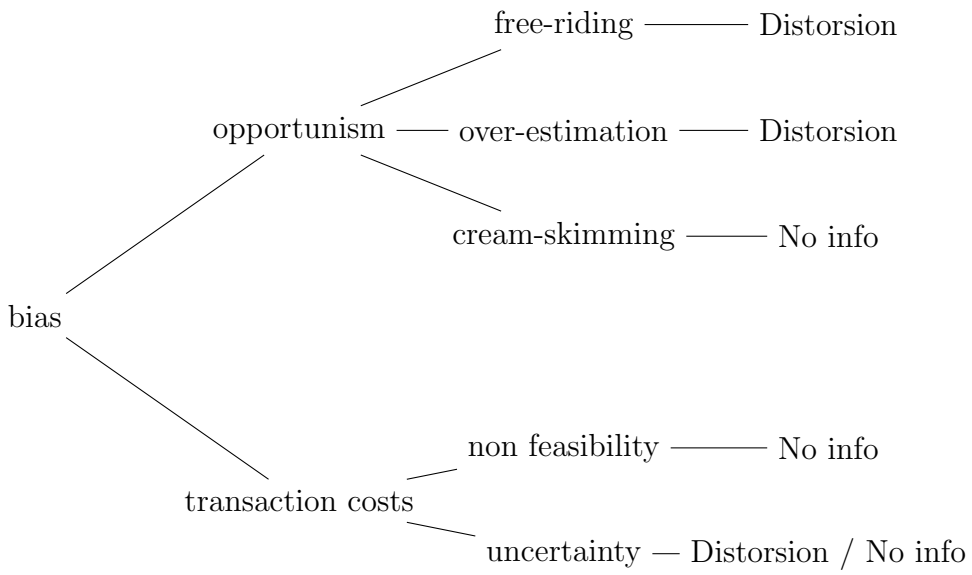

Figure 1: Plausible causes and explanations of disclosure biases under sunshine regulation

3 Methods and Data

3.1 Case selection

To analyze disclosure biases in sunshine regulation, we focus on the case of the water supply sector within the Grenoble metropolitan area (France). Four reasons motivate this focus. First, sunshine regulation plays a crucial role in the water sector (Decker, 2015). The main reasons are the lack of information on the service operation (infrastructure and costs) and the tiny place for market regulation in this sector in comparison to other network industries (Bolognesi, 2018).

Second, in France most of the water utilities regulation is based on contracts. Sunshine regulation complements this central regulatory mechanism by helping contracting parties to adapt (Guérin-Schneider and Nakhla, 2012). 17 performance indicators were implemented in the French water sector in 2007. Table 1 exhibits theirs main characteristics (theme of scope and type), meaning, and name in our study (variable) .[6] The disclosure process is the following. Operators are free to report data to Onema, the regulator,[7] through a dedicated

[6] Canneva and Guérin-Schneider (2011) discuss the process of the reform that lead to this regulatory framework. Especially, they underline the role of performance indicators as a mean to restore the trust of users in operators in consequence of different scandals in the French water sector in the late 1990ies and early 2000ies

[7] Onema was the regulator of the water sector in France until 2017, the Agence Française pour la Bio-

platform, SISPEA. Onema then puts all collected data together and makes it accessible. This dataset has a high number of missing values (Brochet et al., 2016; Chong et al., 2015). Nonetheless, the SISPEA dataset serves for "naming and shaming" purposes (Bolognesi, 2018; Chong et al., 2015). Media use these indicators to highlight water management quality and practices. As an illustration, UFC Que Choisir, a consumer association, has a topic on water and regularly reports on tap water quality and water tariffs.[8] An interactive map allows comparing the quality of tap water in France.[9] In 2013, the association published a report comparing the price of water in each city with more than 60 000 inhabitants.[10]

Third, the Grenoble metropolitan area benefits from favorable conditions such as high water quality, low-cost mobilization and a strong political willingness to modernize water services. These characteristics should lead to underestimating opportunism and symbolic causes of disclosure biases. It is a crucial feature of our research design as we seek to demonstrate, unlike most of the literature, that performance reporting under sunshine regulation may not rigorously represent reality.

Fourth, the French water sector is highly fragmented[11]. The Grenoble metropolitan area gathers many services within the same institutional environment because of its mountainous character. This geographical characteristic constitutes a homogeneous institutional environment with a high number of services which limits exogenous and unobserved divergent impacts. It allows us to control external indirect effects of the institutional environment and makes observations comparable without any sample treatment (Heckman et al., 1997; Leuz and Wysocki, 2016).

---

diversité replaces it. Its main tasks were to monitor water use and water quality as well as to provide local actors with technical support. It is noteworthy that its responsibilities were limited in comparison to Ofwat in the UK.

[8] On the 8^{th} of September 2017, there were 729 results for the request "price water" on the website of UFC Que Choisir (https://www.quechoisir.org/utils/recherche/?keyword=prix+de+l%27eau).

[9] The map is available at https://www.quechoisir.org/carte-interactive-qualite-eau-n21241/.

[10] The report is available at:
https://www.quechoisir.org/enquete-prix-de-l-eau-les-tarifs-des-villes-de-plus-de-60-000\-habitants-n11317.

[11] ONEMA counts 13,806 different water services in France in 2012.

Variable	Theme	Indicator type	SISPEA code	Label
Inhabitants	Customers	descriptive	D101.0	Estimated number of supplied inhabitants
Price	Customers	descriptive	D102.0	Price per m3 incl. Tax. for 120m3
Connection time	Customers	descriptive	D151.0	Maximum connection time
Microbio	Water quality	performance	P101.1	Microbiological compliance rate
Physico-chemical	Water quality	performance	P102.1	Physico-chemical compliance rate
Asset knowledge A	Infrastructure	performance	P103.2A	Asset knowledge and management index (until 2012)
Asset knowledge B	Infrastructure	performance	P103.2B	Asset knowledge and management index
Net efficicency	Infrastructure	performance	P104.3	Network efficiency rate
Unaccounted vol	Infrastructure	performance	P105.3	Linear index of unaccounted volumes
Leakage	Infrastructure	performance	P106.3	Leakage index
Renewal	Infrastructure	performance	P107.2	Network renewal rate
Resource protection	Water quality	performance	P108.3	Water resource protection improvement index
Waivers	Financial management	performance	P109.0	Sum of debt waivers payment to a solidarity fund
Interruptions	Customers	performance	P151.1	Occurrence rate of unscheduled service interruptions
New connect time	Customers	performance	P152.1	Compliance rate of new customer maximum connection times
Debt extinct	Financial management	performance	P153.2	Debt extinguishment period
Unpaid bills	Financial management	performance	P154.0	Rate of unpaid bills
Complaint	Customers	performance	P155.1	Complaint rate

Table 1: Performance indicators of water utilities in France

3.2 Data collection

Our dataset includes four types of data: the revealed performance of services, the observed performance of services, the characteristics of the services, and the features of performance indicators. Information about the revealed performance come from the regulator of the French water sector, the Onema. We obtained access to the performance of the 53 services of the Grenoble metropolitan area in 2011. To our knowledge, a very few studies benefit

from this kind of data.

Data on observed performance have been directly collected to ensure that we do not face the same biases as regulators (Hansen, 2005; Leuz and Wysocki, 2016). This condition is essential although it complicates causality identification compared to natural experiments (Hansen, 2017). It ensures measurement of effect of interest and overcomes any measurement error of our dependent variable. These data have been collected by A. Brochet thanks to direct access to services' data in the context of a three-years research-action about regionalization (Brochet et al., 2016). The Brochet dataset contains information about the characteristics of services that allow us to recalculate free-of-biases performance indicators. It includes the necessary variables to calculate 15 of the 17 existing performance indicators in 2011. [12] The Brochet dataset also provides services and contextual information. Characteristics of the performance indicators come from the French regulator. Our final dataset combines of 795 observations, informing 53 services disclosure of 15 performance indicators. Also, a three-year qualitative monitoring of data has been carried out to prevent errors in the calculation of observed performance and to understand the context of services (see Brochet et al. (2016)).

Data are cross-sectional because the Brochet dataset covers only the year 2011. Nonetheless, regarding external validity, 2011 does not seem to be a non-representative year. It is more than two years away from municipal election allowing no abnormal political capture or third party-opportunism (Le Squeren and Moore, 2016; Shi and Svensson, 2006). To our knowledge, the dataset is new in its structure as it distinguishes and compares revealed and observed performance to assess the quality of the disclosure and identify possible biases. Thus, this study does not face any asymmetry of information while, to our knowledge, previous studies may.

3.3 Dependent variable: disclosure biases

Taking into account the limitations of existing studies (Leuz and Wysocki, 2016; Van Thiel and Leeuw, 2002), we construct our dependent variable by comparing revealed performance (SISPEA data) to observed performance (Brochet et al. (2016)'s data). We calculate an index of disclosure bias as following: $bias_{i,s}^{magnitude} = \frac{sispea_{i,s} - base_{i,s}}{base_{i,s}}$, for each performance

[12] The analysis excludes indicators *Connection time* and *Asset knowledge B* due to a lack of information for the former and because the latter is an update of indicator *Asset knowledge A*, which was used during the period of data collection. We did not recalculate indicator *Resource protection*. Information on this indicator is not provided solely by the water service but depends on interactions with the prefecture, making access to raw data impossible.

indicator i in the service$s$ [13]. We consider a bias occurs when $bias_{i,s}^{magnitude}$ is higher than 1% to exclude rounding errors from disclosure bias. Otherwise they are coded *correct*. The general rule for coding the type of disclosure bias is the following:

- "correct" when $bias_{i,s}^{magnitude} = 0 \quad (\pm 1\%)$, there is no disclosure bias.
- "dist" (for distorted) when $bias_{i,s}^{magnitude} \neq 0 \quad (\pm 1\%)$, there is a disclosure bias as the disclosed performance is different from the observed performance.
- "NR" (for Non Reporting) when $sispea_{i,s} \in \emptyset$ and $base_{i,s} \notin \emptyset$, the information is not disclosed while operators have the data to calculate the performance indicator.
- "NA" (for Not Available) when $sispea_{i,s} \in \emptyset$ and $base_{i,s} \in \emptyset$, the necessary information to calculate the performance indicator is not available and operator does not fill the SISPEA database.

Our dependent variable takes four values depending on the comparison of the disclosed performance and the observed performance: correct, dist, NR, NA. This allows us to identify two types of disclosure biases, *NR* and *dist*, from unbiased disclosure (*correct*) and no disclosure due to a lack of information from the service (*NA*). These distinctions highlight the limitation of the use of an only agents' disclosure based dataset to assess the impact of disclosure process or sunshine regulation on a given output.

Summary statistics, as shown in table 2, indicate that disclosure bias is not noise or should not solely be considered in the error term. In our sample, 69% of performance indicators are subject to biases (Table 2). 28% is a dist type of disclosure bias. 41% of our observation is an NR type of bias. Correct completion represents 15% of the sample as well as NA cases. This means that among the 345 values reported in SISPEA, 64.7% are distorted (223). It is worth noting that reporting was not legally compulsory in 2011 which should contribute to increasing NR type of bias because it may not be a high priority for a service provider to report their performance in the SISPEA database, what we name the full disinterest effect. These figures highlight the need to use disclosed information with care as suggested by Leuz and Wysocki (2016).

[13] if $base_{i,s} = 0$ and $sispea_{i,s} \neq 0$, we calculate $bias_{i,s}^{magnitude} = \frac{sispea_{i,s} - base_{i,s}}{100}$.

Variable	Mean	Std. Dev.	Min.	Max.	N
PI calculated by us	1208.14	11545.48	-29.13	229176	625
PI disclosed in SISPEA	1262.88	9269.21	0	159307	345
magnitude of distorsion	6.72	41.23	-1	529.64	298
dist	0.28	0.45	0	1	795
correct	0.15	0.36	0	1	795
nr	0.41	0.49	0	1	795
na	0.15	0.36	0	1	795

PI for performance indicators

Table 2: Summary statistics of the dependent variable

Figures 2 and 3 show the distribution of the types of disclosure biases, per indicators and services respectively. At first glance, the figures show a significant proportion of disclosure biases, i.e. percents of NR and dist are high. Then, distribution of bias type varies among both indicators and services appears to be not randomly distributed. As an illustration, the performance indicators *Price* (D102.0) and *Unaccounted vol* (P105.3) have the same "profile": a high proportion of dist and NR biases (with dist $\gg$ NR), and frequency of correct and NA types of disclosure bias are almost equal to 0. In the same manner, the performance indicators *Interruptions* (P151.1) and *Complaint* (P155.1) are highly subject to dist biases (about 80% of the total disclosure). From the service perspective (Figure 3), it appears that most undisclosed performance indicators are NR-type biases. It suggests that there is a kind of full disinterest effect due to the non-compulsory character of the regulatory requirement in 2011. Data show that disclosure is significantly subject to biases which confirm the need for direct observational data when assessing sunshine regulation (Leuz and Wysocki, 2016). These data are rare and costly to access. Therefore, the empirical relevance needs at least to be discussed regarding this aspect to avoid misleading conclusions.

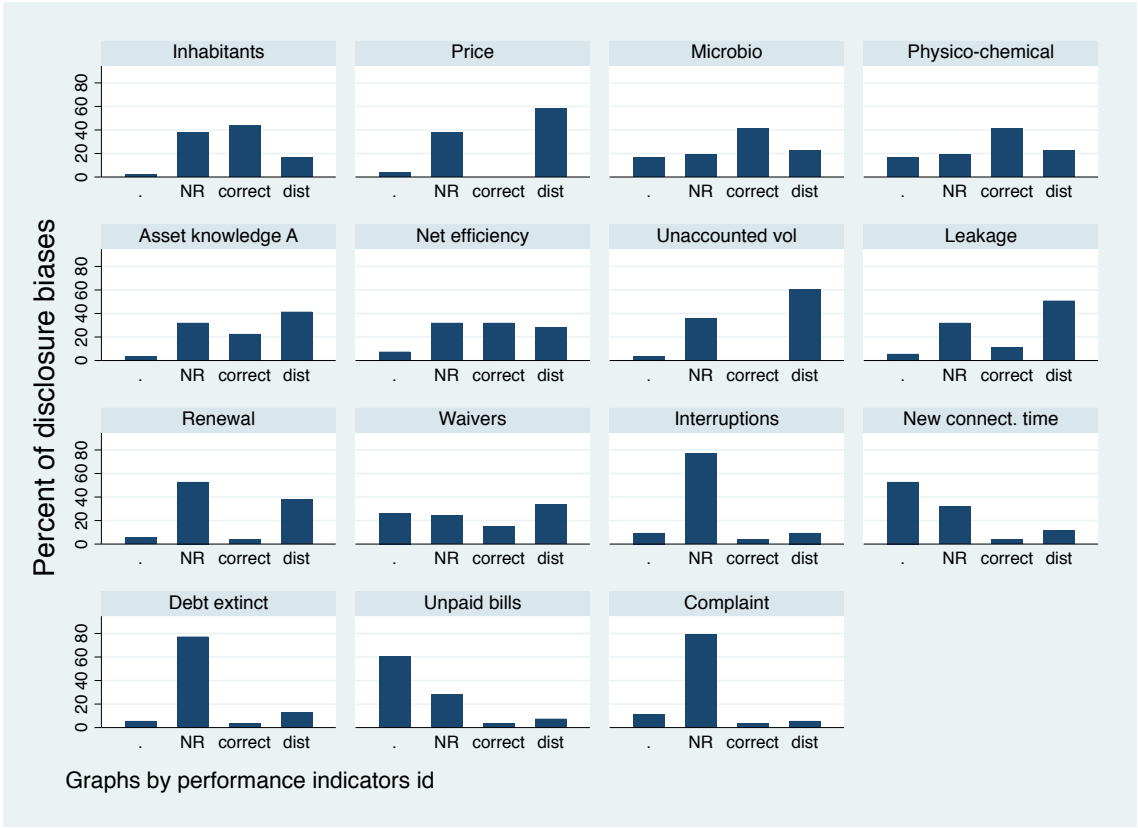

Figure 2: Bias distribution by indicators

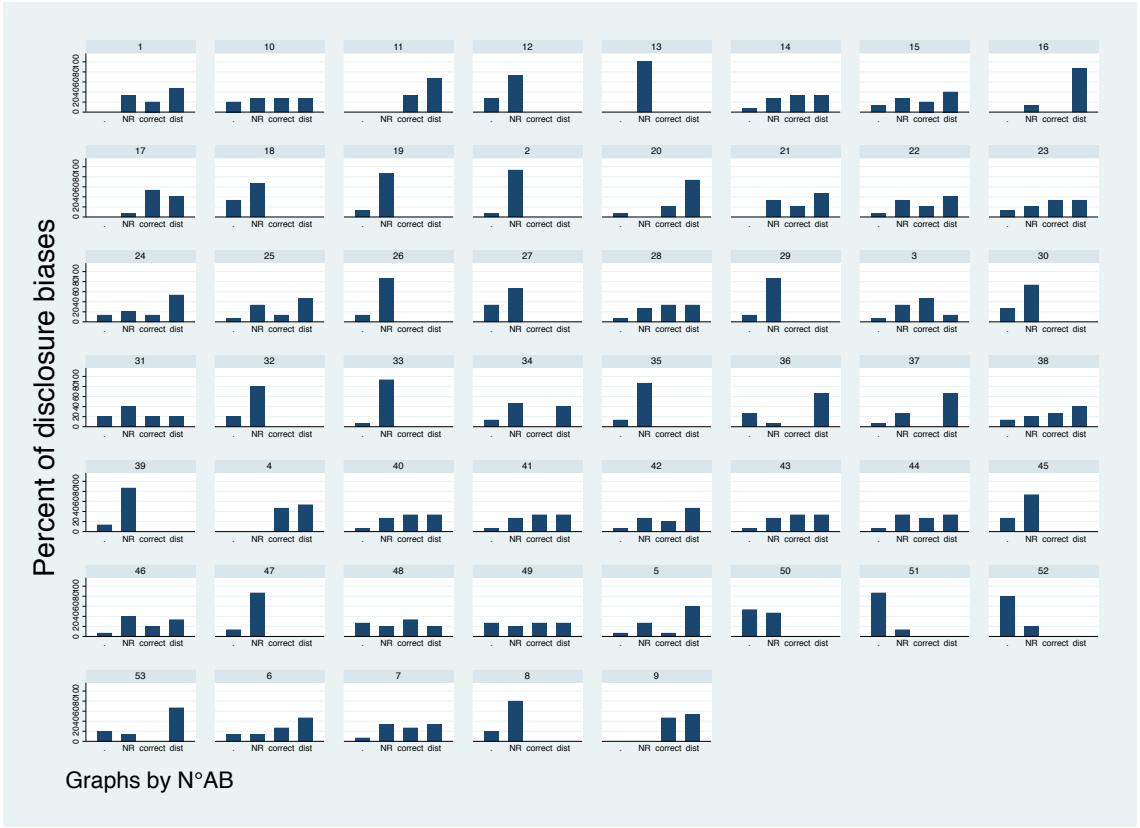

Figure 3: Bias distribution by services

Because 15% of observation are NA-type, the Figure 4 draws the observed propensity of disclosure bias (blue line) and the observed propensity of disclosure bias excluding NA-type observations (red line). Excluding NA-type observations shows the real propensity of disclosure bias as it is the density of disclosure biases within the indicators that could be reported by services. The figure points out that all services exhibit some biased disclosure and that in some cases there is no correct disclosure for the performance indicators reported.

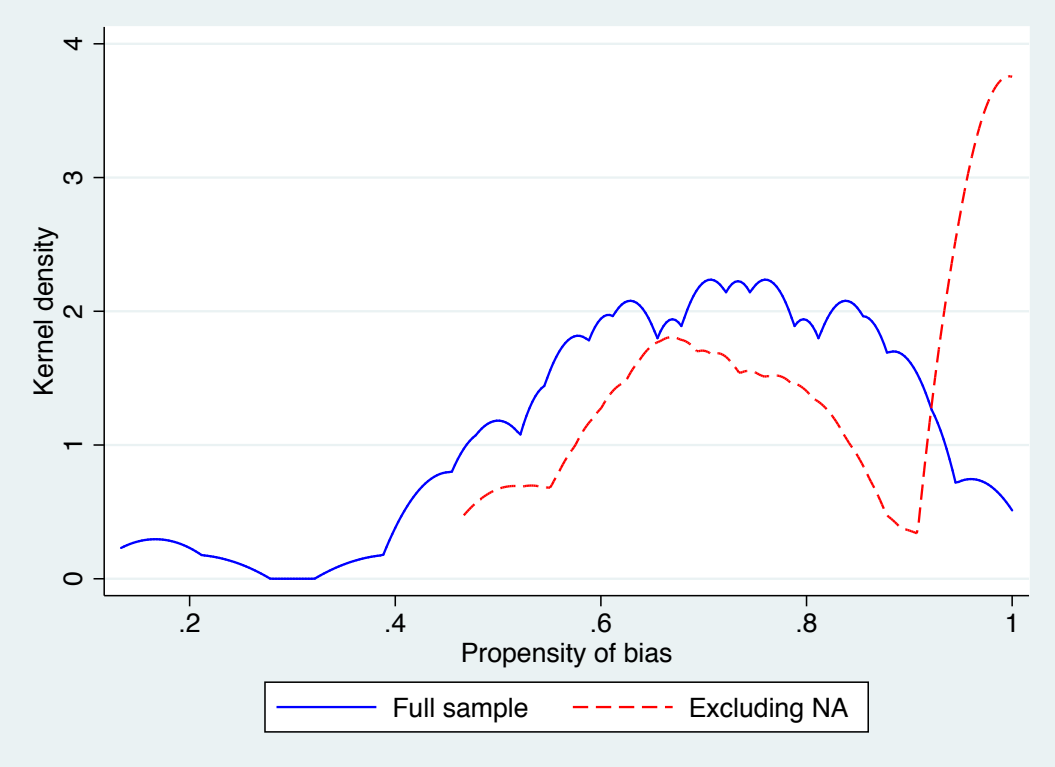

Figure 4: Propensity of disclosure bias

3.4 Independent variables: services and performance indicators characteristics

The dataset contains information on service providers' technical characteristics of supply, organizational aspects of the supply chain and how service providers calculate their performance indicators. Most of these characteristics are coded through dummy variables because they reflect non-ordinal qualitative aspects. Our unit of analysis is the reporting practice of a given indicator i in a given service s. The dataset covers 795 observations, i.e., 53 services reporting on 15 performance indicators. The Table 3 provides the summary statistics.

Many studies highlight how the local technical conditions of supply are pivotal for water services efficiency. They significantly impact on the cost of service delivery (Chong et al., 2015; Marques and Simões, 2008; Pflieger and Ecoffey, 2011). We measure the main characteristics that impact on costs: density of pipes (*mains/km*), the water volume delivered (*m3/km/day*) and the users density (*inhab/km*). We approximate the size of services through the variable *users* corresponding to the number of consumers. This represents economies of scale, and relative level of cost operations.

Regarding organizational aspects, five dummies represent the nature of the service provider. They differentiate between *direct public management*, *leasing contract*, *syndicate*, *service provider* and *others*. We also measure where in the supply chain the operator intervenes. This indicates the level of vertical integration (Gibbons and Roberts, 2013; Lafontaine and Slade, 2007). We code for intervention in *production*, *protection*, *treatment*, *transfer*, *storage* and *distribution*. The variable *cities covered* indicates if the service is inter-communal or not, referring to the regionalization trend in water utilities management (Bolognesi, 2018).

We measure two characteristics of how a service provider must calculates performance indicators. The method of calculating the performance indicators could be: a scoring (*meth_scor*), a direct reporting (*meth_report*) or a formula (*meth_f*). We also consider the number of terms included in the calculation (*nbinput*). In addition, the dummy variable *hinput* codes 1 if this number is high, eg. ≥ 5. This threshold reveals to be a shift in the complexity of calculation.

Considering the regulator purpose, we distinguished between four themes whether the performance indicator relates to customer (*th_cust*), finance (*th_finance*), infrastructure (*th_infrastructure*) or quality (*th_quality*) aspects. *th_cust* is coded 1 for the six indicators *Inhabitants*, *Price*, *Connection time*, *Interruptions*, *New connect time* and *Complaint*. The dummy *th_finance* is coded 1 for the three indicators *Waivers*, *Debt extinct* and *Unpaid bills*. The dummy *th_infrastructure* is coded 1 for the 5 indicators *Asset knowledge B*, *Net efficicency*, *Unaccounted vol*, *Leakage* and *Renewal*. The dummy *th_quality* is coded 1 for the three indicators *Microbio*, *Physico chemical* and *Resource protection*.

Variable	Mean	Std. Dev.	Min.	Max.	N
users	8889.53	25403.79	98	159410	795
cities covered	1.81	4.2	1	31	795
th_cust	0.33	0.47	0	1	795
th_finance	0.2	0.4	0	1	795
th_infrastructure	0.33	0.47	0	1	795
th_quality	0.13	0.34	0	1	795
hinput	0.33	0.47	0	1	795
meth_f	0.73	0.44	0	1	795
meth_report	0.2	0.4	0	1	795
meth_scor	0.07	0.25	0	1	795
production	0.44	0.5	0	1	750
protection	0.42	0.49	0	1	750
treatment	0.44	0.5	0	1	750
transfer	0.48	0.5	0	1	750
storage	0.58	0.49	0	1	750
distribution	0.92	0.27	0	1	750
public direct	0.48	0.5	0	1	750
leasing	0.32	0.47	0	1	750
syndic	0.06	0.24	0	1	750
sce prov	0.14	0.35	0	1	750
other	0.19	0.39	0	1	795
mains/km	50.06	33.59	0	166.51	690
m3/km/day	37.94	56.66	3.56	348.95	705
hab/km	201.55	267.21	30.85	1813.1	705

Table 3: Summary statistics

4 Empirical strategy details

4.1 A three steps approach

To identify service providers' behaviors regarding disclosure in sunshine regulation, our empirical strategy follows three steps. First, we construct the indicator $bias_{i,s}^{magnitude}$ to detect disclosure biases. Then, we qualify the disclosure process and type of bias through four modalities depending on the value of $bias_{i,s}^{magnitude}$, namely correct, NA, NR (not reporting) and dist (distortion).

Second, we run a multinomial logit model to compare the probability of occurrence of each disclosure biases depending on our dependent variables.

The third step consists in robustness checks. They deal with transaction-costs measurement issues, the full disinterest effect that could lead to overestimation, and services

idiosyncracies.

4.2 Model specification

We assess the probability that performance indicators are subject to disclosure biases from the service providers, in a sunshine regulation framework. We assume opportunism and transaction costs explain these biases. Consequently, for the indicator i reported by the service s, the generic model we estimate is:

$$Bias_{i,s} = \alpha + \beta_1.opportunism_{i,s} + \beta_2.transaction\ costs_{i,s} + controls + \varepsilon_{i,s} \quad (1)$$

The dependent variable $Bias_{i,s}$ could take four outcomes: correct, NA, NR or dist. Therefore, we use a multinomial model with the outcome correct as a base category. It means that regression coefficients should be interpreted as the impact of a given independent variable on the probability of observing a specific disclosure biases instead of a correct reporting.

The theory differentiates between different types of opportunism and transaction costs. To identify them we consider together the outcomes and specific dependent variables. Opportunism-related dependent variables are the themes covered by each indicators, the dummies th_cust, $th\_finance$, $th\_infrastructure$, $th\_quality$ and $price$. The complexity of calculating each indicators identify transaction costs minimization, proxies are dummies $nbinput$, $meth\_f$, $meth\_report$ and $meth\_scor$.

Regarding opportunism, we need to identify three behaviors: cream-skimming, over-estimation, and free-riding. Cream-skimming identification differentiates from over-estimation by the expected outcomes, respectively NR and dist. Cream-skimming appears when actors decide to hide information to disclose a more flattering profile while over-estimation consist in reporting false performance (Heckman et al., 1997). Then, the same sets of theme variables are used to identify the behavior. The theme *quality* is the benchmark because in most cases the level is satisfying and there are low incentives for cheating here.

Over-estimation and free-riding behavior lead to reporting false information and should both produce dist-type of disclosure biases. We test a unique configuration to identify free-riding behaviors. The price performance indicator is subject to free-riding because of inconsistencies in the French institutional framework of water services regulation (Chong et al., 2015). Price is used by the regulator to allocate subsidies. If the price is below a certain level depending on service characteristics, the service could not be subsidized. The

rationale is to avoid that the service benefits from national solidarity while its users do not recover the minimum level of cost to allow a "normal" functioning. However, the price is disclosed by the services themselves. Consequently, some services could decide to disclose a higher price than the one they actually charge. That way, they are eligible for subsidies and avoid issues with their users that are also voters (Chong et al., 2015). We integrate the dummy variable called price to identify this effect that we expect to be positively correlated with dist biases.

Regarding transaction costs, our specifications separate the impact of transaction costs in two ways depending on the outcomes, NA in non-feasible cases and NR or dist in strong uncertainty management cases. If transaction costs are too high they could make the disclosure non-feasible. Information is so costly to obtain that agents do not collect it. In that case, we expect to observe an NA type of disclosure bias. At a lower level, transaction costs generate uncertainty in the collection and use of information. Actors approximate the performance indicators (*dist*) or prefer to not report rather than reporting erroneous data (*NR*). In dist and NR types of disclosure biases, transaction costs reflect the complexity of reporting on a given indicator. When dist occurs, it is because agents in charge of reporting make mistakes. It relates to the bounded rationality of actors. When NR biases occur, it is because agents consider the cost of reporting being too high. In all cases, we consider the source of transaction costs in relation to the complexity of calculating indicator. Therefore we use the same proxies: number of inputs and the method necessary to measure the indicators (formula, report or scoring). As a benchmark, we choose scoring, which is more complex than report and presumably less than formula. We then expect negative correlation with $meth\_report$, positive with $nbinput$ and $meth\_f$.

4.3 Endogeneity issues and controls

Our study focuses on actors behaviors in regard to a specific regulatory mechanism. We thus need to control for the context in which services operate (Leuz and Wysocki, 2016). Selecting the case of Grenoble contributes to. The Grenoble Metropolitan area covers a wide range of different services within a common institutional environment, e.g., big cities and small mountain villages. Nonetheless, within this common institutional environment services structures are heterogeneous and could impact on disclosure biases. As an illustration, big and modern services may be more likely to correct disclosure than small services with poor means. We include variables about management and operation characteristics

that proved to efficiently control from impacts of institutional environments, capacities of services and heterogeneous motivations (Chong and Huet, 2009; Chong et al., 2015; Guerrini and Romano, 2014; Pflieger and Ecoffey, 2011; Porcher, 2016).[14] We control for the mode of management if it is public or not (*public*), the size of service (*pop_sce*) , the density of connections (*mains/km of pipes*), volume delivered (*m3/km/day*), and density of population (*hab/km of pipes*). Also, the errors term is clustered by services.

Table 4 presents the baseline models estimating impacts of control variables on the probability of observing a disclosure bias. It suggests that disclosure biases are not significantly sensitive to the technical characteristics of operation, except in NA and NR case. The volume delivered and the size of the services is negatively correlated with the probability of observing NA and NR disclosure biases. The small impact confirms that case selection limits noise from the institutional environment.

	NA	NR	Dist
public	-0.240	-0.152	-0.453
	(-0.48)	(-0.29)	(-1.26)
m3/km/d	-0.0121*	-0.0123*	-0.00469
	(-2.36)	(-2.39)	(-1.26)
mains/km	-0.0128	-0.00862	-0.00739
	(-1.17)	(-0.81)	(-1.06)
hab/km	0.00329	0.00239	0.000710
	(0.60)	(0.50)	(0.29)
pop_sce	-0.000214**	-0.000230***	-0.00000125
	(-2.94)	(-3.46)	(-0.18)
_cons	1.169	2.461***	1.299**
	(1.87)	(4.33)	(2.99)
N		675	
pseudo R^2		0.095	
AIC		1572.9	
BIC		1654.2	

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Baseline models

[14]The literature mainly highlighted the impact of such characteristics on costs, prices or contractual choices. Controls for touristic areas are also used (Chong et al., 2015; Porcher, 2016) but not in this study for two reasons. There are very few touristic areas in our sample, and this variable is used to control from the seasonal variability of operation costs or water demand which should not affect disclosure biases.

4.3.1 Transaction costs measurement

Transaction costs cannot be directly measured, and may face measurement errors in operationalizing. Literature indicates that the time used to gather information, calculate and jump through administrative hoops is a good proxy of private transaction costs (Mettepenningen et al., 2009; Phan et al., 2017). Since we do not have data on time, we use the complexity of each indicator assessment as a proxy. This complexity should encompass different transaction costs such as bounded rationality (cognitive and computational capacity) and access to information. To deal with the risk of measurement error, we run additional specifications with an interaction with the size of the service and organizational characteristics of the service delivery (see Appendix A.1).

4.3.2 Full disinterest effect

In 2011, actors were legally able to not disclose their performance. Consequently, actors could decide not to report because they do not pay interest to this regulatory requirement. It may affect the probability of NR occurrence, non-reporting while actors hold information. Such over-estimation of transaction costs economizing and cream-skimming may skew our results. None of our dependent variables allows measuring this effect. To face this omitted variable issue, we create a sub-sample that is that is free of such full disinterest effect. We replicate the model on this sub-sample, and benchmark results.We expect a weakening of significance levels and a decrease of correlation coefficients the full sample to the sub-sample model.

To delimit the sub-sample, we create the *disinterest* variable as the sum of NA and NR biases on all indicators i per services s, $disinterest_s = \sum_{i=1}^{15}(NA + NR)$. The full disinterest effect occurs when *disinterest* equal 15, i.e., the service does not disclose its performance at all. Figure 5 reports the share of NA and NR biases in each score of *disinterest*. The blue box plots are the sum of NA biases. They show that the number of NA biases is rather stable, between 2 and 4 on average. The red box plots are the difference between NR biases and NA biases. Positive values mean that there are more NR than NA in a given level of disinterest. On average, the difference is positive and dramatically increases from a sum of 9 and 15.

The increase in scores of NR biases to 15 of *disinterest* confirms we accounted for the full disinterest effect. When there is no disclosure, it is not because service provider does not

have the data but because it decides not to report. We assume that the service provider has no interest in performance disclosure and that because of the weak enforcement of the regulatory requirement they do not report anything. Additional evidence is that there is no level of *disinterest* between 9 and 15. This second increase in scores indicates that 15 is a specific case and results from different reasons than lower levels. The structure of $disinterest^{15}$ and the two increases in scores confirm that, in level 15 of *disinterest*, observations represent full disinterest in the sunshine regulation requirements. Further, the sub-sample gathers all the observations subject to this effect within the full sample. Consequently, a sub-sample excluding these observations is free of the full disinterest effect, and of any unexpected noise related to the non-compulsory aspect of the reporting in our explanation of the behavioral causes of disclosure biases in sunshine regulation.

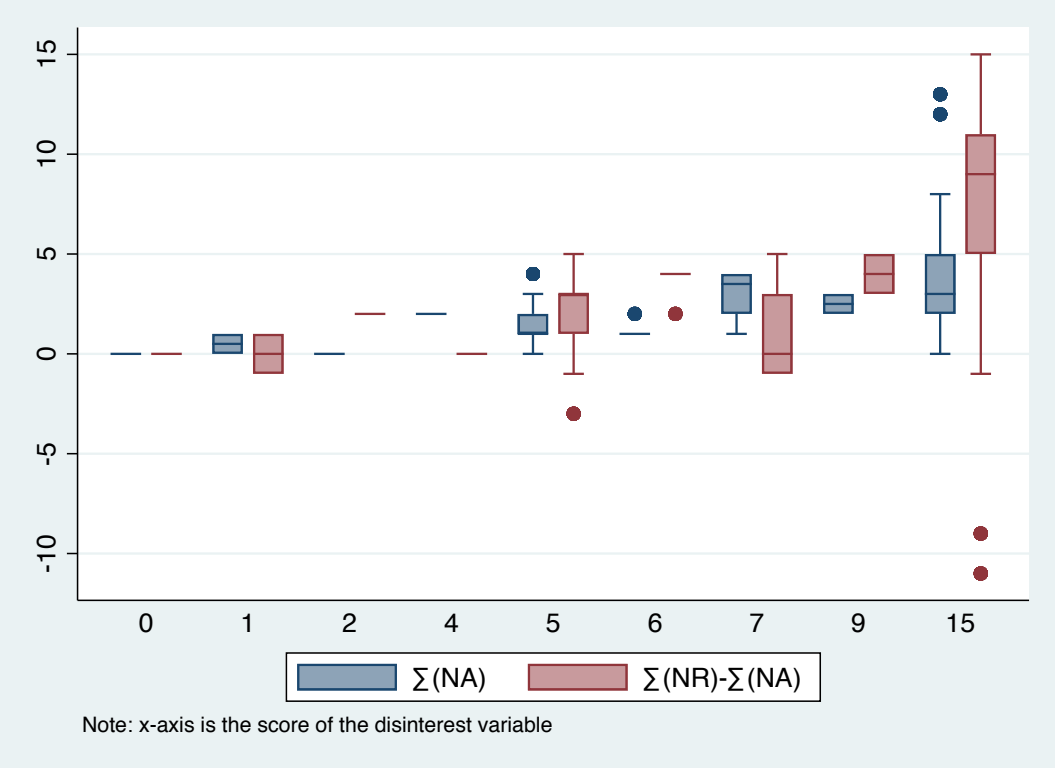

Figure 5: The full disinterest effect

To confirm our results, we compare the full sample estimations to the sub-sample excluding the population with full disinterest effect (Appendix A.2). That way, we control for the impact of the omitted variable.

4.3.3 Services idiosyncracies

Our primary model is a multinomial logit model, fitted by maximum likelihood. The technique estimates more equations than binomial logit models and thus necessitates more observations to be solved. In our case, it constrains our specification as we are not able to add a service fixed-effect to avoid noise from services' idiosyncrasies. Therefore, to see if services idiosyncrasies impact results, we run a logit model for each type of disclosure bias with the same specifications as our multinomial model plus control dummies for service A.3.

5 Results

5.1 Acceptance of hypothesis on opportunism based behaviors

We first estimate the impact of opportunistic and of transaction costs economizing behaviors on disclosure biases in the full sample. The base outcome for the multinomial model is correct. Regarding independent variables, scoring is the base for method variables and quality of water the base for theme variables. Table 5 presents results, which reject the hypothesis of the positive impact of transaction costs economizing on disclosure biases. Conversely, the hypothesis on opportunism is confirmed. These results are robust to full disinterest effects (Appendix A.2) and services idiosyncracies (Appendix A.3).

	(1) TC			(2) Opportunism		
	NA	NR	dist	NA	NR	dist
nbinput	-0.241	-0.0737	0.201*	-0.0594	-0.521**	-0.329
	(-1.37)	(-0.95)	(2.39)	(-0.11)	(-2.71)	(-1.79)
meth_f	13.72***	0.317	1.185	14.17***	-1.226	-0.649
	(11.30)	(0.62)	(1.84)	(7.94)	(-1.84)	(-0.91)
meth_report	14.89***	-0.203	0.615	14.77***	-3.958***	-2.291*
	(9.40)	(-0.28)	(0.68)	(5.46)	(-3.50)	(-1.99)
th_cust				1.166	4.038***	1.332**
				(1.88)	(8.39)	(2.94)
th_finance				2.597***	3.937***	2.266***
				(3.99)	(7.13)	(3.65)
th_infrastructure				-0.131	4.231***	3.464***
				(-0.05)	(4.40)	(3.66)
price				0.0352	14.96***	17.50***
				(0.03)	(33.24)	(31.17)
Control		Yes			Yes	
_cons	-12.26***	2.550*	-0.607	-13.74***	3.034*	1.427
	(-6.17)	(2.50)	(-0.54)	(-4.31)	(2.13)	(1.02)
N		675			675	
pseudo R^2		0.176			0.272	
AIC		1454.7			1300.1	
BIC		1576.6			1444.6	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: The impact of economizing transaction costs and opportunistic behaviours on disclosure biases

Results verify the hypothesis that opportunistic behavior leads to disclosure biases, which confirm the theory. Cream-skimming leads to non-reporting disclosure biases (NR) while over-estimation and freeriding correspond to distortion disclosure biases (dist). Consequently, we accept the hypothesis of opportunistic behaviors.

Regarding cream-skimming and overestimation, it appears that all themes are more subject to disclosure biases than the water quality theme (at 0.001 level). Moreover, it appears

that the effect of cream-skimming is stronger than the effect of over-estimation, i.e., NR column coefficients are greater and more significant than dist column coefficients. In the sub-sample without the full disinterest effect this difference increases A.2. This confirms that disclosure outcomes depend on specific behaviors.

Theory argues organization and contracts frame behaviors by determining their constraints and opportunities (Armstrong and Sappington, 2006; Williamson, 2005). Interestingly, disclosure biases verify this statement. The contractual design of the service impacts the probability of disclosure biases. Figure 6 shows the marginal impact of running the service under direct public management on disclosure biases. Public utilities are more likely to disclose correctly or not reporting rather than private operators. On the other hand, private operators are more likely to disclose distorted performance than public utilities. They are more opportunists regarding free-riding and over-estimation. The behavioral mechanism is that they comply with regulatory requirements by reporting on their performance, which contributes to enhancing their reputation in a context of mistrust (Canneva and Guérin-Schneider, 2011), but they strategically use information asymmetries to provide a more flattering portrait of their activity. Excluding the population with a full disinterest in complying with the regulatory requirement strengthens the statistical significance of the phenomena.

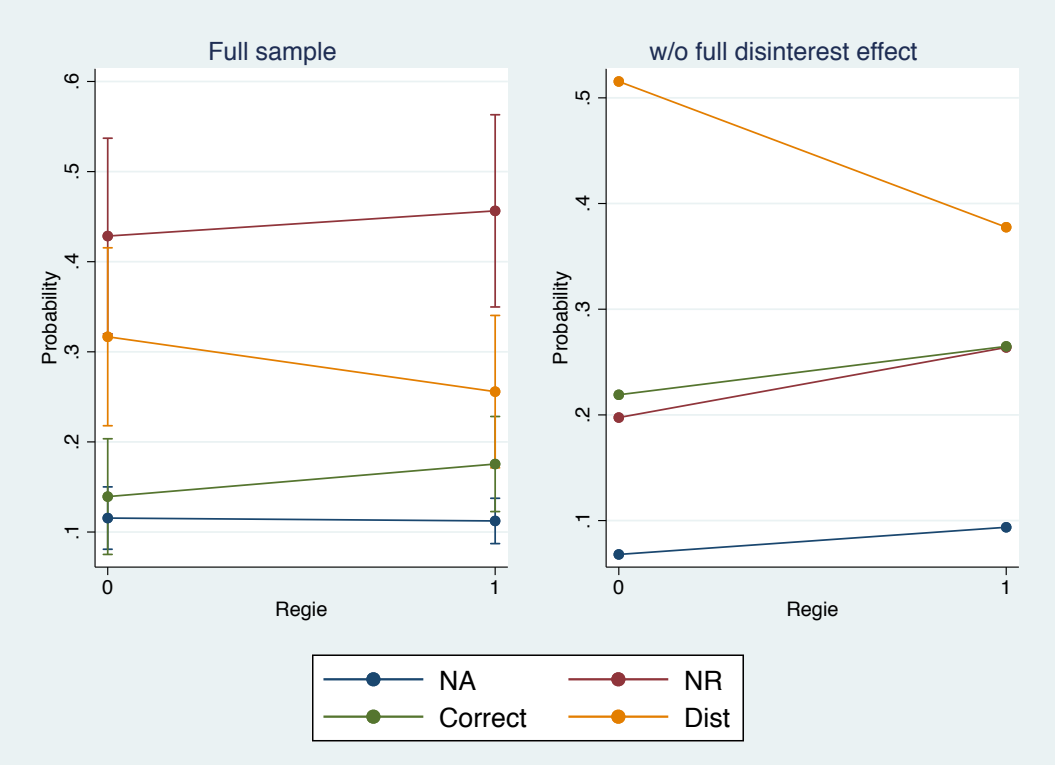

Figure 6: Marginal impact of public direct management on disclosure biases, with 95% CIs

It is noteworthy that the strongest correlation is with the infrastructure theme, underlining to which extent infrastructure management and renewal are critical issues in France and Europe in general (Bolognesi, 2014, 2018). Past underinvestments[15] cause a high level of leakage (about 20% in France) and an accelerating failure rate of the infrastructure. According to Bolognesi (2018), the failure rate is likely to increase from 1% in 2010 to 6.8% in 2040. There is a need for massive investment efforts in renewals. Consequently, public procurement procedures pay high attention to infrastructure management. In a competitive market bidding, it incentivizes operators to draw a flattering portrait on this specific theme or to hide information to potential competitors in order to get renewed on concession, to win new bids or to improve their reputations. For instance, Chong et al. (2015) show that in France the leak ratio is a significant predictor of franchise renewal. Our results show the magnitude of this strategic use of information asymmetries in agency relation. Concerning infrastructure-related performance indicators, the mean of the distortion type of bias is 13.2%.

Freeriding is significant as well. Actors identified the regulatory inconsistencies in the French water sector and game it to both be eligible for subsidies and avoid local political or reputational costs of setting high tariffs as suggested by Chong et al. (2015). Operators report an overestimated price. It allows them to practice a price that meets the expectations of users, whose are reluctant to tariffs increases, and keep being eligible for national financial supports devoted to services modernization. Besides the institutional inconsistency, the structure of the French water sector contributes to explain this freeriding. Water provision is local, with about 13 000 water provision services in 2013, and the price is conjointly determined by the mayor and the operator. Users are the voters, as a consequence, there are low incentives to increase tariffs even. Additionally, the price indicator opens margin for maneuver, meaning that the complexity of the formula is a lever for service providers to satisfy their own interest while complying with the rule. Calculation imposes to refer to a $120m^3$ volume of consumption but households use in mean $93.53m^3$/inhabitant/year (Bolognesi, 2018). For instance, this difference allows operators to over-estimate current prices by using increasing block tariffs, knowing that users will not be subject to the last and expensive blocks as they use less water than $120m^3$.

[15]It is considered that to maintain infrastructure quality, renewals must represent 2% of the network, in France the past level of renewals is about 0.6% per year. At this rhythm, the complete renewal of the French water supply infrastructure would last 166 years while the life expectancy of a water network is 60 years.

5.2 Reject of the hypothesis on transaction costs based behaviors

Results indicate that there are transaction costs in reporting performance indicators, but we found unexpected estimates. Regarding these results, we rejected the hypothesis of transaction costs economizing. Unlike theory would predict, there are no convincing correlations between the complexity of performance indicators and disclosure biases.

NA cases confirm that transaction costs occur in the disclosure of performance indicators. Correlation coefficient of reporting methods that implies direct measurements on the field or complicated formula are positive, with high and highly significant coefficients ($p < 0.001$); while the number of inputs is non-significant. This means that some indicators necessitate measures that are costly to acquire or to deal with, making the reporting not feasible. This also illustrates the discrepancy between the current structure of services (old services) and the structure hypothesized by performance indicators construction (a modern service) (Brochet et al., 2016; Lewis, 2015; Renou, 2017; Tsanga-Tabi and Verdon, 2014). In that perspective, scoring is most straightforward to do and thus results in fewer disclosure biases.

Results are unexpected in the NR case. On the one hand, NR biases are less likely to occur with direct reporting than with scoring, which confirms the theory. On the other hand, results show that the more there are inputs in the calculation the less likely is NR bias, which is the opposite of theoretical predictions. Finally, there are no significant differences in the impact of formula and scoring methods. The impact of complexity is ambiguous. Results on the relation between distortion biases (dist) and transaction costs are in line with the previous outcome (NR). As the identification becomes more precise, impact of the inputs number tends toward being negative and its significance decreases. [16] Like in the NR case, direct reporting is negatively correlated to distortion disclosure biases (at the 5% level).

To emphasize this unexpected and ambiguous results, figure 7 graphs the marginal effect of using a high number of inputs or formula on the probabilities of the disclosure biases. It draws that with a high number of inputs (> 4) the likelihood of NR bias decreases while the probability of observing correct reporting goes up. The graph on the right-hand side illustrates the unclear impact of using a formula on the chance to find a disclosure bias. It appears that dist is lowly sensitive to $meth\_f$ while correct and NR outcomes tend to be less likely if calculation proceeds through a formula. As the confidence interval is large the significance of the links remains low. The case of NA outcomes denotes because the relation

[16] at the level $p < 0.1$ the negative impact of *nbinput* would have been accepted.

is strongly and undoubtedly positive.

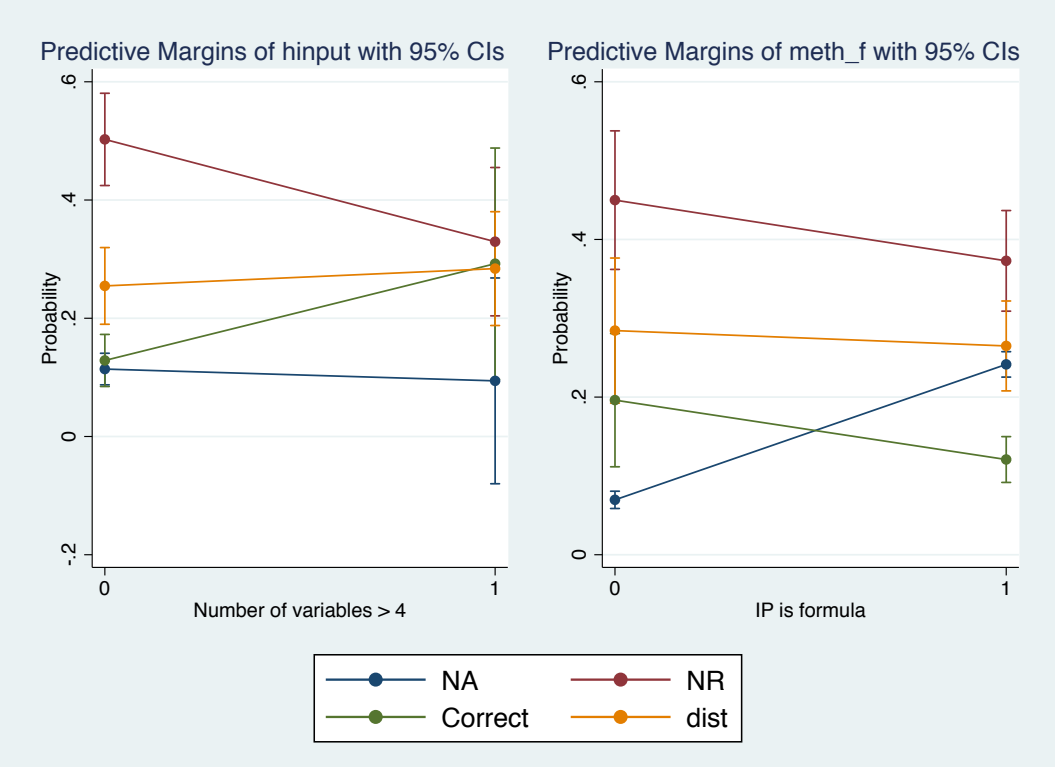

Figure 7: Marginal effects of a high number of inputs and using a formula on disclosure biases

To better understand the behavior behind these unpredicted results, we track for organizational sources of transaction costs, especially the size of the service and its vertical integration (Appendix A.1) (Gibbons and Roberts, 2013; Lafontaine and Slade, 2007). There is no significant link between vertical integration, i.e., outsourcing and unbundling, and disclosure biases. This additional specification does not change conclusion from the primary model. By being a proxy of water services capability, the size could predict water provision outputs (Chong et al., 2015; Porcher, 2016). We investigate the marginal impact of the size on disclosure (Figure 8). According to transaction costs theory, the smaller the service, the less able service providers are to disclose without biases. These results are unexpected. We can't affirm that small services are bad operators. Interestingly, the probability of NR outcomes significantly increases in large services, and decrease in the case of dist.

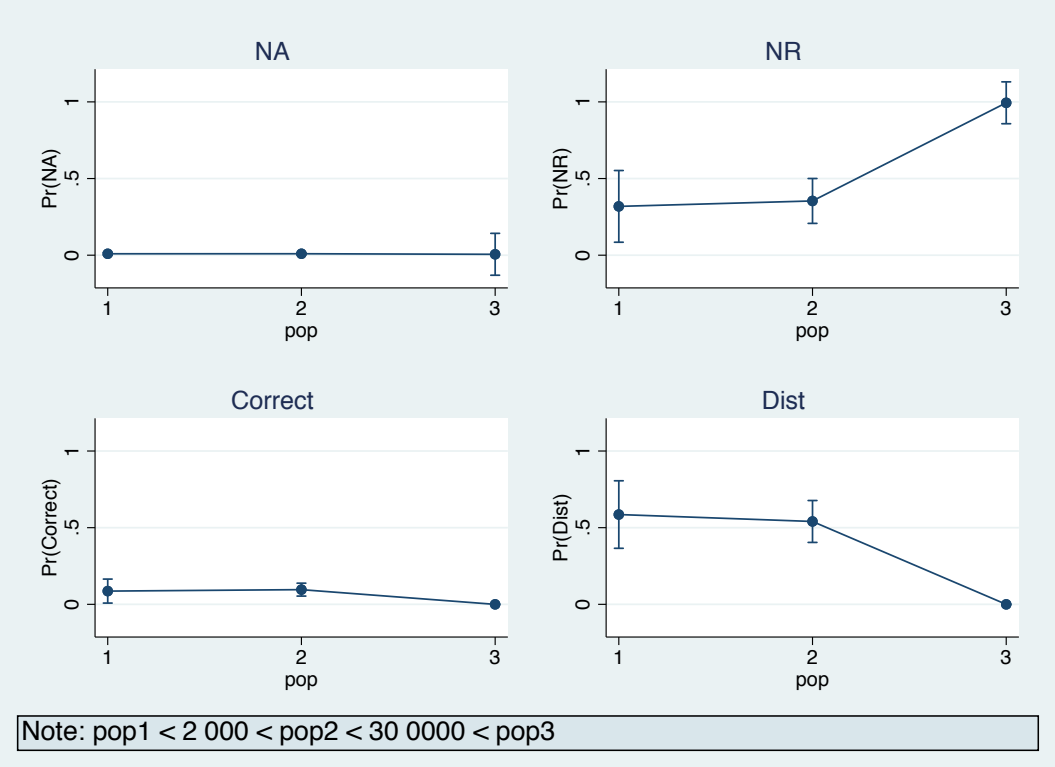

Figure 8: Marginal effects of the size of service on disclosure biases

This last observation paves the way toward an interpretation of the ambiguous effect of transaction costs. Actors use transaction costs as a margin for maneuver rather than trying to minimize them. Large services are more able to deal with information, and they have more capabilities to use margins for maneuver. This results in a decrease of dist biases and an increase of NR biases (Figure 8). This finding challenges the traditional explanation of the role of transaction costs.

The opening of margins for maneuver through transaction costs explains the change of the *nbinput* correlation coefficients when adding opportunistic behavior specifications. In the first block, the results fit with the theory (significant positive impact on dist) while in the full model this effect disappears and tend to be negative, and effect on NR becomes negative and significant. At the same time, direct reporting becomes negative and significant as expected. Consequently, we claim that opportunistic behaviors are rooted in transaction costs structure. Actors strategically use this margin for maneuver.

Results confirm that opportunistic behaviors lead to disclosure biases. Findings of the effect of transaction costs economizing on disclosure are counter-intuitive. It appears that transaction costs are present, but the economizing behavior is not a trigger of disclosure

biases. Rather, transaction costs create windows of opportunity to actors.

6 Discussion and conclusion

In this paper, we analyzed the behaviors that lead to disclosure biases in sunshine regulation using the case of performance indicators implementation in the French water sector. It contributes to explaining the discrepancy between empirical evidence (Gerrish, 2016; Witte and Saal, 2010) and increasing use of sunshine regulation (Baldwin et al., 2012). Additionally, we provided original insights on the analysis of information asymmetries in principal-agent relationships.

To perform the analysis, we draw insights from the Leuz and Wysocki (2016)'s highlights of current limitations in the assessment of disclosure and performance, especially the risk for measurement errors of the dependent variable. We constructed a unique dataset of 795 observations that allows for comparison of revealed to observed performance of water services. This empirical strategy prevents our results from any use of information asymmetry by agents and enables both robust identification of agents' behaviors and classification of disclosure biases. To accurately identify effects, we distinguish from three forms of biases: voluntary non-reporting, distorted reporting and non-feasible reporting.

We observed that disclosure biases are frequent. The sample counts 28% of distorted reporting (excluding rounding errors) and 41% of non-reporting. This result has strong methodological implication for studies focusing on sunshine regulation or disclosure. Research design must consider dataset construction carefully to avoid non-representative results and misleading conclusions. In particular, we recommend using directly measured data when possible or adopting a strategy to control for these disclosure biases.

We tested the impact of opportunistic behaviors and transaction costs minimizing on disclosure biases. Results indicate opportunism is a trigger of disclosure biases, which confirm theory. However, we rejected the hypothesis of transaction costs. Indeed, the calculation complexity of performance indicators has ambiguous links with the likelihood of disclosure biases.

Opportunism takes three forms: cream-skimming, over-estimation, and free-riding (Heckman et al., 1997; Laffont and Martimort, 2009). The free-riding form is interesting as it points out institutional inconsistencies in the regulatory frameworks. In parallel, we show

that there are transaction costs related to indicator measurement but that they are not decisive in predicting biases. With the two others explanatory behaviors, we can hypothesize that these transaction costs create rooms for maneuver, then it is the actors' interests that trigger whether they will bias or not the disclosure. This result has two policy implications. First, it suggests looking at the design of the reporting process rather than the indicators calculation to observe a potential impact of transaction costs on disclosure. Calculation might not be the key issue; the organizational structure of performance management could be. Second, reducing the complexity of the performance measurement should reduce the rooms for maneuver that facilitate opportunistic behaviors. Besides, it is noteworthy that sensitivity tests for transaction costs (A.1) highlight that very small services ($\leq$5000 users) encounter difficulties in reporting their performance to SISPEA because of a lack of means. This suggests that using the same set of indicators, whatever the services structures, is misleading in appreciating performance. There is a need to fit the policy instrument with the technical characteristics (Lascoumes and Le Gales, 2007; Lewis, 2015).

An alternative explanation to theoretical prediction could be that complexity creates rooms for maneuver in interpreting formulas or calculation, but the trigger of disclosure biases is the opportunistic behavior, i.e. the way agents decide to use the existing room for maneuver. Further, these rooms for maneuver create high incentives to bias disclosure while the complexity of indicators measurement does not limit the ability to report correctly. Reports may be not consistent with each other, because actors could use definitional drifts to report in ways that are favorable to them over years (Bevan and Hamblin, 2009; Christensen et al., 2006). In other words, the strategic explanation is more accurate in our case than the cognitive one, as we approximate them.

A Appendix: Robustness checks

A.1 Tracking transaction costs

We found that transaction costs economizing is not a direct cause of disclosure biases, rather a room for maneuver. As this result is counter-intuitive, we explore sensitivity to specifications with two additional specifications of transaction costs economizing.[17] Firstly, we look at the conjoint effect of the complexity of performance indicators measurement i and the size of the

[17]Other existing proxies are proxies of complexity such as the cost of contracts, the condition of operation (Porcher, 2016) or asset specificity scored through surveys (Brown and Potoski, 2003). For a general discussion on measuring transaction costs see Sykuta (2010); Wang (2003).

service s. The rationale is that the smaller the service, the more it is impacted by transaction costs. We add intercepts of high number of inputs, $hinput = 1 \quad if \quad nbinput \geq 4$, and size categories ($pop1 \leq 2000$; $2000 \leq pop2 \leq 30000$; $30000 \leq pop3$). Secondly, we consider the vertical integration of the service. Since the beginning, transaction costs analysis focuses on the make-or-buy decision of firms making vertical integration a central feature of analysis (Lafontaine and Slade, 2007; Williamson, 2005). We integrate two dummies to inspect this relation: outsourcing and bundling both coded 1 if this form of organization occurs. We run a multinomial model with error clustered by services. Results confirm that transaction costs economizing does not impact as theory predicts (Table 6). Especially, the intercept *hinputpop1* is not significant as well as vertical organization variables. These checks consolidate our results.

	(1)			(2)		
	NA	NR	dist	NA	NR	dist
th_cust	0.755	3.331***	1.085*	0.919	3.422***	1.109*
	(1.31)	(8.69)	(2.45)	(1.41)	(8.29)	(2.50)
th_finance	2.197***	3.277***	1.997***	2.297***	3.430***	2.031***
	(3.70)	(7.56)	(3.47)	(3.56)	(7.23)	(3.51)
th_infrastructure	0.693	3.392***	2.854**	0.925	3.466***	2.879***
	(0.57)	(4.03)	(3.28)	(0.77)	(4.13)	(3.35)
price	1.503	15.41***	17.35***	1.522	16.03***	17.99***
	(0.87)	(18.81)	(19.03)	(0.87)	(19.38)	(19.81)
hinputpop1	-0.291	-0.609	-0.122	-0.384	-0.837	-0.215
	(-0.14)	(-0.56)	(-0.11)	(-0.18)	(-0.77)	(-0.20)
hinputpop2	-1.799	-2.375*	-1.269	-1.888	-2.308*	-1.241
	(-1.09)	(-2.45)	(-1.52)	(-1.16)	(-2.35)	(-1.48)
hinputpop3	-11.17***	-13.68***	-1.939**	-10.51***	-13.41***	-1.911**
	(-4.81)	(-7.32)	(-2.60)	(-4.14)	(-5.19)	(-2.63)
meth_f	13.90***	0.437	0.376	14.58***	0.448	0.379
	(15.96)	(1.18)	(0.82)	(16.93)	(1.20)	(0.82)
meth_report	14.73***	-1.324*	-0.718	15.48***	-1.316*	-0.732
	(14.12)	(-2.50)	(-1.11)	(15.26)	(-2.47)	(-1.12)
outsourcing				1.274	0.579	-0.153
				(1.44)	(0.57)	(-0.48)
bundling				-0.924	-0.865	-0.359
				(-1.71)	(-1.50)	(-1.01)
Control op		Yes			Yes	
_cons	-13.45***	0.389	-0.411	-14.56***	0.291	-0.214
	(-10.53)	(0.49)	(-0.48)	(-10.97)	(0.33)	(-0.23)
N		675			645	
pseudo R^2		0.236			0.244	
AIC		1369.9			1306.8	
BIC		1532.5			1472.2	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Additional identifications for transaction costs

A.2 Robustness to the full disinterest effect

The reporting of performance indicators was not compulsory in 2011.[18] NA and NR disclosure biases could result from agents' full disinterest in the regulatory measure.

Table 7 shows that results in the sub-sample are consistent with those of the full sample. The full disinterest effect does not affect the previous conclusion. There are no noticeable changes in the NA case. However one can underline that opportunism gets both more robust.

	(1)			(2)		
	Transaction costs			Opportunism		
	NA	NR	dist	NA	NR	dist
nbinput	0.0311	-0.488***	0.217*	0.302	-1.046***	-0.330
	(0.13)	(-3.65)	(2.44)	(0.62)	(-3.38)	(-1.71)
meth_f	14.85***	12.52***	1.250	15.73***	10.44***	-0.668
	(8.94)	(12.28)	(1.86)	(7.50)	(5.34)	(-0.89)
meth_report	17.38***	11.31***	0.682	17.28***	6.662**	-2.334
	(7.68)	(8.67)	(0.72)	(5.65)	(2.68)	(-1.93)
th_cust				17.15***	20.22***	1.299**
				(20.00)	(48.78)	(2.84)
th_finance				18.51***	20.07***	2.224***
				(24.37)	(45.12)	(3.55)
th_infrastructure				14.49***	20.00***	3.476***
				(6.53)	(19.88)	(3.56)
price				-0.590	14.65***	18.11***
				(-0.49)	(16.07)	(31.15)
Control		Yes			Yes	
_cons	-16.40***	-10.04***	-0.659	-33.97***	-24.70***	1.510
	(-6.22)	(-7.08)	(-0.57)	(-10.21)	(-9.07)	(1.03)
N		435			435	
pseudo R^2		0.193			0.363	
AIC		914.5			743.2	
BIC		996.0			849.2	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Results of estimations in the sub-sample

[18] it became mandatory in 2014.

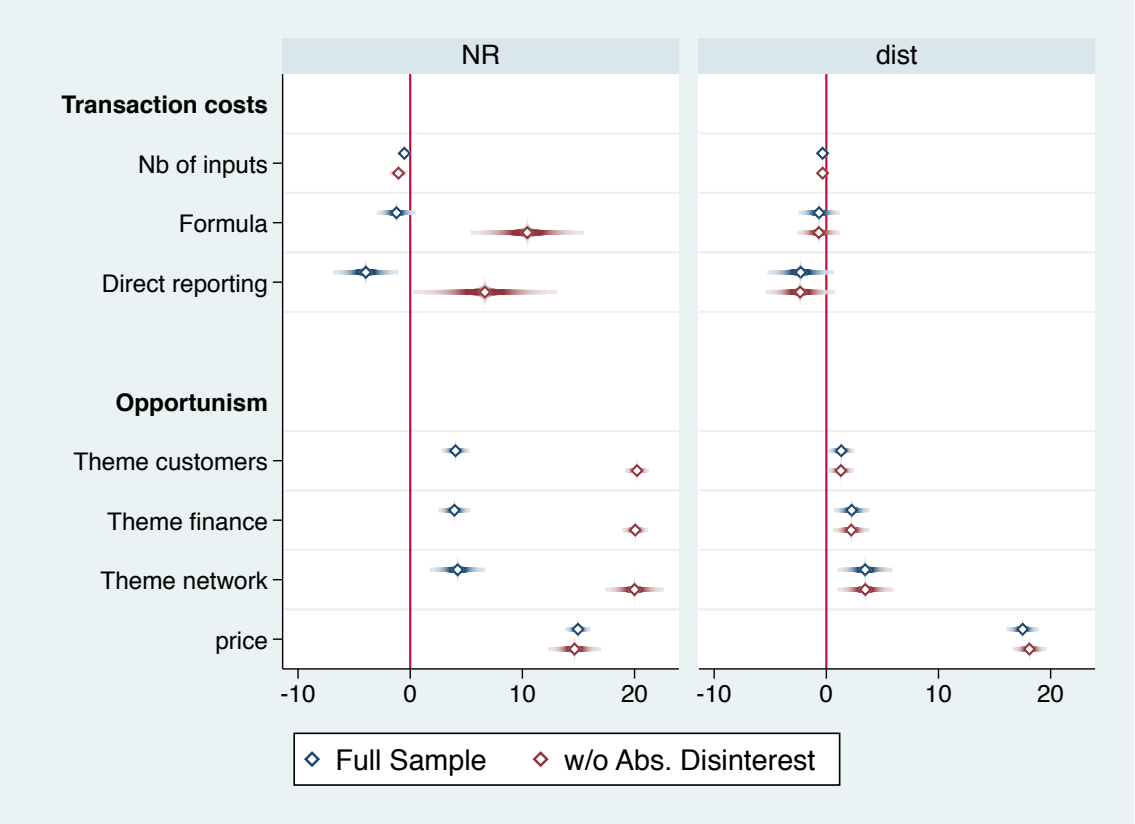

Figure 9: Comparison of estimates w/ or w/o full disinterest effect

The figure 9 offers a benchmark between our two models for NR and dist disclosure biases. It shows that there is no significant change in the dist case. Some changes come up in the NR case, but they do not imply any modifications in our conclusions. A switch of the sign of transaction costs economizing proxies occur. The use of formula and direct reporting become positively correlated with NR type of biases in the sub-sample. The impact of opportunistic behavior gains in strength.

Regarding transaction costs, we reject the hypothesis of the positive impact of the calculation complexity on disclosure biases like in the full sample model. The number of inputs stays negatively correlated with the probability of NR biases (at $p < 0.001$ level), in opposition to theoretical prediction. Some changes occur. The direct reporting method turns to be positively correlated with NR biases (at $p < 0.001$ level), as well as formula (at $p < 0.01$ level). These results are not consistent with the theory. The significance of correlation with distortion biases decreases under the 5% level but maintains close to the acceptance level (8.8% for *nbinput* and 5.3% for *meth report*). Signs do not turn. Impact of transaction costs remains ambiguous in the sub-sample.

Figure 10 draws the marginal effect of transaction costs proxies on disclosure biases. The positive relation with NA outcomes clearly appears. Then, we see that NR is negatively linked with a high number of inputs and positively with the use of a formula.

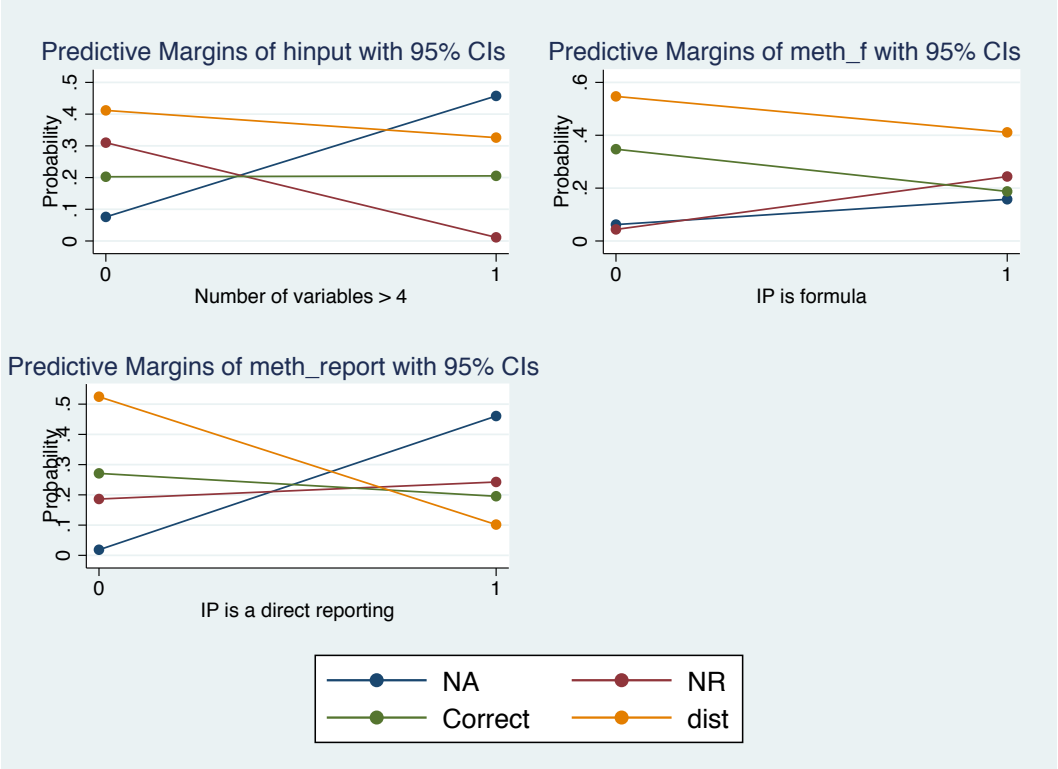

Figure 10: Marginal effects of a high number of inputs and using a formula on disclosure biases

These results support our interpretation of the role of transaction costs as a room for maneuver. Their direct impact remains ambiguous and unexpected while the identification strategy seems to be efficient. In the same time, the effect of opportunism becomes stronger. Transaction costs occur, and they offer rooms for maneuver to actors in the way they disclose their performance.

A.3 Control of services peculiarities

These robustness checks consist in controlling impacts of services peculiarities. We run a logit model for each disclosure biases outcome with the same specifications as our main model. Results are consistent with the main model results. They conduct to reject the transaction costs economizing hypothesis and confirm that opportunistic behaviors motivate disclosure biases (Table 8).

		NA	Distorsion	Non-reporting
Transaction costs	nbinput	0.384	0.117	-0.395***
		(1.20)	(1.27)	(-4.17)
	meth_f	15.12	0.458	-1.189
		(0.02)	(0.72)	(-1.81)
	meth_report	18.40	0.235	-4.904***
		(0.02)	(0.28)	(-5.61)
Opportunism	th_cust	-1.904***	-0.994*	5.116***
		(-3.34)	(-2.22)	(9.26)
	th_finance	-0.302	-0.190	3.398***
		(-0.61)	(-0.45)	(7.00)
	th_infrastructure	-3.788*	1.262*	3.302***
		(-2.23)	(2.51)	(5.85)
	price	0	5.188***	-2.398***
		(.)	(4.58)	(-3.96)
Control op		Yes		
Control service		Yes		
	_cons	-15.73	-0.965	0.191
		(-0.02)	(-0.22)	(0.11)
	N	532	435	600
	pseudo R^2	0.333	0.317	0.444
	AIC	381.3	479.9	555.3
	BIC	569.5	626.6	762.0

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: Logit estimations of each disclosure biases

References

Armstrong, M. and D. E. M. Sappington (2006, June). Regulation, Competition and Liberalization. *Journal of Economic Literature 44*(2), 325–366.

Bakker, K. (2010, August). *Privatizing Water: Governance Failure and the World's Urban Water Crisis*. Ithaca, NY: Cornell University Press.

Baldwin, R. and J. Black (2008, January). Really Responsive Regulation. *The Modern Law Review 71*(1), 59–94.

Baldwin, R., M. Cave, and M. Lodge (2012). *Understanding Regulation: Theory, Strategy, and Practice*. OUP Oxford.

Bevan, G. and R. Hamblin (2009, January). Hitting and missing targets by ambulance services for emergency calls: effects of different systems of performance measurement within the UK. *Journal of the Royal Statistical Society. Series A, (Statistics in Society) 172*(1), 161–190.

Bevan, G. and C. Hood (2006, August). What's Measured Is What Matters: Targets and Gaming in the English Public Health Care System. *Public Administration 84*(3), 517–538.

Bevan, G. and D. Wilson (2013, July). Does 'naming and shaming' work for schools and hospitals? Lessons from natural experiments following devolution in England and Wales. *Public Money & Management 33*(4), 245–252.

Bolognesi, T. (2014). The Results of Modernizing Network Industries: The Case of Urban Water Services in Europe. *Competition and Regulation in Network Industries 15*(4), 306–334.

Bolognesi, T. (2018). *Modernization and Urban Water Governance: Organizational Change and Sustainability in Europe*. London: Palgrave Macmillan UK.

Brochet, A., T. Bolognesi, and Y. Renou (2016, December). Caractériser l'étendue des résistances locales aux indicateurs de performance des services d'eau. Le cas de l'agglomération grenobloise. *Développement durable et territoires. Économie, géographie, politique, droit, sociologie 7*(3).

Brown, T. L. and M. Potoski (2003, October). Transaction Costs and Institutional Explanations for Government Service Production Decisions. *Journal of Public Administration Research and Theory 13*(4), 441–468.

Canneva, G. and L. Guérin-Schneider (2011, December). National monitoring of water utility performance in France. *Water Science & Technology: Water Supply 11*(6), 745.

Chong, E. and F. Huet (2009, September). Yardstick Competition, Franchise Bidding and Firms' Incentives to Collude. *Review of Industrial Organization 35*(1-2), 149–169.

Chong, E., S. Saussier, and B. S. Silverman (2015, August). Water Under the Bridge: Determinants of Franchise Renewal in Water Provision. *The Journal of Law, Economics, and Organization 31*(1), 3–39.

Christensen, T., P. Lægreid, and I. M. Stigen (2006, July). Performance Management and Public Sector Reform: The Norwegian Hospital Reform. *International Public Management Journal 9*(2), 113–139.

Coase, R. H. (1960). The Problem of Social Cost. *The Journal of Law & Economics 3*, 1–44.

Decker, C. (2015). *Modern Economic Regulation: An Introduction to Theory and Practice*. London: Cambridge University Press.

Dosi, G. and M. Egidi (1991, June). Substantive and procedural uncertainty. *Journal of Evolutionary Economics 1*(2), 145–168.

Gerrish, E. (2016, January). The Impact of Performance Management on Performance in Public Organizations: A Meta-Analysis. *Public Administration Review 76*(1), 48–66.

Gibbons, R. and J. Roberts (2013). *The handbook of organizational economics*. Princeton University Press.

Guerrini, A. and G. Romano (2014). *Water Management in Italy*. Cheltenham: Springer.

Guérin-Schneider, L. and M. Nakhla (2012, February). Emergence of an innovative regulation mode in water utilities in France: between commission regulation and franchise bidding. *European Journal of Law and Economics 33*(1), 23–45.

Hansen, B. E. (2005). Challenges for econometric model selection. *Econometric Theory 21*(01), 60–68.

Hansen, B. E. (2017). *Econometrics*. Madison: University of Wisconsin.

Heckman, J., C. Heinrich, and J. Smith (1997). Assessing the Performance of Performance Standards in Public Bureaucracies. *The American Economic Review 87*(2), 389–395.

Künneke, R., J. Groenewegen, and C. Ménard (2010, September). Aligning modes of organization with technology: Critical transactions in the reform of infrastructures. *Journal of Economic Behavior & Organization 75*(3), 494–505.

Laffont, J.-J. and D. Martimort (2009). *The theory of incentives: the principal-agent model*. Princeton: Princeton University Press.

Laffont, J.-J. and J. Tirole (1993). *A theory of incentives in procurement and regulation*. Cambridge: MIT press.

Lafontaine, F. and M. Slade (2007). Vertical Integration and Firm Boundaries: The Evidence. *Journal of Economic Literature 45*(3), 629–685.

Lannier, A. L. and S. Porcher (2014). Efficiency in the public and private French water utilities: prospects for benchmarking. *Applied Economics 46*(5), 556–572.

Lascoumes, P. and P. Le Gales (2007, January). Introduction: Understanding Public Policy through Its Instruments—From the Nature of Instruments to the Sociology of Public Policy Instrumentation. *Governance 20*(1), 1–21.

Le Squeren, Z. and J. Moore (2016). The Political Cycles of Public Contract Renegotiation Evidence from the French Parking Industry. *Discussion Paper Series chaire EPPP* (2015-1).

Leuz, C. and P. D. Wysocki (2016). The Economics of Disclosure and Financial Reporting Regulation: Evidence and Suggestions for Future Research. *Journal of Accounting Research 54*(2), 525–622.

Lewis, J. M. (2015, March). The politics and consequences of performance measurement. *Policy and Society 34*(1), 1–12.

Marques, R. C. and P. Simões (2008). Does the sunshine regulatory approach work?: Governance and regulation model of the urban waste services in Portugal. *Resources, Conservation and Recycling 52*(8–9), 1040–1049.

McCann, L., B. Colby, K. W. Easter, A. Kasterine, and K. V. Kuperan (2005, March). Transaction cost measurement for evaluating environmental policies. *Ecological Economics 52*(4), 527–542.

Mettepenningen, E., A. Verspecht, and G. Van Huylenbroeck (2009, July). Measuring private transaction costs of European agri-environmental schemes. *Journal of Environmental Planning and Management 52*(5), 649–667.

Meyer, M. W. and V. Gupta (1994). The Performance Paradox. *Research in Organizational Behavior 16*, 309–369.

Pawson, R. (2002, September). Evidence and Policy and Naming and Shaming. *Policy Studies 23*(3), 211–230.

Pflieger, G. and F. Ecoffey (2011). The Cost of Urban Sprawl and its Potential Redistributive Effects: An Empirical Cost Assessment for Water Services in Lausanne (Switzerland). *Environment and Planning A 43*(4), 850–865.

Phan, T.-H. D., R. Brouwer, and M. D. Davidson (2017, March). A Global Survey and Review of the Determinants of Transaction Costs of Forestry Carbon Projects. *Ecological Economics 133*, 1–10.

Pollitt, C. (2013, October). The logics of performance management. *Evaluation 19*(4), 346–363.

Porcher, S. (2016, October). Neither Market Nor Hierarchy: Concurrent Sourcing in Water Public Services. *Journal of Public Administration Research and Theory 26*(4), 800–812.

Propper, C. (2003, June). The Use and Usefulness of Performance Measures in the Public Sector. *Oxford Review of Economic Policy 19*(2), 250–267.

Renou, Y. (2017, June). Performance indicators and the new governmentality of water utilities in France. *International Review of Administrative Sciences 83*(2), 378–396.

Romano, G. and A. Guerrini (2011, September). Measuring and comparing the efficiency of water utility companies: A data envelopment analysis approach. *Utilities Policy 19*(3), 202–209.

Ross, S. A. (1973). The Economic Theory of Agency: The Principal's Problem. *The American Economic Review 63*(2), 134–139.

Shi, M. and J. Svensson (2006, September). Political budget cycles: Do they differ across countries and why? *Journal of Public Economics 90*(8–9), 1367–1389.

Shleifer, A. (1985). A Theory of Yardstick Competition. *The RAND Journal of Economics 16*(3), 319–327.

Simon, H. A. (1982). *Models of bounded rationality: Empirically grounded economic reason*, Volume 3. MIT press.

Speklé, R. F. and F. H. M. Verbeeten (2014, June). The use of performance measurement systems in the public sector: Effects on performance. *Management Accounting Research 25*(2), 131–146.

Sykuta, M. E. (2010). Empirical Methods in Transaction Cost Economics. In *The Elgar Companion to Transaction Cost Economics*, pp. 152–160. London: Edward Elgar Publishing.

Tsanga-Tabi, M. and D. Verdon (2014, March). New public service performance management tools and public water governance: the main lessons drawn from action research conducted in an urban environment. *International Review of Administrative Sciences 80*(1), 213–235.

Van Thiel, S. and F. L. Leeuw (2002, March). The Performance Paradox in the Public Sector. *Public Performance & Management Review 25*(3), 267–281.

Wang, N. (2003). Measuring transaction costs: an incomplete survey. *Ronald Coase Institute, Working Paper 2*.

Weber, L. and K. Mayer (2014, July). Transaction Cost Economics and the Cognitive Perspective: Investigating the Sources and Governance of Interpretive Uncertainty. *Academy of Management Review 39*(3), 344–363.

Williamson, O. E. (2005). The Economics of Governance. *The American Economic Review 95*(2), 1–18.

Witte, K. D. and D. S. Saal (2010, June). Is a little sunshine all we need? On the impact of sunshine regulation on profits, productivity and prices in the Dutch drinking water sector. *Journal of Regulatory Economics 37*(3), 219–242.