Studienabschlussarbeiten
Faculty of Social Sciences

UNSPECIFIED

Wising, Joakim:

Public Procurement and Social Goals

Master Thesis, Summer Semester 2021

Faculty of Social Sciences
UNSPECIFIED
UNSPECIFIED

Ludwig-Maximilians-Universität München

https://doi.org/10.5282/ubm/epub.77968

LMU

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

GESCHWISTER-SCHOLL-INSTITUT
FÜR POLITIKWISSENSCHAFT

**Münchener Beiträge
zur Politikwissenschaft**

herausgegeben vom
Geschwister-Scholl-Institut
für Politikwissenschaft

2021

Joakim Wising

Public Procurement and Social Goals

Masterarbeit bei
Dr. Xavier Fernández-i-Marín
2021

Abstract

Public sector purchasing is being called for to be better utilized to achieve goals related to sustainable purchasing. Most previous research have been focused perceptional and/or qualitative studies identifying barriers and enablers of sustainable public procurement within a public organization. One such study identified an overarching concept supposedly affecting the uptake of sustainable public procurement within a country, namely a "policy context". The purpose of this study is to test the findings of previous research on a more general level through a quantitative examination of the previous expectations. This study offers a new conceptualization of a "policy context" measurable through a quantitative-text-analytical method for scoring a country's policy context in a way sensitive to time.

Three hypotheses were formulated and tested. Expectation was that if a country's policy context was framing public procurement in more economic/efficiency terms, the organization would rely more on a lowest price criteria. Expectation was that if an organization was for profit, it would rely less on lowest price criteria in purchasing, another expectation was also that in countries where an entity that coordinates public purchasing exist, the organizations would be less reliant on lowest-price criteria. A measurement of Denmark, Sweden and Finland's respective policy contexts were made. The measurements and hypotheses were tested against the TED.csv data set through a mixed-model logistic regression.

The results show that the measured policy context did not have significant effects on the tendency to apply lowest-price criteria in the chosen cases. An organization being for profit had significant effects, but more research on the topic is needed to draw conclusions. Coordinating entities did have an effect on an organization's tendency to apply other criteria than lowest price in their purchases.

Keywords: Public procurement, Sustainability, Transparency in Measurement, Quantitative text analysis, Mixed-effect modeling.

Contents

1 Introduction

Goal number 12 in the 2015 sustainable development goals reads "Responsible consumption and production", with the mission statement "ensure sustainable consumption and production patterns". The mission 12 has ambitious goal targets to be achieved by at least 2030. One of these targets reads "promote public procurement practices that are sustainable, in accordance with national policies and priorities".

Public procurement is a term used for the acquisition of goods and services by governments or public sector organization, and it is of no surprise that these procedures were considered important enough to be given their own SDG indicator. Reason being is that public procurement is the largest business sector in the world (Hawkins, Gravier, and Powley 2011), and in public purchase procedures, the procuring entity has the opportunity to set demands on the potential contractors. Using public purchases to achieve certain social goals is nothing new, the mechanism of public contracting was used by the British government to address the issue of disabled ex-servicemen returning from the war, already in 1926 (McCrudden 2004). With the advent of climate change and other environmental disasters caused by unsustainable consumption and , it is of no surprise that the biggest business sector in the world is being called for as one tool to help stave off the effects of looming environmental destruction. Public organizations are not to be treated as a monolith, however. They differ between countries, regions and forms, and thus reasonably differ in their ability and tendency to apply criteria in their purchasing. Research on the subject have tried to understand what are the barriers and enablers of applying sustainable criteria in organizations, and have identified several affecting factors on very granular levels. What is lacking is an overarching study where country and organizational specificities are compared to see how this affects an organizations tendency to apply criteria in their purchases, testing the theorized mechanisms identified by previous literature on a larger scale.

1.1 Problem Statement

Within the European Union in 2018, 14 percent of GDP was spent by public agencies on the procurement of goods, services and works (Grandia 2018, p. 375). In these procedures, the procuring actor has opportunities to set demands on the potential contractors. Using public procurement procedures to achieve social goals is nothing new (see McCrudden 2004). Despite this, Grandia (2018) tells us that this is a relatively understudied phenomena in public administration, despite recent movements in the field of "sustainable public procurement". The studies that have been done on public procurement (see for example, Gormly (2014) & Grandia and Voncken (2019)) have been using perception data such as interviews or surveys of individuals in public organizations. Scholars in the field are calling for a movement away from perception data (ibid., p. 13). Further, with the lack of larger studies on sustainability-criteria in public procurement, little is known about what affects an organization's tendency to apply these when procuring goods and services, even though some variables are hinted at in previous research (Tkachenko, Yakovlev, and Rodionova 2018), (Nijboer, Senden, and Telgen 2017). There is thus a need for a theory-testing study that moves away from perceptional, to observational data. This is not only pertinent for scientific reasons – understanding how we can better utilize public procurement as a policy tool is an essential part for achieving social goals, such as the agenda 2030 goals (Agenda 2019), (United Nations, Sustainable Development goals 2012).

Public organizations are known to have more ambiguous elements than private ones, such as ideology and objectives that affect the organization in different ways (Botti and Monda 2019, p. 1). McCue, Prier, and Swanson (2015) are able to show that public procurement systems around the globe are being called for to balance the inherent tensions existing in purchasing. The researchers state that little is known about the various values and goals that underlie the public sector procurement process (ibid., p. 178). To understand the trade-offs that purchasing professionals make – the perception data used by previous public procurement literature have used is apt. The purpose of this study is however to understand public procurement trade-offs not as isolated events within an organization, but as a general phenomenon. Consequence of this ambition is that we need to leave individuals behind and look at larger contexts as influencing the decisions of the purchasing professionals. Brammer and Walker (2011) theorize that a nation's "policy context" have large effects on a nation's public organizations tendency to apply sustainability criteria. A "policy context" can be understood as what values are espoused in relation to the policy in question, however no quantitative measurements of such a policy context exist. Can we understand the values that surrounds public procurement in a country as affecting the application of sustainability criteria? Previous research on organizational values and their effects suggests this is not an unreasonable assumption, (see (Oldenhof, Postma, and Putters 2014), (Peng, S. Pandey, and S. K. Pandey 2015), (Bao et al. 2012), (Elkington 1994)). Further, we know that public organizations take political cues from governments (Terman 2015), (Ringquist, Worsham, and Eisner 2003), which affect they way in which they conduct their "bureaucratic policy-making" as well as what questions are considered salient. Gormly (2014) findings also show that governments play an important role in what values are supposed to be realized through public procurement. Despite these findings, procurement has mainly been studied in disciplines such as economics and law (Knebel et al. 2019). With the role of the government as steering and setting the goals for the bureaucracy - we need ways to understand the impact of "the political" on sustainable public procurement.

While previous research have identified ability, motivation and opportunity (see for example Grandia and Voncken 2019), (Gormly 2014))as important conductors for applying sustainable public procurement (SPP) criteria. Yet, with the relatively small sample sizes, interview studies or perceptional data used by previous research, we know little about how this looks on a larger level or even the effects of government giving political cues (Terman 2015). These issues together with the general lack of theory and theory-testing studies within public procurement (Grandia 2018) (Koala and Steinfeld 2018), creates a need for a large sample-size study looking deeper into the relationship between values, country- and organizational characteristics and public procurement.

1.2 Purpose and questions

The purpose of this study is to elevate the understanding of sustainable public procurement procedures to a more general level through a quantitative analysis exploring the relationship between a measured policy context, country- and organizational characteristics and choices of criteria in public procurement procedures.

Through the following research questions we shall get closer to the previously stated purpose.

- How can we conceptualize and measure a country's "policy context?"
- How can we understand the relationship between our measured "policy context", country- and organizational characteristics and criteria-choice in public procurement procedures?

Through using the Tenders electronic dataset (TED) (TED 2010-2019) from 2010 to 2019 which contains data from over 1830000 public purchases within the European Union, and homing in on the Nordic countries within the European Union & organizations who are represented in the data, looking at how often they apply other criteria to their procurements in contrast to how often they rely solely on lowest price. Comparing these observations to that of a nation's "policy context", that is, a measurement of values espoused by the government and parliament in relation to public procurement, through a text-analytical model -with inspiration from previous research in value-measurements such as (Dolan et al. 2013), (Esteve, Grau, and Valle 2013). With these steps, the study will move from perception data to observational and taking into account how a general politics might affect these relationships, something that previous research in sustainable public procurement lack (Knebel et al. 2019). This study will be able to offer a more general explanation than previous research regarding the ifs and whys some organizations are better than others in their efforts of achieving social goals with public procurement.

2 Sustainable and green public procurement, *"the missing multiplier"*

While sustainability has been a buzzword within the environmental-policy-discourse for quite some time now, sustainable public procurement is a relatively new term, however "green public procurement" has been used in the pre-2010's (see example (Parikka-Alhola 2008)). The difference between sustainable and green, according to the European commission is the inclusion of social and economic factors, beyond the environmental impact (European Commission, Sustainable GPP 2021). The European commission states that achieving the appropriate "balance" between these terms is key for achieving sustainable public procurement. These three pillars are also referred to the "triple-bottom-line", we shall return to this concept later (see under section "Measuring values"). The importance of sustainable practices, names aside, is only becoming more and more important, however. It is to no surprise that several international organizations with a focus on environmental and social issues have dabbled in guides and regulations to facilitate states' sustainable purchases, from the European commission (European Commission 2017a) , to the UN incorporating it into their sustainable development goals, specifically number 12.7 "promote public procurement practices that are sustainable in accordance with national policies and priorities" (United Nations, Sustainable Development goals 2012). For more examples, see (WTO, Nordic council (who aptly named sustainable public procurment the "missing multiplier" Nordic Council (2021)), World bank, transatlantic networks etc).

In some research, a focus has been on the growing divide between lip-serving vs pragmatic action (Seele 2016), and these international organizations are taking steps towards a further implementation of sustainability in procurement. Knebel et al. (2019) argues that sustainable public procurement has become a topic which gather several different disciplines, from law, business and sustainability, with the overarching purpose of coming up with practical solutions to the sustainable development challenges present, just one example would be previously named development goal 12.7. Previous research have reasonably focused in on one country and/or sector to focus in on one particular legal framework (I will argue that this is a problem if we want an abstract understanding of sustainable public procurement as a phenomena, we cannot and should not lock ourselves to one framework, see section "summing up and moving forward").

Worth noting is that according to Knebel et al. (2019)'s study, there is a seeming lack of studies coming from the perspective of political science, which is surprising considering how environmental issues are highly salient political issues. This means that there is pressure on governments to perform, and McCue, Prier, and Swanson (2015) finds that sustainable public procurement is being used to measure performance of governments. Thus there is a need for a study where "the political" is taken into consideration, and not only focus on procurement from an economic or law perspective.

2.1 Why are public procurement procedures worth studying?

Public procurement is distinctly different from private purchasing. We know that public organizations have more complex goals and act under different constraints and rules (Rainey, Backoff, and Levine 1976)(Botti and Monda 2019). These goals could be political goals, through giving contracts with conditions that the contractor to fulfill some demands. These goals, McCrudden (2004) tells us, have spanned from hiring disabled soldiers returning from world wars, to attempts of solving racial issues in post-apartheid South Africa.

More recent studies of public procurement as a way of achieving social goals can be found in the "sustainable public procurement" (or SPP, for short) literature. The definition of sustainable procurement instead reads:

> The acquisition of goods and services in a way that ensures that there is the least impact on society and the environment throughout the full life cycle of the product (Grandia 2018, p. 373).

Within the EU, which gives member states a minimal harmonized level of legal framework see Directive 2014/24/EU (2014), we have seen increasing movements for organizations to incorporate these issues into their purchases (Nawrocka

2008) (Grandia 2018). With this increase of sustainability issues in public procurement, there are increasing calls for studies regarding this subject, with the goal of understanding the variation that we see between countries and organizations (ibid., p. 373). What are the potential barriers or drivers behind the degree of implementation of SPP applied in procurements?

In their 2019 study Grandia, et. al. focuses on the relationship between ability, motivation and opportunity and the way these factors affect types of "sustainable public procurement (SPP)". To measure this, they employ an online survey - asking the bureaucrats subscribed to a public procurement newsletter to gauge their "knowledge" and "skill" about sustainable public procedures. Then, for measuring implementation of these SPP, they asked the bureaucrats on a 5 point likert scale "[...] how often you actually applied them [SPP] in your procurement projects?" (Grandia and Voncken 2019, p. 7). The researchers did find significant results through use of perceptional data, that ability, skills & commitment (measured through the surveys) did have significant effects on how often the bureaucrats applied these sustainability criterion in procuring (also measured through survey data). Parallel to these developments, another field of literature focusing on the social part of sustainable public procurement. Vluggen et al. (2020) introduced the concept of "social return on investment" as a way to understand the social aspects of sustainable developments in public procurement, which they argue are understudied in comparison to the environmental aspects. The findings of the study stresses the importance of clear public procurement policies for affecting the application, another important finding was the organizational incentives (or rather, lack thereof) was a major player (ibid., p. 239).

McCue, Prier, and Swanson (2015) offers an insight on what mechanisms might be at play here. Public procurement's three previously mentioned "pillars" does create tension within the purchasing organization as well as the purchasing individual, being placed at the forefront of the balancing of this tension. We know the profit motive is not always the primary driver of action within public bodies, however economic values are always a factor in any organization. These dilemmas are the focus of McCue, Prier, and Swanson (ibid.)'s study -

through the use of expert interviews in different organizations across the US, the authors identify 5 dilemmas that purchasing professionals face - and what strategies are adopted in solving these. Most interesting for our study is the recurring problem of identifying what is "best value". The authors frame this dilemma in terms of a principal agent problem - while there is responsibility of the individual to make sure that money is being spent in good fashion, what is best practice for the organization is not necessarily best for the individual (while McCue, Prier, and Swanson (2015) mentions corruption, I do think just relying on lowest price criteria could save time while not "best value"). To counteract these, guidelines and policies are put in place to offer the individual guides in how the they ought to understand what is "best value". A remaining problem is still that it is not obvious how to interpret best value, how do we balance the three "pillars" of sustainable public procurement, and where do public organizations look to get guidance in how to understand "best values"? The discussion will return to these questions under section "bureaucratic response to political cues". Next section will focus on how previous literature have tried to understand public procurement in comparative fashions.

2.2 Comparative Public Procurement

Both Grandia and Voncken (2019) & Vluggen et al. (2020) focus on municipalities or their social enterprises within one country (The Netherlands) as the procuring actors, where they do find significant results. The question remains however, how far can we draw these findings. There have been surprisingly few comparative public procurement studies within the European union (Grandia 2018, p. 374) - meaning that we do not know how well these findings hold up in an international or even cross-organizational context.

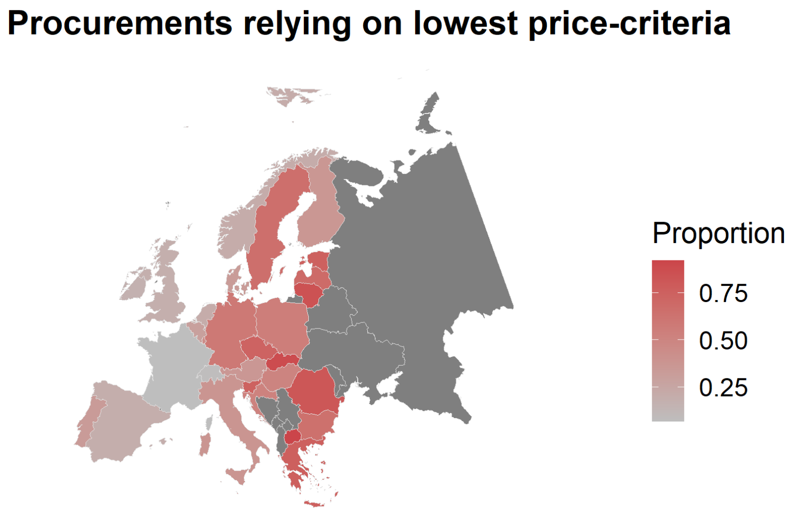

Figure 1: Map of Europe with the proportions of procurements from 2010-2019 that relied on lowest price criterion. Self produced with data from *Global Administrative Areas*, University of California (2021) and TED (2010-2019)

2.3 Transnational comparative procurement

The lack of transnational comparative studies, Patrucco, Luzzini, and Ronchi (2017) tells us,

> is due to the management strictly depends on country characteristics, and that the results are hardly generalizable (ibid., p. 245).

meaning previous literature had its focus on single organizations and countries, avoiding comparative procurement studies. I agree with Grandia (2018, p. 374), that the time is right for such studies. Within the EU, there is a harmonization of public procurement procedures, see for example Directive 2014/24/EU (2014) that member states had until 2016 to transpose to national law. Further, within the EU, the commission has a common "public procurement strategy" (European Commission 2017b) aiming to, among other things, increase the uptake of innovative, social and sustainable procurement. There is a need to study precisely the reason for previous literature's avoidance of comparative procurement, namely country characteristics. "How an organization being in another country affect its' tendency to apply other criteria in their purchases?" is an important question to answer. Refer to figure one, how come France is closing in on every procurement applying a criteria, whereas North Macedonia is seemingly relying heavily on a lowest price criteria? (Note that it is all procurments in the TED dataset from 2010-2029.) (TED 2010-2019)

There have been a selected few cross-national procurement studies. Meaning studies that compare either policy, or outcomes in procurement with the explicit purpose of controlling for country-unique variables.

An early study is Parikka-Alhola (2008) who looked at 30 of "calls for tenders" (calls for tenders could be understood as a simple notice, "the organization would like to buy X") in Sweden and Finland - to see how they differed in the weighting of environmentally friendly furniture procurement versus price. The conclusions drawn were mostly focused on the seller side, unfortunately.

Brammer and Walker (2011) conducted a survey study across different countries. The data was slightly skewed, with difficulties reliably finding public procurement professions from different countries to answer the surveys. The questions that they seek to answer is "how do regions differ in their approach to sustainable procurement?" & "What enablers and barriers to SP exist?". How these findings are interesting for this study, in line with the stated research questions is by the professionals perceived barriers and enablers of sustainable procurement. Brammer and Walker (ibid.) find that financial constraints were the most cited barrier to applying SPP-criteria in their procurements, it is perceived as simply more expensive to procur sustainably than not - sustainability comes with a premium. These findings are echoed in Tkachenko, Yakovlev, and Rodionova (2018) where they find that organizations with hard budget constraints are more "efficient" (defined in economic terms) due to financial constraints (ibid., p. 1157). This does not mean that relying on a lowest price-criteria is the most efficient option, however (see the discussion under "bureaucratic response to political cues).

The second most important barrier was a lack of information. One Swedish respondent stating that people in charge of purchasing does not have a clear "mandate" to apply ethical criteria. The respondent further emphasized that it would probably be better if top management was more engaged - so that it is not up to the individual procuring professional (Brammer and Walker 2011, p. 467). That idea that the personal commitment of procurement professionals helps in the application of sustainability criteria is echoed in Grandia and Voncken (2019) & (Drumwright 1994).

The authors conclude that indeed, national and international SP-policy environments are the largest determinants of the uptake of sustainable procurement - and there is a significant variation between regions and countries. The authors offer no closer explanation or examination of these variations or variables supposedly affecting procurement outcomes (i.e. application of sustainable criteria), instead offering somewhat vague policy-advice for governments. The strengths of this study is the thorough survey application, breaking comparative procurement ground - however it lacks a granularity and specificity regarding relationships, something that is difficult with survey data.

Another related study was conducted by Nijboer, Senden, and Telgen (2017), looking at if there are any evidence of "cross country learning" within the field of public procurement. Through an extensive literature review - and interviews with procuring professionals, there does not seem to be any evidence of cross-country learning. The authors conclude that countries do however implement similar policies, but the lessons learned remains implicit (ibid., p. 463). Perhaps we can see evidence of this in our chosen three countries (see section "countries chosen"), where all three countries, within years of each other, developed "national procurement strategies", as to guide their procuring actors towards more sustainable and innovative procurements.

2.4 Cross-organizational comparative procurement

What does organizational form matter for how an organization procurs? It is established in previous research that public organizations have more ambiguous elements than private ones (Botti and Monda 2019). Then what happens if the organization is stuck in an "in-between" - say a state owned enterprise (SOE) - having to abide by public procurement law, yet intended to make money? Even disregarding country characteristics that previous literature argues is the most permeating factor for procurement outcomes (see Patrucco, Luzzini, and Ronchi 2017, p. 245) - how does such organizational characteristics affect the way in which an organization chooses to purchase? Are there barriers or enablers in stemming from organizational characteristics that are present in one kind of organization, but not in others?

Gormly (2014) study focuses on public procurement in specifically commercial semi-state bodies. Through an extensive interview study, the author is able to identify several variables and factors that the respondents perceived to be the largest barriers in applying SP-criteria in their procuring processes, one main finding is the important of how to achieve "value" when using sustainable procurement. Often, "value" was defined in measures of "cost of a product/service", similar results to Brammer and Walker (2011). The researcher, while focusing on SOEs avoids making claims or questioning the respondents regarding the puzzle if these barriers are unique to SOEs? A missed opportunity, especially since the creation of state owned enterprises from agencies, due to the supposed "performance improvements" see Nelson and Nikolakis (2012) is a continuing trend (Österberg 2020), and we need to step up our efforts in procuring sustainably if we are to meet Agenda 2030 goals (Agenda 2019).

Gelderman, Semeijn, and Bouma (2015)'s study instead focuses on the procurement procedures of local governments and recognizes differences between organizational forms - as well as the potential gain of understanding municipal and regional procurement. The author conducted a multitude of interviews with representatives of different municipalities, executive with political responsibilities as well as procurement managers acted as respondents. For us, the most interesting findings are that there are different perceived barriers depending on what roles the respondents play in the procurement process. Second, the authors identify factors at play in "successful" and "unsuccessful" sustainability initiatives within the municipalities. Some of these perceived barriers, as well as factors at play in initiatives, were institutional in character. While influenced by principal-agent theory, institutional in character meaning here that they can be traced to the organization. Two enlightening examples being lack of appropriate monitoring systems and as how much political involvement was allowed. The authors does admit however that a major weakness of their study was the inability to establish relationships between sustainable criteria application and the the characteristics of an organization (Gelderman, Semeijn, and Bouma 2015, p. 86) - as this is difficult to do with the authors chosen approach. They call for future research to fill this gap.

Soudek and Skuhrovec (2016) offers us an example of a cross-institutional comparison on how procedural characteristics have an effect on procurement procedures. The authors main question is how procedural characteristics affect the final price in public procurements? The authors did a quantitative study, divided the observations depending on if they had been procured by SOEs, Public bodies, Regional authorities or central state authorities and found that electronic auctions with more "bidders" resulted in lower prices. Interestingly enough, according to their findings, there does not seem to be a relationship between the category of institution and the price of the procurement. This is contrary to Tkachenko, Yakovlev, and Rodionova (2018) findings which found that organizations that were exposed to competition as "income earning units" were more efficient in their procurements than those who were not. The difference in findings could be the way in which the authors define "efficiency".Soudek and Skuhrovec (2016) focusing solely on price on a homogeneous group of goods - whereas Tkachenko, Yakovlev, and Rodionova (2018) allows for more complexity in measurement. This paper will leave the issue of whether or not being an income earning unit or not is economically more beneficial in procuring open. We will instead be focusing on whether or not different categories of institutions are better or not in applying criteria in their procurement procedures, as is hinted in the previously discussed perceptional SPP-literature.

2.5 Summing up - and moving forward?

Key points to take away from this section is this. The most studies done on public procurement have been using perceptional data, interviews and surveys. Through these studies, we know certain things - public procurement is not just "buying stuff" - it has large potential (European Commission (2017b), Mc-

Crudden (2004), Grandia (2018), Agenda (2019)) and the outcome is seemingly affected by many factors.

We learned from Parikka-Alhola (2008), Brammer and Walker (2011) & Nijboer, Senden, and Telgen (2017) that in what country the procurement takes place has an effect on the outcome of the procurement, whether sustainability or social goals are being applied. None of them, however, delve deeper into the relationships between what "country characteristics" (recall the quote in Patrucco, Luzzini, and Ronchi (2017) stating that previous literature had opted out from making cross-national comparisons, due to "country characteristics") supposedly affecting the outcomes.

Further, most cross-country studies mentioned have been based on perceptional data - gathered through interviews or surveys - which have gotten us significant results, however perceptional data is limited in several ways. First, one of the studies recognize how results can be skewed due to bias in respondents. Who answers the questionnaires & act as interview subjects could lead to biases in results. I believe this is a significant strength in Gelderman, Semeijn, and Bouma (2015)'s study, where the biases' of the respondents were interesting results when compared, pointing towards differences in perceptions of procurement depending on what part of an organization a respondent is part of. Such granularity is out of the scope of this study. Problems do however arise from the use of perception data, especially in studying public matters. Rainey and Bozeman (2000), offers us an early critique of perception based studies in relation to public administration studies. They argue that that in the context of comparing public to private (as is often the case in the studies of public organizations, even in more contemporary studies see (Tkachenko, Yakovlev, and Rodionova 2018) & (Bilodeau, Laurin, and Vining 2007)) - the issue of social reflexivity is pertinent. Social reflexivity refers to a two-directional relationship between questions and answers. In the context of Rainey & Bozeman's critique, studying the differences between public and private organizations, will necessarily draw forth differences between them as the questions asked highlights the differences - public bureaucrats will lean towards the more social desirable answers. This is not completely unrelated to well-studied phenomena of "lipsyncing" or "window-dressing" see (Fang, Liu, and Zhou 2020). This is further argument as to why there is a need for studies utilizing actual procurement data in outcomes, moving away from perceptional.

What can be learned from the cross-organizational studies? Previous research have pointed towards there being certain differences stemming from the characteristics of an organization. From the literature cited in this study, there seem to be several ideas that points towards affecting outcomes in public procurement. First, it is the idea of values espoused by an organization. Gormly (2014) told us that a main challenge of procuring actors within SOEs is to achieve an often vaguely defined "value" - most often understood by the bureaucrats as economic value, which does not necessarily take away from sustainability, as "sustainability" could be part of the most economically advantageous offer. Several authors have however identified an issue with this explanation - it is that purchasing "sustainable" is often perceived as coming with a premium, it is seen as more expensive to procur sustainably (see Prier, Schwerin, and McCue 2016) (McCue, Prier, and Swanson 2015), (Brammer and Walker 2011). Understood together with other factors, I believe we have identified a multiplying effect. If the organization only having one "bottom line" (i.e. focused solely on economic factors), procuring sustainable is seen as coming with a premium - we should see an organization being more prone to rely on "lowest" price in choosing contractor. In regards to economics and organizational form, we have conflicting results from previous studies. Tkachenko, Yakovlev, and Rodionova (2018) tells us that organizations that are exposed to market competition are more efficient in their procurements - however Soudek and Skuhrovec (2016)'s result show that there is no effect on price regarding the organizational form of the procurer, but rather other factors are playing in (such as amount of bidders). We do not know, however, what this mean for sustainable procurements - are some organizations more prone to applying sustainability criteria due to "soft" - budget restraints making them less efficient (as defined in economic terms) as Tkachenko, Yakovlev, and Rodionova (2018) shows however better in terms of quality and sustainability? We do know that there is consensus in the research literature that applying criteria is more efficient, can we thus expect, in line with Tkachenko, Yakovlev, and Rodionova (2018) findings, that SoE's apply more criteria than non-SOEs?

Regarding the values espoused in relation to purchasing - Brammer and Walker (2011) told us that a large barrier was an unclear mandate of applying "ethical criteria" in procurement. If the values of the organization of solely focused on the economic bottom line, then the "ethical mandate" of the procuring professional would be even more difficult to apply. When there is a clash of values, the procuring professional will fall back on the organization values. Thus, organizations which focuses on the economic bottom line should be more inclined to rely on lowest price criteria. This is not due to the fact that lowest-price criteria being more efficient, on the contrary, applying criteria means possibility to take other factors than lowest price in, often leading (in the long term) to more value for money. However, there is recurring findings that "procurement is seen as "buying with a premium" (Prier, Schwerin, and McCue 2016). With this in mind, we expect contexts where procurement is spoken about in "enviro-social" terms, we should see more criteria being applied, whereas if economic-pragmatic values in terms of efficency etc, we should see more reliance on lowest-price criteria. In line with the stated purpose of this study to try to get at a more general understanding of public procurement, we will not be focusing on values espoused by single-organizations. Instead, we will assume that an organization's values relating to public procurement, mirror that of the "political", and we have good theoretical reason to believe so (see section "Bureacratic response to political cues"). Through this assumption, this study will be able to make larger comparisons and draw more general conclusions than previous literature.

What is there to be done? The previous research discussed here have been focused on barriers and enablers on a relatively small scale, being focused in single countries, single organizations or even individuals within organizations. Through the data they have used, mostly perception or small observation - they have studied these issues very "close" (i.e. on individual or organizational levels). What is left to be done, is to test these on a larger scale. There is also a need to understand "the political" in relation to the practice of procuring as most previous research is from law and/or economics (Knebel et al. 2019), how does the "political" who sets the goals of bureaucracy, affect its practices? Through using the TED.csv dataset and a fresh conceptualization of Brammer and Walker (2011)'s "policy context", the study will be able to see whether previous studies identified barriers and enablers break through on a larger scale. We will be sacrificing the granularity and the "experience-close" (See, Schaffer (2015) for a longer discussion regarding "experience near" and "experience far" conceptions) for a large-scale theory-testing study where the expectations, mechanisms and approach is formed through the extensive experience near-reserach done by previous authors. Through the conceptualization and measurement of a policy context in a comparable way, this study will be able to break out of country borders to cross-compare between countries and take steps towards answering the question, "what is the relationship between a country's policy context, country- and organizational characteristics and their procurement practices?".

3 Conceptual background, mechanisms and a general framework of sustainable procurement

This section will go through how to go about making sense of a comparative study of sustainable procurement outcomes on a more abstract level.

Gelderman, Paul, Brugman, et al. (2006) developed a framework for understanding the compliance with EU-rules. In it, four variables are depicted, that according to the authors affect the organizations compliance with EU-tendering rules. Interestingly enough, while the outcome is different from our research purpose, the variables that the authors identify as important could be interpreted as the same. Indeed, unfamiliarity with sustainable rules and regulations was identified as a significant barrier by Grandia and Voncken (2019) among others. Recall the respondents that Gormly (2014) interviewed stating that they avoid applying sustainability criteria due to the perceived costs, among other inefficiencies. Further, Yang et al. (2019) found organizational incentives playing a role in the application of sustainable public procurement, there is indeed significant overlap between this framework developed and the purpose of our study. It is thus apt to adapt this framework to the purpose of our study as to un-

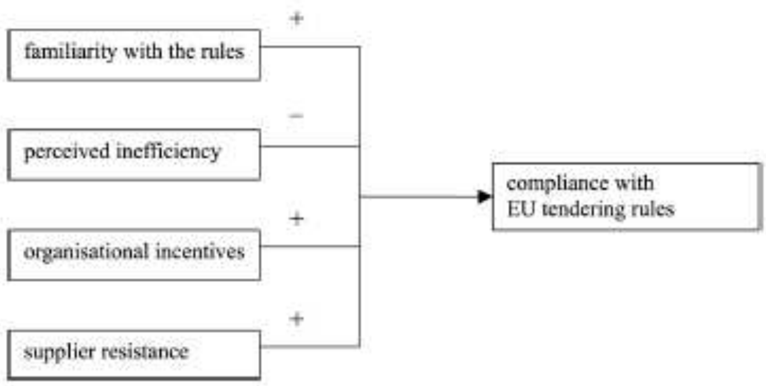

Figure 2: Gelderman et. al's framework for understanding compliance with EU tendering rules, (Gelderman, Paul, Brugman, et al. 2006, p. 706)

derstand how organizational differences affect the application of sustainability criteria in procurement. It is however not as easy to apply it without adaption to our study, as we are doing a transnational comparison. Thus, we must pose the question, how can we fit countries in to this framework?

Recall Brammer and Walker (2011) study regarding international comparative public procurement. The authors introduce an interesting concept that will be of importance to us. The framework that the authors employ make use of a concept called "policy environment" or "national policy context". The concept of policy environment has been employed elsewhere and it gives us a good insight to how we might go about seeing the relationship between a country's policy and the outcome it has on sustainable procurement. The authors suggest that sustainable procurement outcomes stems primarily from the pressure on the organization to apply sustainability criterias in their procurements - and that these pressures stems from the character of the policy environment.

3.1 Getting closer to a policy context

Brammer and Walker (ibid.) does not offer any closer definition of a nation's policy context, despite it being a significant part of their model, influencing every variable that supposedly affect sustainable outcomes in public procurements.

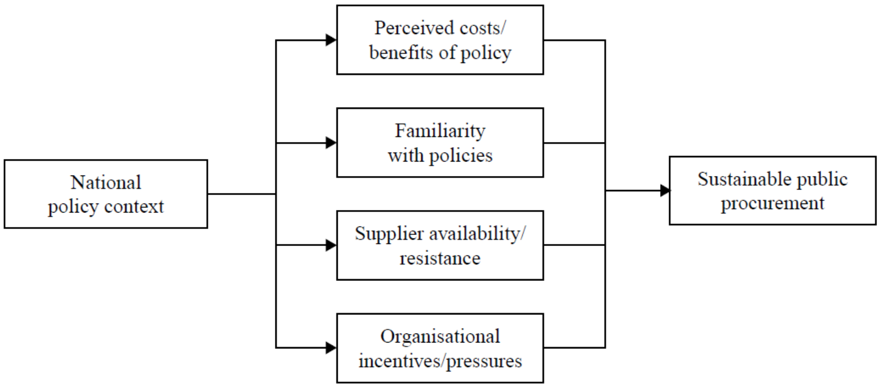

Figure 3: Brammer & Walker's framework for understanding what shapes sustainable public procurement, (Brammer and Walker 2011, p. 456)

In their previously named "flavours of policy contexts" the authors offer us examples of "guiding principles", defined by the government that give the public body's their "scope" about what principles to take into account when procuring. What follows in an example from the UK

> Consistent with a focus on "sustainable efficiencies" through a focus on whole life costing, "best value of money" gives scope to public bodies to take social and environmental policy objectives into account in their procurement activities (Brammer and Walker 2011, p. 458).

This means that rather than being focused on the legal and/or institutional context and quirks of the country, the author's conceptualization of policy context instead focused on softer matters. Softer matter such as values espoused by government through different forms that sets the objectives in their procurement goals. This is in line with other authors employment of the concept. Petrie (2013) for example, albeit in a different context, implies that a policy context is about "what is being communicated". In the same collection, the authors show how the policy context during the New labour government (1997 - 2001) employed austerity politics through the communication of different values in such messages as "responsible citizen's" etc (Kather 2013, p. 104). In the chapter, the authors discuss policy context in different times, implying that policy-contexts change and morph - affecting outcomes differently. Thus, I argue that Brammer and Walker (2011)'s rigid example of a policy context as something that is fixed is far too simple to understand a policy context - we need to make "policy-context" sensitive to time.

What this means for is is that we cannot focus on the institutional and legal framework of a country, but instead, we need to look elsewhere. Previous research have been doubtful of comparing results between countries, example is Patrucco, Luzzini, and Ronchi (2017) stating that "results are hardly generalizable" (p. 245). If we want to test this framework, in line with Grandia (2018) & Koala and Steinfeld (2018)'s calls for theory testing studies within the studies of public procurement, we need to somehow measure values espoused as well as adding time as a dimension of our measurement. Indeed, "what is being communicated" in a policy context is seemingly as important as the legal and institutional framework (Petrie 2013). Through the conceptualization and measurement of a policy concept, we will be able to compare results between countries, indeed, the concept of policy context allows us to "break out" from studying single cases, towards a comparative crossnational research designs.

Through scoring a policy context, using years as an indicator in a way that makes them comparable this study will improve Brammer and Walker's model through giving a more thorough examination of the "national policy context" - of a country. We shall also test our measurement using data from tender electronic's dataset (TED 2010-2019) to see how the measured policy context affects criteria application. Brammer and Walker (2011)'s study came out in 2011, when only one third of the EU countries had adopted a general strategy for sustainable procurement, which has changed until then. The time is apt for another study employing the concept of policy-context, but improving it through making it sensitive to time. Further, we should see movement towards more sustainable through the sustainable development goals introduced by the UN in 2012, where sustainable public procurement was identified as one of the sub-goals (United Nations, Sustainable Development goals 2012).

Now, how can a policy contact be measured? To answer this, a look into previous literature on the measuring of values is needed.

3.2 Previous research on "measuring values"

The "triple bottom line" is discussed heavily in various sources as a way to incorporate sustainability in public purchasing. What was first developed by Elkington (1994) as part of an accounting framework, has now become popular under several names in relation to sustainable public procurement. Within the European context, it is spoken of in terms of "the three pillars of sustainability" (European Commission 2017a). Or in other contexts, the "three P's" (people, planet, profit) (Utz n.d.). Despite the difference in names, the values they cover are the same, economic, environment and social responsibilities. While it is popular in the accounting literature, a buzzword within the sustainability literature, it does not necessarily help in the measurement. Instead, it is worth looking into the previous research literature on the measurement of values. It will become apparent that a popular way of measuring values does not differ significantly from the "triple-bottom-line" mentioned in the sustainable procurement-discourse, see for example (European Commission 2017a).

Measuring a value is a contested and controversial task. It has been done in different ways and in different universes. Within the cross-cultural management-literature, a popular way to structure the values was introduced by Jones (2007), where four key-dimensions were identified to comprise a culture. In later works - more dimensions were added, and as more dimensions were added, critique also increased. Capell et al. (2013) argues that within the "value-mapping"-literature, researchers attempt to improve measurements of the construct by adding more value dimensions for analyzing culture. Capell et al. (ibid.) argues that this is a mistake, that instead we should have parsimony as the guiding value when creating the constructs in line with the saying :

> "too many trees, there is a risk of not seeing the wood, which suggests that increasing the number of dimensions used to describe a culture, ends up damaging the integrated quality of the construct we want to measure (ibid., p. 506)

Pointing towards a successful value-measurement is one that is flexible and simple - should ideally be applicable to many different levels of analysis. For their study of mapping the values of public service organization in old and new EU member states, Capell et al. (2013) thus opt for the tri-axial model for structuring the values of an organization. This model of values assumes a universe of three axes consisting of:

(1) economical-pragmatic This axis refers to values that guides elements such as Competitiveness, Discipline, Economic growth, Efficiency, Quality, Productivity, Results, etc.

(2) ethical-social values associated with the ES-axis are Commitment, Consciousness, Generosity, Respect for people, Respect for natural environment, Sharing, Transparency, etc

(3) emotional-developmental Here we find values connected to Autonomy, Creativity, Initiative, Enthusiasm, Joy, Passion, Playfulness, etc. (Brillo and Silva 2020, p. 145), (Capell et al. 2013, pp. 510–5011)

The overall interaction and combination of these three axis is what constitutes the cultural "profile" of the measured unit, be it a company, agency or family. Through measuring these, we can get an organizational "profile", for example 80 % EP, 10 % ES, & 10 ED % values. When faced with a task or challenge, the organization will approach it in a pragmatic way, focusing on the economical results, more willing to sacrifice other values for the maximization of outcomes that are valued placed on the EP-axis (ibid.).

Now, how is the tri-axial model of values interesting for this study? As previously discussed, there is a need to delve deeper than previous research have done on the topic of "policy context" as laid forward in Brammer and Walker (2011). Sadly, the authors were only able to give us a "taste"(p. 457) - whereas this study will need to *score* a country's policy context. Through looking at the way the political discussion regarding public procurement is framed against the backdrop of the tri-axial model, will get us closer to inferring a policy context. This, in turn, will get us a better understanding of how sustainable public purchases might be shaped. The policy context is important in several ways, as we shall see in the section below - where the theorized mechanism is gone through in more detail.

The chosen measurement method allows measurement on a two-dimensional axis (see section "latent semantic scaling"), corresponding to two out of three "pillars" - or axes - of sustainable public procurement. The environmental and social will be at one end of the spectrum, whereas economic will be on the other. What this means is that while we take inspiration from the tri-axial model of value measurement and the value-measurement research, letting it us guide is in what indicator of values to look for - the actual measurement is not wholly true to the model as theorized by (Brillo and Silva 2020) (Capell et al. 2013). As both the environmental and social "pillar" of sustainable procurement fits under the E-S axis. Staying true to parsimony, and only measuring what is valuable to measure was the guiding principle in the decision to "cut" a dimension was taken. While emotional-developmental values is important in other contexts as identified by several authors (Brillo and Silva 2020) (Capell et al. 2013), I deem it better to focus on the EP-ES, a two-dimensional measurement instead to better capture the phenomena.

Now, let us turn to what mechanisms might be at play - how might a policy context affect whether an organization chooses to apply criteria in their purchases?

3.3 Bureaucratic response to political cues

There have been several published studies see for example, (Terman 2015), (Terman 2014), (Ringquist, Worsham, and Eisner 2003), on the mechanisms by which bureaucratic institutions interpret and implement political cues coming from the political. Embedded in the "bureaucratic response"-literature is the idea that "bureaucracies care" - they are concerned with the diverse preferences of different political actors, having the wrong reaction could have unfavourable consequences (ibid.), (Terman 2015)). Thus, several studies have looked at signals sent by government, defined in different ways, be they laws or political salience of issues (Ringquist, Worsham, and Eisner 2003). The same with bureaucratic responses, be they regulatory outputs or other.

In Ringquist, Worsham, and Eisner's 2003 literature review, the authors state that there is remarkable consistency in the results of previous research. The salience of an issue attracts political attention, which in turn directs attention to authorities that have capabilities to act upon issues with political salience. The authors claim that this mechanism is independent of what assumptions you make regarding the politicians, rational choice or not (Ringquist, Worsham, and Eisner 2003, p. 145). It is thus reasonable to assume that the way in which an issue is framed within a debate, (what this study calls a "policy context") have effects on the bureaucracy. The division between government and parliament is interesting due to the institutional framework of affecting bureaucracy is reasonably different if part of executive- or the legislative branch. There is thus a need to look at from where the salience is coming from, the legislative or executive branch?

> "we expect that salience [of an issue] will attract members of Congress to the activities of a particular agency and that agencies working in highly salient policy areas will experience more attempts at congressional influence" (ibid., p. 153)

The same mechanisms should go for parliamentary systems. Couple this theorized mechanism with McCue, Prier, and Swanson (2015)'s statement that public procurement is being used as a way to gauge government performance, and Brammer and Walker (2011) theorized policy context as shaping the sustainable procurement in a country, we should see an effect on an organization's tendency to apply criteria in their purchases if we are somehow able to measure how public procurement is being discussed. Especially since Terman (2015) found in their study that public purchasing and contracting are usually highly salient policy areas (p. 721). This should also be affected by how much control over the bureaucracy the political has, as political control increases, there should reasonably see higher effects on criteria applied. In the analysis, we shall also thus test for the political interference in bureaucratic matters by "the political", (see Cooper (2021) for a discussion on the dichotomy between "bureaucracy" and "political".)

Our theorized mechanisms are echoed in the sustainable procurement-literature as well. Prier, Schwerin, and McCue (2016)'s concretely summarizes the effect of government on the application of criteria.

> "if governments operate within an environment where they are responding to citizens who have a preference for sustainable public purchasing, they would have strong incentives to develop a reputation for sustainable buying. In effect, governments would try to differentiate themselves from other governments by signaling they engage in SPP to enhance their reputations in this policy area." (Prier, Schwerin, and McCue 2016, p. 315)

My argument is that through mechanisms such as those introduced in Terman (2015) and Ringquist, Worsham, and Eisner (2003), governments are trying to enhance their reputation with the help of the bureaucracy, as McCue, Prier, and Swanson (2015) also identified.

There is a looming problem, however. Prier, Schwerin, and McCue (2016) tells us that despite the research-consensus that applying MEAT-criteria (that is, "most economic advantageous tender" which allows purchasing with other considerations than price) instead of lowest price is more "efficient". There are still considerable disincentivizing structures built within the public sphere and bureaucracy (such as the inherent risk-aversion for uncertain outcomes, no meritocratic gains within the organization but large risks etc. For a classic discussion on bureaucratic incentives, see (Downs 1967)). Further, there is the idea that sustainable methods of purchasing is considered more expensive (see section "finding sustainability").

Thus, we expect, with the advent and salience of sustainability-issues we expect governments to try to enhance their reputation within the green/sustainable public procurement - area and try to incentivize bureaucracies to not rely on lowest price-criteria. However if not explicitly coated in green/sustainable and social terms within the political debate, we should expect bureaucracies to a larger extent rely on lowest-price criteria for two reasons. First if it is spoken only in terms of economics and efficiency, where focus on price and convenience takes precedence, we can infer that there is less political interference. Second, with the conception that applying sustainable criteria is more expensive (Brammer and Walker 2011), (McCue, Prier, and Swanson 2015), coupled together with the disincentives identified in Prier, Schwerin, and McCue (2016) and risks of adverse effects of not "listening" to the political (Ringquist, Worsham, and Eisner 2003), bureaucracies will rely more on lowest-price criteria.

Through looking into the literature on public procurement and the relationship between the political and the bureaucratic, a mechanisms through which a "policy-context" should be affecting the application of criteria in public-sector purchasing have been identified. It is thus time to discuss how we can conceptualize and measure a policy context.

3.4 Conceptualizing, recognizing and transparency in measuring a "policy context"

For the study, we are looking to see the choice of criteria in public procurement as more than within-organizational-isolated phenomena. Through the literature review, we have identified several potentially interesting variables that we can use to model the choice of criteria in public procurement and to look for variation between countries. One variable that we have put special weight upon throughout the review is that of "policy context" - it thus makes sense to ponder for a while, the conceptualization and operatioanlization of such a concept that, in theory, supposedly has large influence on an organization's sustainable purchases. How can we make sure that the measurement is a good one, especially one that is as vague and fleeting as a "context"? Hooghe (2017)'s discussion on "production transparency" tells us the importance of having a transparent measurement of political concepts, as well as a guide to how to achieve a fair production transparency (ibid., p. 7).

Throughout the study, in relation to measuring a policy context, we shall return to Hooghe (ibid.)'s process of achieving production transparency, moving from broad theoretical strokes, to particularities in our social reality. In this section, I wish to discuss the first three steps in the process - making sense of how previous social scientists have understood the concept, specifying the measurement concept, and unfolding the concept into dimensions.

We saw that previous social scientists (Brammer and Walker 2011) and (Kather 2013) had understood a policy context not as legal frameworks, but instead of what values were espoused in relation to the concept they were studying supposedly having social and political outcomes. In line with previous literature, we shall also focus in on values. Defining the "policy context" of public procurement as "how public procurement is defined, discussed and framed in political discussions and documents".

In specifying the measurement of a concept, we decided to structure it differently than previous literature, instead making use of the value-measurement research previously done, taking inspiration from the tri-axial model. In specifying the measurement, we also identified the need to to measure the values in a contested manner, i.e. in relation to each other - we will thus have a two-dimensional measurement, where we place the observation on a scale from the economic-pragmatical to environmental-social. We have thus started on the third step of moving on to empirics, the dimensions of the measurement. We identified a potential flaw in previous conceptualization of a policy context, namely that is not sensitive to temporality - the measurement need to be sensitive to time. Further, through Ringquist, Worsham, and Eisner (2003), we saw that it is of importance to divide the measurement along the lines of executive and legislative. These dimensions can be independently assessed but also aggregated to a total score (per year/per branch/ or a total score). We shall return to the last three points of production transparency in another section.

4 Hypotheses

With this deep dive into previous literature on public procurement, policy contexts and bureaucratic response, we are ready to formulate the hypothesis regarding how a country's policy context, country - and organizational characteristics affects an organization's tendency to rely on lowest-price criteria in their purchases.

The first hypothesis pertains to how the measured policy context affects an organization's tendency to rely on lowest price criteria, through the mechanisms discussed discussed in the previous sections.

H1: Countries and years where the policy context scores heavier toward the economical rely more on lowest-price criteria in their purchases.

H1b: Government polarity score have larger effects on the organization's tendency to apply criteria than parliamentary score.

H1c: Countries which have higher political interference in bureaucratic matters are more sensitive to the relationship between measured policy-context and outcome.

The second hypothesis pertains to the discussion about differences between organizations. Soudek and Skuhrovec (2016) and Tkachenko, Yakovlev, and Rodionova (2018) concluded that there should be differences in the "efficiency" and prices between an income earning units and non-income earning units - one stating that the hard budget constraints causing more efficient purchases (ibid.), and there is consensus within the literature that applying criteria is "more efficient" (Flynn 2018, p. 5). There is thus an expectation that:

H2: For profit organization rely less on lowest-price criteria compared to non-for profit organizations.

The third hypothesis is about Grandia (2018), Gormly (2014) and Brammer and Walker (2011) findings that a lack of knowledge and information regarding procurement was identified as a barrier to applying criteria in purchasing. The chosen countries (see section "countries selected") have in recent years established entities which purpose is to coordinate public purchases, supplying information and already written criteria. A coordinating entity both lowers the threshold of applying criteria, as well as reducing risks. There is this an expectation that:

H3: When an coordinating entity exist, organizations are less reliant on using lowest price criteria.

5 Data and measurements

It was concluded in the last section that the study will need to focus on the values to understand a policy context. For the measurement of values - it is of importance to situate values in competition with one another. The chosen measurement method allows us to put values into competition with each other through the use of "word-embeddings" (see section "measuring using latent semantic scaling").

Why is it worth scoring a government on the a value-model? Recall the framework employed by Brammer and Walker who argue that the national policy context is of great importance when a public organization decides whether or not to apply sustainability criteria in the purchasing. While the "objective" policy context is of importance, Brammer and Walker argues that it is not only the objective (that is, laws and regulations) that matter - the bureaucratic policy also needs to "fit" within the framework set by the context, and I argue that the values espoused in the context shapes the room for maneuver of the different bureaucratic policies - through both informal (political cues) and formal (regulations, laws and regulatory letters) mechanisms (see previous section "bureaucratic response").

5.1 Analysing a policy context

Here I shall go through the last three steps of Hooghe (2017) of achieving transparency in measurement. Under the next section, two points will be dealt with. The third: "Variable indicators" will be expanded upon, as well as the the rule used for ambiguous cases will also be discussed below.

5.1.1 Government and parliamentary documents

For the measurement of a country's "policy context", I downloaded documents from the parliament website of each of the countries chosen, they are as follows: Finland, Sweden and Denmark (see section "Countries chosen" for a deeper discussion on why these are suitable cases). All three of them had an option to search for terms to show documents that mention this specific word or collection of words. I searched for the word "public procurement" in the respective languages. I limit the search time-wise from 2009, the reason for this limit is that the TED-dataset (see section "Finding sustainability in procurement") contains observations from 2010 to 2019. In the collection of documents for analysis, I did gather from more years, however only 2010-2019 will be analyzed.

The Finnish parliament (Eduskunta) gave me about 500 hits for the terms with documents that contain the phrase "public procurement" (swedish: "offentlig upphandling") within the time period. Keep in mind that while Finnish is the main spoken language in Finland, Swedish is an official language and most if not all official documents are also translated into Swedish. When gathering these files, I divided them into years as well as if the document was produced by government or parliament. Government documents in the Finnish context are as follows: law proposals by the government (regeringens propositioner), government minister's messages and accounts(statsrådets meddelanden and redogörelser) as well as reports from the different government departments (berättelser). For parliamentary, documents coming from committées ("utskottens utlånanden" and "utskottens betänkanden", questions posed to government minister(interpellation), as well as answers from the parliament on government proposals. (see *Eduskunta, Riksdagsärenden och riksdagsdokument* (2021) for more information)

Table of distribution of the documents.

	Government	Parliament
2009	33	11
2010	45	24
2011	13	6
2012	28	7
2013	24	12
2014	39	22
2015	12	8
2016	30	17
2017	38	16
2018	46	27
2019	17	4
2020	29	18
2021	17	7

Figure 4: Table of how the Finnish documents that will be analysed are distributed. While there is data from 2020 and 2021, only the policy context up to 2019 will be analyzed.

The Swedish parliament (Riksdagen 2021) gave about 3000 results containing the phrase ”offentlig upphandling”. For government documents, I coded government law proposals (”propositioner” and ”skrivelser”), government and ministry reports (offentliga utredningar, departmenetsserier) and texts from different government committés. For parliamentary I coded questions to ministers (interpellationer), parliamentary law-proposals (motioner) as well as different documents produced by parliament committées.

	Government	Parliament
2009	35	138
2010	59	89
2011	98	122
2012	94	112
2013	155	137
2014	62	90
2015	90	123
2016	136	108
2017	136	133
2018	58	93
2019	32	70

Figure 5: Table of how the Swedish documents that will be analysed are distributed

The Danish parliament gave about 2000 documents results containing the phrase within the time period. In the Danish context, I coded document as government if they were: law proposals (lovforslag) as well as related documents (preparatory documents), government documents (”Redegørelser”) as well as answers from government ministers (Udvalgsspørgsmål). For the parliamentary - I coded questions to ministers (spørgsmål), committé documents (kommissionsforslag), as well as debates within the parliament (referater) (Folketinget 2021).

The year the document was produced and the class of the document is the two indicators of what dimension is being measured. The time dimension is straightforward, we do however miss monthly and daily indicators focusing in on years. The ”branch”-dimension does contain some borderline cases. There are two bordeline cases of coding I want to discuss here, that of debates in the Danish parliament, as well as interpellations in the Swedish and Finnish parliemnt.

	Government	Parliament
2009	39	3
2010	41	2
2011	120	4
2012	264	1
2013	277	2
2014	220	3
2015	213	2
2016	162	3
2017	116	6
2018	76	4
2019	87	3
2020	272	16

Figure 6: Table of how the Danish documents that will be analysed are distributed, only the data up until 2019 will be analyzied. Please note the relative lack of documents on the parliamentary side. One document can have large effect on the parliamentary score.

The Danish parliament "Folketinget" provides transcribed debates that takes place between MPs. It is not uncommon for these to also contain answers from minsters or MPs from governing parties - meaning that the answers could be a mixture between executive and parliamentary. These debates, however, have been coded as "parliamentary"-branch despite. Reason being is number's game, it is indeed more likely for members of parliament to use these debates rather than minister's as ministers usually have more channels to get their messages out. The second borderline case are that of interpellations in the Swedish and Finnish parliaments. They contain both a question, voted forward by parliament to a member of government. These contain a written question - as well as a spoken response (transcribed) from a live answer. Using the same principle as with the Danish debates, the government answers are *usually* more long winded than the question, meaning that the opportunity for the minister to say more things regarding the subject is higher. Thus - it was coded as government. The reason for this not being a problem in the Danish case was that the Danish-parliamentary website divided them up to "question to minister" and "answer by minister" making the coding straightforward. Indeed, the "who says more in the context" - was the guiding heuristic when coding these documents - a simple but effective rule for ambiguous cases as step 6 of Hooghe (2017)'s guide.

The last two steps will be discussed after the chosen measurement method, these are the scoring of cases in their respective dimensions, as well as scoring of cases.

5.1.2 Measuring using Latent Semantic Scaling

For the measurement of this, I have chosen an approach as laid forward by Watanabe (2020a) in his 2020 article. Latent semantic scaling is an semi-supervised text analysis technique.

First discussed will be the promises and pitfalls of treating text as data. Then we will move on to what it has been used for in previous research, last, a dive into latent semantic scaling and discuss why it is a fitting approach for helping us get closer to better conceptualization of Brammer and Walker (2011)'s proposed concept of policy context.

5.2 "Text as data" or "How to Automatically Analyse Text Responsibly"

Politics often happens in written and spoken words, indeed language is the tool through which politics is communicated, conflicts arise and resolve. In all areas of politics, there is usually written evidence, even political debates (a very spoken political phenomena) are often transcribed into text. It is to no surprise, then, that scholars of political science recognized textual analytical methods as an important part of understanding political phenomena. There is however one recurring problem, there is too much text to manually go through them all. Scholars have thus developed methods of analysing large volumes of text with the help of computers. Automated content methods have made the *"impossible possible"*(Grimmer and Stewart 2013, p. 268), a systematic analysis of large scale text collections. Grimmer and Stewart (ibid.) however state that due to the complexity of language, fully automated analysis will never replace the thoughtful reading of a human. Due to just this complexity, the authors state that the validation of the approaches are of utmost importance. In their 2013 article (ibid.) the authors lay forward four basic principles for employing a quantitative text analysis in the best way possible. Unexpectedly, the first principle they say "all quantitative models of language are wrong - but some are useful" (Grimmer and Stewart 2013, p. 270). What they mean by this is that analysing text as data is not the way language should be understood nor is it a good way to understand language - however it is very useful because it allows us to make inferences.

Looking at the first principle of employing text as data for understanding the concept of "policy context". This approach indeed prohibits us from making certain statements such as the precise nature of the policy context of a nation. It will, however, allow us to make inferences - it is useful.

The second principle, the authors call "Quantitative methods augment humans, they do not replace them". The employed model of measurement requires fine tuning, theoretically informed decisions and a good understanding of what kind of texts (data) are being dealing with.

Principle three states that "there is no globally best method for automated text analysis". There is indeed a need for a discussion of why this method is a good choice for understanding a country's policy context.

The last principle - and arguably the most important the authors aptly name "validate, validate, validate". We need to avoid blindly employing models and trusting the results - there is thus need for a validation of the output, i.e. asking ourselves the question "does this output make sense?" and testing it in various different ways. This study will employ two validatory tests, one cross-lingual and one measurement.

To make sure that we are employing text as data model responsibly, we will go through each one of these principles under the section "latent semantic scoring", see how it applies to the case, how well the chosen method fits the data and how well it works as a measurement of policy context.

The most common approach within the social sciences, due to its' technical simplicity, are different dictionary approaches. A dictionary approach uses the rate of which chosen terms appear in a text to classify or score documents.

In theory, these terms connotate a concept of interest - something that is being measured. Dictionary approaches are necessarily measuring latent concepts (Grimmer and Stewart 2013)

A popular dictionary approach is scoring documents relating to a sentiment - so called "sentiment analysis". In such approaches, a dictionary consisting of words such as "good" and "bad", as well as synonyms of these. These words are given a sentiment score, in effect, a dictionary is divided in two. With this very simple approach, you can see whether a document scores positively or negatively - a simple but powerful technique. Such approaches have been applied with fruitful results on party politics through their manifestos (see Laver, Benoit, and Garry 2003).

Watanabe (2020a)'s tells us, however, that a dictionary approach is completely dependent on the availability of dictionaries pertaining to the studied phenomena in the language suitable with the data. Most dictionaries are unsurprisingly only in English - as it is one of the most used languages in the world. Thus several researchers have been working on dictionary availability, translating or comprising similar dictionaries in different languages (see Proksch et al. 2019) & (Matsuo et al. 2019). New dictionaries are made by adding or removing words from existing dictionaries, always starting with a "base" - due to the cumbersome nature of creating a new dictionary, trying to capture thousands of words that pertain to the concept one is trying to capture is often very resource demanding.

In our case, employing a dictionary approach would encounter two large problems. First - we are trying to capture a phenomena where no known dictionaries exist - that is, a dictionary corresponding with terms corresponding to the values on the discussed two-dimensional scale. Second, we are trying to do it in two different languages spanning three different countries and 10 years.

Instead - we shall be opting for an approach and method developed by Watanabe (2020a) named "Latent semantic scaling".

5.3 Latent semantic scaling

Latent semantic scaling (LSS) is an semi-supervised machine learning technique. Text-analytical techniques are labeled unsupervised or supervised depending on if they require manual coding of documents or not. A popular unsupervised method is "Wordfish". An example of Wordfish is action can be found in Gross and Jankowski (2020) where the authors use a wordfish approach on German local party manifestos, and it seemingly places the parties on a surprising "national" vs "regional" dimension. Wordfish algorithms are however always subject to ad-hoc interpretation of the results.

Latent semantic scoring does not require any manual coding of documents, it is however a semi-supervised technique. The one large strength of the LSS is that it allows the user to define the dimensions on their own. Indeed, through the use of "seedwords" LSS is able to filter out irrelevant dimensions or categories in documents - allowing for the analysis of large and noisy collections of texts. For the measurement of a policy context - the seedwords will be chosen with reference to the social-environmental(ES) or the economical-pragmatical (EP). Positive values will be Economical-Pragmatical, and negative Environmental-social, they will thus be in competition as previously discussed. Each seed word is intended to capture the attitude or behaviour towards the discussed subject - in our case public procurement. The possibility of seed-words and other reseracher-controlled factors is what makes the model "semi"-supervised.

How then, will the model know where to measure these? Through the use of "model terms" which I supply. These model terms consist of words that are relevant to the subject, public procurement, guiding the model *where* to measure the sentiment.

> "computes polarity scores of words by accurately estimating the semantic proximity of words by SVD (singular value decomposition [auth. note]) of a document-feature matrix, in which original documents are segmented into sentences to capture similarity of words in immediate local contexts." (Watanabe 2020a, p. 7).

In other words - it treats the data not as individual words, but as sentences.

The model then looks in these sentences which always contain the model terms, for us terms relating to public procurement, to see what other words appear close to the chosen seed words. These "close words" are then given their own score based on their closeness to the seed-words. A word's "polarity score" is given by

$$g_f = \frac{1}{|S|}\sum_{s \in S} \cos(v_s, v_f) p_s$$

Then, through the words given polarity scores, we can then calculate a document's polarity score through how often they appear in the documents. The polarity score of a document is given by

$$y = \frac{1}{N}\sum_{f \in F} g_f h_f$$

Where hf is the frequency of words and N is the total number of words in the document, excluding words that are not in the model. Note that the document's polarity score are recentered around the mean = 0, and SD = 1 to easier interpret this. As the polarity-scores will be divided on a yearly basis, the interpretation of this number is relatively straight forward. Please understand negative values as public procurement being discussed in terms of social and environmental-values (that is, EP-values, borrowed from the tri-axial and triple-bottom line- literature, see (Esteve, Grau, and Valle (2013) & Elkington (1994)), whereas positive values, more economic and efficiency is espoused in relation to it. Around zero, it is best to understand it as a balance between how it is discussed or it is terms that the LSS-model did not pick up on. These scores can later be used for future regressions as an operationalization of a country's policy context. (See section "analysis").

5.4 The four principles of responsible text analysis and the chosen case

In this section, we shall go through the four principles of applying automated textual analysis (Grimmer and Stewart 2013, p. 269) responsibly and discuss its' applicability to the case.

Recall the previous discussion on the difficulties with finding dictionaries relating to new concepts as well as in different languages. Through the employment of the LSS-model instead, we will only need to supply a handful of carefully chosen model words and scored seed-words. Few selected words are easier to validate in all languages, and is preferable than translating whole dictionaries, as seen in Proksch et al. (2019) & Matsuo et al. (2019). Especially when the dictionaries would pertain to a previously unstudied (dictionary-wise) phenomena as this is. Indeed, the technical usefulness of this model is overwhelming. How we will make use of this model is as a measurement from Brammer and Walker (2011) borrowed concept of "policy context". Defining the concept of procurement policy context as "how public procurement is defined, discussed and framed in political discussions and documents", and operationalizing it as "the polarity scores of political documents, when defined through the weighing of seed-words on the EP - ES dimension" - I believe that the model, while wrong, is the best approach. Through the reasons stated above, I do think we reasonably uphold the first principle.

Principle 2 & 3 - Models do not replace humans and there is no universally best model.

Another strength with the LSS-model over a fully automated one such as wordfish, (see for example Gross and Jankowski (2020) and Lo, Proksch, and Slapin (2016)), is that there is more researcher agency. Through the selection of model- and seed words we are in effect forcing the documents on to the chosen dimension. It is thus very important that this dimension is theoretically rooted. In the chosen case, the theoretical root is the tri-axial value model & the three bottom-lines of sustainability (see, (Esteve, Grau, and Valle 2013), (Elkington 1994)) that has been employed in both public- and business administration. We do, as previously mentioned, sacrifice one of these dimensions, a compromise between what is gained through the model and theory. A more conservative approach would be to search government documents that has "public procurement" in the title to make sure that the documents only pertains issues related to public procurement and public procurement procedures and scoring these depending on what terms relating to a value on connected to one of the axis using a dictionary approach. The dictionaries would be constructed with terms connected to values as shown by Capell et al. (2013) & Brillo and Silva (2020) and translated to the language of the documents in question, all the terms would have been chosen by the me. Strengths of this approach would have been more research agency over the dictionaries, as well as the incorporation of all three of the value axes. The weakness would be the limitation of data, the same search method but only title yields less than a third of the results on the Danish parliament's website, however the biggest weakness (or rather, the biggest strength of my chosen approach) is that of language as previously discussed (see section "text-as-data". While multilingual dictionary - approaches have yielded impressive results (Proksch et al. 2019, see), a direct translation of words used in a dictionary approach introduces certainty, while the translation might be the same, it is not certain if the *meaning* of the word is the same. Opting for the latent semantic scaling is indeed a stronger approach in this case. The chosen approach also picks up environments *where* public procurement is discussed as well, which tells us about how public procurement is seen, in terms of environmental social - or economic pragmatical?

5.4.1 Principle 4: Validation of the model

The fourth, and arguably, the most important dimension is that of validation (Grimmer and Stewart 2013). I will test the model in two different ways. Through applying the model on data (texts) where we know the outcome of the application, and if the results fit with the expectations.

To check for cross-lingual validity, I will employ two LSS-models with the chosen seed words in Danish and Swedish on two corpus (a corpus is a collection of texts) of texts that are professionally translated. I will use the same documents from the European union, but translated into Danish and Swedish to see whether they get similar polarity scores - as well as how they are distributed in the document. If they do, the seed scores given to words that the LSS-model has calculated makes sense, leveraging professional translations.

To check whether or not the seed scores given to words actually correspond to the ES and EP-dimension, I will employ the same model to two sets of different texts. One being a company or agency's annual report. Annual reports are a company's way to disclose their finances with a heavy economic focus - meaning that they should score more on the economic-side of the measurement. To contrast this, I will employ the same model to a company's responsibility and/or environmental report - these should in effect score higher on the ES-dimension. This should give a substantial validation that the seed-scores given to words and document actually score what they are intended to score, that is words and documents along the chosen economical-pragmatic or environmental-social dimension.

To check for cross-lingual validity, I downloaded two official documents from the European Union, translated to Swedish and Danish and applied the LSS model. I translated the seed-words through the Neural-machine translation service Deep.l (2017) - but manually cross-checked the model-word so that they overlap. From the following density plot - we see that the values when predicting the polarity scores of the documents overlap very well, meaning that most values significantly overlap when predicting documents. The word seeds are very similarly distributed in the four documents. as well More precisely, what this shows is the sentiment scores of the words and their appearances in the two documents, applied to the same documents used to calculate the seedwords.

The difference is that when calculating seedwords, the model sees the the documents as collections of sentences, one unit of analysis would be a sentence (Watanabe 2020a, p. 9). For then predicting the polarity score of the document, however, I restructure it back to a "normal" document feature matrix, where the "atom" are words instead of sentences. Then, using R's "predict()" function which invokes a specific method for this certain kind of LSS-model (Watanabe 2020b). Since we used four texts for this validation, when summarizing, I group them by the document as to get one score for each document.

The polarity score of the Danish document directive 2014/24 turned out to be: -0.0002395335, 2014/25: -0.0004769248. For the Swedish document directive 2014/24: 0.0010308155, and the document 2014:25 turned out to be:

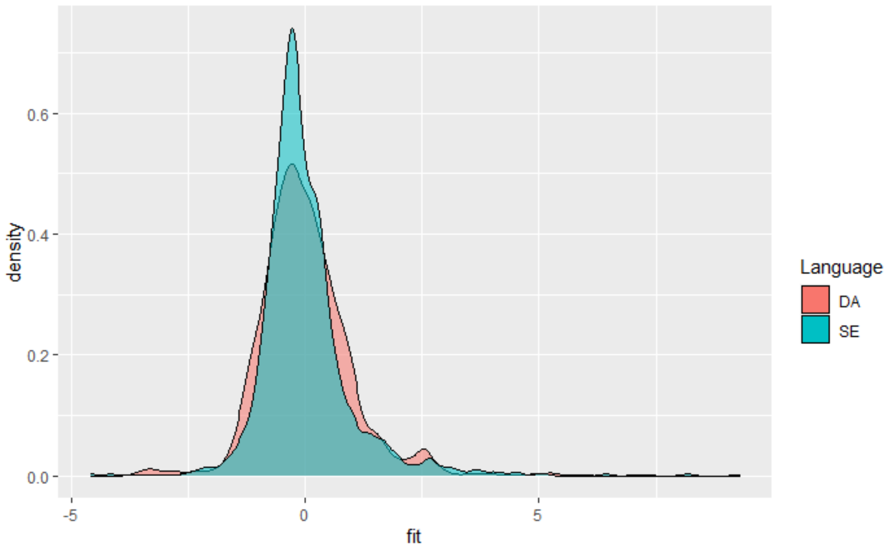

Figure 7: Density plot of the polarity scores in the four documents. Note the significant overlap. Both languages are centered around 0 with about the same distribution.

0.0005230424. There is only a minuscule difference between the polarity scores of the documents - meaning that applying the model on different languages does not have significant effects on outcome of the documents polarity scores. Paying careful attention to the seed-words and model word, the employment of this model will have comparable results across three countries, and two languages with the help of the 4 professionally translated documents.

The second validation test for this model is whether or not it properly measures on the chosen two dimensional scale. For this I have taken part of Händschke et al. (2018)'s large collection of corporate annual financial reports, and social and/or environmental responsibility reports. Creating a collection with reports from two companies, covering several years. For this we will need to use a model-word which is present however unrelated to the dimension as to not affect the polarity scores. As the documents are in English, the chosen model word and seed-words will be translated into English. We know that the chosen method of translation seed-words only has minor impact from the previous validation test. The model word chosen for this test is "contract" - as it a word that is a highly loaded word, on both of the dimensions. It is as natural think about "sustainability-contract" as "business-contract". We expect that annual reports (AR) will score higher on the economic pragmatical dimension(positive values), whereas the corporate social responsibility (CSR) reports will score more towards the ES dimension (negative values). See figure 8 for the chosen seed words, and figure 9 for a visual representation of the results structured by year.

word
sustainab*
respect
environ*
green
recycl*
carbon*
warming
respons*
financial
econom*
growth
price
efficiancy
competit*
results
profit

Figure 8: Seed words chosen for validity test 2, the model word chosen is "contract". The "*" operator makes the terms pick up different endings. Sustainab* will pick up on both "sustainability" and "sustainable"

What follows is a graph showing the trend of the polarity scores of the documents (n = 10) - structured after the year they were released.

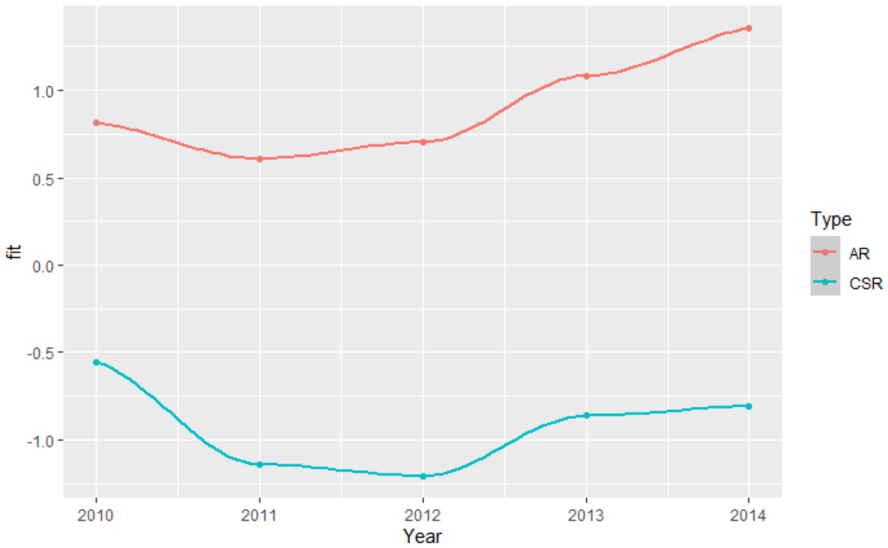

Figure 9: Plot of the polarity scores of the documents, divided if they are a Corporate social responsibility report (CSR), or Annual report (AR).

We see that the model correctly scores the annual reports more on the "E-P" side of the spectrum, whereas the CSR clearly score on the more "E-S" - dimension.

Through applying the model to two different cases, where we knew what the outcome was supposed to be, we have successfully validated the model from two different angles relevant for the study. Both the intra-lingual, answering the question "can we trust the polarity scores of documents in two different languages?", and "is the model able to properly place documents on our chosen two-dimensional scale?"

5.5 Scoring through latent semantic scaling

Hooghe (2017)'s two steps that are yet to be discussed is the scoring of cases within their dimensions, as well as the total scoring of the case. The earlier discussed polarity score, measured on the EP-ES scale, surrounding the "public procurement debate" in relation to all other words in the document will be the score that we give to the cases. The scores will be divided by year and branch, meaning the measurement will be applied to government / parliamentary documents separately (per year) once we have the polarity scores of the words through the LSS-model applied to all of the documents. The scores given here can then be summed together to get total yearly score regarding if public procurement was discussed mostly in EP-terms or in ES-terms. The largest obstacle of this approach is however the information used to score these. The dictionary created (i.e. the word scores that comes out of the application of the LSS-model) is out of my control. While we provide it with the seed words and the model terms, it is not possible for us to control every score for every term, especially not for such large data. This means that while we know where and how the information comes to us, the exact process of these scores is for us "black boxed". There are ways scholars are trying to go inside this black box of dictionary approaches, see Deng et al. (2019) for example, this is however unavailable for the chosen method. While black boxed, through the validating tests gone through, both the cross-lingual and cross-dimensional, two tests where the outcome was known beforehand and tested the model, and the model did confirm these expectations. In terms of measurement validity, I believe the chosen method is the most fitting one. In terms of measurement and production transparency, however, the model leaves a lot to be asked in Hooghe (2017) in the fifth step, "the scoring of cases".

We have, with the help of Hooghe (ibid.), gone through the six steps of how to move from a theoretical concept to an empirical measurement in a transparent way. This transparency allow others to probe at the validity of our measurements in ways that I might have missed - it is also a prerequisite for the possibility of replication of the theorized "policy context".

Now that I have conceptualized the concept of "policy context", discussed its definition and how I will go about measuring it, it is time to move on answering the second overarching research question. How can we understand the relationship between the measured policy context, and the choice of criteria - application in public procurement? For this, we will need data on the criteria choices in public procurement. This study will utilize the "Tender-electronic Daily's dataset" to answer this question.

5.6 "The Tender Electronic Daily" - dataset

The "tender electronic daily" is the online version of the "supplement to the Official Journal" of the European Union, which publishes hundreds of thousands procurement award notices a year (About TED - TED Tenders Electronic Daily n.d.). It should be noted, however, that most of what is being published on the TED are above the procurement thresholds. "Above procurement threshold" means calls for tenders whose monetary value exceeds a certain amount that governed by EU-law and is considered to be of "cross-national-interest" (TED 2010-2019) It is nonetheless considered good practice to publish even those calls for tenders below the threshold- meaning that while over the threshold calls are guaranteed, there are a substantial number of those under the threshold as well (ibid.).

For the purpose of this study, I downloaded the contract award notices CSV datasets and combined the data from 2010 to 2019 through matching columns, filtering out other countries in line with the chosen cases (see section - Countries selected). Procurements notices are sometimes divided into several lots, while referencing the same Contract award notices (CAN) are divided into several rows. For the purpose of this study - the lot division is unnecessary, as we are interested in the criteria the procuring organization puts on the potential contractor. To rid this redundancy problem, I used the "TED notice URL" to filter out the lot divisions - leaving me with a data set with only unique CANs. The most important columns are the ones containing information on criteria - what criteria are considered in the selection process, as well as their weights. In making this data meaningful for this study, I coded observations that rely on a lowest price criteria through the help of the "criteria code" column, and "price weight"-column. The column "Criteria code" contains a variable that can either take the form of "L" (for lowest price) or "M"(For: most economically advantageous tender) for their award criteria. "L"means that the procuring actor will choose the tender with the lowest price. If the price weight was "100", it was coded as relying on "lowest price" as well as if the award criteria was "L". There were no ambiguous cases in the data.

5.7 Finding sustainability

How can we know if a procurement is sustainable or not? The TED-dataset offers us several pieces of information regarding the observed procurement. The most important column, however, is the criteria code. It tells us whether the procuring actor applied criteria or relied on the lowest price in purchasing the good or service. In other words, it shows whether or not sustainability or quality is taken into account when procuring. Indeed, through looking at the outcome, "are criteria being applied or is the procuring actor only choosing by price" we are getting an insight to if sustainability was taken into account. This is the outcome that we shall be looking for in the data - because if the procuring actor relies on lowest price - it does mean that sustainability was not taken into account. Within the context of public procurement, two assessment methods are generally talked about. There is consensus that MEAT or "most economically advantageous tender" - criteria (rather than lowest price) is considered "better" in terms of value for money, and the EU uses the application of criteria in procurement as an indicator for "performance" of procurement in the EU. The goal with these measurements is to minimize the proportion of lowest-price purchases (Flynn 2018, p. 5). While the research literature agrees, there is still a conception among procurement professionals that applying social and environmental criteria is considered more expensive (Brammer and Walker 2011), (McCue, Prier, and Swanson 2015) and thus without explicit encouragement from management, professionals might be wary to apply such criteria. .

While the TED-dataset is the best collection of data on purchases in Europe, there are remaining problems. Since the TED-dataset is wholly reliant on the procuring organization reporting back through public procurement standard forms (SIMAP Standard form 2014). They are filled out by the contracting bodies - and it is not unusual for them to be incorrectly filled out, meaning one have to be very careful in analyzing and drawing conclusions from the data (TED CSV Advanced Notes 2020). For a longer discussion on the potential pitfalls, see section "limitations of the study".

For this study - the issue that this incorrect reporting brings is: how will we know what kind of criteria are being applied to the procurement? The idea behind the standard form is for the contracting body to write what criteria were taken into account when contracting - as well as their weights, we can see this in the two columns "Award criteria" & "Criteria weights". Most of the times, this information is missing due to either being ignored or inputted incorrectly as it on the SIMAP Standard form (2014) is a free-hand column, same as for the criteria weights. In an ideal scenario, had these been filled out correctly, we could have coded sustainability criteria on their own, technical quality-criteria on their own and answer more fine grained questions such as the effect of a strategy on one of these specific outcomes. Due to missing or incorrect inputs, drawing these conclusions would only be possible in a handful of cases, and they would be require extensive manual coding; which would not be possible because if we go back far enough, the data on the procurement is no longer available where they once were published.

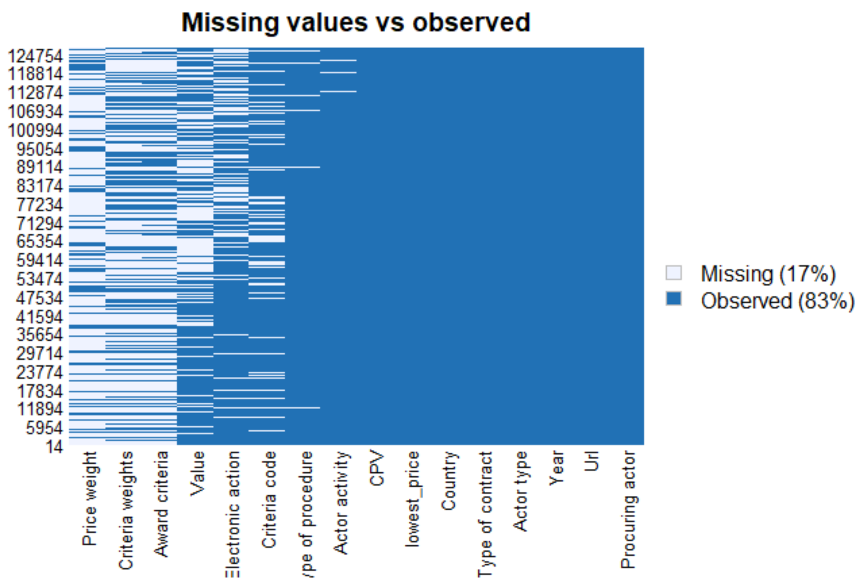

Figure 10: Map of missing values per column in the TED.CSV dataset, columns have been renamed and filtered to only contain the Nordic countries, 126 644 observations.

In figure three, we see indeed that the two columns with the most missing values are "Price weight", "Criteria weight" & "Award criteria". At the cost of granularity in output, we shall opt for a much simpler however safer route. Using the data from the criteria code - we are able to see if the procuring actor used other criteria than lowest price. Important to point out is that with this method, we do not know whether these criteria are sustainability related or not - introducing an uncertainty about whether it is an apt method for studying the "sustainability" outcomes of procurement. Turning this around, applying criteria is identified as a pre-requisite for achieving any social or sustainable goals according to several authors and institutions (see Stake (2017), European Commission (2017b), Nordic Council (2021)), studying obstacles that are in the way of an organization's applying criteria in their purchasing is as important. Through re-phrasing the questions like this, we can start to understand why this is relevant for furthering the understanding of how we can best achieve sustainable outcomes in procurement.

The strengths of this approach is several. First it is the obvious reporting strengths being a binary box in the standard form (SIMAP Standard form 2014) - it is easy to report meaning that the frequency of answer throughout the year is higher but most importantly, it is difficult to report incorrectly. The method introduces uncertainty regarding if the criteria is related to sustainability, but reduces uncertainty regarding the problems of reporting. This means that the analysis can use more observations to test the from theory inducted hypotheses. While we lose outcome granularity in one way, being able to utilize the "grandness" of the TED-CSV dataset (that is, one variable that has been consistent throughout the dataset's existance), we are able to make better comparisons, not only time-wise but also organization wise.

5.8 Countries selected

If we are to understand how values espoused by government and parliament affect whether a purchasing organization decides to apply sustainability criteria in their procuerments or not, it makes sense to control for other factors. In this study, I thus opt for choosing cases that are within the same "administrative tradition". One especially interesting commonality that will help is an administrative norm is political loyalty if we want to understand how a government can affect the procurement- policy context. Thus, ideally, the cases should also have similar political systems.

I have decided to choose three different countries that match up well for our study. Greve et al. (2020)'study focuses on the "Nordic adminstrative model" and an effect this has on administrative reform. The authors study the administrative reform-strategies adapted by the different countries, while not irrelevant for our study, this is not the reason for mentioning this. The authors discuss the similarities and differences of the Nordic countries - and they conclude that the administrative traditions match up well, where political adherence is the main administrative norm. Further, they share similarities in other characteristics such as the relationship between the "public and the private", development level, economic structure etc. There are notable differences, however, such as Iceland & Norway not being part of the European Union. For our study, as the EU regulates much, for example through Directive 2014/24/EU (2014), of the procurement framework, this could possible be one intervening variable, that I am unable to control for. Further, as the Tender Electronic Daily's dataset is wholly focused on European countries, it would mean less observations for countries outside of the EU that could lead to inference issues. I have thus opted to exclude non-EU-members from analysis, meaning that Norway and Iceland will not be covered in this study. What is left is Sweden, Denmark and Finland that will be used to test whether the country's policy context affect how often public organizations apply other criteria than lowest price.

Why the "most similar" - logic is fitting in the chosen case is because we need to control for the bureaucracies' relationship to the political. A common understanding is that the administration lies outside of regular politics however the relationship is close. As Woodrow Wilson noted in 1886:

administrative questions are not political questions. Although politics sets the tasks for administration, it should not be suffered to manipulate its offices (Wilson 1886, p. 210).

If we believe this dichotomy between the political and bureaucracy, then looking at the relationship between them is a worthwhile effort if we want to properly understand public procurement.

Regarding the "Nordic administrative tradition" and politicization of the bureaucracy Cooper (2021)'s study offers us a theoretical insight. A administrative tradition is defined as a

more or less an enduring structure pattern in the style and substance of public administration in a particular country or group of countries. The nature of an administrative tradition is manifested in both a country's formal and informal institutions.

Cooper (ibid.)'s study offers us an insight into the relationship between the "political" and the administration in different administrative traditions. The authors employ an expert survey regarding the interference of the political into the administration and run several analyses. The results show that the Nordic administrative model consistently show up as one of the least-interfered administration out of the chosen cases. These results are echoed in Lember, Kattel, and Kalvet (2013) chapters on Sweden and Denmark in relation to public procurement. Why choosing the least-interfered with administrative-traditions, if we want to understand the relationship between a country's measured policy context and the outcome of criteria application in procurement? It is a valid question, but there are strong reasons to do so. If we are to trust Cooper (2021)'s results, with three different least-likely cases and if we confirm a relationship between a policy context and criteria application - we will be able to draw strong inferences to other cases. Indeed, Levy (2008)'s well cited paper on case studies talks about the aptly named "Sinatra-inference" where if a causal - mechanism "can make it here, it can make it anywhere" from Sinatra's song "new york, new york". The Nordic countries within the European union thus make sense to study, as if we are able to confirm a relationship theorized by Brammer and Walker (2011)'s framework of sustainable procurement, this mechanism should also exist elsewhere.

5.8.1 Sweden, quick facts and LSS score

Sweden has been known to be a leading country when it comes to using procurement as a policy tool for boosting innovation within different sectors. The country employs, what the authors of Lember, Kattel, and Kalvet (2013) calls a "no-policy-policy". This chapter was, however written before the "national procurement strategy" of 2016 was implemented. Thus the truthfullness of the "no-policy-policy" is questionable. However, the authors do state that the Swedish "institutional set-up" with state-authorities relatively high freedom from the the executive branch. Such freedoms include the illegality of minister's to overrule an authorities' decision - including procurements. A national procurement agency was established in 2015 with the explicit purpose of "strengthen the strategic weight of public procurement" (". Upphandlingsmyndigheten n.d.). From 2015 and onwards, Sweden will thus be coded to have a coordinating entity. The national procurement strategy that was set in motion in 2016 and implemented in 2017 takes aim on seven goals. There are several goals, from procedural, such as having a procurements procedures that are legally sound, to social goals such as fostering innovation, environmental & other social issues. The strategy is mostly aimed at state authorities - but state that it is a prompt to all procuring actors to create their own governing documents for attaining these same goals. In 2019, 68 % of procurements within Sweden was done at a municipal level, whereas national level was 19 % (*Statistik som utvecklar den offentliga affären — Upphandlingsmyndigheten* 2021).

The choice of seedwords correspond to the "bottom lines", as is prominent in the accounting literature, as well as laid forward by the European Commission in their definition of "sustainable procurement" as well as identified by (McCue, Prier, and Swanson 2015) as having an effect on procurement professionals when purchasing. As LSS requires us to have two dimensions, the social and environmental will be on the negative side, whereas the economical will take the positive values. The seed words are connected to these values, as when wanting to talk about the importance of the environmental baseline, "green" is most likely used, and is thus seeded accordingly, for example. Find below a textplot with the seedwords highlighted together with other words scored.

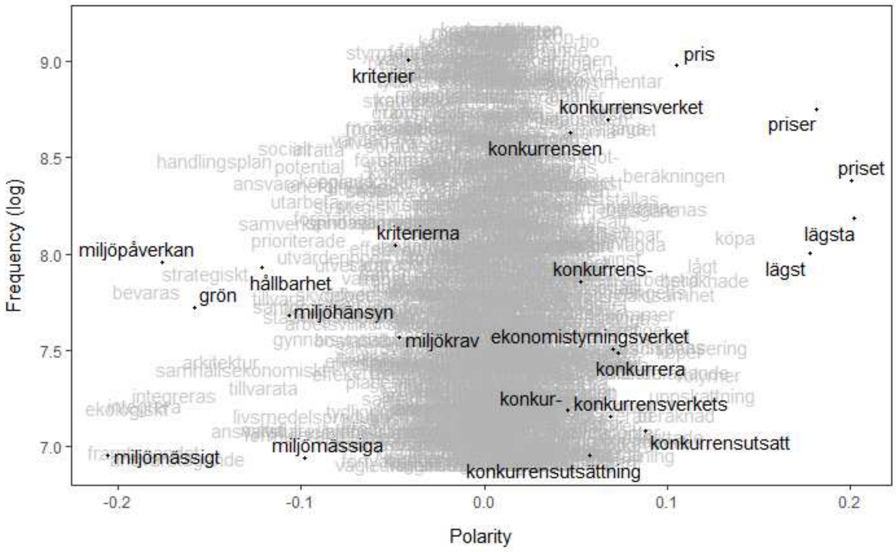

Figure 11: Text plot with the seedwords highlighted. Note that the closer words are together on the x-axis, the more similar they are in polarity score. Y-axis shows the frequency.

The polarity score of procurement in the Sweden follows below.

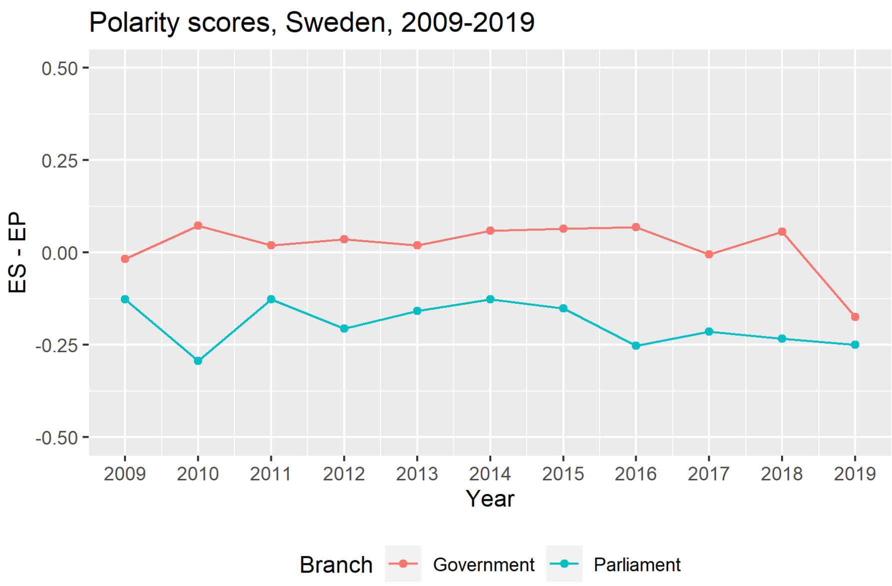

Figure 12: Plot of the polarity scores in Sweden grouped by year, and divided by executive / legislative. The Y-axis is structured from ES (Environmental-social) - and EP (Economic-pragmatical)

We see that the executive branch have been speaking relatively neutral about procurement, centering around zero. This does not necessarily mean that procurement is coated in terms that to not pertain to any specific value, but it might mean that in documents throughout the year, the government is balancing the values in different documents. The legislative branch, however, is slightly more focused on coating procurement with ethico-social terms. This is somewhat expected as legislative-branch tend to be less neutral as being in opposition is as much about critique of executive as it is presenting alternative policies, see (Garritzmann 2017), (Brack and Weinblum 2011).

5.8.2 Denmark, quick facts and LSS-score

Denmark, in contrast to Sweden, has historically not focused much on using public procurement as a policy tool (Lember, Kattel, and Kalvet 2013, p. 109), it was instead focused on cost-effectiveness and efficiency. The authors state that recently (at the time of writing), Denmark's public administration have been implementing policies to better help foster innovation and other social goals when purchasing. In 2012, procurement in Denmark went up to equivalent of 16 % of GDP, and in similarly to Sweden, local- governments and public corporations represent the largest share of purchases in the country (Lember, Kattel, and Kalvet 2013, p. 113). Denmark has a non-profit company, owned by the state that streamline and develop public purchasing, much like the Swedish equivalent. It was established in 1994 (Facts about SKI (2021)), all relevant years for this study, Denmark will thus be coded to have a coordinating entity. Below follows a textplot with the chosen Danish seed-words, recall that these seed words are related to a "baseline" (McCue, Prier, and Swanson 2015).

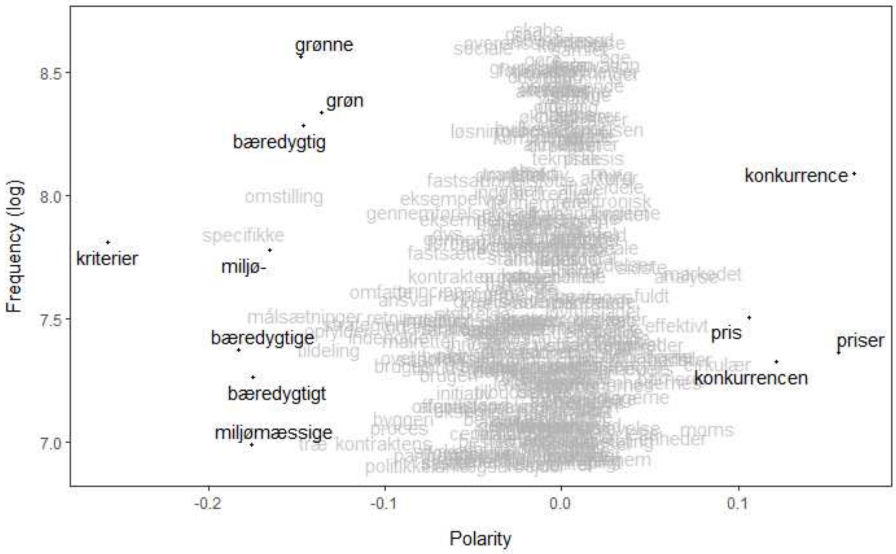

Figure 13: Textplot with the seedwords highlighted. Note that the closer words are together on the x-xis, the more similair they are in polarity score. Y-axis shows the frequency.

The polarity scores of procurement in the Danish debate follows below.

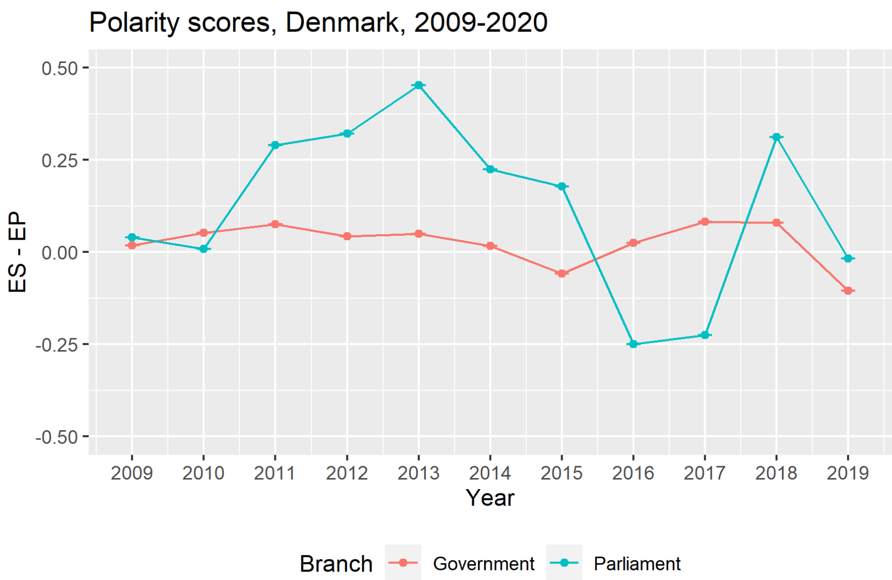

Figure 14: Plot of the polarity scores in Denmark grouped by year, and divided by executive / legislative.The Y-axis is structured from ES (Environmental-social) - and EP (Economic-pragmatical

We see, that the executive side, once again is more balanced when it comes to what terms are used in relation to public procurement. The parliamentary does larger movements over the spectrum. Recall the table with how the documents were distributed, there was a general lack of documents from the Danish parliament - meaning that the score of just one document can have large effect, as we see in the plot above.

5.8.3 Finland, quick facts and LSS-score

Finland did not have any common framework strategies for public procurement up until late 2020, where the ministry for agriculture presented a national strategy for procurement. The national strategy was aimed towards municipal and regional purchasing, but also including state authorities and agencies. For Finland, we will thus not see any effects of the introduction of a streamlining, central "knowledge"-base. There is, however, a competence center, that states it's 2018 goal to increase the uptake of social- and environmentally friendly procurement (*Om kompetenscentrumet — Hankintakeino.fi* (2018)). It will be interesting to see in relation to Grandia and Voncken (2019) & Brammer and Walker (2011)'s findings that lack of knowledge was one barrier for organization's to apply criteria in their procurements, did the establishment of these organization help the authorities in each country to apply more criteria when purchasing?

Below follows a textplot for the Finnish polarity scores, the seed words are highlighted (the same seed-words were used in the Swedish and Finnish scoring)

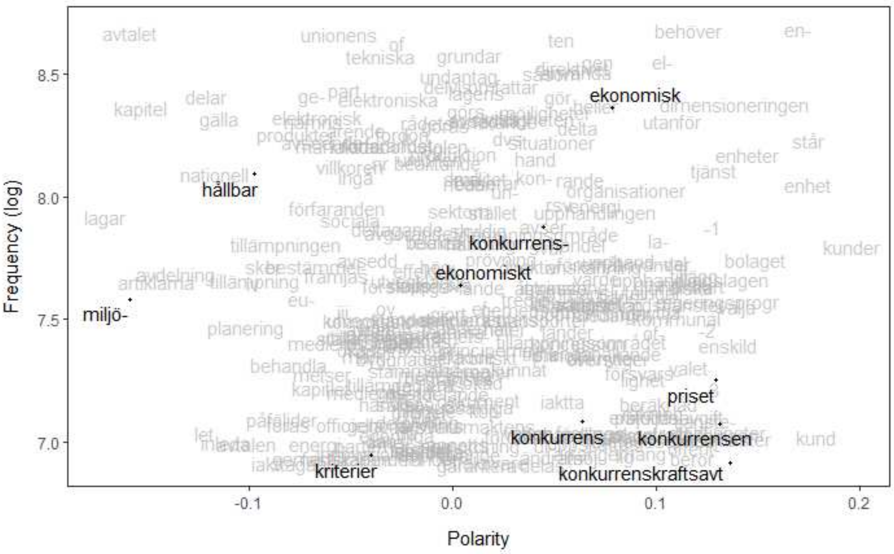

Figure 15: Textplot with the seedwords highlighted. Note that the closer words are together on the x-xis, the more similair they are in polarity score. Y-axis shows the frequency.

The polarity scores of procurement in the Finnish debate follows below.

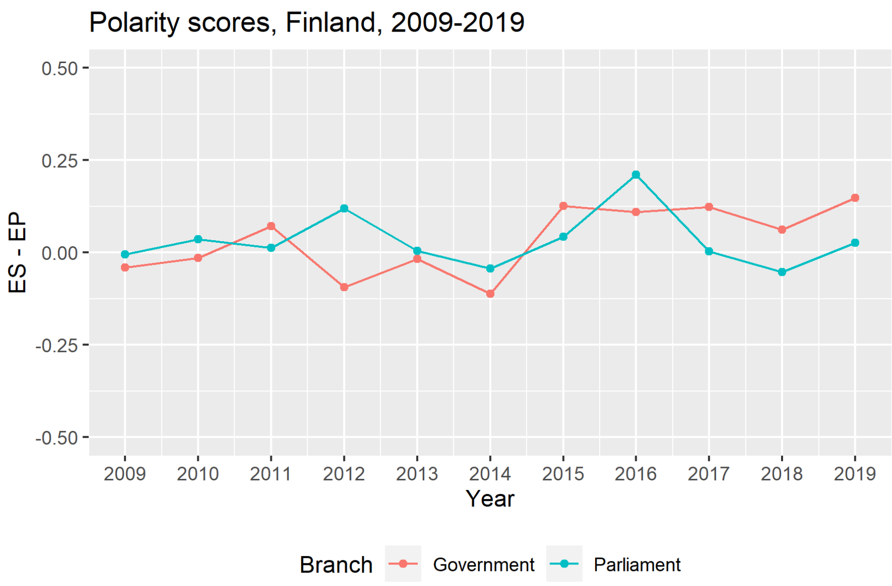

Figure 16: Plot of the polarity scores in Finland grouped by year, and divided by executive / legislative. The Y-axis is structured from ES (Environmental-social) - and EP (Economic-pragmatical

Here, we see that the movements across the spectrum does not seem to be specifically contingent upon being part of either the executive or the legislative branch.

5.9 Public organizations chosen

Gormly (2014)'s study conducted interviews with bureaucrats from several commercial semi-state bodies (i.e. state owned enterprises) regarding their relationship to SPP. They focused on the problems for practicing SPP (see introductory section for a discussion and definition). Gormly did identify certain barriers that they argue are unique to semi-state commercial bodies, suggesting that we should see variation between commercial and non-commercial actors. Recurring themes in the conducted interviews was that of sustainability from an economic perspective, that is, most bureaucrats considered sustainability in procurement as solely in economic terms, saving money in the long term (ibid., p. 416).

These results could be understood in the light of Tkachenko, Yakovlev, and Rodionova (2018), where they found that state income-earning units had more "efficient" outcomes (defined in economic terms, such as time, deadlines held, etc, (ibid., p. 1162)). They explain this effect in a two-fold way. First it is the hard budget constraints of the income earning units seemingly creating more efficient procurements. The authors also find that harder controlling mechanisms seemingly counteracts this effect on organizations with soft budget constraints, leading the authors to conclude that having income earning unit exposed to competition is an incentive for what the authors define as efficient procurment, whereas these conditions need to be "artificially created" by control-mechanisms.

With the help of these authors, we see that there should be a variation between income earning units, and non income earning units - as well as between state-bodies and semi-state bodies. For this, I have chosen two organizations within each of the chosen country, organizations that in theory would purchase similarly. The chosen organizations are the traffic administration of each country, as well as the company (semi-state body) that runs the largest airport within the country. This means that we have both state bodies, as well as a semi-state body. There are slight variation within the chosen group, however. The Danish company allows for private ownership, where the Danish state has 39.2 % of the shares, the rest are private interests (CopenhagenAirports 2021). While in the selection of countries, a selection-motto was as little variation as possible, within this group, having this variation is not a problem, as it allows us to test whether allowing private ownership or joint-ownership helps or harms the application of criteria. For the Swedish and Finnish airport-companies Swedavia & Finnavia, they are both fully state owned however income earning (Finavia annual report, (Näringsdepartementet 2020)), meaning that we can infer variation in criteria applications in the case of copehagen airport to allowing private interests owning parts of the company. These organizations will be contrasted against the traffic administrations of the countries - Swedish Trafikverket, Finnish Väylävirasto & Danish Banedenmark and Vejdirektoratet. In Sweden (Trafikverket 2021) & Finland (Trafikledsverket 2021) both railway and road administration is gathered under one organization, whereas in Denmark these are split in to two (Vejdirektoratet 2021), (Banedanmark 2021). Further, the municipality of each capital city was also chosen to see how the measured variables affect the purchases of non-state organizations. Municipalities are also heavily discussed entities in the literature, Gelderman, Paul, Brugman, et al. (2006) & Vluggen et al. (2020) are two studies focusing on specific characteristics of local level purchasing. It is also of societal interest, as municipalities constitutes a large share of public purchases being made, example in Sweden year 2019, 7 out of 10 purchases were made by a municipality (Upphandlingsmyndigheten 2019)

Why is this worth investigating? Recall McCue, Prier, and Swanson (2015)'s study regarding the dilemmas that procurment professionals are facing. One recurring theme throughout the text, (and in a way - throughout the whole measurement of a policy context) it is how to identify and define "best value". Thus, McCue, Prier, and Swanson (ibid.), when looking at the procurement-process as a principal-agent problem, they argue that a company uses texts and messages as

> codes of ethics are developed to regulate public procurement and typically apply to all expenditures. They often include internal policies and statutes that describe the appropriate conduct for all public employees, but sometimes even when faithfully following these codes of conduct, the "best" value option may not be entirely apparent or clear because of competing goals and competing principals /.../ (ibid., p. 189)

Through looking at how public procurement is framed from the outside, we can infer how the term is coated from the inside. This also ties in to our framework, borrowed and adapted from Brammer and Walker (2011). Recall the "organizational incentives & pressures", these should also be understood from the inside, i.e. if the organization talks about public procurement in E-S terms, the procurement professional will have more incentives to purchase sustainably. If the "the political" coats public purchasing in E-S terms, there is higher pressure on the organization to work towards sustainable purchasing, taking these political cues (see section "Bureaucratic response").

We will also be looking at a selected few municipalities. Gelderman, Paul, Brugman, et al. (2006)'s study focused in on the municipal and local government, stating that we do know little about the sustainability choices of municipalities. The municipality of each country will also be chosen to see whether or not a national policy context, the implementation of a national procurement agency and/or strategy (recall that Grandia and Voncken (2019) & Brammer and Walker (2011) identify lack of information as a barrier to applying criteria in purchases) has significant effects on the choices of criteria application at a local level. Further, it will be tested to see whether "answering to government" will have an effect.

6 Analysis

6.1 Choice of analytical method.

Throughout this introductory section, we shall discuss what mode of analysis fits our case. We will end up with a mixed-effect logistic regression, the outcome being: "does the organization rely on a lowest price-critereon?" a binary response, yes or no.

6.2 Major assumptions of logistic regression.

Our outcome variable is "does the purchase rely on a lowest-price criteria?", meaning that it is binary. Thus one of the most important assumptions of logistic regression is met. Second, we must deal with outliers and missing data in our chosen predictor variables. Missing data was omitted from the analysis. For the outliers in our continuous variables, I calculated a z-score from the standardized vectors and removed any observations that lied outside of 3 standard deviations from the mean.

Second, to see whether it is worth applying a mixed-effects model, we need to look at the intra-class correlation. Intra-class correlation (ICC) is to be understood as the proportion of variance that is explained by how we group the observations (Gelman and Hill 2006, p. 438). For this particular test, I ran the model without any predictor variables, instead just using potential observational groupings. The groupings are: Country and year together, and the individual organization. The ICC-score ranges from 0 to 1. First I ran to get the ICC-score of just our level-two variable, country per year, this yielded the score 0.12 . This means that while the groups are significant, there is still a amount of variation that need to be explained by other variables.

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: lowest_price ~ (1 | CountryYear)
   Data: polarityscore_long

     AIC      BIC   logLik deviance df.resid
 14890.0  14904.7  -7443.0  14886.0    11628

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.8304 -0.8584  0.5463  0.8360  1.6939

Random effects:
 Groups      Name        Variance Std.Dev.
 CountryYear (Intercept) 0.4565   0.6756
Number of obs: 11630, groups:  CountryYear, 29

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.08663    0.12715  -0.681    0.496
# Intraclass Correlation Coefficient

     Adjusted ICC: 0.122
  Conditional ICC: 0.122
```

Figure 17: The regression output from the model run with only our chosen groupings together with the ICC

Had the ICC been closer to, or negligible from 0, it would mean that the groupings does not add any explanatory power - meaning that a complete pooled logistic regression would have been preferable. However, we do see that our chosen groupings (*k's*), as theorized by previous literature are meaningful, we shall thus opt to run the regression with these groupings.

6.2.1 Aggregation versus disaggregation.

To understand the strengths of an hierarchical model in this case, it is worth to quickly discuss the concepts of aggregation vs disaggregation. As previously discussed, we have a lot of observations from the TED-dataset (TED 2010-2019). From this dataset, we can get a lot of information, and we can understand this information to be ordered in levels. Example at the "lowest" level is our outcome "did the organization rely on lowest price criterion?", as well as the organization itself, that can be said to belong to a group, say municipalities or SOEs. What "group" an organization belongs can affect the outcome. At the higher levels, we have what country the organization is in, and countries can also said to be groups that differ in different ways. A difference in our case is our measured policy-context. Previous research tells us that both of these "levels" and groups supposedly affect the outcome of the procurement as we have discussed in other parts of the paper, we thus say that the outcome is "nested" within these levels (Gelman and Hill 2006, p. 4). Previous to the development of hierarchical modeling, there were two ways to deal with nested data, disaggregation and aggregation. Disaggregation means ignoring the differences and variances between levels and groups and treating all data "the same". This method brings the problem of potentially being unable to uphold the assumption of independence of observations required for regressions. Aggregation, on the other hand, means ignoring the within-group variations, and treating, in our case, all organization as the same which might lead to wrong statistical conclusions. Instead, we can employ hierarchical modeling, allowing for within group variation, as well as group variation(ibid.).

6.3 Description of the data

For the independent variables, there is the polarity score of each of the country, structured by year as well as branch of government, be it executive or legislative, this is the operationalization of a nation's policy context as have been discussed in earlier parts of the texts. Adding categorical variables of whether or not there is a procurement agency or organization with the explicit mission of streamlining and providing information about sustainable purchasing. For the "individual level", there is whether the organization is for profit or not as well as whether or not the organization answers to government or not.

How a policy context affects an organization could contingent on the degree of politicization, it thus important to account for in our model, policitzation of the bureaucracy in the Nordic administrative tradition is discussed under section "Countries selected".

To control for this, I added a variable "political interference", which is an expert perception data from the Quality of Government expert survey data - set (Dahlberg et al. 2021). This score is significant in two ways. The logic behind this is that if a bureaucracy is heavily politicized, it would mean a higher political control over bureaucratic decisions, institutions and policy-frameworks. Important to keep in mind that the survey was conducted in 2014 and the score for each country will be used as a variable for each year to see whether this has an effect on the criteria-choices of organizations. Variables of political control, security of tenure, merit based recruitment as well as political impartiality is also included from the 2020 dataset (Dahlberg et al. 2021) however not utilized in our analysis.

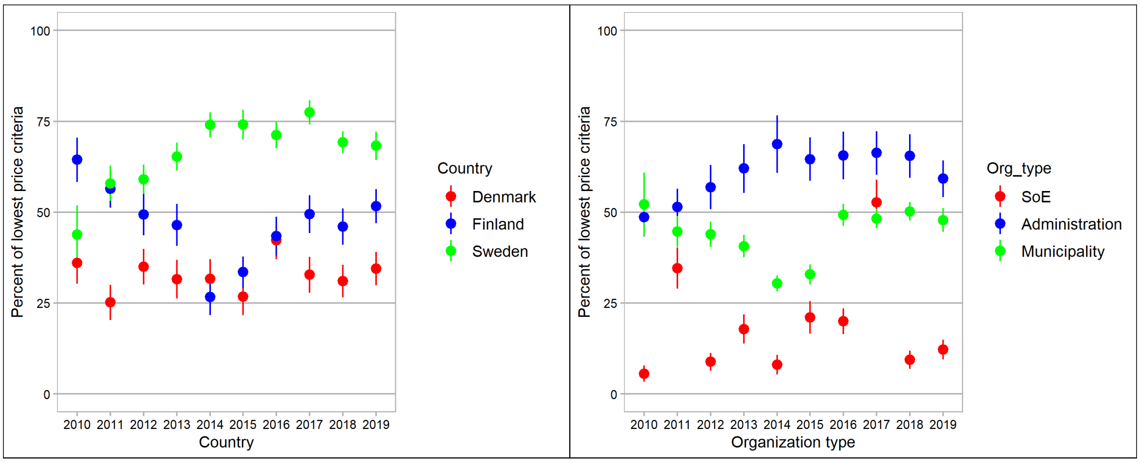

Figure 18: Description of the data by two different groupings. Percentage of procurements relying on lowest price by year.

We do see that generally, Danish organization have relied less on lowest price criteria, Sweden relies more heavily as whereas Finland scores in-between. Our other grouping, seeing percentage of procurements relying on lowest price criteria is generally lower in State owned enterprises than they are in administration and municipalities.

```
'data.frame':	11630 obs. of  18 variables:
 $ CountryYear           : Factor w/ 29 levels "2010 Denmark",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Branch                : chr  "Government" "Government" "Government" "Government" ...
 $ Year_Score            : num  0.0525 0.0525 0.0525 0.0525 0.0525 ...
 $ Year_Score_total      : num  0.0512 0.0512 0.0512 0.0512 0.0512 ...
 $ Patronage             : num  -1.05 -1.05 -1.05 -1.05 -1.05 ...
 $ Merit                 : num  1.19 1.19 1.19 1.19 1.19 ...
 $ Political_interference: num  -0.937 -0.937 -0.937 -0.937 -0.937 ...
 $ impartiality          : num  1.03 1.03 1.03 1.03 1.03 ...
 $ Procuring actor       : chr  "Københavns Kommune Kultur- og Fritidsforvaltningen Københavns Ejendomme"
Miljøforvaltningen; Center for Anlæg og Udbud" "Københavns Kommune; Teknik og Miljøforvaltningen; Center
 $ Actor type            : chr  "R" "1" "3" "3" ...
 $ lowest_price          : Factor w/ 2 levels "No","Yes": 1 2 1 1 1 2 2 1 2 1 ...
 $ CPV                   : chr  "45" "45" "71" "90" ...
 $ Criteria code         : chr  "M" "L" "M" "M" ...
 $ actorID               : Factor w/ 9 levels "1","2","3","4",..: 8 5 8 8 5 5 8 3 6 6 ...
 $ Org_type              : Factor w/ 3 levels "SoE","Administration",..: 3 2 3 3 2 2 3 1 2 2 ...
 $ coord_ent             : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
 $ answ_gov              : Factor w/ 2 levels "0","1": 1 2 1 1 2 2 1 2 2 2 ...
 $ forprofit             : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 2 1 1 ...
```

Figure 19: A glimpse at the dataset with the existing variables. Several are from the Dahlberg et al. (2021)'s QoG-dataset but not utilized in the analysis.

Figure 19 shows us the structure of the dataset used for running the models.

6.4 Specyfing the model, levels and units

With the scoring of the ICC's (see figure 17) we saw that employing an hierarchical model is worthwhile. To specify this model, we first need to decide how to group our observation, as well as what effects are to be seen as fixed, and what effects are to be seen as random. The purchases (which is also the outcome in our study) are done by the chosen organizations which constitutes the model's "lowest" level. Understanding what makes an effect "fixed" or "random" is not an easy task, as there are several meanings attached to the terms by researchers (see p. 20 of Gelman (2005) for a deeper discussion). How I decided on what effects are mixed or random is on the rule of thumb set forward by Searle, Casella, and McCulloch (1992). That is, an effect is to be considered as random if it is interest in the underlying population, and fixed if they are interesting in themselves. Thus, in the first level, fixed variables are "for profit", organization type, as well as whether they answer to the government (see section "bureaucratic response to political cues" for why this is interesting). For the second level, a country group. Here the the variables that will be seen as fixed effects are the quality of governance indicator from (Dahlberg et al. 2021) political interference. As to make the model time sensitive, the time-variable is nested "within" the country. This means that the data is structured per country and year, with 2 variables varying per year. The two varying are the polarity score measurements as well as whether the country had a coordinating entity during the year. Refer to figure 18 for a visual representation of how the data is structured.

After decided on what random variables to assign, I started adding prediction variables, starting from level one (organization itself). I used Gelman and Hill (2006) tip on "fit many models" (P. 547) and starting simple, I thus started on level one and added one variable at a time. The reasons for omitting or keeping variables is discussed further down.

Whether the organizations "answers to government" did not have any effect on the model when included, so I decided to omit it. The same with what type the organization was, an SoE, administration or municipality, this was also giving the model singularity issues. Whether or not the organization was for profit yielded significant results. The implications of this result is found under section "implications of results"

For our second level variables, I ran two models, first with the polarity score per branch and then with the total score. Neither of these yielded significant results this does not mean that it is not worth keeping them in the model, however (see Gelman and Hill (ibid.)) as they did affect results slightly. For the sake of model-parsimony, I decided to keep the aggregated total score in the model centered with mean 0 and SD 1, in line with Gelman and Hill (ibid.) principle. The model showed that there was indeed slightly less effect from the parliament's polarity score, see section "implications of results" for a discussion. For the quality of governance-indicator from (Dahlberg et al. 2021) the relevant one "political interference" scaled in the same way as the polarity score per year. Recall that this variable is based on a 2014 expert survey on the politicization of a country's bureaucracy, thus it is not a repeated measurement. We can assume that it does not fluctuate too much in our chosen countries as they are stable and with similar bureaucratic structures, meaning that the score given to a country is most likely similar between the 2014 scoring, in 2010 when our observations start and 2019 when it ends. This measurement did reach statistical significance, implications are discussed under section "implication of results".

I also checked the interaction effect between our measured polarity score and political interference - this did not reach statistical significance. Yet, there is strong theoretical reasons for why political interference to be included. Recall McCue, Prier, and Swanson (2015) discussion on how public procurement often is used to measure the integrity of governments in power, political interference in bureaucratic matters should thus be checked for, and will also be in the model.

7 Results

What follows below is the output from the fitting of the model to our data.

7.1 Model fit

```
Generalized linear mixed model fit by maximum likelihood (Adaptive Gauss-Hermite Quadrature, nAGQ = 0)
 Family: binomial  ( logit )
Formula: lowest_price ~ forprofit + Year_Score_total + coord_ent + Political_interference *
    Year_Score_total + (1 + coord_ent + Year_Score_total || CountryYear)
   Data: polarityscore_long
Control: glmerControl(optimizer = "bobyqa")

     AIC      BIC   logLik deviance df.resid
 14454.3  14535.3  -7216.2  14432.3    11619

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.3362 -0.7792  0.4827  0.8277  4.0963

Random effects:
 Groups        Name             Variance  Std.Dev.  Corr
 CountryYear   (Intercept)      0.000e+00 0.0000000
 CountryYear.1 coord_ent0       1.522e-01 0.3901348
               coord_ent1       5.006e-01 0.7075443 1.00
 CountryYear.2 Year_Score_total 4.873e-07 0.0006981
Number of obs: 11630, groups:  CountryYear, 29

Fixed effects:
                                        Estimate Std. Error z value Pr(>|z|)
(Intercept)                             -0.81801    0.27058  -3.023   0.0025 **
forprofit1                              -1.80008    0.09696 -18.566  < 2e-16 ***
Year_Score_total                         0.57648    4.53079   0.127   0.8988
coord_ent1                              -0.45069    0.10660  -4.228 2.36e-05 ***
Political_interference                  -1.04976    0.24040  -4.367 1.26e-05 ***
Year_Score_total:Political_interference -0.63232    5.64299  -0.112   0.9108
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) frprf1 Yr_Sc_ crd_n1 Pltcl_
forprofit1  -0.021
Yer_Scr_ttl -0.366 -0.009
coord_ent1  -0.139  0.026  0.162
Pltcl_ntrfr  0.949  0.000 -0.361 -0.216
Yr_Scr_t:P_ -0.246 -0.014  0.956  0.114 -0.269
```

Figure 20: The regression output from the model run

Recall that the outcome in this binary logistic regression is "relying on lowest price" where lowest price = 1. Here we will go through each of the fixed effect coefficients and how to interpret them. The results shown in figure 19 are on the logit-scale, to make them more interpretable in terms of odds-ratios instead some manipulation is needed. Luckily through what Gelman and Hill (2006) calls the "divide by 4 rule", interpretation of coefficients is not too difficult.

> As a rule of convenience, we can take logistic regression coefficients (other than the constant term) and divide them by 4 to get an upper bound of the predictive difference corresponding to a unit difference in x. This upper bound is a reasonable approximation near the midpoint of the logistic curve, where probabilities are close to 0.5. (ibid., p. 82)

While they are not exact estimates, but an acceptable approximation of the coefficient effects on the outcome if it close to the midpoint of the logistic curve.

An odds-ratio is the relationship between the variable and the outcome, in our case, relying on a lowest price criteria. An odds ratio of 1 means the outcome is equally likely, if it is less than 1 it is expected to have lower odds of the outcome, more than one, higher odds of the outcome (ibid., p. 82).

7.2 Interpretation of coefficients

The intercept under the "fixed-effects" corresponds to an organization that is non-profit, within a country where public procurement is spoken about in completely neutral terms, where no coordinating entity exists and rated by the expert's in Dahlberg et al. (2021)' quality of government data set as "0". This hypothetical organization has a log odds of -0.20, meaning that the odds of this organization to rely on lowest price is -0.20 than relying on a lowest price.

As for the "forprofit" variable, if you look closely at figure 18, you see that it is a factor with two levels, meaning that it takes on two values within the data, the organization is either for profit or not. In the regression table we see that there is a "1" after - meaning that the reference group are those organizations that are not-for profit. We see that the estimate is -1.8, this means that the odds of an organization to not rely on lowest price criteria are -0,45 than those in the reference group (that is, non- for profit organizations).

The measured policy context - operationalized as the year polarity score for a country. As this is a continuous variable, it can takes on many different values. Doing a logistic regression, we have assumed that it has a linear relationship with our outcome variable - we thus interpret this score as: for every point increase in polarity score (recall that *increase* in the polarity score means that public procurement is spoken about in more economic than environmental-social terms within governemnt and parliament for the given year) the odds that an organization relies on lowest price criteria increases by 0,14. Note that this relationship is not statistically significant, meaning that we cannot say that there is a relationship between these to variables, but it is kept in the model for reasons discussed above in section "Specyfing the model, levels and units".

The variable "coordinating entity" is similar to that of "for profit", it is a categorical variable with two levels. The estimate means that the odds of an organization to not rely on lowest price criteria are -0,11 in comparisons to the reference group (where no coordinating entity existed in the country).

For political interference, this is a continous like our measured policy context. For every unit increase in political interference, the odds of an organization relying on lowest price decreases by -0.26.

The last variable shown is the interaction effect between our measured policy context and political interference, this did not reach statistical significance, meaning that we cannot say there is a relationship between this and our outcome variable. The reason for why it is included in the model is discussed in previous section "specifying the model".

The figure below summarises all of this information, it is created with the "sjPlot" package, Lüdecke (2015).

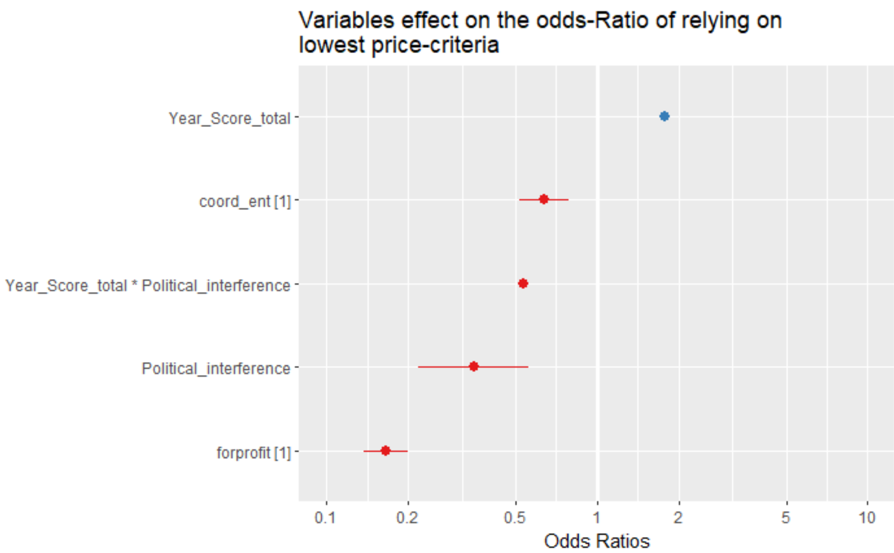

Figure 21: Caterpillar plot with with the effects of variables tested on an odds-ratio scale. Recall that year score and the interaction effect were not significant.

As for the random effects, we should understand the variance within the group of country and year in the coordinating entities. We see that there is less variance when there is none, and more where it does exist. This implies that the coordinating entities have different impact on the application of criteria in different countries, this result is discussed under "implications of results".

7.3 Hypothesis 1: How does the measured policy-context affect public procurement?

Recall that we first ran the model with both executive and parliamentary score. Neither reached statistical significance, but as expected, the executive score had a slightly higher estimated effect. The division was thus dropped from the model, however the total polarity score per year was kept in the model for reasons discussed in section "specifying the model). As for our first hypothesis, we are unable to state any relationship between our measured policy context and that of applications of criteria in public procurements. This is further discussed under section "implications of results".

We saw that the effect of the parliamentary score was slightly less than its' legislative counterpart. Explanation could be that government have the goal setting authority for administrative units, see (Latham, Borgogni, and Petitta 2008). The "answers to government"-dichotomous variable did not have a significant impact. Recall that the only entities which did not answer to government was local level governments.

For the H1c - we expected countries with a higher political interference in bureaucratic matters to be more sensitive to the relationship. It seems as if political interference did give us significant results, implying that there is a relationship between Dahlberg et al. (2021)'s 2014 expert score, and the application of criteria in procuring.

7.4 Hypothesis 2: How does being for profit affect criteria choices?

We expected for-profit organizations to rely less on lowest-price criteria, and our expectations held true. This reinforces Tkachenko, Yakovlev, and Rodionova (2018)'s findings that being exposed to competition could improve the effectiveness of procurement (potential pitfall of this particular result is discussed under "limitations of the study".)

7.5 Hypothesis 3: How does a coordinating entity affect criteria choice?

We expected the existence of a coordinating entity increased an organization's tendency to apply criteria in their purchases. Our expectations were right, for the countries and years where a coordinating entity existed, the organizations we looked at were indeed more likely to apply criteria in their purchases. Denmark was coded to have had a coordinating entity for all the years studied. Sweden from 2015 (". Upphandlingsmyndigheten n.d.), Finland 2017 and forward (JHNY 2021).

8 Discussion

In this section, we shall discuss what implications the results of our study has for previous literature on public procurement, bureaucratic responses as well as other

8.1 Implications of results

8.2 The policy context and political interference

Recall our earlier discussion on the bureaucratic response to political cues, specifically Ringquist, Worsham, and Eisner (2003), Terman (2014) & Terman (2015) that argued "bureaucracies listen" Our measurement of a country's polarity score did seemingly not have any effect on how often our chosen organizations applied lowest price criteria in their purchases. What does this mean for the theorized bureaucratic response to political cues?

The lack of significant impact of the measured policy context on the tendency for organizations to apply criteria in their purchasing could be an effect of political window dressing, a concept that is popular within the international relations-literature (see, for example, (Dosch 2017)). "Window dressing", can be explained in plain terms as "just talking, no action", whereby a politicians seek to gain political points among the electorate on salient issues, however not formally acting upon them. Implications are somewhat more serious for the "bureaucracies listen"-literature. The measurement did pick up on how public procurement was framed and spoken about, and even if there is "window-dressing"-involved, could this lack of response cast shadow on the "bureaucracies listen"-literature?

Ringquist, Worsham, and Eisner (2003) tells us that the use of legislation and formal attempts at influence is an effect of high political salience, implying that there are more informal levels and channels of influence before reaching the formal stage. I see three explanations for the lack of relationship between these two variables. Either sustainable public procurement has not reached the stage of having such high salience, due to its' relatively complex nature, value trade-offs (McCue, Prier, and Swanson 2015, p. 187) and regulatory frameworks that it does not have the public's eye focused on it. This would put the theorized mechanism of governments wanting to improve their reputation in the policy area out of play. Meaning that the salience of public purchasing that formal interference is warranted, has not yet been reached.

Second, it could be that the "Bureaucracies listen"-literature, where parts of the theoretical background of the theorized mechanism stems from, have been developed in a US-context. While I see no reasons for why it should not apply in a European context as well, we should be mindful the effects of applying concepts without considering the effects of context.

Third, it could be a bias due to the chosen cases. Recall the discussion on the "Sinatra inference" (If I can make it here, I can make it anywhere) in the section "case selection". The Nordic countries were chosen due to their, compared to other administrative tradition, relative independent bureaucracy. They were chosen for good reason, as there was a strong theoretical background for the mechanism. Had the relationship been confirmed there would have been strong inferential conclusions about the other contexts. However, the pitfall of such a design is that it does not work the other way around, i.e. the lack of such mechanisms in other contexts (Levy 2008). Thus, future studies should return to this issue with a larger span of countries to get larger variation in measurement to understand the relationship better. Answering to government variable did not have any significant effects either. I believe this is for the same reasons outlined above, this might be about to change, however due to the development of national procurement strategies making it easier for a government to benchmark performance in sustainable procurement (see section "limitations of the study".)

It seems as if the Dahlberg et al. (2021)'s measurement of political interference did show significant results on the tendency of organizations to apply criteria in their purchases. As this was just one measurement for each of the countries during all years, and they were scored similarly. Thus, to draw any meaningful conclusions, we would need either more measurements over time, or a larger set of countries to properly understand the relationships, this result is most likely spurious - it can however serve as a pointer for future studies to explore further. The interaction between this and the measured policy context did not reach any significant results.

8.3 For profit and coordinating entities - the effect of country and organizational characteristics

Being for profit and having coordinating entities did have significant effects of an organization's tendency to apply criteria in their purchases. I am wary about drawing conclusions regarding if we can trust the results about the for-profit characteristics for reasons discussed below. However, having a coordinating entity seems to be significant for both reducing the risk for bureaucracies (Prier, Schwerin, and McCue 2016) in implementing SPP-criteria, as well as reducing the threshold for doing so(Grandia 2018), we thus have good theoretical reason to believe this result, coordinating entities tend to increase the application of criteria other than lowest price in procurement.

A small amount of studies have been done regarding the cross-country learning in relation to public procurement. Nijboer, Senden, and Telgen (2017) could not find any evidence of cross-country learning existing in this area, meaning we are losing efficiency in procurement if countries are "reinventing the wheel". With studies such as this one, pointing towards practices that are indeed good for the implementation of sustainability in public sector purchasing, we would be on a good way to changing Nijboer, Senden, and Telgen (ibid.)'s disappointing results.

The results show that there is indeed a large variation within the country and years where a coordinating entity exists - this implies that the coordinating entities are doing differently well. I have two spontaneous reasons for this. First, it is the form of the organization. Sweden has a full administrative agency, whereas in Denmark and Finland, they are more informal. Second it is once again the issue of time - as the Swedish and Finnish are relatively new, it is probable that they are still establishing the frameworks and networks needed to have a real impact. More research on this is in the future needed to fully draw any conclusions.

8.4 Limitations of the study

One of the largest limitations of the study is that of limited data, before we draw any conclusions, it is important that we discuss in ways the results might be skewed. While the TED-dataset supplies us with a large amount of observations, we had to limit the study to studying whether the purchase applied other criteria than "lowest price" or not (see section "Finding sustainability"). This means that we are unable to know whether the criteria applied is related to social and or environmental, or general "quality-criteria" as labeled in the procurement literature. Relying on lowest price does however certainly mean that no sustainability-concerns were taken into consideration when purchasing, meaning that we at least can be certain about the lowest-price outcome. We should thus be careful about using these findings to draw final conclusions regarding the sustainability-outcomes of the purchases studied. What we can be certain of is that organizations cannot rely on lowest price crietria if they are to use procurement in a sustainable way. Framed in this way, the question being "how can we make organizations less reliant on lowest-price criteria?", the conclusions and findings do have an impact for the understanding of sustainable public procurement. Another issue pertaining to the data is that of the value-threshold. While it is considered "good practice" (TED 2010-2019) to publish all purchases made by the organization on the TED-website, only the purchases above the value threshold are required to be. This means that the data and thus results could be skewed in some ways, that other practices are more common when the purchase is below the threshold, i.e. when there is less money involved.

Finally, regarding the time aspects. While the years and countries having a coordinating entity did have a significant effect on an organization's tendency to apply other criteria than lowest price, the entities are continuously creating copy-paste criteria for different areas of procurements (see for example the Swedish procurement agency's website ". Upphandlingsmyndigheten (n.d.)). Thus, expect the application of criteria to be increasing as the threshold to apply sustainability criterias for certain purchases is steadily decreasing. Since both the Finnish and Swedish organizations are relatively new (2015 and 2017 respectively) we have probably not seen the full effect of these coordinating entities. This could also explain why there is large variation within the country/year grouping where a coordinating entity exists, as they continually expand cope-paste criteria to apply, and thus lowering the threshold to apply criteria as time goes on. Another aspect regarding time is the implementation of National procurement strategies implemented in later stages of the 2010's (or in 2020 s in Finland, meaning we have no observations from post-strategy procurement). Sweden's strategy, Upphandlingsmyndigheten (2016) for example, can be understood as giving the political a baseline from which it can benchmark the purchases of its bureaucracy, which I believe will increase the tendency to apply sustainable criteria in bureaucracies for two reasons. First, understood in light of Tkachenko, Yakovlev, and Rodionova (2018), it is another control mechanism which is theorized to increase the efficiency in purchases. Second, in light of Ringquist, Worsham, and Eisner (2003), it is an incentive for bureaucracies to "listen", as it easier enables strategies such as "naming and shaming", among other (see Franklin (2015) for more on naming and shaming). We have probably not seen the full effects of these strategies as they get institutionalized, meaning that future studies could well return to these countries to see the effects of such measures. My belief in the improvement is based on Raymond (2008)'s finding that benchmarking within public procurement was found as very important in the implementation of procurement policies, thus a clearer benchmark should reasonable lead to a higher uptake.

Another limitation is the amount of organizations chosen. The choices were motivated by a "least-likely most-similar"-logic, and thus while the number of observations (i.e. number of purchases) are satisfactory, the number of actors should be expanded in future studies. An example of a possible skewing effect of this is that the only "for - profit" enterprises are the SoE's that govern airports within a country. While it yielded statistical significance, we should be careful to draw conclusions that for profit organizations are necessarily more prone to applying criteria when purchasing - as the "for profit" in this case only studied 3 specific for profit organizations in relation to 7 other non-for profit. A potential trap door is that airport companies might have some traits that other for profit organizations lack that is actually driving this tendency to apply criteria, rather than being "for profit". However, there are theoretical reasons to believe that this might be the case from previous research, see Tkachenko, Yakovlev, and Rodionova (2018).

Lastly, it is this analysis has not taken into account what area the organization is purchasing in, i.e. the CPV-codes, see (Regulation (EC) No. 213/2008). It is reasonable to believe that some areas it is more common to apply criteria than others. The reason for why this was not taken into account is because this study does not have the reason to understand *why* some differ from others. While we could have kept CPV-codes as a discrete variable, no previous literature was found explaining why some areas were more prone to apply than others. While it might have improved model fit to the data, it would not help us in understanding public procurement on a more general level due to the fact that there is no general theoretical explanation as to why.

9 Conclusion

Public procurement is the largest business sector in the world (Hawkins, Gravier, and Powley 2011), it is thus potentially a very powerful policy tool if utilized correctly. We knew from previous literature that lack of knowledge and incentive are two potential barriers, making them rely on lowest price criteria when purchasing goods and services. Previous studies had made use of very granular data, where interview and surveys comprised the lion's share of material studied. The purpose of this study was to make use of observational data to study the relationship between country and organizational characteristics and criteria application as to expand our knowledge about public procurement.

It was theorized that a nation's "policy context" should have large effects on an organization's application of sustainable criteria. Through a fresh conceptualization of such a "policy context" and other characteristics that differ between countries, I compared three countries and set out to see whether these differences had an affect on nine chosen organization's application of criteria in their purchases. A conclusion was made that a policy-context could be understand in how an issue was framed in political debate, and we applied a semi-supervised text-analytical model to get the score of a country on a two-dimensional scale, using a large corpus of texts created specifically for this study. The dimensions went from procurement solely spoken about in terms of social and environmental values to only in economic efficiency terms. It was divided by year and whether the text was produced by the executive or the legislative. Other things taken into account at the country level was whether there was a coordinating entity as well as the level of political interference in the bureaucracy. For the organizational level, the organization's characteristics were also scored, whether they answered to the government or were for-profit. To see whether there is a relationship between these variables and criteria choice, we decided to employ a mixed-effect logistic model, the level one was the organization themselves, the level two variables were the countries coupled with year-indicators. The relationship was studied with the help of the Tenders electronic daily's dataset on contract award notices from 2010-2019.

The analysis showed that there was no significant effect by the measured policy context and the application criteria. More research into this is needed for us to be able to draw conclusions whether how the issue is framed in debate and how this affect the outcome of criteria choice. We did however see effects on whether there was a coordinating entity that reduces the risks and lowers the threshold for organizations to apply criteria when purchasing. Further the results indicated that being for-profit does increase an organization's tendency to apply criteria when purchasing, however with the relatively small sample size, we should be careful about drawing conclusions and more research is needed on whether being exposed to competition is beneficial for criteria application or not.

With this study, one of very few studies in comparative public procurement, several steps have been made towards increasing the understanding of public procurement as a general phenomena, hopefully bringing us closer of how to better use public procurement as a policy tool in achieving the Agenda 2030 and other social goals. Not only this, recall Patrucco, Luzzini, and Ronchi (2017)'s somewhat pessimistic quote in the beginning of section 2.3 - that we cannot compare as the results in procurements as they "depend strictly on country characteristics" (p. 245). We have done, not only this, but also taken steps towards comparing exactly how these country characteristics affect the tendency of organizations within a country to apply criteria in their purchases. The conceptualization of a "policy context" in a way that was cross-country comparable laid forward in this study hopefully opens up the way for future research to expand the cases to test the mechanisms laid forward.

10 Bibliography

References

About TED - TED Tenders Electronic Daily (n.d.). *About TED - TED Tenders Electronic Daily*. `https://ted.europa.eu/TED/misc/aboutTed.do`. (Accessed on 05/17/2021).

Agenda, Sammanfattning av (2019). *2030 och Sverige: Världens utmaning–världens möjlighet*.

Banedanmark (2021). *Borger — Banedanmark*. `https://www.bane.dk/`. (Accessed on 07/07/2021).

Bao, Yuanjie et al. (2012). "Understanding the motivational mechanisms of value congruence". PhD thesis. Universitat Ramon Llull.

Bilodeau, Nancy, Claude Laurin, and Aidan Vining (2007). ""Choice of organizational form makes a real difference": the impact of corporatization on government agencies in Canada". In: *Journal of public administration research and theory* 17.1, pp. 119–147.

Botti, Antonio and Antonella Monda (2019). "Goal Ambiguity in Public Organizations: A Systematic Literature". In: *International Journal of Business and Management* 14.7.

Brack, Nathalie and Sharon Weinblum (2011). """ Political Opposition": Towards a Renewed Research Agenda". In: *Interdisciplinary Political Studies* 1.1, pp. 69–79.

Brammer, Stephen and Helen Walker (2011). "Sustainable procurement in the public sector: an international comparative study". In: *International Journal of Operations & Production Management*.

Brillo, João and Erivelton Silva (2020). "Tri-Intersectional Model of Leadership by Values: A Fuzzy Multi-Criteria Decision Making Open Technology Assessment System". In: *Chinese Business Review* 19.5, pp. 141–156.

Capell, Ben et al. (2013). "Mapping values in old vs new members of the European Union: A comparative analysis of public sector cultures". In: *Cross Cultural Management: An International Journal* 20.4, pp. 503–527.

Cooper, Christopher A (2021). "Politicization of the Bureaucracy across and within Administrative Traditions". In: *International Journal of Public Administration* 44.7, pp. 564–577.

CopenhagenAirports (2021). *Share information.* https://www.cph.dk/en/about-cph/investor/share-information. (Accessed on 07/08/2021).

Dahlberg, Stefan et al. (2021). "The Quality of Government Basic Dataset, Version Jan21". In: *University of Gothenburg: The Quality of Government Institute, http://www. qog. pol. gu. se doi* 10.

Deep.l (July 2017). *Deep.l translator.* URL: https://www.deepl.com/translator.

Deng, Qi et al. (2019). "Inside the black box of dictionary building for text analytics: a design science approach". In: *Journal of international technology and information management* 27.3, pp. 119–159.

Directive 2014/24/EU (2014). *Directive 2014/24/EU of the European Parliament and of the Council of 26 February 2014 on public procurement and repealing Directive 2004/18/EC Text with EEA relevance.* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex\%3A32014L0024.

Dolan, Simon L et al. (2013). "Mapping values in old vs new members of the European Union". In: *Cross Cultural Management.*

Dosch, Jörn (2017). "The ASEAN economic community: Deep integration or just political window dressing?" In: *TRaNS: Trans-Regional and-National Studies of Southeast Asia* 5.1, pp. 25–47.

Downs, Anthony (1967). "Inside bureaucracy". In:

Drumwright, Minette E (1994). "Socially responsible organizational buying: environmental concern as a noneconomic buying criterion". In: *Journal of marketing* 58.3, pp. 1–19.

Eduskunta, Riksdagsärenden och riksdagsdokument (2021). https://www.eduskunta.fi/SV/valtiopaivaasiat/Sidor/default.aspx. (Accessed on 07/15/2021).

Elkington, John (1994). "Towards the sustainable corporation: Win-win-win business strategies for sustainable development". In: *California management review* 36.2, pp. 90–100.

Esteve, Marc, Monica Grau, and Ramon Cabrera Valle (2013). "Assessing public sector values trough the tri-axial model: empirical evidence from Spain". In: *Cross Cultural Management.*

European Commission (2017a). *Buying green! A handbook on green public procurement 3rd Edition.* https://ec.europa.eu/environment/gpp/buying_handbook_en.htm.

— (July 2017b). *Public Procurement.* URL: https://ec.europa.eu/growth/single-market/public-procurement_en.

European Commission, Sustainable GPP (2021). *Sustainable GPP - Environment - European Commission.* https://ec.europa.eu/environment/gpp/versus_en.htm. (Accessed on 07/08/2021).

Facts about SKI (2021). *Facts about SKI - SKI.* https://www.ski.dk/videnssider/facts-about-ski/. (Accessed on 07/07/2021).

Fang, Hanming, Chang Liu, and Li-An Zhou (2020). *Window Dressing in the Public Sector: A Case Study of China's Compulsory Education Promotion Program.* Tech. rep. National Bureau of Economic Research.

Flynn, Anthony (2018). "Measuring procurement performance in Europe". In: *Journal of Public Procurement.*

Folketinget (2021). *Dokumenter / Folketinget.* https://www.ft.dk/da/dokumenter. (Accessed on 07/07/2021).

Franklin, James C (2015). "Human rights naming and shaming: International and domestic processes". In: *The Politics of Leverage in International Relations.* Springer, pp. 43–60.

Garritzmann, Julian L (2017). "How much power do oppositions have? Comparing the opportunity structures of parliamentary oppositions in 21 democracies". In: *The Journal of Legislative Studies* 23.1, pp. 1–30.

Gelderman, Cees J, W Th Paul, Marc J Brugman, et al. (2006). "Public procurement and EU tendering directives–explaining non-compliance". In: *International Journal of Public Sector Management.*

Gelderman, Cees J, Janjaap Semeijn, and Frank Bouma (2015). "Implementing sustainability in public procurement: The limited role of procurement managers and party-political executives". In: *Journal of public procurement.*

Gelman, Andrew (2005). "Analysis of variance—why it is more important than ever". In: *The annals of statistics* 33.1, pp. 1–53.

Gelman, Andrew and Jennifer Hill (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Gormly, Joey (2014). "What are the challenges to sustainable procurement in commercial semi-state bodies in Ireland?" In: *Journal of Public Procurement*.

Grandia, Jolien (2018). "Public procurement in Europe". In: *The Palgrave handbook of public administration and management in Europe*. Springer, pp. 363–380.

Grandia, Jolien and Dylan Voncken (2019). "Sustainable public procurement: the impact of ability, motivation, and opportunity on the implementation of different types of sustainable public procurement". In: *Sustainability* 11.19, p. 5215.

Greve, Carsten et al. (2020). "Unpacking Nordic administrative reforms: Agile and adaptive governments". In: *International Journal of Public Administration* 43.8, pp. 697–710.

Grimmer, Justin and Brandon M Stewart (2013). "Text as data: The promise and pitfalls of automatic content analysis methods for political texts". In: *Political analysis* 21.3, pp. 267–297.

Gross, Martin and Michael Jankowski (2020). "Dimensions of political conflict and party positions in multi-level democracies: evidence from the Local Manifesto Project". In: *West European Politics* 43.1, pp. 74–101.

Händschke, Sebastian GM et al. (2018). "A corpus of corporate annual and social responsibility reports: 280 million tokens of balanced organizational writing". In: *Proceedings of the first workshop on economics and natural language processing*, pp. 20–31.

Hawkins, Timothy G, Michael J Gravier, and Edward H Powley (2011). "Public versus private sector procurement ethics and strategy: What each sector can learn from the other". In: *Journal of business ethics* 103.4, pp. 567–586.

Hooghe, Liesbet (2017). *Measuring international authority*. Vol. 3. Oxford University Press.

JHNY (2021). *Kärnkompetens inom upphandling: Vad är Rådgivningsenheten för offentlig upphandling? — Kommunförbundet*. https://www.upphandling.

`fi/karnkompetens-inom-upphandling-vad-ar-radgivningsenheten-offentlig-upphandling`. (Accessed on 07/09/2021).

Jones, Michael L (2007). "Hofstede-culturally questionable?" In:

Kather, Gesa (2013). "Community Cohesion Policy in a Northern English Town: An Ethnographic Approach". In: *Controversies in Policy Research.* Springer, pp. 98–118.

Knebel, Sebastian et al. (2019). "9.5 Trillion USD for Sustainability: a Literature Review on Sustainable Public Procurement". In: *National Research Project NRP* 73.

Koala, Kouliga and Joshua Steinfeld (2018). "Theory building in public procurement". In: *Journal of Public Procurement.*

Latham, Gary P, Laura Borgogni, and Laura Petitta (2008). "Goal setting and performance management in the public sector". In: *International Public Management Journal* 11.4, pp. 385–403.

Laver, Michael, Kenneth Benoit, and John Garry (2003). "Extracting policy positions from political texts using words as data". In: *American political science review*, pp. 311–331.

Lember, Veiko, Rainer Kattel, and Tarmo Kalvet (2013). *Public procurement, innovation and policy: International perspectives.* Springer.

Levy, Jack S (2008). "Case studies: Types, designs, and logics of inference". In: *Conflict management and peace science* 25.1, pp. 1–18.

Lo, James, Sven-Oliver Proksch, and Jonathan B Slapin (2016). "Ideological clarity in multiparty competition: A new measure and test using election manifestos". In: *British Journal of Political Science* 46.03, pp. 591–610.

Lüdecke, Daniel (2015). *Package 'sjPlot'.*

Matsuo, Akiko et al. (2019). "Development and validation of the japanese moral foundations dictionary". In: *PloS one* 14.3, e0213343.

McCrudden, Christopher (2004). "Using public procurement to achieve social outcomes". In: *Natural resources forum.* Vol. 28. 4. Wiley Online Library, pp. 257–267.

McCue, Clifford P, Eric Prier, and David Swanson (2015). "Five dilemmas in public procurement". In: *Journal of public procurement.*

Näringsdepartementet (2020). *Redogörelse för bolag med statligt ägande.*

Nawrocka, Dagmara (2008). “Inter-organizational use of EMSs in supply chain management: some experiences from Poland and Sweden”. In: *Corporate social responsibility and environmental management* 15.5, pp. 260–269.

Nelson, Harry W and William Nikolakis (2012). “How does corporatization improve the performance of government agencies? Lessons from the restructuring of state-owned forest agencies in Australia”. In: *International Public Management Journal* 15.3, pp. 364–391.

Nijboer, Kimberly, Shirin Senden, and Jan Telgen (2017). “Cross-country learning in public procurement: An exploratory study”. In: *Journal of public procurement.*

Nordic Council (2021). *Methodology for mapping The Sustainable Public Procurement and the Sustainable Development Goals.*

Oldenhof, Lieke, Jeroen Postma, and Kim Putters (2014). “On justification work: How compromising enables public managers to deal with conflicting values”. In: *Public Administration Review* 74.1, pp. 52–63.

Om kompetenscentrumet — Hankintakeino.fi (2018). `https://www.hankintakeino.fi/sv/vad-ar-keino/om-kompetenscentrumet`. (Accessed on 07/15/2021).

Österberg, Emma Ek (2020). “5 Att styra med upphandling–staten som marknadsaktör”. In: *Statlig förvaltningspolitik för 2020-talet.*

Parikka-Alhola, Katriina (2008). “Promoting environmentally sound furniture by green public procurement”. In: *Ecological economics* 68.1-2, pp. 472–485.

Patrucco, Andrea Stefano, Davide Luzzini, and Stefano Ronchi (2017). “Research perspectives on public procurement: Content analysis of 14 years of publications in the journal of public procurement”. In: *Journal of Public Procurement.*

Peng, Shuyang, Sheela Pandey, and Sanjay K Pandey (2015). “Is there a nonprofit advantage? Examining the impact of institutional context on individual–organizational value congruence”. In: *Public Administration Review* 75.4, pp. 585–596.

Petrie, Stephanie (2013). “Policy Contexts, Consequences and Controversies”. In: *Controversies in Policy Research.* Springer, pp. 1–12.

Prier, Eric, Edward Schwerin, and Clifford P McCue (2016). "Implementation of sustainable public procurement practices and policies: A sorting framework". In: *Journal of Public Procurement*.

Proksch, Sven-Oliver et al. (2019). "Multilingual sentiment analysis: A new approach to measuring conflict in legislative speeches". In: *Legislative Studies Quarterly* 44.1, pp. 97–131.

Rainey, Hal G, Robert W Backoff, and Charles H Levine (1976). "Comparing public and private organizations". In: *Public administration review* 36.2, pp. 233–244.

Rainey, Hal G and Barry Bozeman (2000). "Comparing public and private organizations: Empirical research and the power of the a priori". In: *Journal of public administration research and theory* 10.2, pp. 447–470.

Raymond, Jeanette (2008). "Benchmarking in public procurement". In: *Benchmarking: An International Journal*.

Riksdagen (2021). *Dokument & lagar - Riksdagen*. `https://www.riksdagen.se/sv/dokument-lagar/`. (Accessed on 07/08/2021).

Ringquist, Evan J, Jeff Worsham, and Marc Allen Eisner (2003). "Salience, complexity, and the legislative direction of regulatory bureaucracies". In: *Journal of Public Administration Research and Theory* 13.2, pp. 141–164.

Schaffer, Frederic Charles (2015). *Elucidating social science concepts: An interpretivist guide*. Vol. 4. Routledge.

Searle, Shayle R, George Casella, and Charles E McCulloch (1992). *Variance components*. Vol. 391. John Wiley & Sons.

Seele, Peter (2016). "Envisioning the digital sustainability panopticon: a thought experiment of how big data may help advancing sustainability in the digital age". In: *Sustainability Science* 11.5, pp. 845–854.

SIMAP Standard form (2014). *SIMAP - Standard forms for public procurement*. `https://simap.ted.europa.eu/web/simap/standard-forms-for-public-procurement`. (Accessed on 07/15/2021).

Soudek, Jan and Jiřı Skuhrovec (2016). "Procurement procedure, competition and final unit price: The case of commodities". In: *Journal of public procurement*.

Stake, Johan (2017). "Evaluating quality or lowest price: consequences for small and medium-sized enterprises in public procurement". In: *The Journal of Technology Transfer* 42.5, pp. 1143–1169.

Statistik som utvecklar den offentliga affären — Upphandlingsmyndigheten (2021). `https://www.upphandlingsmyndigheten.se/kunskapsbank-for-offentliga-affarer/statistik-som-utvecklar-den-offentliga-affaren/`. (Accessed on 07/16/2021).

TED (2010-2019). *TED csv dataset (2010-2019).* `https://open-data.europa.eu/cs/data/dataset/ted-csv`.

TED CSV Advanced Notes (2020). *European Commission, DG GROW G4TED CSV open data Advanced notes on methodology Version 0.91.* `https://data.europa.eu/euodp/repository/ec/dg-grow/mapps/TED_advanced_notes.pdf`. (Accessed on 07/15/2021).

Terman, Jessica (2014). "Evaluating political signals: The nature of bureaucratic response in minority preference purchasing". In: *The American Review of Public Administration* 44.5, pp. 522–549.

— (2015). "A state-level examination of bureaucratic policymaking: The internal organization of attention". In: *The American Review of Public Administration* 45.6, pp. 708–727.

Tkachenko, Andrey, Andrei Yakovlev, and Yuliya Rodionova (2018). "Organizational forms and incentives in public procurement: Natural experiment at a large public sector organization in Russia". In: *International Journal of Public Administration* 41.14, pp. 1157–1168.

Trafikledsverket (2021). *Vårt arbetssätt - Trafikledsverket.* `https://vayla.fi/sv/trafikledsverket/arbetssatt`. (Accessed on 07/07/2021).

Trafikverket (2021). *Trafikverkets uppdrag - Trafikverket.* `https://www.trafikverket.se/om-oss/var-verksamhet/trafikverkets-uppdrag/`. (Accessed on 07/07/2021).

United Nations, Sustainable Development goals (2012). *Promote public procurement practices that are sustainable in accordance with national policies and priorities.* `https://indicators.report/targets/12-7/`. (Accessed on 07/10/2021).

University of California, Berkely (2021). *GADM database of Global Administrative Areas*.

Upphandlingsmyndigheten (2016). *Nationella upphandlingsstrategin — Upphandlingsmyndigheten*. https://www.upphandlingsmyndigheten.se/forbered-organisationen/nationella-upphandlingsstrategin/. (Accessed on 07/11/2021).

— (2019). *Ökat intresse för offentliga upphandlingar*. URL: https://www.upphandlingsmyndigheten.se/kunskapsbank-for-offentliga-affarer/statistik-som-utvecklar-den-offentliga-affaren/okat-intresse-for-offentliga-upphandlingar/#7_av_10_upphandlingar_g%5C%C%5C3%5C%B6rs_av_kommunerna.

Upphandlingsmyndigheten, "Kriterier" (n.d.). *Hitta hållbarhetskriterier — Upphandlingsmyndigheten*. https://www.upphandlingsmyndigheten.se/kriterier/. (Accessed on 07/07/2021).

Utz, Clayton (n.d.). *PPPs - closing the gap in the triple bottom line: people, [planet,] profit - Lexology*. https://www.lexology.com/library/detail.aspx?g=6999ab30-38aa-40f0-b919-fb230e067c4f. (Accessed on 07/18/2021).

Vejdirektoratet (2021). *Om os — Vejdirektoratet*. https://www.vejdirektoratet.dk/sektion/om-os. (Accessed on 07/07/2021).

Vluggen, Rob et al. (2020). "Social return on investment in the public sector". In: *Journal of Public Procurement*.

Watanabe, Kohei (2020a). "Latent Semantic Scaling: A Semisupervised Text Analysis Technique for New Domains and Languages". In: *Communication Methods and Measures*, pp. 1–22.

— (2020b). *Model for Semisupervised Text Analysis Based on Word Embeddings*. https://rdrr.io/github/koheiw/LSS/.

Wilson, Woodrow (1886). *The study of public administration*.

Yang, Shijing et al. (2019). "Research on developers' green procurement behavior based on the theory of planned behavior". In: *Sustainability* 11.10, p. 2949.