Diversity in Stakeholder Preferences Regarding EU Policy:

The Effect of Survey Elements within Processes of Open Public Consultation

Rebekka Strandmann Hanssen

Master's thesis

Spring 2022

Department of Comparative Politics
University of Bergen

Abstract

Since the "European Governance A White Paper" launched in 2001, the European Union has worked to improve accountability and transparency within the organization. Preferences voiced by the public are one of the most important means of ensuring democratic policy formulations. Previous research on participation and voicing preferences through different means for influencing EU policies has indicated different expectations for what affects and motivates diversity in voiced preferences. In the policy formulation stage, open consultation venues are often used to receive intel and inputs from diverse sets of stakeholder groups. Additionally, open public consultations have been found to attract broad stakeholder representation, indicating higher possibilities for observing diversity in stated preferences.

This thesis analyzes how diverse received inputs are within open public consultation surveys and tests how different survey elements affect the observed diversity. Clustering methods have been applied to analyze how diverse stated preferences are. By mapping groups of similar preferences, the result shows that most surveys receive from 2 to 7 observed preference groups, clearly indicating diversity. This aligned with previous findings of more diverse stakeholder participation through open consultation venues.

For testing effects on the observed diversity in preferences, issue salience, mentioning targeted stakeholders, complexity within the survey, and complexity within policy issues have been collected and applicated through negative binomial regression and multiple linear regression. The number of observed preference groupings decreases when targeted stakeholders are mentioned in the open public survey. Simultaneously, issue salience increases observed preference groups and the size of respondents observed within the preference groups. Neither issue complexity nor survey complexity affects diversity in stated preferences.

The stated effects of issue salience and mention of targeted stakeholders indicate support for previous research, which has found different stakeholders types to voice diverse preferences.

Acknowledgments

Sidan 2017 har eg vert ein av mange Samanliknande politikk studentar. Gjennom møte med mange fine sjeler, vil eg her nytte anledninga til å særleg takke Ingrid Wågen og Maren S. Haukeland for å ha vore faglege støttespelarar og særs nære vennar gjennom alle dei fem åra på Universitetet i Bergen.

Further in the line for thanks, I want to thank my supervisor. Writing this master's has given me an association of being Luke Skywalker inside the Dagobah swamps. Being able to succeed with writing this master, or completing this Jedi training, is thanks to my supervisor, Reto Wüest (also referred to as Yoda in this universe). Thank you, Reto (Yoda).

Vidare, på nynorsk, ynskjer eg å takke Adriana Bunea. Dette både som sjef, med innspel på innhald og litteratur, og som støttespelar. Mange takk til deg for å rekkje ei ekstra lang hand til meg i dei turbulente tidene, både i form av utvida fristar med jobb, men også for å sikre at innhaldet gjev meining.

For digresjonar, diskusjonar, råd og gode samtalar vil eg takke gjengen på Sofie Lindstrøms hus. Dykk er ein fasinerande komponert gjeng, som har gjort det mogleg for meg å trivast på sal. Eg ser tilbake på lange dagar, endelause veker og mange gode stundar delt på lunsjrommet. Samla har vi delt nedtur og oppturar. Mine kjære samansvorne 2022-masterstudentar, dykk er fantastiske individ.

Eg må også sjølvsagt takke mamma og pappa. Takk til mine foreldre med kontinuerlig mangel på innsyn i kva eg held på med, men med evig tru på at eg lykkjas med alt eg gir meg ut på. Eg sett enormt stor pris på den endelause støtta og dei styrkande orda dykk har gitt meg i alle dei håplause stundar gjennom masterprosessen.

Heilt til slutt må ein ekstra stor takk gå til min kjære ven Nora Elvestad Reinsnes og Michael Alvarez for bistand med korrekturlesing og opponering på samanheng. Dykk er heltar!

Bergen, 22.juni 2022

Table of content

List of Figures

List of Tables

List of abbreviations

CLIMA	Directorate-General for climate action
DG	Directorate-General
EC	European Commission
ENER	Directorate-General for energy
ENV	Directorate-General for environment
EMPL	Directorate-General for employment, social affairs and inclusion
EU	European Union
GROW	Directorate-General for internal market, industry, entrepreneurship and SMEs
HOME	Directorate-General for migration and home affairs
IG	Interest group
MOVE	Directorate-General for mobility and transport
NGO	Non-governmental organization
OPC	Open public consultations
PC	Public consultations
IRR	Inter Ratio Residuals
PCA	Principal Components Analysis

1. Introduction

Public preferences are essential to investigate and understand how political decision-making is shaped. In the context of Europe, the European Union (EU) makes policies that affect more than just one nation. Due to being the largest single markets in Europe, understanding how the consultation-venues effects the diversity of stakeholders stated preferences stands to high importance. Through various consultative and participatory venues, the EU receives valuable insight and policy inputs from stakeholders. The open public consultations[1], established for receiving inputs and policy relevant information, are stated to be one of the most advanced open public consultation regimes in the world (Rodrigo and Amo 2006, 1). By using collected feedback from open public consultations on policy initiatives, this thesis aims to assess preference diversity, and explain how surveys used in open public consultations affects the potential for diversity of received inputs from stakeholders. Accordingly, the thesis aims to answer the following research questions:

"1) How diverse are the opinions stakeholders express in EU open public consultations?
2) How can we explain differences in the diversity of expressed opinions across consultations?"

Studying diversity in stakeholders preferences within the EU open public consultations is increasingly essential to understand participation within the European political sphere and to assess the scope of legitimacy the EU policies possess (Majone 2002, 320). The EU uses various consultation tools to address different stakeholders' responses. According to Hermansson (2016), stakeholders are defined as "(…) any individual or organization including firms, interest groups, trade unions, NGOs and sub-national governmental bodies that has an expressed interest in the policy outcome and participate in the consultative process." Given stakeholders' differences in national and economic sector background and organizational structure, they are deemed important for legitimizing policies being put forward.

As stakeholders' input and preferences are important for legitimizing EU policies, there is a need to understand what motivates stakeholders to participate, and potentially reveal what hinders them in participating. If you do not participate in public policymaking via consultations,

[1] Public consultation (PC), also referred to as "Open public consultations" (OPC) due to its similarity of structure and "open" for self-selected participation.

you cannot expect to influence European policymaking. It is shown that a wide variety of stakeholders participate, and we can assume they participate based on their potential for influence on policymaking (Klüver 2013b). Therefore, motivation to participate must be derived from previous experiences of preference attainment – getting preferred outcomes on policies. Whereas motivation for preference attainment drives participation, the venue being used to consult will as well be deterministic for how diversified the inputs are. Stakeholders' motivation to participate will impact the expectations for inputs, and how diversified these inputs are in open public consultations. Understanding the diversity existing in stated preferences via open public consultation, and which factors impact the received plurality of stated preferences, is essential for understanding how legitimate EU's policy proposals are.

1.1 Why Study Diversity of Stakeholders' Preferences Expressed in Open Public Consultations?

In the EU system of governance, stakeholders are invited to participate in the policy formulation process of supranational policies. Considering that stakeholders are of diverse types, ranging from individual citizens and civil societal organizations to financial companies and trade associations, their preferences they have on EU policy can vary significantly. In addition, as stakeholder represent different interest, they can vary in both access to recourses and technical competence. Recognizing that politics in the EU affects more than just the member countries, it is fruitful to understand how to ensure broad input-representation among stakeholders, and which elements affect input representation.

Amongst different consultation venues used by the EU, open public consultations are among the most accessible and least costly venues for stakeholders to provide inputs on policy relevant issues. Open consultations offer everyone the chance to state inputs and indicate the preferred policy options on new policy initiatives or evaluation of existing policies. In the context of consultations, the European Commission's (EC) lack of resources and expertise is frequently highlighted as reason for increased inclusion of stakeholder inputs (Arras and Braun 2017; Dür and De Bièvre 2007). Thus, formulating the policy, the EC depends on input from the society to build and maintain input, throughout and output legitimacy. Proposing and revising existing policies, the EC strengthen their proposals internally by having societal inputs supporting their policy propositions (Bernauer and Gampfer 2013, 440).

Whereas the EU depends on stakeholders' input to derive policies, stakeholders gain influence by expressing their preferences on policy issues. As the EU provides different venues for consultations, allowing a variety of stakeholders to express their preferences, stakeholders have their self-driven motivation to take advantage of these venues. Building on the insight gained from research on lobbying in the EU, stakeholders participate with the ambition of getting their policy preferences translated into policy outputs and outcomes (Löfgren and Lynggaard 2015). Despite not always achieving their preferred outcomes, they still take advantage of consultation venues in the hope to align the result with their most preferred outcome. Stakeholders' motivation to participate in consultations is thus understood to be driven from previous experience of some degree of alignment between stakeholder's expressed policy preferences and policy outputs. Additionally, acknowledging differences between stakeholders, motivations can range from agendas within corporations and interest groups to personal preferences and strong beliefs among citizens.

Seeing consultations as an opportunity for both the EU and stakeholders to exchange information and influence, understanding the plurality of received preferences is necessary. As Bunea (2015, 50) states, "mapping preferences is essential for understanding the aggregate constellation of demands formally expressed at EU level (…)". This highlights the importance of gaining insight into preferences and understanding how diverse they might be.

Investigating stakeholder preferences are thus seen as fruitful for researchers to gain insight and knowledge of EU policymaking. Through a broad approach to open consultations, large amounts of preferences can be collected, which allows for making assessments of how diverse preferences received are and how much difference the EU encounters in its goal of legitimizing its policy. While research has been done focusing on preference alignment and understanding lobbying success, little is known so far about which factors lead to more or less diversity in preferences within consultation.

1.2 A Novel Contribution

In this thesis, I contribute to our understanding of which elements within surveys used in open consultations impact diversity in stakeholders' expressed preferences on EU policy. Mapping stakeholders' preferences in open consultations across various policy issues allows for an assessment of the broader picture of preferences and extend the knowledge of how diversity in

stakeholder preferences can be impacted. Gaining an overview of preference diversity received through open public consultations leads to new insights into the plurality of preferences the EU receive through policy inputs from the broader public. With the utilization of clustering methods, I am able to collectively address preferences within the received responses to OPC surveys and map preference clusters which provide insight into how diverse preferences received from stakeholders are. Addressing preferences as groupings based on similarities within responses allows mapping and comparing preferences across policy areas. To gain insight into what might affect the diversity observed from the clustered preferences, regressions for evaluating survey elements impact are applied.

Existing studies rarely address the diversity observed amongst the stated preferences as much of previous research has focused on how stakeholder groups' preferences occur and how well they get translated into policies (Chalmers 2018; Pagliari and Young 2016). As the EU facilitates broad representation through consultations, a renewal of the understanding of which factors impact the likelihood of preference diversity is needed. Jeremy Richardson (Lowery et al. 2015), states "(...) that the 'heavenly choir' has mostly been drowned out by the voices of the masses, (…)" referring to the rise of new stakeholder groups gaining more access and more contribution than previously when business interests dominated the influence in the general political policymaking structures. This gives reasons to believe there is new knowledge to gain by exploring preferences voiced during EC consultations, and there is potential for diversity due the presence of a new "masses" (Lowery et al. 2015, 1224).

With the expectation of diversity in opinions and contributions from stakeholders, few attempts have been made to address diversity while not looking at the stakeholder specifically. Mizruchi (1990) highlights the mistake many researchers address as "assumed" preferences. Instead of focusing on which policy area they address and "theorization" of assumed interest, he emphasizes the comprehensive need to explain differences of empirically represented preferences to truly gain insight into differences. He additionally notes: "An objective definition enables us to avoid the often untenable assumption that similar motives produce similar behavior and that different motives produce different behavior," (Mizruchi 1990, 29). Whereas understanding how to facilitate preference diversity, this thesis aims to provide extended insight into how diverse stated preferences in OPC surveys truly are and how the survey used in OPC can affect the observed preference diversity.

Stakeholder preferences are thus seen as among the most elementary scopes to gain resourceful insight for EU policymaking. Through a broad approach to open consultations, large amounts of preferences can be collected, which allows for making assessments of how diverse preferences received are and how much difference the EU encounters in its goal of legitimizing its policy.

Van Hecke, Bursens, and Beyers (2016, 1434) highlight the importance to learn how, why, and under which conditions stakeholders use open consultations to upload preferences. In addition, critique is raised to the limited scope for understanding participation in open public consultations. By mapping diversity in stated preferences and focusing on the survey elements' effect on the distribution of preferences, this thesis will provide new insight into the effect surveys used in open public consultations have on received preferences.

1.3 Structure of the Thesis

In the extension of the given background on the European Union's increased focus on stakeholder involvement, chapter 2 addresses previous research on stakeholder influence and participation. Consequently, I review literature describing the different consultations instruments used by the EC to get public and external input on policy proposals and existing policies. Various venues for consultation raise diverge opportunities for stakeholders to express their preferences. Additionally, the different venues for consultations allow the EC to adjust the response, knowing their need for expertise or familiarity with raised issues. Finally, insight from existing research on stakeholder participation and thresholds for representation is presented.

In chapter 3, I present the theoretical framework of the thesis. This by first introducing the main concepts relevant for exploring preferences expressed by stakeholders in open public consultations. Secondly, I present the main theoretical arguments explaining how and why stakeholders participate and which elements impact their preferences. Considering the importance of understanding how and when stakeholders state preferences, give reasons for addressing components of the survey design which might affect their inputs. Thirdly, and lastly, I use these main theoretical arguments to structure the hypotheses presented accordingly.

In chapter 4, the data is presented. The data collection is thoroughly described, as data have been collected from open public consultation surveys. These questionaries used in the open public consultations have been human-coded to enable a more precise dataset for the abstraction of stated stakeholder preferences. After explaining the dataset function and the data-wrangling process it has been through, I describe the dependent, explanatory and control variables. Before elaborating on the methods used, I also discuss the problems human coding faces and measures taken to avoid dataset errors.

In chapter 5, I present my methodological approach. This accounts for the usage of clustering to obtain the dependent variables and regressions for testing for survey elements' effect on diversity in stakeholder preferences. All methodological applications are discussed as different tests have been run to determine which regression models to apply.

Chapter 6 entails the results from the analyses. The answer to diversity in consultations is discussed with results from clustering, and which effect survey elements have on diversity is determined. Negative binomial and multiple linear regression have been applied to the dataset to determine the effect of the explanatory variables on preference diversity.

Chapter 7 provides the discussion part of the thesis. First, I answer the research question by addressing the main findings from the analysis. In doing so, the implications for the hypothesis will be revised. Several evaluations have been included throughout the thesis, and I also elaborate on some of the most recurring challenges regarding the applications and methodological choices made.

In chapter 8, the conclusion is presented. Here, the conclusion for the research question is stated. Diversity in stated preferences is sufficiently supported as most surveys receive inputs indicating more than two preference clusters. Based on the findings of the analyses, the salience a survey experiences and whether they determine to target stakeholders are shown to be most prominent in affecting diversity in preferences received through OPC surveys. Lastly, I present notes for further research.

2 Previous Literature

This chapter will present an overview of existing research on stakeholder involvement in EU consultations. Through various consultations organized by the EC, stakeholders are invited to participate in the formulation of EU policies. These consultations are structured differently, providing stakeholders with different opportunities to voice their preferences. For instance, whereas some consultations involve closed roundtable discussions with invited stakeholders, others employ open public surveys for targeted stakeholders or for all interested parties to participate.

I start by providing a review of the increased focus on receiving policy inputs from stakeholders in EC consultations. By advocating for consultations with different stakeholders, the preferences may vary according to stakeholder origin, stakeholder type, consultation venue, and policy area. Though few scholars have centered their research on diversity in voiced preferences among stakeholders, research implies a greater likelihood for diverse preferences among stakeholders as they take advantage of different consultation venues, have different barriers for participation, and vary in their field of interest (Hermansson 2016). Investigating the relationship between stakeholders and the EC shows a vital exchange as influence and knowledge becomes trades for both parts (Rodrigo and Amo 2006, 2-3). However, due to limited research on diversity in stakeholder preferences, not much is said about the factors within the venue for consultation and how they impact the potential for stakeholders to differ in their inputs.

2.1 What Types of Consultations Does the European Commission Use, and When?

Previous research on stakeholder participation can be organized into three main strands that focus on the participation of stakeholders and their use of venues to express their opinions on EU policies. These three strands either see participation determined by the EU institution that holds the consultations, characteristics of the policy being consulted on, or stakeholder characteristics (Fink et al. 2021, 201). It is vital to understand the EU's role as a supranational actor starting on the institutional level before presenting previous research within the different channels existing for inputs. Background on the EU and its mechanisms will also give extensive insight into why mapping diversity in stated preferences among stakeholders is fruitful for understanding stakeholder influence in the EU research field.

Since the beginning of 2000, the EC has implemented several strategies to involve the public in their policy-making process (European Commission 2021b, 8). Accordingly, the initiative "The Better Regulation Agenda" was set to improve structural mechanisms for including public voices in the EU and increase the EU's internal transparency after assessing a White Paper on European Governance, launched in 2001 (European Commission 2001). While aiming to improve their accountability in public, the main focus has been to increase transparency in work within the European Commission (EC), as well between the three "main" institutional bodies making EU policy (the Commission, the Council, and the European Parliament). The EC is responsible for "planning, preparing and proposing new EU laws and policies," along with the responsibility to frequently propose improvements and evaluations of EU laws where it is deemed necessary (European Commission 2022). While the EC consists of non-elected executives, the Council and the European Parliament consist of elected officials. In detail, the Council consists of government ministries from every EU country, whereas the European Parliament consists of elected representatives from every EU country (Thomson and Hosli 2006, 393). When adopting new laws or amendments, the Council must approve the policy, making them the institution with the highest legislative power.

With the role of forming and proposing new EU policies and laws, the EC has the monopoly for informal agenda setting at the supranational level (Majone 2002, 324). Given their autonomy for proposing policies, the EC ultimately stands freely to dismiss alterations raised by other EU institutions if they see it conflicting with public interests. Though an unequal opportunity to put forward new regulations, the Commission alone is not equipped to know the potential impact caused by every policy they propose. Previous research frequently points to this aspect as one of the significant implications for the EC's authority. With the embodied role of proposing new policies to the different bodies within the EU, the EC faces obstacles as their administrative infrastructure is not adequate to know everything (Majone 2002; Fischer, Leydesdorff, and Schophaus 2004). The lack of knowledge and familiarity is relevant when discussing revisions of already implemented policies and understanding the consequences and implications they might cause.

In order to propose functioning policies, the EC is dependent on the information provided by those affected by it. Knowingly facilitating public input, the EC sees stakeholder preference and expertise as necessary in policy formulations as "(…) good policymaking involves those

affected by the decisions(…)" (European Commission 2021b, 4). The importance of inputs from stakeholders has caused increased prioritization of extensive consultations, both with independent experts, the public, and public officials, as it improves the legitimacy for policy proposals. Scharpf (2006, 850) argues that the EC's role as information gathers is utterly reliant on inputs from different societal groups. In his study of the joint-decision mechanism within the EU, the entire supernational body depends on the support from the different actors they regulate. The continuous exchange of expertise and information for influence can also bring restrictions for the EC. Relying on the expertise can also enable the Commission to put forward intended policies, which puts experts, the public, institutional stakeholders, and non-institutional stakeholders in a more regulative position. Scharpf (2006, 859) reviews this as a restraining mechanism, which he arguably means creates better conditions for policies that get put forward in the hierarchical machinery, a "win-win solution".

In a general notion, for the EU to sustain trust, it depends on experiences, values, and information provided by those affected by the issue raised. Active inclusion of stakeholders, including experts of different subject matters, citizens, and institutional actors, is deemed essential – primarily when their administration cannot provide sufficient knowledge alone. Nevertheless, consultations should not impose unnecessary burdens that can alter their efficiency (European Commission 2021b, 10). Balancing both forms of accountability issues from other EU legislative bodies and the public might not always yield satisfactory outcomes for every actor involved. The value of inputs gained for the EC might be questioned when not receiving adequate information or intended preference.

Facilitating participation can at times become fruitful for both the EC and stakeholders, but stakeholder participation can also become traps. As Yackee (2015, 431-432) acknowledges, stakeholders providing inputs might not result in their voices being influential for the policy output. Through her studies, she finds that decision-makers are "dependent on information," but the information given might not become realized. The potential situations created by lacking acknowledged influence can create dissatisfaction among stakeholders. Not getting their preferences met might result in less participation of stakeholders in the consultations. As for the EC, this means less expertise on different matters, which is vital for their legitimacy. Despite potential traps, the EC still aims to achieve broad representation and give "acknowledgment" to influential actors across the scope. As shown by Fischer, Leydesdorff, and Schophaus, "the Commission commits itself to enhance openness, participation, accountability, effectiveness

and coherence" (2004, 208). Building on the importance of stakeholders' motivation for influencing in order for the EC to access information and sustain legitimacy within the EU, research on how diverse the input gained through consultation is deemed necessary in order to understand the dynamics between inputs and what actually becomes implemented in the policy output (Rauh 2021, 4). Assessing what affects the potential for gaining inputs from different stakeholders will derive insight into more than just motivation for participation, hence what motivates stakeholders to express their preferences.

2.1.1 Different Venues for Stakeholder Consultations

Through increased focus on stakeholder involvement in the policymaking process, the EC has introduced more opportunities for participation through different consultation venues. One such venue is online consultations, which reduces costs and burdens for stakeholder participation as it is digitalized and accessible on the internet. Pisano et al. (2015) explain how the European Commission structures its role as a policymaker by addressing different mechanisms to improve its legitimacy while efficiently facilitating stakeholder participation. In their study, they define stakeholder participation as (2015, 5):

> *"(...)the inclusion of various stakeholders that can affect, or are affected by, the results of policy-making and decision-making processes. In general, a number of institutions and actors are invited to participate in such processes, for instance, civil society organizations/ NGOs, business representatives, social partners (i.e. trade unions, chambers of commerce, etc.), sub-national authorities, academia and individual citizens."*

By acknowledging the variety of actors referred to as stakeholders, the Commission has adopted diverse strategies to gain desired expertise and broad representation. In other words, the EC aims to make it easier for stakeholders to participate, regardless of their geographic closeness to Brussels or financial resources. In return, the EC gain access to more targeted expertise and information.

Fraussen, Albareda, and Braun (2020) identify different opportunity structures around the various consultation approaches when mapping these different consultation channels. Their analysis presents three distinctions for structuring consultation: Open, Hybrid, and Closed. The

three distinctions vary in their consultation tools and accessibility for stakeholders to participate. Open consultation involves “the utilization of tools that provide unlimited “self-selected” involvement to everybody who wishes to contribute—from private citizens to interest groups, firms and public institutions(…)” (Fraussen, Albareda, and Braun 2020, 476). Whereas open consultations allow every stakeholder to voice their preferences on specific policy issues, closed consultations facilitate targeted access for certain stakeholders, commonly organized through expert group meetings and workshops. Accordingly, hybrid consultation combines both open and closed consultation approaches (Fraussen, Albareda, and Braun 2020, 477). When looking at frequencies of used consultations, they find that the public officials used only one consultation tool, not both open and closed, for 50% of the regulations accounted for in the analysis. Knowing the access stipulated through the different consultations, the open consultation tools were remarkably less favorited, only constituting 7% of the consultations included in the analyzes (2020, 483).

In contrast to distinguishing between these channels of consultations based on the degree of accessibility, Stirling (2006, 2008) studies consultations by addressing how EU policymakers target their inputs from different stakeholders. Due to various tools being available for addressing which expertise to consult, he focuses on the consultation approaches. The three approaches are referred to as instrumental, substantive, and normative. The instrumental approach is focused on the "policymaker" and how they, through efficient discussion about the details of a policy, will be able to diffuse conflicts, "create ownership," legitimatize- and improve policy outcomes (2006, 96).

Whereas the instrumental approach limits the stakeholders' participation to indicate preferences on a policies' details, the substantive approach allows more involvement. Thus, non-expert stakeholders are included to ensure that the policymakers do not miss important information; the involved actors are thoroughly selected to supplement the feedback given by the experts. In this approach, the policymakers acknowledge the impact-related scope that policies can create and therefore want a broader representation of stakeholders to note and highlight, avoiding potential conflicts before implementation (2006, 97). The last approach, the normative approach, is grounded in the democratic ideal, obtaining maximal participation. Considering that targeted stakeholders are included in both the instrumental and the substantive approach, the normative approach allows all affected parts within societies to make comments and raise

questions on policy formulations before implementation (Stirling 2008, 270). In this sense, the normative approach is based on self-selected participation.

Whereas Stirling (2006, 2008) investigates how various motivations among policymakers impact stakeholders' potential for influencing policy formulations, Fraussen, Albareda, and Braun (2020) explain the accessibility structure for how the policymakers structure the consultation to serve their intentions. The two different approaches help explain the usage of tools that either limit or enrichen accessibility to consultations. Their explanations of the various venues can be illustrated briefly as open consultations, allowing self-selected participation, which resembles the normative approach from policymakers. In the same way, the hybrid consultation structure will adopt substantive approaches based on its combination of closed and open consultation tools. Lastly, the closed consultation structure facilitates an instrumental approach, targeting expertise inputs based on policymakers' preferences. Collectively, the different observations ground reasons underlying the expectations for who and how many stakeholders can be expected to participate. By applying more open consultation tools, broader representation from different stakeholder groups is expected, and with higher chances for multiple stakeholders, more diverse preference representation is assumed. For assessing diversity in stated preferences, this indicates higher chances of observing preference diversity in open public consultations.

Elaborating further on how consultations are structured, Arras and Braun (2017) interview 12 policymaking agencies within the EU. They find that EC agencies with regulatory competencies most frequently uses public consultations as input instrument. Following the categorization of consultations presented by Fraussen, Albareda, and Braun (2020) and Stirling (2006, 2008), open public consultations are characterized by lower barriers for access. With "low barriers," the policymakers intends to gather as much information as possible from the broadest possible audience, allowing participation from every "self-selected" stakeholder (Beyers and Arras 2019; Quittkat 2011; Pijnenburg 1998). Additionally, it is highlighted that open public consultations are assessed as one of the most cost-efficient due to being available on online platforms (Arras and Braun 2017, 1263). Arras and Braun's interviews with policymaking agencies indicate that policymakers find it beneficial to consult stakeholders early. They also state that all policymakers interviewed stress the importance of gathering inputs from different "relevant stakeholders" due their expertise and familiarity with consulting topics (2017, 1268).

Quittkat (2011), on the other hand, studies how the format of consultations impacts stakeholder participation. Using available data from open online consultations, she finds evidence that the consultations' format considerably impacts which type of stakeholders participate in consultations. The consultations can either be standardized, semi-standardized, or open. Standardized formats use targeted consultation strategies, whereas open formats facilitate self-selection for stakeholders to participate. Her findings suggest that more diverse stakeholder types use open consultations as different associations tend to participate in consultation on topics related to their field of expertise. For example, Health Associations shows the highest participation in consultations on public health (Quittkat 2011, 667). The variation in participation by public stakeholders differs due to their preference in policy areas (2011, 653-654). It is notable that she also finds "the more open the format, the lower participation (...)" and "(…) the more concrete and technical the issue, the less the Commission is prone to consult the wider public(…)"(Quittkat 2011, 670). In other words, her findings suggest that more open consultation formats attract broader participation, but still, fewer participants than in more standardized consultations. Open public consultations would in this instance be categorized as an open, yet standardized, consultation format as its standardized for how to approach yet open for receiving inputs from stakeholders interested to state their preference.

Various reasons can explain the broader, yet less, participation through open formats. One, showing that the technicality of a policy impacts the EC's usage of consultation, giving reasons for using targeted consultations as their need for expertise and opinions are more concentrated (Quittkat 2011, 670). Another indication for lower participation in open format consultations is stakeholders' preference in the consulting topic (Quittkat 2011, 665). Stakeholders tend to have concentrated expertise and agendas they work by. For instance, civil organizations need greater internal agreement within the organization to voice the organizations preferences. As this expertise, and organizational structure, vary between stakeholders, it is less likely to get representation of the same stakeholder across all policy areas.

Analyzing how different consultation formats and structures affect stakeholder participation, previous research indicates that whether consultations are open, targeted, or mixed types impacts the overall expectation for inputs on stakeholder preferences (Dür and De Bièvre 2007, 4). Taking lessons from previous research on implications and expectations for participation in the different venues used by the EC, the structure of the different consultation venues implies

different expectations for stakeholder participation. Hence, which format and consultation structure the policymakers use significantly impact which preferences get voiced.

Despite the findings indicating less expertise through open consultation formats, the overall response is still beneficial for the EC. Thus facilitating open consultations giving more stakeholders feedback opportunities will give broader responses bringing more nuances into the light (Quittkat 2011). Considering that the EU is a supranational system of governance, the broader approach gained through usage of open public consultations for collection of inputs can be seen as suitable due to the policies' impact on diverse nations and actors.

2.2 Which Stakeholders Participate in Which Types of Consultation?

In previous research, much focus has been devoted to explaining the institutional design of consultations and how they function as gatekeepers for the represented views (Beyers and Arras 2019; Pagliari and Young 2016). Further, scholars have analyzed how certain stakeholder groups have been included and which form of consultation they prefer to use. The overall tendency is that business interest dominates much of the consultation channels. In contrast to other stakeholders, they have the technical competence and financial means to involve themselves in the different consultations (Rasmussen and Carroll 2013; Klüver 2013b). Knowing that open consultations are preferably used as a supplement to targeted consultations, open consultation is still, as described in the previous section, characterized as less demanding to use compared to other consultation venues. Additionally, open public consultations facilitates a broader representation of diverse interests from the public. Given the low cost of participating, especially when online tools such as the *Have Your Say* portal are employd to enable the public to participate, mapping the preferences of different stakeholders represented in the overall consultation will be fruitful.

Before elaborating on how different stakeholders have been documented to participate, understanding how and when consultations are used will be necessary. As Grimaud (2018, 64-65) categorize it, the policy cycle can be seen as consisting of five stages; 1. Agenda setting, 2. Preference formation, 3. Decision making, 4. Policy implementation and 5. Policy evaluation. Within each stage, different consultation approaches are seen more frequently used than others.

In the first stage, “fixed use” and “all-around use” consultations are commonly organized. “Fixed use,” also referred to as open access-one-off meeting, indicates Commission usage of public and open characterized consultations to draw maximum attention and to mobilize for societal support. Consultation tools such as online open public consultations are frequently used in this part of the cycle (Van Ballaert 2017, 409). Whereas “fixed use” consultations are open, the “all-around” consultations are referred to as restricted access-consecutive meetings, where specific stakeholders are invited to participate in constructively problem-solving or consensus building. Examples of these consultations are often associated with targeted stakeholder meetings (Van Ballaert 2017, 409). More closed consultation venues are practiced for the next step, preference formation. In these “custom use” consultations, more restricted venues such as seminars and workshops are used to gain more specific and technical inputs, not gaining legitimacy. For the decision-making stage, combinations of open feedback on roadmaps and impact assessment are used. Lastly, the policy evaluation stage often consists of ex-post evaluations consultation venues where feedback can be provided and often results in ex-ante assessment for new public actions. The use of consultations might vary with which DGs are organizing the consultations (Van Ballaert 2017, 419).

As the different stages of the policy cycle differ in their need and opportunity for stakeholders to participate, Grimaud (2018) studies the impact of participation and functions of these different stages. Looking at the three main decision-makers in the EU (EC, EP, and the Council), fewer actors are granted involvement as further in the policy cycle a policy gets. Based on timing as an essential mechanism for regulating participation, Grimaud (2018, 84) states that the public gains on channelizing their preferences in the earlier stages, highlighting the policy-formation stage as most beneficial.

Supplementing Grimaud (2018) findings, Bernhagen, Dür, and Marshall (2015) study the role of policymakers' "friendliness" to assess lobbying success in consultations. "Friendly DGs" is defined as Directorate-General (DGs) which ideologically aligns with the stakeholders. Through the analysis, "friendliness" is stated to increase the likelihood of participation of interest groups who agree with the policymakers (Directorate-General) and are specially invited to consult on specific policy proposals. The DGs require external expertise, and therefore they tend to require the assessments from stakeholders they have maintained more regular interactions with the DGs. This continuous interaction gives "friendly" organized interests an advantage in being included when consultations are less open (2015, 573).

Extending the insight on participation in EU consultations, Pakull, Marshall, and Bernhagen (2020) investigate the potential advantages of the different venues, recognized as consultations at different policy stages, which can be addressed as most successful for stakeholders. Their findings support the previous indication outlined above. Organizations with less technical competence and resources are considered more successful in influencing the earlier stages than those stakeholders with more technical competence. During open consultation, which is more frequently used in earlier venues, intentions are to gain as much information as possible. Having more technical resources does not create more influence in open stages. Ergo, highly technical organizations gain more influence when invited to consult through other venues (Pakull, Marshall, and Bernhagen 2020, 522). Therefore, organizations, such as business interests, are more likely to participate when they get privileged access to consultation. This relationship can be explained due to the EC's increased focus on the inclusion of non-business organizations through facilitating less costly participation in general for stakeholders.

Gornitzka and Sverdrup (2010) study the consultation with stakeholders in expert groups consultations. DGs use expert groups to target experts' inputs for elaboration on policy formulation options. Expert groups are organized to include experts within consulted policy areas. As for the EC's role in introducing policy initiatives, the expertise gained through expert groups consultation assures valid policy formulations. As open consultation venues encourage self-selected participation among stakeholders, giving the EC no reassurances for gaining needed expertise. Expert group consultations are less frequently used compared to the other consultation venues organized by the Commission (2010, 49).

Discovering that most participants in these targeted consultations usually represent governmental stakeholders and scientists clearly illustrates a paradigm for business interest. As business interests are not usually being consulted in these venues, Gornitzka and Sverdrup (2010, 56) find that business interest benefit from participating in the earlier and more open venues. Despite suggesting contradicting findings to Pakull, Marshall, and Bernhagen (2020), gaining insight into participation within expert groups helps shed light on how and where stakeholders participate.

2.3 How Does Stakeholders' Influence on Policy Differ by Type of Consultation?

Within the EU's increased focus on opportunities and to lower the threshold for the participation of stakeholders in different venues of the policy cycle, the better regulation agenda has been a leading force in maneuvering how to increase access for the public (European Commission 2021a). As encountered by Lowery et al. (2015), the EC is dependent on the external expertise from different stakeholders as they are under-resourced in the range of functions they do in the policy-making stages. To address the various aspects affected by policies adequately, there is a need for communication and exchange (Lowery et al. 2015, 1271).

Whereas scholars have been looking at how lobbying can create different "sides" among stakeholders, either by coalitions or by potential accident (Bunea 2014b), raising their voice has been based on either one of two strategies. Outside lobbying pays attention to the usage of public mechanisms, creating a public "outside" pressure on policy-makers (Dür 2008b, 561). Making "noise" can create external pressure which in turn might help increase the likelihood of obtaining a preferable outcome on policies. In contrast to outside lobbying, inside lobbying happens within the venues through different consultations and traditional lobbying channels (Dür 2008b, 567). Inside lobbying has been stated to be used less frequently by stakeholders voicing "diffuse interest" (Beyers 2016, 217). "Diffused interest" is commonly used to describe preferences which can consist of contradicting opinions within a stakeholder group. It is noteworthy that there can be collective, and therefore more coherent, preferences or interests, just as there can be more complex and ultimately more diffuse ones.

Strategies and influence differ as stakeholders vary in their recourses, availability, and technical competence. Whereas some might exert influence through consistent participation, others base their influence on the knowledge they can offer. As mentioned previously, interest groups are among the few stakeholders that take advantage of external venues, whereas they can "make noise" to get their preferences heard. For instance, if they do not get heard nor access to consultations their voice can be publicly displayed through demonstrations, media coverage, or rallies (Dür 2008a, 1222). Although interest groups can use different means for influence, they have shown to be more aligned with policy proposals compared to business interests. Reasons why interest groups more often support new proposals relate to their wish for generalized regulations (Dür, Bernhagen, and Marshall 2015, 952). Despite their efforts, interest groups are shown to be able to influence only technical aspects of policy proposals, not the core issue (Dür 2008a, 1219).

In analyzing the variety of interest groups that participate in financial regulatory policymaking at the EU level, Pagliari and Young (2016) provide evidence speaking for the perception of business dominance in EU consultations. Studies show that business interests, such as financial organizations, business associations, and companies, have emerged as the most active participant (Klüver and Zeidler 2019; Lowery et al. 2015; Chalmers 2014). Looking at preference attainment, they extract information about stakeholder preferences for financial consultations. One of their findings shows a more coherent formulation between financial industries, in addition to being more coherent with the final policy proposal. Compared is preferences voiced by civil societal interests, such as NGOs, consumer protection representatives, and organized research groups, more diverse and less aligned with the final policy proposal (2016, 329). Civil societal interests seem to suffer in preference attainment due to lack of mobilization, thus indicating variation among preferences voiced during financial consultations.

Despite the EC's increased efforts to lower cost of participation for stakeholders, there are still signs that business interest dominates in consultations. Interest groups and societal organizations represent diverse masses of people. As they consist of more "diffuse" interest, they has shown to give stakeholders representing more coherent interest a better advantage in influencing policies (Dür and De Bièvre 2007, 6). Bunea (2013, 566) finds that the stakeholders who experience influencing policymaking, representing a median position, more specifically aiming to maintain the status quo, influence policy outcomes more successfully. As preferences and influence are difficult to measure, Bunea (2013) uses preference attainment to indicate stakeholders' success. Preference attainment refers here to whether the outcome of the policy is aligned with stakeholders' expressed preference for the policy. Business interests are experiencing greater success because of their efforts to "represent concentrated interests" (Bunea 2013, 567).

Further, Hermansson (2016) studies which function increases lobbying success. On this notion, lobbying success is transferred with policy success as to whether a policy proposal from the EC successfully gets transferred into policy. Addressing the three questions raised in the research, whom you know, what you know, and what you own, the overall results indicate that knowledge offered by the stakeholder is the most important prediction for experiencing lobbying success, i.e., influence (Hermansson 2016, 194). However, "(…)the consultation also significantly favor

wealthy business interests" as access through a direct approach often relates to owning an office in Brussels (Hermansson 2016, 192). Due to business interest perceived highly technical expertise, they tend to have relevance within some policy issues in almost all proposals, giving them an advantage for pursuing inside lobbying on the policy level. In terms of representation advantages, findings on the issue-level suggest otherwise. The business interests mainly yield advantages associated with resources, making them less likely to be successful on the issue-level, creating a bias towards their resources (Hermansson 2016, 195). Additionally, positive relations regarding stakeholders supporting the EC's policy proposal have an increasingly higher percentage to result in policy success.

Although resources and preferences alignment are highlighted as important factors explaining success in influencing across venues in the EC, Beyers and Braun (2013, 110) find evidence indicating relations and "ties" as influential for interest groups' access. Interest groups are, as previously shown, highlighted to be "favored" to include in expert venues due to their expertise. Their relevance might additionally be due to their potential for gaining new insight. Access gives better conditions for influencing, and interest organizations strengthen their possibility to be targeted for consultations by having informal ties with other organizations. Beyers and Braun (2013, 115) analyses show that these "ties" should not include close relations but rather "weak ties". Explained, not interacting too often gives opportunities for supplementing new information when first interacting. Maintaining "weak ties" with several actors helps gain more insight, allowing higher chances for access (Beyers and Braun 2013, 116). Ultimately, ties to distant stakeholders and policymakers show high importance for interest groups' possibility to influence.

When organizing consultations, the EC prioritize to facilitate for approaches based on their need of representation and stakeholders' advantages. Nevertheless, there are diverse strategies for influencing and getting preferences heard. As shown, resources such as expertise and alignment help give advantages when stating preferences through the various venues in the EC. The consultations, especially the open consultations where everyone can participate, are by scholars emphasized to consist of contradictions and competing views (Grant 2015, 4). For Truman, this contradiction is a natural tool which can create alliances and "groupings" (Grant 2015, 5-6). Accordingly, Beyers and Braun (2013) sees these groupings as efficient for interest groups, mainly focusing on ties and familiarities which can help stakeholders gain access. Despite

gaining access, Truman highlights approaching consultations collectively as the best opportunity for stakeholders to influence overall policymaking (Grant 2015, 5-6).

2.3.1 Lobbying Strategies as Tools for Influence

The trade between access to expertise and the opportunity for stakeholders to voice their preferences in the EU has been a growing topic within the research fields of political science (Lowery et al. 2015). Within the different venues and institutional mechanisms explained earlier, they all facilitate diverse sets of lobbying mechanisms. One must pay attention to lobbying in order to properly understand this trade.

As a means of influencing, lobbying has gained increased focus in the past years (Pagliari and Young 2013; Hermansson 2016; Klüver 2013a). While acknowledging the competition within stakeholder participation, different systems have been revealed. Lobbying as a function is by Klüver (2013b, 4) conceptualized as "(…) an exchange relationship in which the European institutions' trade influence for information, citizen support, and economic power." In such exchanges, it is sometimes expected to be contradicting opinions, and knowing how others positions themselves might create strategic options for collaboration (Klüver 2013b, 95).

Assessing how lobbying is used as a tool for influence raises expectations regarding preference alignments within stakeholder groupings and diversity between different stakeholder groups. As Bernhagen, Dür, and Marshall (2015, 573) states, "the context argument starts with the insight that different policy-making venues are more favorable to some organized interests than to others and that groups adjust their strategies accordingly." This is further strengthened as Holzer (2008, 61) highlights how NGOs often share lobbying strategies, allowing for adopting similar strategies and sharing of ties. In contrast, when NGOs increase their possibilities for policy alignment they experience more competition due to the growing audience perceived to participate in EC consultations (Lowery et al. 2015, 1216). With growing participation from a range of stakeholders, lobbying strategies imply more similar approaches among stakeholders within the same stakeholder groups which share policy preferences.

Drawing from a growing audience of stakeholders, which more frequently seek out alignment both in terms of policy output and with more like-minded stakeholders, this implies expectations for more diverse preferences with a higher degree of density in consultations.

Messer, Berkhout, and Lowery (2010, 185) indicate an additional notion for the growing coherent approach among lobbying strategies, finding evidence for long-term consultation to receive more density among social participants and providing expectations for the density preferences encountered.

3 Theoretical Framework

This chapter will provide insight into theoretical arguments, giving reasons for exploring diversity in stakeholder preferences expressed in open consultations. Building on insight gained from previous research on lobbying success, we know there are differences in how civil society groups and business groups operate when lobbying and attempting to influence decision-makers. Additionally, the EC uses different venues and strategies to gain insight and expertise from different stakeholders. Despite the efforts done, few attempts have been made to address and explain how eliminants within the consultation effects the diversity within received preferences. Knowing that preferences represented in open consultations are received through stakeholders' self-selected participation, I aim to present observations explaining why stakeholders participate. Explaining participation, and how different elements within the survey affects participation, helps derive expectations about the level of diversity observed in stakeholder preferences. Whereas limitations can occur due to modes of consultations, the policy issue can also impact which preferences are observed.

Firstly, I present previously explored features shaping stakeholders' preferences. Elaborating on previous empirical evidence will help explain how stakeholders position their inputs during consultations and where to expect them to be present. Secondly, I will elaborate on external limitations impacting participation. Already addressed, both the consultation and policy issues affect stakeholders' expectations for stating preferences. Lastly, I will present the hypotheses derived from explanations from previous research.

3.1 Conceptualizing Stakeholder Preference

In the context of understanding why assessing diversity in stakeholder preferences are considered valuable in political science and the context of the European Union, elaboration about the increased means of influence gained through voiced preferences is necessary.

To state preference in the context of the embedded democratic ideal, participation is crucial and should be possible for everyone concerned with political regulations in a society (Kohler-Koch

2010, 100). In other words, participation and preference are directly linked as participation is based on expressing preferences and expressing preference direct gives higher credentials than through representation when considering democratic functions.

Within Europe, several input options exist for civil society to state their preference on political issues. Previous research shows bias in preference representation in the EU due to the overrepresentation of financial interests. With hopes to alter this tendency, the Commission has been increasingly facilitating broader representations, securing opportunities for even more stakeholders within different fields of competence to engage (Kohler-Koch 2010, 110).

Still, "the organizational capacity of societal interests is distributed unevenly and, due to the economic origin of the EU, market-related associations have a long history of a strong presence in Brussels" (Kohler-Koch 2010, 110). By allowing diverse sets of stakeholders the opportunity to state their preferences, the EU gets inputs vital for their legitimization.

Influence is additionally crucial for the representation of stakeholder interest. Stakeholders who prioritize participating in open public consultations have the goal of being able to influence EU policies. This motive is vital for the EU institutions to consider as they need interaction to maintain autonomy. Influence is defined by Dür (2008b, 561) as "an actor's ability to shape a decision in line with her preferences, or, in other words, a causal relationship between the preferences of an actor regarding an outcome and the outcome itself". The facilitation of receiving preferences does not imply similar preferences. Stakeholders vary broadly and might change their preferences due to interactions with other actors or gaining new insight (Dür 2008a, 1219).

3.2 Factors Determining Stakeholder Participation

Stakeholders vary in their organization, competence, and size. As of February 19th 2022 more than 13000 stakeholders have registered in the Transparency Register (European Union 2022), showing participation of diverse set of actors in the different venues for influence. While it is voluntary for stakeholders to register information regarding their organization or firm in the Transparency Register, it is not necessary to register for participating in the open venues in the EU. However, participating in open public consultations allows for increasingly more prominent population of stakeholders to have the opportunity to participate through open public

consultations versus other types of consultation such as targeted consultations. Indicating high levels of participants increases chances for variety between their stated preferences. When discussing stakeholders, scholars tend to distinguish them based on public, societal, or business interests.

Whether being an EU member, or an associated external partner, public institutions tend to participate on behalf of a country's interests. Though public actors can arguably be said to speak on behalf of civil society, it does not include particular interests within society. Societal actors such as trade unions, interest organizations, NGOs, academia, and civil individuals, are often observed to provide representation for more specific interests (Pagliari and Young 2016; Klüver 2013b). These actors all represent distinct interests and are often referred to as *interest groups* (IG). Whereas both public and societal stakeholders might be associated with being related, societal organizations are frequently working towards both national and supranational institutions. In contrast, private stakeholders are often classified as *business interests* and are frequently composed of private businesses, trade associations, and professional associations.

3.2.1 Public interests

The input of various stakeholders representing public interest is prioritized in consultations. Their specialized expertise and representation of public opinions make them desirable stakeholders for the EC (Persson 2007, 234). Most of these stakeholders encountered as IGs are acknowledged by their targeted expertise within a specific field. With the specialization and knowledge in specific political areas, interest groups are often assessed as experts on political issues.

Among interest group stakeholders, they have one thing in common: representing the public interest. They desire to influence politics regarding what best serves their and public interests through participation. For example, public interest tends to engage in consultations regarding budgets as they will have substantial impacts on public funding (Rasmussen and Carroll 2013). National trade unions pursue participation based on what policy is considered close to their program and interests. While they engage in consultations affecting their interest together, they still may vary in their standpoints (Larsson 2015, 106). As public interest might vary regarding what is happening in the world and what their members are engaged in, variation in preferences is assumingly relevant for other stakeholders representing public interests. Interest organizations, for example, have multiple interests making it harder to clearly state preferences.

Research has shown that citizen organizations often support new regulations which allow more generalized regulations across country borders (Dür, Bernhagen, and Marshall 2015, 952).

Public organizations can be organized in different ways, and citizens stand freely to approach issues according to their preferences. In contrast, public authorities represent national interests, making them accountable for conflicting political interests. Taking into account both political interest and national interests creates restrictions and diffusion in assessing issues (Yackee 2015, 249). Interest organizations must also consider what their members prefer, having citizens as regulators. As "diffuse interests are difficult to mobilize," it might be more challenging for stakeholders representing public interest to coordinate their preferences, hens gaining power through collectively impact politics (Beyers 2016, 217).

For instance, despite having more challenges collectively stating preferences, organizations working within environmental causes are seen to interact with other organizations more frequently. This is often an exchange due to a lack of resources that other organizations and actors have (Bouwen 2004). As stakeholders representing public interests tend to vary in how they are structured and participate, there are indications confirming the interaction between stakeholders sharing the same core beliefs (Dudley and Richardson 1999, 228).

3.2.2 Business interests

Business interests are those interests representing financial interests. In the case of business interests, they have dominated consultations from the beginning of 1985. From 1985 to 1995, the EU increasingly focused on entrepreneurship, commercialization, and privatization, ultimately attracting business interests (Fischer, Leydesdorff, and Schophaus 2004, 201). Their expertise is often related to financial competencies and skills. Therefore their presence is observed in consultations across several policy issues (Dür, Bernhagen, and Marshall 2015).

Through previous research, business interests are highlighted as resourceful stakeholders as they manage to keep a broader view over issues being consulted. Golden (1998, 257) explains this as they "utilize much more sophisticated monitoring techniques than the smaller advocacy groups." By having more sophisticated techniques for insight, they tend to have better opportunities for stating their preferences across consultations. In addition to monitoring consultation opportunities within the EU more efficiently, business interest tends to have more

resources. Resources such as specialized employees and more financial freedom give business interests advantages such as being more "present" in Brussels (Bouwen 2004b, 201).

Business interests often benefit from their structural organization as they consist of hierarchies and have more concentrated interests (Dür and De Bièvre 2007, 6). Having hierarchical structures allows them to focus on specific information across different issues. Historically, business interests have dominated financial issues (Pagliari and Young 2016; Pagliari and Young 2013). Business interests prefer more privileged access to consultations as they have more technical expertise and resources (Pakull, Marshall, and Bernhagen 2020, 525). Privilege assess is understood as being invited to consult, rather than participate based on self-interest.

In addition, business interests tend to gain influence in more technical and targeted consultation venues. Despite being highly internally organized, business interests find it harder to get access in targeted consultations as the EC increasingly focuses on including public actors. When stating preferences on financial issues, they tend to influence the most when speaking "with a single voice" (Chalmers 2018, 391). Whereas interest organizations and business interests both tend to speak in "unified voices" or share constitutional ties, but with different motives, gives reasons for predicting the presence of preference diversity within OPC consultations.

3.2.2.1 The Heavenly Choir

Golden (1998) raises awareness about a “heavenly choir” among lobbyists due to who states their preferences and who gets heard. This “heavenly choir” is stated to dominate much of participation, ultimately giving business interest a “stronger voice” (Golden 1998, 257). Building on the perception of dominance by business interest, which Golding referred to as “whether they are corporations or trade associations, they deem to assess more “sophisticated monitoring techniques” that raise advantages when consultations arise. Having resources in that aspect gives business stakeholders benefits in mainly two ways; understand the technical complexity of the policy, and secondly, address the consultations while they are open.

Despite business interests’ clear advantage in regards to participation, business interests differ in advocating their cause. Thus, Dür, Bernhagen, and Marshall (2015) highlight that business interest do not act as a unified actor. As intentions for participation can vary, “the type of issues on the political agenda of the EU” steer how business interests position themselves in the consultations as they often desire solutions closer to the status quo (Dür, Bernhagen, and

Marshall 2015, 957). Non-business interests, in contrast, often prefer policy changes that regulate for more harmonized standards on higher levels within Europe. As non-business interest does not necessarily speak in unified voices, this gives reasons for more variation in preferences.

3.3 Consultation Elements Observations Impacting Stakeholder Participation

In the case of preferences, external observations refer to limitations or tools not explained stakeholder characteristics. As Fink et al. (2021, 199) highlight, the literature on stakeholder participation within the EU often concentrates on three main functions; *institutions, policy design, stakeholder characteristics.* They all create different limitations and opportunities for stakeholders to participate. Institutions, such as the EC, create certain restrictions through the different venues for consulting. As the consultation limits or enriches the variety for expressed preferences, the policy design creates limitations. The policy design is adjusted as to which issue is being consulted. Therefore, one can arguably address policy issues as a limiting force itself.

3.3.1 Consultation Venues Effect on Participation

Venues used for consultations create different obstacles for expectations to which stakeholders will participate. For instance, the period in which a venue is open for receiving feedback. Open consultations are restricted to at least be "open" for feedback for two months (Quittkat 2011, 665). Only allowing feedback in a timespan for two months creates challenges for less organized stakeholders which might not have the resources to prioritize formulating a response.

Additionally, the stage of the policy cycle might have consequences for which stakeholders are granted opportunities to state preferences (Pagliari and Young 2016). Notably, the policy cycle's earlier stages are favorable due to their possibilities for changes. Stakeholders benefit from participating in the policy formulation stage. It is also beneficial for the EC to get insight from stakeholders in the policy formulation stage due to their lack of insight and expertise.

As for the EC's need for expertise, venues can vary in their extent of accessibility for stakeholders to state their preferences. The EC can target which stakeholder preferences they receive by using different consultation tools such as targeting stakeholders and organize conferences for discussion (Van Ballaert 2017, 408). By appliance of such tools, the EC gets

desired insight through different structures, including only relevant preferences. Despite this being reasonable for the policy being issued, it still bears consequences for lack of representation of preferences affected by the policy.

When formulating policies, addressing broader audiences to receive inputs from stakeholders or the public helps indicate if the policies are needed or appropriately adjusted. Using consultations such as open consultations in policy formulation stages ensures that the Commission receives competence and insight from different stakeholders which helps minimize technical challenges. As well as for the EC to gain insight and competence, it is documented higher levels of "pay-off" for stakeholders to voice in consultations happening in the policy formulation stages (Bernhagen, Dür, and Marshall 2015, 571).

Drawing on the necessity for expertise, venues do not have to be accessible for all stakeholders. The EC can use targeted tools to get desired insight (Fraussen, Albareda, and Braun 2020, 474). Despite this being reasonable for the policy being issued, it still bears consequences for lack of representation of preferences affected by the policy.

3.3.2 Policy Issue's Impact on Participation

The policy issues the EC consults on severely affects the type of stakeholder who are expected to participate. Previous literature highlights technical complexity and salience as the most prominent causes for limiting stakeholders' participation in consultations (Fink et al. 2021, 216).

When referring to technical in this regard, advanced wording and technical terms are creating difficulties for stakeholders with less expert knowledge (Fink et al. 2021, 215). Creating higher barriers for understanding the context of the issue being consulted can potentially reduce interest in participation. By using too complex wording, the consultation will result in a skewed representation of different stakeholders (Rasmussen and Carroll 2013; Pagliari and Young 2016; Fink et al. 2021).

On the other hand, simplifying the information to consult issues can create too few barriers. When policy issues are salient, essential knowledge and expertise can get lost in the masses as they tend to gain too much input (Pagliari and Young 2016, 314). *Salience* is thus defined as

"the attention paid to one issue by stakeholders, as indicated by the number of organizations expressing a preference on that issue" (Bunea 2013, 556). As implied, experiencing too high degree of salience can cause stakeholders to abstain from stating their preferences as they will not be able to influence due to the size of responses. Being "drowned in the masses" is highlighted by Todorova (2020, 52) as she finds EU advocates to have 0% for fully attaining indented goals, whereas 77.8% succeed with attaining nothing or only slightly influence. Despite this, Todorova (2020, 52) remarks, "if policy issues are salient, a wide variety of interest groups are working on these issues, such as business groups, trade unions, and non-governmental organizations," implying broad attraction from different stakeholders, and most likely different preferences.

3.4 Hypotheses

Through the literature presented, differences between stakeholders and consultations have been shown. Drawing on the most noticeable difference between public interests and business interests, hypotheses can be formed about the diversity of stated preferences in open consultations based on their striking differences regarding preference position, competence, and resources. However, as previous literature indicates differences between stakeholders' preferences based on organizational structures, other elements can affect diversity in stated preferences at the consultation level itself. Bunea (2015, 64) finds preferences to vary within cases, which gives reason for studying issue-level characteristics. As this thesis intends to address diversity in open public consultations; technicality, salience, and policy area are stated to affect participation. With the intention to analyze stated preferences within surveys used in OPC, in addition to test for effects on diversity in stated preferences, the hypotheses have been focused on survey-level elements.

Dudley and Richardson (1999) state that core beliefs have impacted how stakeholders address each other, and how they potentially align when replying to consultations. Additionally, Bunea (2014a, 14) finds that groups sharing an organizational tie have a 70 percent increased probability of sharing verbal behavior, strengthening the belief of more aligned preferences between similar stakeholders. For example, stakeholders representing public opinions have more aligned preferences with proposed changes, whereas business interests prefer the status quo. The chances of receiving inputs from different stakeholders associated with different interests will increase the higher salience a policy issue gains. As previously shown, salience

attracts broad forms of representation, assumingly for various stakeholder groupings. Salience can be measured as a survey characteristic, giving reasonable indications for predicting a diverse set of preferences. Accordingly, one can therefore expect to see more diversity and more expression of different preferences the more publicly known an issue is:

H_1: Open public consultations on more salient issues are associated with more diversity in expressed stakeholder preferences than open public consultations on less salient issues.

Despite expecting stakeholders sharing ties to state similar preferences and for different stakeholders groupings to establish differences, Binderkrantz et al. (2022, 17-18) show that open consultations are less attractive for stakeholders to use. For instance, more targeted approaches to consulting have been of greater interest to business interests. Although open public consultation surveys are available for all interest actors, some may still mention specific stakeholders. This can be seen as a means for gaining technical competence and insight from expert stakeholders for the given policy issue. These might be favored by specific stakeholders, such as business interests, but meanwhile, reduce the possibility of participating from other than specified stakeholders. Targeting stakeholders gives reasons to believe in reduced participation in OPC surveys. With reduced participation, less diverse stakeholders will participate. Despite not guaranteeing similar preferences from the same stakeholder groupings, targeting stakeholders are predicted to have a negative impact on general participation and the level of preference diversity:

H_2: Open public consultations that target specific stakeholder types are associated with less diversity in stated stakeholder preferences than open public consultations that refer to "all" stakeholder types.

As shown in the theory section, both the stage in the policy cycle and the venue format impact which stakeholders can be expected to state their preferences. As I will be using data collected from open consultations arranged in the policy formulation stage, I won't be able to explore how stakeholders' expressed preferences vary across different stages of the policy cycle. The policy issue can be predicted to limit general participation. Stakeholders tend to vary in forms of technical competence and resources. Hence if the issue being consulted on is highly complex, consisting of technical wording, it is less likely to receive diverse inputs. This implies a negative

association between issue complexity and stakeholder diversity. Accordingly, the following hypothesis is presented:

H3: Open public consultations on more complex issues are associated with less diversity in expressed stakeholder preferences than open public consultations on less complex issues.

In line with hypothesis 3, concerning issue complexity, is survey design complexity associated with the need for more resources. In the case of addressing survey design, survey design complexity is understood as the complexity related to how the survey is composed. If surveys contain long questions, it might be harder to follow through on the survey. In the same vein, if a survey consists of many questions, it might appear harder to complete for less resourceful stakeholders, such as individual citizens. As the design of a survey will impact the chances for stakeholders to finish through, the following hypothesis is presented:

H4: Open public consultation that use more complex surveys are associated with less diversity expressed stakeholder preferences than open public consultations that use less complex surveys.

Response to open public consultation surveys have been collected to test the hypothesis. The following chapter thoroughly describes how the data used in this thesis have been collected. For assessing diversity in stated stakeholder preferences, preferences within the responses to open public consultation surveys have been clustered. Using clustering methods enables to assess preference diversity in the means of groupings based on similarities of responses received. Additionally, the utilization of explanatory variables is described. These have, in addition to the preferences, been collected based on information gathered from the surveys.

4 Data

As shown by previous research on stakeholder consultations in the EU, various consultation venues can be used for stakeholders to express their views to influence the EU's policy. In this thesis, open public consultation surveys will be used as data. Through previously presented literature, OPC has been proclaimed to be a venue allowing participation from the wider public (Austgulen 2020, 774). It is additionally assessed to have the lowest thresholds for participation, and the results from the consultations are available digitally. Using OPC surveys submitted from stakeholders across different legislative proposals allows for the adoption of correct methods to provide a picture of preference diversity within the survey collected. Additionally, the surveys are available digitally.

In this chapter, I first elaborate on collecting data and how the data wrangling has been conducted. As this has been the most time-consuming part of the thesis and all data is singly-human-coded, a thorough revision of the collected data and their operationalization is necessary.

4.1 How to Approach Diversity in Stated Preferences

In the research discipline, stakeholder preferences in consultations in the EU have been increasingly researched. In the context of previous research on stakeholder preferences, the approach mainly consists of preference attainment (Bunea 2013; Löfgren and Lynggaard 2015; Bunea 2017) or preference alignment (Bunea 2014b; Chalmers 2018). Through this, indications of how different stakeholder groups tend to position themselves within consultations and how well they succeed in influencing based on their preferences have been the main focus. Understanding stakeholder preferences can be tough as there are several different types of stakeholders, and not even the same stakeholder groups can be expected to state similar preferences (Binderkrantz et al. 2022, 17-18).

I aim to investigate stakeholder preferences, but not in the same vein as much of existing research has done. Instead of looking at specific preferences, the position of stakeholder-group preferences, or how successful preferences get translated into the final output, this thesis intends to empirically the degree of diversity in the preferences revealed in OPC surveys and how the elements of the surveys affect this diversity. I only measure effects of survey-related elements instead of looking at explanatory variables on the stakeholder level, such as organizational

structure and resources. Nor will variables that can explain activity and relationship between stakeholders and the European Commission, such as the Brussels office and count of presence in other consultation venues, be included. By looking at the survey-level, and how this level potentially regulates or creates diversity in stated preferences, the variables used will be somewhat different from previously used variables to explain differences. Variables related to the structure and functions of surveys are exclusively collected to explain diversity in stakeholder preferences.

Specifically, preferences in the case of this thesis are addressed as responses to surveys in OPC. Through clustering methods, preferences within the received responses to OPC surveys have been collectively addressed, giving an overview of groupings of similar preferences among the attributions for the given surveys. Emphasizing that I am not looking at preference but rather the diversity in preferences allows for assessing preferences in a pluralistic term and treating them as different groupings of interest instead of individual preferences.

4.2 Dataset: Surveys in Open Public Consultations

Survey data collected through open public consultations have facilitated a low degree of thresholds for participation. With a low degree of thresholds for participation, it is more likely to receive inputs from different stakeholder groups. At the same time, previous research shows that open public consultations experience a lower degree of participation compared to targeted consultations (Quittkat 2011, 670). With the intention of mapping diversity among responses, OPC surveys still provide the best conditions for doing so. Internally in survey data, the questions and the design of the surveys have proved to have a great impact on the outcome of inputs received. This makes surveys a good starting point for investigating how diverse preferences within responses from stakeholders are.

The open public consultation surveys used in the dataset have been hand-picked based on the following six criteria:

- Available as an Excel file or as a CSV file.
- EU documents dated between November 1, 2014, to May 1, 2021.

- Documents fall into the "Preparatory documents" section within the Eur-Lex classification system, allowing for insight into stakeholder states' preference for policy initiatives for legislative proposals[2].
- Legislative proposals by the EC.
- Proposals for binding secondary EU law[3].
- Documents authored by the following selection of DGs:

Table 4.1 *Selected DGs.*

ENV	Environment
CLIMA	Climate Action
GROW	Internal Market, Industry, Entrepreneurship and SMEs
EMPL	Employment, Social Affairs and Inclusion
HOME	Migration and Home Affairs
MOVE	Mobility and Transport
ENER	Energy

The DGs, shown in Table 3.3.2.1, are selected due to their difference in the policy area, the possibility of generating interest from a range of stakeholders, and their frequency of proposing policy initiative proposals (Rauh 2021, 14, Appendix B).

Further, as for the collected open consultation surveys available for the 43 proposals of the EC, the questionaries within each OPC survey for the legislative proposal have been assessed. This process has been part of a three-part collaboration where two coders have parallelly labeled questions into categories indicating the type of information respondents are asked to give[4]. Alongside the categorization, the third person was responsible for cross-checking the labeling of questions and validating potential disagreements between the two coders. Applying a three-step coding procedure for validating the correct information gained from answers to the questionaries reduces the chances of abstracting other information than intended.

[2] See https://eur-lex.europa.eu/content/help/eurlex-content/documents-in-eurlex.html. For detailed overview of policy initiatives encountered, see Appendix A.
[3] There are four possible document types in Sector 5 - Preparatory documents: PC for legislative proposals by the Commission (COM documents), etc.; DC for other COM documents (green and white papers, communications, reports, etc.); SC for SWD documents (staff working documents, impact assessments, etc.); JC for JOIN documents (adopted jointly by the Commission and the High Representative).
[4] See Appendix B. for overview of labels used for categorizing the questions used within the surveys.

When conducting the categorization, necessary reductions within the dataset were made. The removal of surveys from the dataset was mainly due to two reasons; the first was related to the questionary within the surveys, the second was the presence of NAs. Surveys only consisting of questions that provide information regarding stakeholder assessments, evaluations, or stakeholder behavior, hence not giving insight into policy-implying preferences, have been removed from the dataset. The second reduction due to NAs was specific for the case where one legislative proposal as it entailed more than 75% NAs in both its arranged OPC surveys. The database for the dataset has been reduced from consisting of 43 to 33 policy initiative proposals. The remaining 33 proposals[5] provide in total 54 open public consultation surveys, which is the foundation of the analyses.

The reduction of legislative proposals entailing either no preference questions or too many NAs has no significant adverse effects on the presented results. As the remaining dataset consists of N=54 surveys representing 33 legislative proposals, assessing diversity in stated preferences will be entirely doable.

After categorizing the questions, every question providing preference indications have been abstracted, creating the data corpus. All responses for each survey have been character variables that either have been nominally scaled or ordinally scaled. Responses consisting of nominal and ordinal variables can provide sufficient insight for indications of preferences (Agresti 2013, 2). As both ordinal and nominal variables consist of multiple-choice, the respondents' submissions indicate comparable preferences. In addition, by usage of Gower distance when applying the clustering method, there are possibilities for cross-comparison of preferences across all surveys as their re-coded as stringed factors.

Notes must be made, as the ordinal and nominal variables are expected to give indications for preferences and not clearly stated preferences. Therefore, the methodological approaches for addressing preference diversity across the surveys have been thoroughly investigated and applied - this will be more thoroughly explained in the methods chapter. The three-part collaboration for the labeling process has also been necessary as the selection of questions has

[5] See Appendix A. for overview of the consultation title and belonging DG.

been human coding. The cross-checking of the coding ensures the most accurate information provided by the questions.

4.3 Variables

After categorizing and reviewing the survey content, the data basis has been decisive for how to select variables. In order to ensure appropriate insight through computing preference diversity at the survey level, two dependent variables have been selected from the gathered data. Explanatory variables are based solely on survey-related elements which can impact the outcome of preference diversity. Whereas the explanatory variables have been aimed at survey-specific items, the control variables included are founded in the theoretical review for elements that may have a general prediction for the masses who choose to respond to the surveys.

4.3.1 Dependent Variables

A deductive research approach is used to investigate diversity in stated preferences and what affects diversity. For deductive approaches, the adoption of conditions from previous research is essential. The deductive approach allows for testing assumptions and findings in other empirical cases (Gerring 2012, 173). In order to ensure correct measurement, careful consideration for how to approach the phenomenon, dependent variable (Y), is crucial. As I will not be looking at individual stakeholders' preferences but rather at preference diversity within each survey from 33 different policy initiatives, diversity in stakeholder preferences is divided into two dependent variables. Preference is abstracted from the survey responses, creating limitations for methodological approaches as it is not generalized nor has been done similarly before.

The first dependent variable is the ***the number of clusters (k)***[6], also understood as groupings of similarities within responses between participants in the open public consultation survey. Addressing each individual stakeholder's preference would demand a much more thorough and increased data resource. Elaborating on mapping diversity in stated preferences are, it is considered efficient to address preferences collectively as it will illustrate diversity based on groupings of similarities within questionaries giving indications for preferences. Defining several groupings based on similarities within their responses gives an overview of "diverse preference" groupings within the survey. The variable numbers of clusters (k) range from 1 to

[6] This is more thoroughly described in the following chapter 5. Methodological approach.

15, indicating the optimal predicted number of clusters within the response for each survey. As it is accounted for, the number of clusters (k) is a count variable.

Whereas the preferences are based on measurements from an individual level, there is no guarantee for capturing all preferences when grouping them based on similarities. In order to evaluate the size of the groupings created, an additional measurement needs to be supplied. Therefore, adding a second dependent variable, the ***weighted proportion of assigned responses within clusters (k)***, is necessary. In order to control for the assigned preferences within each cluster across the surveys, the weighted average of the proportion of assigned responses within clusters has been calculated. As given by the name for the variable, it is created by calculating the weighted size of assigned responses by the proportion of total responses of clusters. The variable weighted proportion of responses in (k) is a discrete variable, as the value is the size of each cluster, divided by the added size of all clusters, then multiple with the ranking of cluster size. In the case of this thesis, the weighted proportion of responses in (k) ranges from 1.00 to 32136.4: indicating the size of responses included in the N of clusters for each survey.

By considering how many clusters are suitable based on input to the OPC survey, combined with the weighted average of the distribution of responses for each cluster, they provide a comprehensive insight into the distribution of preferences. The inclusion of both measures also allows for testing effects on how many different preference groupings are entailed in a survey and encounters potential effects on the size of the groupings. In sum, the two variables comprehensively represent preference inequality in surveys.

4.3.2 Explanatory Variables

Salience is referred to as the publicity a proposal has been getting. In previous research, salience is often used as an indication for explaining how publicly known a legislative proposal is and how much public attention it is getting (Todorova 2020, 50). The degree of salience will help predict how publicly known a political proposal is and how much attention it has received. The more attention a proposal gets, the higher chances for the Commission to receive input. In the analyses, salience is the total response submissions for each survey. In this thesis, the variable salience is an ordinal variable, ranging from 2 to 8749 survey responses.

Target stakeholder indicates whether the Commission has intended to specify the desire for inputs from simple stakeholder groups. In previous research, targeting stakeholders have been used to indicate what kind of interests the Commission particularly wants input on behalf. In this thesis, targeted stakeholders are coded as a dichotomous variable. Upon careful review of information available in the portal for submission of OPC surveys[7] and in Roadmaps for the proposal, the surveys have either been given the score 1 if specifying stakeholders and 0 if this is no specification or mention of "all stakeholders." The variable targeted stakeholder is a dichotomous variable. The target stakeholder variables can additionally provide indications for diversity expectations in stated preferences based on whether the desire with the consultation has been to get the broadest possible input or from more specific stakeholder groups.

Survey design complexity is, in this analysis, divided into two measures: the count of how many questions each survey entails and the second as the mean of words used in all questions within the survey. The complexity of a survey has been addressed in previous research, and measurements for survey complexity have mainly been used to indicate how technical the wording or the consulting proposal is (Beyers and Arras 2019, 591). Due the scope of this thesis, survey complexity primarily indicates the design of the survey. As mentioned, this has been divided into two, of which the length of the questions and a total of N of questions could give insights into the complexity of the survey. It is assumed that surveys consisting of more extensive questions can be perceived as somewhat more demanding to complete than surveys consisting of shorter and fewer questions. The variable ***survey complexity: mean question length*** is the mean of the number of words used in the questions for each survey. Therefore, the mean of words used in each question is a discrete variable, ranging from 8.057 to 50.389 mean words used in the survey questionaries. The variable ***survey complexity: total N question*** is a count variable, ranging from 16 to 317.

Issue complexity is understood as the issue's complexity and potential technical terms used within the survey. As both Fink et al. (2021, 216) find technical linguistics to impact participation from stakeholders negatively in consultations, issue complexity is assessed as an essential independent variable to include. Issue complexity has, through text mining, been extracted by human evaluation of not frequent, rare words used within the questionaries in the survey. By creating a corpus from all survey questions and then testing the corpus on all text

[7] Have Your Say (online platform for submission): https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives_en

within the survey questionaries, words with low frequency have been addressed and found to be of a more "complex" art. Issue complexity is a continuous variable scaled between 0 and 1. Each word included within the questions gets a score based on its frequency. As words appear frequent, they are assigned a score closer to 0. The closer a score is to 1, the less frequently the combination of words used within the survey is. The mean for all words entailed within a survey is used for comparison, indicating the total score on issue complexity.

4.3.3 Control Variables

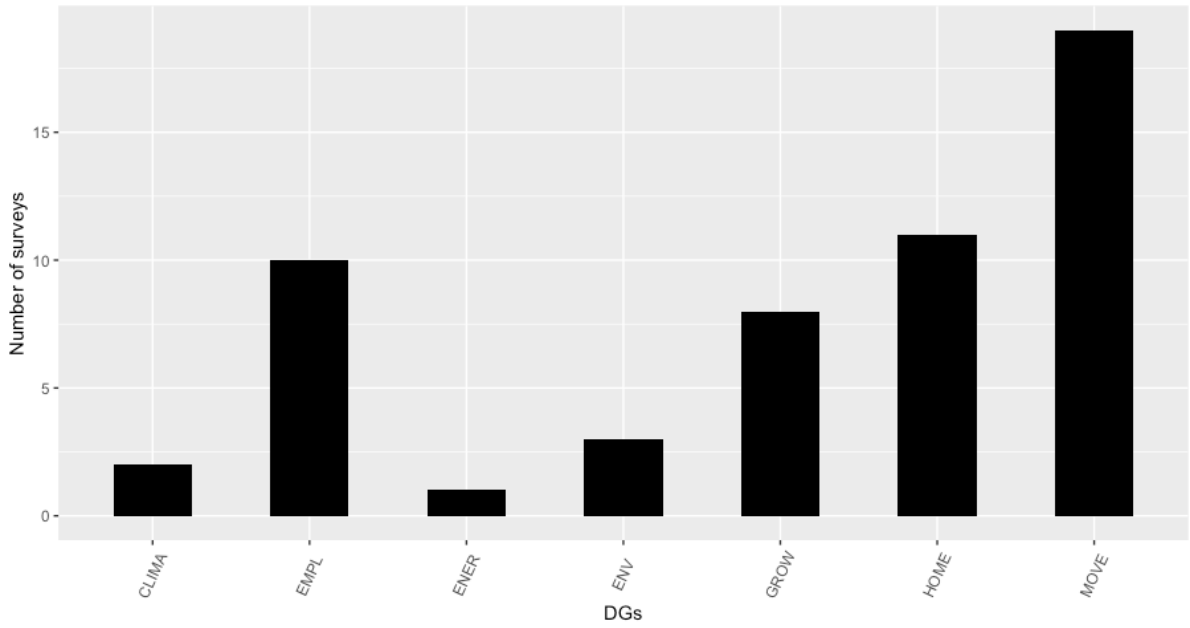
The assumption for the inclusion of control variables within a regression is to "(…) decrease the aggregate bias on the coefficient of interest for every additional relevant control variable that we include" (Clarke 2009, 49). The theory presented has addressed patterns for how different stakeholders relate to consultations. Overall, expectancies for diversity are greater when attracting responses from a broad scope of stakeholders. Whereas a statistical approach often reduces the scope for explaining relations by only viewing effects between selected variables, it is addressed necessary to include control variables. As Ragin (1989, 59) explains it, "(…) the effect of a control variable is its average effect on the dependent variable, across all cases, net of the effects of other variables. The subtraction of effects central to statistical control is a purely mechanical operation predicated on simplifying assumptions". Based on presented research, the control variables included in the analyses found in general expectancy for "controlling" the effect of the explanatory variables. The two control variables are additionally expected not to explain the same relations as the explanatory variables.

The drafted policy initiatives can either be drafted for new legislative proposals or amendments to already existing legislative policies. As for the ECs need for external expertise and inputs, OPC surveys can be expected to facilitate reaching inputs from the broadest possible audience when creating drafts for new legislative proposals (Van Ballaert 2017, 409). When amending proposals, it can be assumed that the EC instead targets their audience more and facilitate less for reaching a similar extensive audience as for new proposals. Having explanatory variables focused on complexity in survey design and issue forms, the variable measuring proposal status is assumed to control their effects across the surveys. Including the policy-proposals status will additionally provide insight into the effects of diversity in preferences across drafts for new legislative proposals or drafts for amending legislative proposals. In this case, the variable ***proposal status*** is a variable indicating the drafted initiatives status as to existing legislation.

The status is either new or amending, making the policy status a dichotomous variable. In the dataset, policy status is coded as 1, indicating an OPC survey for a new proposal, or 0, indicating an OPC survey for an amending proposal.

Previous literature indicates stronger relations between different stakeholder groups and which policy area they choose to participate in (Rasmussen and Carroll 2013, 453). As for including variables for salience and issue complexity, the policy area is assessed to be vital as an explanatory variable for controlling effects on diversity in stated preferences. Knowing which type of policy area is seen to stipulate for participation from the broadest range of stakeholders helps increase expectations for receiving diverse preferences. The variable ***policy area*** consists of the seven DGs responsible for drafting the proposals being consulted. The policy area consists of a factor variable, whereas the number between 1-7 indicates the DG responsible for drafting the legislative proposal and receiving inputs through the OPC survey[8]. Including policy area as a control variable will also provide insight into how OPC surveys impact diversity in received inputs between the different EC DGs.

Figure 4.1 *Total survey distribution by DGs.*



Note: Figure 4.3.3.1 Shows total distribution of OPC surveys included in the dataset, and which DG have been responsible for drafting the proposal.

[8] DGs categorization: 1 = ENV, 2 = CLIMA, 3 = GROW, 4 = EMPL, 5 = HOME, 6 = MOVE, 7 = ENER.

4.3.4 Descriptive Statistics of Variables

The variables included in the dataset consist of different units of measurement. The two dependents are a count (the optimal number of (k)), and a ratio variable (the weighted average proportion of responses in (k)). Furthermore, the different explanatory and control variables consist of different measurements. More specifics about the measurement levels can be seen in Table 4.2.

Table 4.2 *Descriptive Statistics of Variables*

Variables	Mean	Std.Dev.	Minimum Value	Maximum Value
Dependent Variables				
Optimal N. of (k)	3.944	2.756924	1	15
Weighted Proportion of Assigned Responses within (k)	124.6	5588.463	1.00	32136.4
Explanatory Variables				
Salience	1002.1	2083.037	2	8749
Targeted Stakeholders	0.5926	0.4959656	0	1
Survey Complexity: Mean Question Length	24.915	7.932275	8.057	50.389
Survey Complexity: N. Questions	100.74	67.84421	16.00	317.00
Issue Complexity	0.8651749	3.688234e-06	0.8651665	0.8651817
Control Variables				
Proposal Status	0.6111	0.4920756	0	1
Policy Area	4.574	1.512238	1	7

4.4 Limitations of Human Coding

Most of the included variables are drawn from indications either for stakeholder relations or pure assumptions. This might not be sufficient for capturing all possible implications explaining diversity in stakeholder preferences. In general, the inclusion of variables has often been highlighted as a critique of quantitative analyses (Gerring 2012, 89-90).Whereas studying the societal phenomenon, there is not possible to control for effects not included as variables in the quantitative analyses. Additionally, data is needed, which might not always yield the most resourceful measures for explaining a phenomenon.

The selected explanatory variables in this thesis have been assumed to explain how stakeholder preferences differ within OPC survey responses. All variables have been single-human-coded, which implies collecting he data from raw materials. Applying the usage of single-human-coded variables causes a narrower implementation and a more thorough explanation of how they are structured. When using data collected from the EU, Bunea, Ibenskas, and Binderkrantz (2017, 348) note that the technicality within the data might create a risk of errors. Additionally, it is crucial to understand that the human coding of variables requires many resources, which limits the feasibility of conducting a large-N analysis of raw EU consultations. Hence, it explains why this study only analyzes preference diversity within 54 surveys.

4.4.1 Efforts to reduce limitations

This thesis adopts an explanatory approach by drawing on previous research and findings that can provide predictions for the effects surveys used in OPC have on the diversity of inputs from stakeholder preferences. The aim is to assess diversity in preferences and test the potential effects survey-level elements have on the observed diversity in preferences received in OPC. An explanatory approach allows for testing hypotheses and provides insight for further research on diversity in preferences stated in consultations within the EU.

All included variables collectively bear relevance for all elements within every survey encountered in the dataset. The explanatory approach allows for testing the specifically selected variables without encountering too unpredicted a lack of variables. As surveys can only consist of a limited amount of elements, the chosen variables can at least predict some of the encountered associations between diversity in stakeholder preferences and survey-level items.

This thesis, as previously mentioned, does not intend to explain diversity in stakeholder preferences other than within surveys. Narrowing the scope only to investigate survey elements effect allows the analyses to provide insight on how surveys affect the responses gained through open public consultation surveys. Hence, accounted relations between the selected variables will provide insight into preference diversity within consultations and give indications, which can be fruitful for new research on stakeholder preferences and EC consultations.

4.4.2 Reliability and Validity

When conducting research, there are high standards for reamplification and accuracy. High reliability implies the analyses be easily repeated and produce similar results. Validity refers to how precise the analysis measures what it intends to measure (Gerring 2012, 82-83). Combined, both validity and reliability are assessed as highly important in research as knowledge is generated and needs to be precise. It should be open and transparent about how it has been executed.

The dataset used in this thesis consists of human hand-coded variables. It will be possible to find the surveys used as all OPC surveys are available digitally on the EU's website. The biggest challenge for the reliability is related to the process of categorizing variables. As the variables have been subjectively assessed, hence have been coded based on judgment calls. Despite problems with replicating the exact data extraction, the three-party collaboration allows for more trust in the questions extracted and that they are related to particular preferences. Hence, the reliability might be questioned, but the validity is estimated to obtain high standards.

5 Methodological Approach

The application of methodological approaches is driven by what is intended to investigate. The research method usually involves finding evidence for a causal mechanism that explains a phenomenon. Gerring (2012, 200) defines a causal mechanism as the path, process, or chain between diverse elements explaining how X possibly affects Y. Uncovering connections and how one can put things in the context of others is the very root of research. Figuring out what may affect Y leads to applications of a method that will test the effect X have. The method chosen to map the relation between X and Y determines how the relationship is assessed. As the causal mechanism is grounded in a belief about a relationship between X and Y, the contextual factors must be considered as they may affect the effect aimed to measure (Falleti and Lynch 2009, 1143).

This chapter explains which methodological approach has been employed to answer (1) how much diversity in stakeholder preferences exists in OPC surveys and (2) test whether different explanatory factors impact the diversity in stated preferences on a survey level. I begin by explaining why a multimethodological approach is applied in the thesis. Following, I assess the methods used and their lacks and possibilities: *clustering, negative binomial regression, and linear regression.* Additionally, remarks regarding the correlations and alternative methods are discussed.

5.1 Why Consideration of Methodological Approach

Combining methods has become increasingly used in research. Thus, addressing the proper method for the case is essential (Gerring 2012, 383). As the natural world never stops evolving, and we tend to study real world happenings or interactions. The natural world never stops evolving, and we tend to study real-world happenings or interactions. Applying different approaches to investigate different functions is "necessary to deal effectively with the full richness of the real world" (Mingers and Brocklesby 1997, 492).

To investigate the effect survey-level elements have on how diverse stated stakeholder preferences are in open public consultation surveys, the first step is to assess the diversity in stated preferences. It is reasonable to expect that there will be some causal relationship between the different survey elements (X) and how this affects diversity among stated preferences (Y). Bearing in mind the size of the EU, and that the EC recieves different inputs through different

venues, the thesis will be focused on a survey level. The thesis intends to investigate diversity in stated preferences from stakeholders participating in OPC surveys; the application of a broader approach to stakeholders has been addressed as necessary. Instead of looking at individual stakeholders, the adopted methodological approach allows for mapping diversity in the stated preferences across various policy initiatives in OPC surveys. By applying different quantitative approaches for the creation of variables and for analyzing effects regarding elements within an open public consultation survey, regression models have also been applied. Combined, it allows for assessments of diversity stated in OPC surveys and evaluation of survey elements that impact preference diversity.

More specifically, cluster analysis has been used to measure diversity within stated preferences, forming the two dependent variables: the optimal number of (k) and the weighted proportion of assigned responses within (k). The optimal number of k is referred to the number of observed preference groupings within the survey. The weighted proportion of assigned responses within (k) is an average of sizes for the derived clusters based on the assignment of respondents within the clusters. Additionally, text mining has been used to abstract the explanatory variable "Issue complexity." All words used in questions within a survey have been collected and formatted as a corpus, scoring the "rarity" of words used within each survey to address less frequently used words. Words scored with lower values are assumed to be more technical. Applications of different regression analyses allow for a more comprehensive explanatory approach which is needed as the number of clusters (k) and weighted average of the respondent proportion of (k) in order to be able to test the effect of the independent variables on stakeholder diversity. Combining both regressions will allow a unified perspective on the effect of the independent variables (X) on the two dependent variables (Y).

5.2 Measuring Diversity in Stated Preferences

Clustering as a methodological approach has emerged as one of the leading machine learning techniques within multiple analysis (Kettenring 2006, 4). Clustering is a machine learning method that constructs groupings of objects so that the groups obtained are as homogeneous as possible, meanwhile being as different from one another as possible. Classifying societal observations has been approached by finding similarities or dissimilarities, especially when identifying new objects or phenomena (Xu and Wunsch 2008, 1). Using clustering to extract the dependent variables allows for collectively addressing diversity in stated preferences

through the creation of groups within the data. More specifically, clustering algorithms are adapted to find similarities within existing data, generating an outcome of clusters (k) based on accounted similarities. In the case of this thesis, the clusters are interpreted within OPC survey responses for creating groupings based on similar responses. The number of k gained through the application of cluster algorithms indicates the number of groups of similar observations.

Clustering is defined as "(…) an aggregate of points in the test space such that the distance between any two points in the cluster is less than the distance between any point in the cluster and any point not in it" (Xu and Wunsch 2008, 4). For clusters to interpret the inputs used for creating k of similarities, both the distance measure and selection of the clustering algorithm severely affect the potential outcome. As Xu and Wunsch (2008, 6-7) highlight, all clustering algorithms are connected to a proximity definition of proximity measure, giving reasons for careful interoperation of both distance measure and algorithm to apply to the data used. The characteristics of the data and the intention for the application of clustering determines which approach to apply for both algorithm and distance measure.

As clustering allows for the generation of groupings based on inter-similarity, the clustering method has been conducted to extract information at a survey level for differences in response gained through OPC surveys. With the intention to assess diversity in stated preferences across the surveys collected, the clustering method allows finding nearby responses and distinguishing between the different groupings generated. Adapting OPC survey response to clustering allows empirically addressing diversity as an entity for analyses.

5.2.1 Gower Distance Measure

In order to measure the distance between observations, a distance measure must be defined. The distance measure defines how to approach the similarities/ dissimilarities within the sample. Among the distance measures used in clustering, the Euclidean distance and Manhatten distance are the most frequently used in previous research (Boehmke and Greenwell 2019). The data in this thesis consists of both nominal and ordinal measures. Hence they are mixed data, which the Gower distance can interpret. It translates proportionality between elements into scaled vector points so that the sums of squares are equal to the latent roots, allowing for assessing the distance between groupings of similar preferences (Gower 1971, 860). Scaling the samples from 0 to 1, the Gower distance rates and calculates the Manhattan distance for ordinally coded

variables, whereas turning each level of a nominal variables into binary columns before interpreting the dice coefficient. The dice coefficient interprets the distance of similarities between the different levels given within the variable. Being able to interpret distance measures accounting for both ordinally and nominally scaled variables makes the usage of Gower distance a suitable distance measure for the data in this thesis.

The usage of Gower distance is assessed to be intuitive and follows quite simple adaptions for giving a final calculation of distance between the clusters. Although it is adaptable for usage on mixed data, it hinders the distance measure from being sensitive to non-normality and outliers within continuous variables (Xu and Wunsch 2008, 29).

The Gower distance is additionally sensitive for NAs as the distance measure will not provide calculations for datasets entailing NAs. During the data wrangling, there was a need for a threshold for including questions entailing responses with NAs. When removing question alternatives such as "Do not know," which does not provide any preference indications, the total of NAs among the responses grew. In order to use as much as possible from the questionaries within the surveys in the dataset and to use Gower distance, the removal of questions with too much NA has been deemed necessary. In the case of the thesis, a threshold was set for questions containing more than 75% NA in the responses. When applying the 75% threshold for NAs, several questions was removed. Only two surveys was deemed necessary to remove due to entailing more than 75% of NAs within responses to all questions. The remaining surveys still entails sufficiently enough questions in order to interpret clustering on the responses.

5.2.2 Partition Based Clustering Algorithm

Before applying a clustering algorithm, the first thing to note is how to distribute the data into the clusters. Among several different distribution approaches which can be adapted to the data, the most commonly used is partitioning-based clusters (Boehmke and Greenwell 2019). Partitioning clustering divides data points into clusters based on their characteristics. For this clustering method, specifying a pre-set number of clusters is needed.

K-Means clustering is among the most used unsupervised clustering methods. K-Means partitions the observations into distance groups. Each cluster has a center, called centroid, to

which observations get assigned (Xu and Wunsch 2008, 69). Calculating selections creating high inter-class similarities also implies creating the greatest possible distance between another grouping of inter-class similarities with other clusters. The K-Mean clustering allows for the collection of objects without pre-based intentional categorization. It estimates the number of homogeneous groups within the data, with each group (k) being as different from the other as possible. Despite its functionality, K-mean takes little to account for outliers, and one must pre-specify the number of (k) before applying the algorithm to the data. When encountering mixed data, alternative algorithms can be employd.

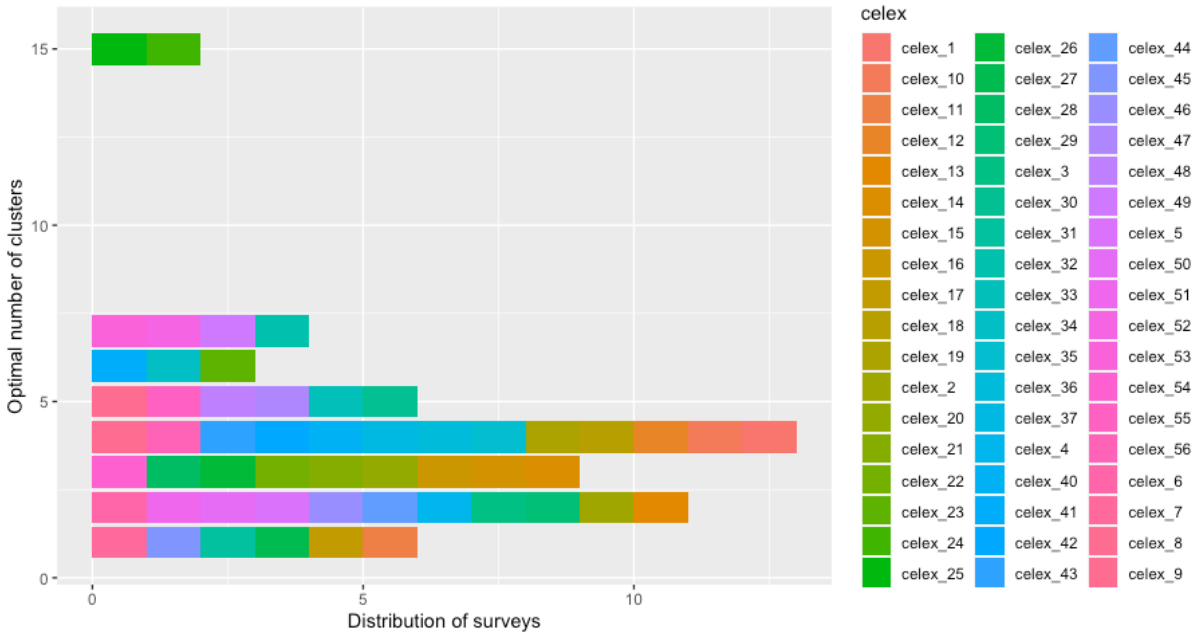
In cases where the data is mixed and deciding a prefixed number of (k) becomes more complex, other algorithms for clustering than K-Means can be used. The clustering algorithm PAM, which stands for partitioning around medians, has been used in this thesis. PAM, like K-Means, creates groupings by assigning observations to a given number of k clusters based on a high degree of intra-class similarity. The main difference between K-Means and PAM's calculation for distribution is their choice for determining the centroids. Whereas K-Means uses the mean and becomes much less robust to outliers, the PAM algorithm partition is based on medians (Boehmke and Greenwell 2019).

Using response data consisting of ordinal and nominal variables with appliance of the Gower distance, the PAM algorithm succeeds in comparing both types of variables. Overall, the data is mixed and, therefore, gives higher probabilities for entailing outliers, making PAM more efficient in acquiring groupings based on distance and similarity. Additionally, in Reynolds, Richards, and Rayward-Smith (2004, 117) comparison of different cluster algorithms, they find PAM to obtain the most efficient calculation of clusters when k =2 to 7. PAM is also noted to generate better results for overall silhouette width, an interpretation of clusters within the data, for values of k and the sum of the distance between medoids. Their main adverse finding regarding PAM's calculations is the time PAM uses for calculating the k's compared to other methods.

Whereas PAM spends considerably more time in the calculation of k in larger datasets, the CLARA, *clustering large applications*, performs the same cluster steps as both K-Means and PAM but "(…) in less than $\frac{1}{5}$ of the time!" (Boehmke and Greenwell 2019). As mentioned in the previous subsection of the thesis, the PAM algorithm has been applied for all surveys but

with the expectation of 2. As seen in Figure 5.1, showing the distribution of k across surveys, two surveys were assessed as too “large” for the PAM to run: entailing an optimal number of clusters equal to 15 k. The CLARA generates several random sample schemes with the appliance of the PAM algorithm (Xu and Wunsch 2008, 216).

Figure 5.1 *Distribution of Optimal Number of (k) for Each Survey Included*



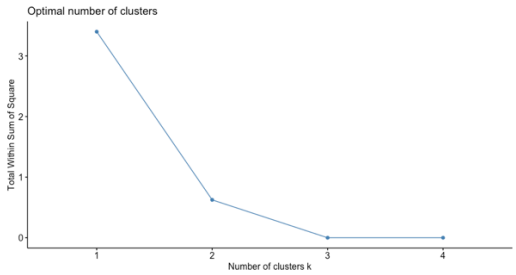
Note: Celex_x represent the number of surveys included in the dataset. The celex for”_38” and “_39” was removed due exceedingly high numbers of NA within all questionaries. The different colors represent the each survey encountered in the dataset and the numbers of clusters (n= 54).

5.2.3 Selecting Number of (k): The “Elbow” Method

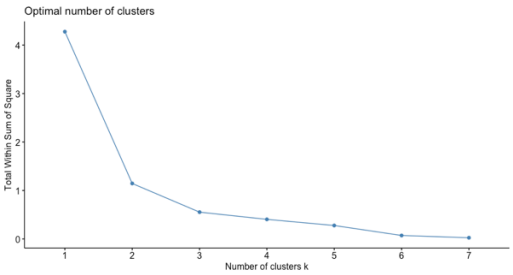
In order to determine the optimal number of k for each survey, the "elbow" rule has been used. This has been done by using the "fviz_nbclust" partitioning function and the "wss" estimate. The wss, within-sum og squares, estimate creates sums of squares to measure the variability of observations within a cluster. Generally, clusters with a small square sum are more compact than clusters with a large square sum. The "elbow" method is one of the most frequently used methods for determining (k) when a specified number of clusters is needed to be generated. Using the fviz_nbclust, with the “wss” estimate produces a plot of the within-sum of squares for the data. The observed dent within the curve is addressed as the “elbow”. Deciding when you can see an "elbow" can sometimes be challenging, which is one of the downsides of the elbow method. In Figure 5.2, four illustrations of different outcomes from usage of fviz_nbclust on the surveys included in the dataset are shown. The four surveys outline different outcomes

for estimating the optimal number of (k) due size of potential for prediction. For plotting an estimate, the size of the encountered data limits the max (k) of predictions possible for the plot. For data consisting of few entities, deciding how many clusters to compute as optimal is extensively harder than with data allowing for higher values. For visualizing and estimating the optimal number of (k), the elbow method has been interpreted as the best fit due to the time needed to produce predictions for (k) for all 54 surveys.

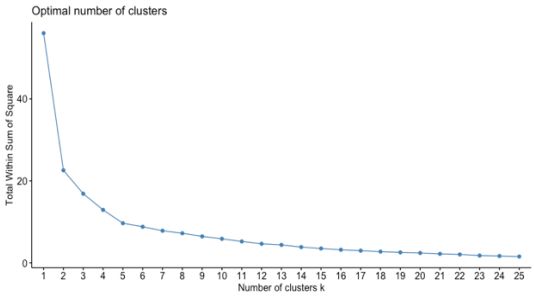
Figure 5.2 *Illustrations of Results from "The Elbow Method" with fviz_nbclust*



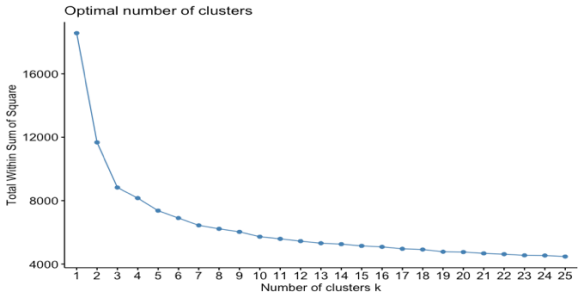
Celex_3



Celex_14



Celex_30



Celex_49

Note: These are samples illustrates challenges regarding determining the number of k for each survey. As shown, some surveys encountered data for estimating plots with maximum of 25 clusters, whereas others were limited to estimate plots for a maximum of 4 clusters.

5.3 Regression Assumptions

Based on the distribution of units in the dataset and the variable type for the two dependent variables, it is considered most appropriate to perform two separate regressions, one for each of them.

The first dependent variable, *the optimal number of (k)*, show for each survey how many clusters of similar observations have been determined to be optimal based on the elbow method. This variable is a count variable, consisting of independent observations for each survey. For deciding which regression to use for analyzing the cluster variable, an assessment of count regressions will be presented.

The second dependent variable, *the weighted proportion of assigned responses to clusters (k)*, shows the weighted average of the proportion assigned of responses to cluster for each survey. The variable is a weighted calculation for the cluster size of each survey. This has been calculated by dividing each cluster size by the total size for all clusters and multiplying each by their ranked size. Finally, all weighted proportions have been added, giving the weighted proportion size to clusters on the survey level. Linear regression has been assessed as suitable for appropriately adapting a regression to the weighted cluster proportion due it being a continuous variable (Kellstedt and Whitten 2018, 129).

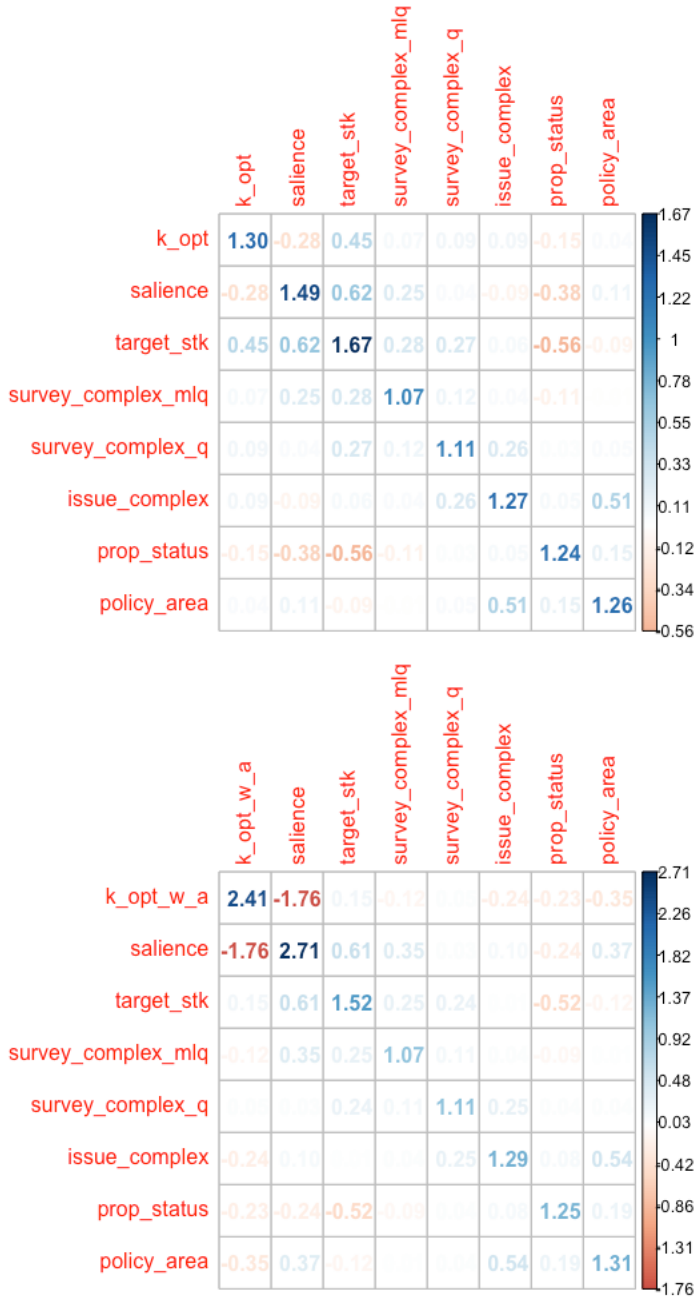
5.3.1 Multicollinearity

Before running a multivariat regression, there are some prerequisites for the data that must be considered. Multicollinearity is one of them. Multicollinearity indicates correlation between the variables included in the data set. This could make the results misleading and reduce the reliability of statistical inference, i.e., becomes difficult to precisely estimate coefficients, which results in large uncertainty around your point estimates (Kellstedt and Whitten 2018, 264-265).

One of the ways to check for multicollinearity in multiple regressions is to calculate the VIF, the variance inflation factor. The results gained from the VIF test indicate what percentage of the standard error square for each coefficient (Kellstedt and Whitten 2018, 266). As a rule of thumb, the VIF results should not exceed 5, indicating a high degree of correlation between the explanatory variables. For the dataset used in this thesis, both application with the optimal N. (k) and the weighted proportion of assigned responses within (k), the VIF score is under 2:

highest for the optimal n. (k) = just above 1.4, and highest for the weighted proportion of assigned responses within (k) = just above 1.5[9]. An additional approach to control for multicollinearity is creating a correlation plot for the variables.

Figure 5.3 *Plot for Correlation for both Dependent Variables*



Note: Correlations for all variables, both for negative binomial regression and linear regression. First graph is for all variables in the negative binomial regression, the second for the variables in the linear regression.

[9] VIF tests for both dependent variables can be seen in Appendix C.

As seen in Figure 5.3, there is a notably low degree of correlation between the explanatory variables in the dataset. The correlation plot and the results from the VIF test show no sign of correlation or multicollinearity within the dataset.

5.3.2 Count Variable Regression

When doing a regression using a count variable as the dependent variable, which is the case for the *optimal number of (k)*, Poisson regression is considered as the easiest to run (Agresti 2013, 115). Using count data does not follow a normal distribution assumption, assumed for standard models, as the lowest possible value equals 0 counts (Finch, Bolin, and Kelley 2014, 126). For the data to properly fit the Poisson model, the variance is presumed to be equal to the mean, also referred to as equidispersion (Hilbe 2011, 64; Cameron and Trivedi 2005, 670). This is rarely the case for much societal research as the variables can entail variance which exceeds the mean. When encountering overdispersion or under dispersion, the standard errors estimated for the model are compromised, leaving the model's standard errors to be smaller or higher than they would be within a real-world population. To assess which count model gives the best fit for the data, there is a need to control for dispersion, multicollinearity, and linearity within the variables.

Table 5.1 *Variance and Mean on the Dependent Variable; Number of Observed (k)*

VARIANCE	**MEAN**
7.600629	3.944

Note: Noting the difference in variance and mean to be quite big indicates for further investigation of overdispersion, so no modelling for underdispersion will be encountered.

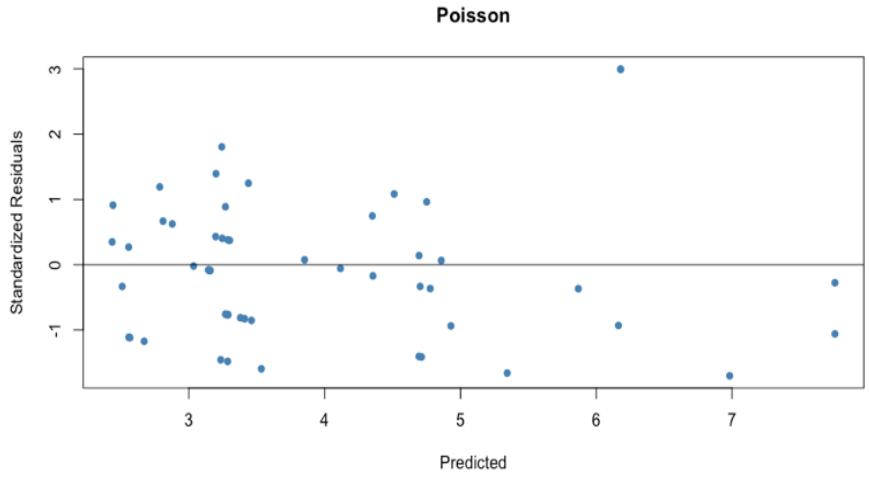
5.3.2.1 Overdispersion

In the case of fitting a count regression model for the variable *number of optimal (k)*, the *dispersion test* available from the "AER" package in R has been used. The dispersion test calculates the difference between the mean and the variance for the data. If the score equals 1, Poisson is assessed as a good model fit for the data. For the data encountered in this thesis, the dispersion shows 1.198, indicating overdispersion as the variance exceeds the mean, shown in Table 5.1. This calls for the use of either negative binomial - or quasipoisson regression, which is recommended when experiencing overdispersion (Finch, Bolin, and Kelley 2014, 128-129; Hilbe 2011, 64).

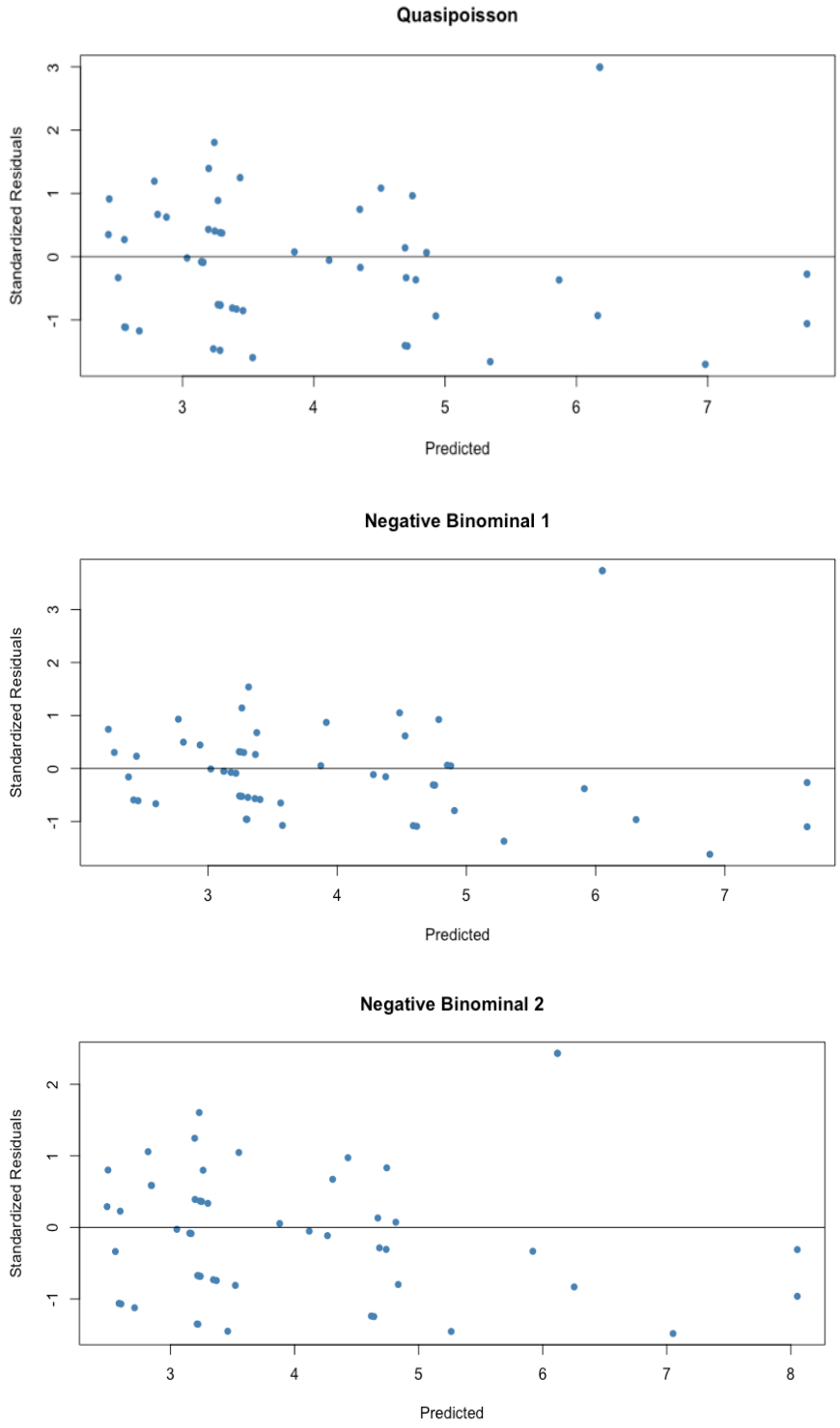
The Quasipoisson model is similar to the Poisson model with the expectation of restricting the dispersion to 1. This will only impact the standard errors for the parameter estimates without changing the values for the estimate of the coefficient (Finch, Bolin, and Kelley 2014, 128). The negative binomial model allows the conditional variance of the dependent variable to exceed the conditioned mean, bringing more flexibility when fitting the data to the model (Yang and Berdine 2015). The negative binomial model comes in two different formats. The most frequently used negative binomial model, NB2, takes the quadratic mean for the variance, allowing for necessary flexibility when applied to different types of count data. The other negative binomial model, NB1, uses a linear variance function that holds the dispersion constant (Cameron and Trivedi 2005, 676-677).

To evaluate overdispersion within the model, test for predicted residuals, model-fit, and Likelihood Ratio have been conducted on the dataset using the Poisson, NB1 model, and NB2 model[10]. Both the Poisson and NB2 models have been applied using the glm package, whereas the NB1 model has been applied using the gamlss package. Assessing the residuals plot predicted for the three models, Figure 5.4, the residuals for the NB2 are smaller, ranging from -1 to 2, compared to Poisson, NB1 and Quasipoisson, which ranges from -2 to 3. This indicates a better model fit using NB2 model regression on the dataset.

Figure 5.4 *Residuals Plot Applied on the Full Dataset*



[10] Quasipoisson has additionally been included when assessing predictions for residuals.



Note: Illustrated, the prediction for standardized residuals for the whole dataset is identical for Poisson, Quasipoissn, Negative binomial 1 and Negative Binomial 2.

When running the Likelihood Ratio Test for the Poisson and the negative binomial models, the NB1 model receive a p-value of 1, whereas the NB2 model get a p-value of 0.4943436. Due to the p-value for both negative binomial regressions are over the standard threshold of 0.05

indicates that there are no significant differences between the three models' applications for the data. Despite this, both the indication for overdispersion and the residuals plot favor running a negative binomial regression, NB2, for the count variable *number of optimal (k)*.

Table 5.2 *Model-fit Comparison of Poisson and Negative Binomial*

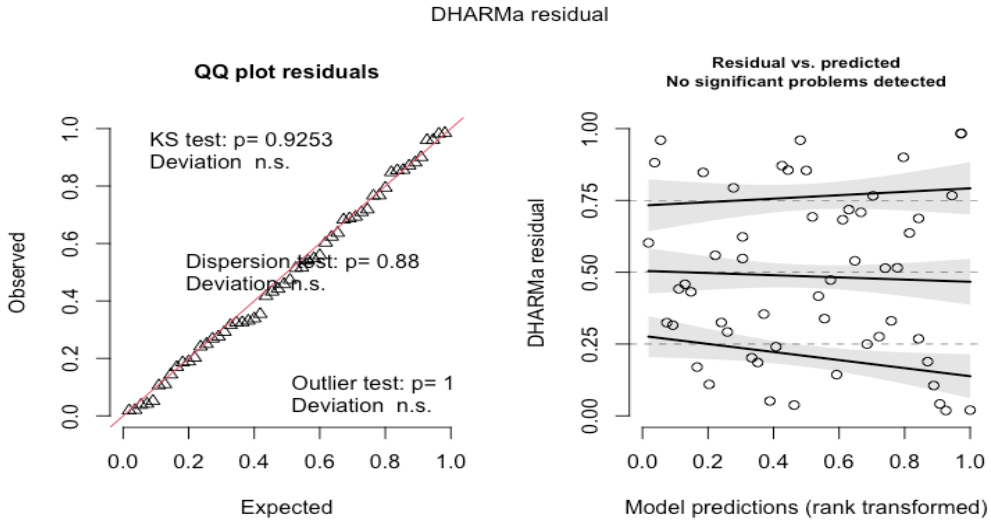
	AIC	*BIC*	*Log Likelihood*
Poisson	239.9851	265.8419	-105.9925
NB1	264.8016	292.6474	-118.4008
NB2	***239.7573***	***267.6031***	***-105.8787***

When testing for the goodness of the model fit on the entire dataset, both shown by BIC and Log-Likelihood, NB2 shows the best results: this by having slightly lower values for BIC than Poisson and obtaining the highest value for Log Likelihood. Additionally, the AIC value for NB2 is the lowest, making the NB2 model the best model for explaining the variation within the data. The NB2 model performs slightly better than the Poisson, but it significantly outperforms the NB1 model. Summed is the NB2 model assessed as the best fit for running a regression on the variable *number of optimal (k)*.

5.3.2.2 Linearity

For running a negative binomial regression, which by now contains the best results, it is assumed that linearity among the residuals is included. When plotting the NB2 into a quantile-quantile (QQ) plot, the plotted residuals are supposed to fit the predicted distribution line. As seen in Figure 5.5, the residuals for the NB2 model deviate to some degree from the expected distribution. In the model to the left, the KS test shown indicates the fit of the data compared to the correct distribution for the residuals. The P-value for the KS test shows a value of 0.9253, which is above 0.05, thus indicating that the null hypothesis for normal distribution is not rejected. This gives reasonable indication to assume that the model's parameters are linear to a sufficient enough degree. Additionally, influential cases are checked for, also known as outliers. The outliers' p-value is not significant, indicating no expectancy for observation disproportionality (Kellstedt and Whitten 2018, 258).

Figure 5.5 *Residuals Test For NB2*



Note: This residual test is plotted by usage of the "simulateResiduals" from the DHARMa package. In the model to the left, a qq-plot of estimated parameters for the multiple regression using all variables are presented. Additionally, is information regarding the correct distribution: the KS test, the dispersion for the data, and outliers included. To the right is a model showing residuals plot for the model.

5.3.3 Linear Regression

Using a continuous variable as dependent variable, linear regression is often assessed as a good model for the application. Multiple linear regression provides the opportunity to estimate the effect of several independent variables on the dependent variable. The linear model makes several assumptions for the relationship between the variables included in the model. Among the most critical assumptions, there should be a linear relationship between X and Y (Finch, Bolin, and Kelley 2014, 18-19). Despite not expecting a perfect linear relationship within the real world, linearity is among the most natural assumptions for using the linear model. In addition, the multiple linear regression assumes the variance in Y is held constant across X, the distribution of data (Y) is normally distributed, and the residuals for the variables used are independent of each other (Finch, Bolin, and Kelley 2014, 3-4).

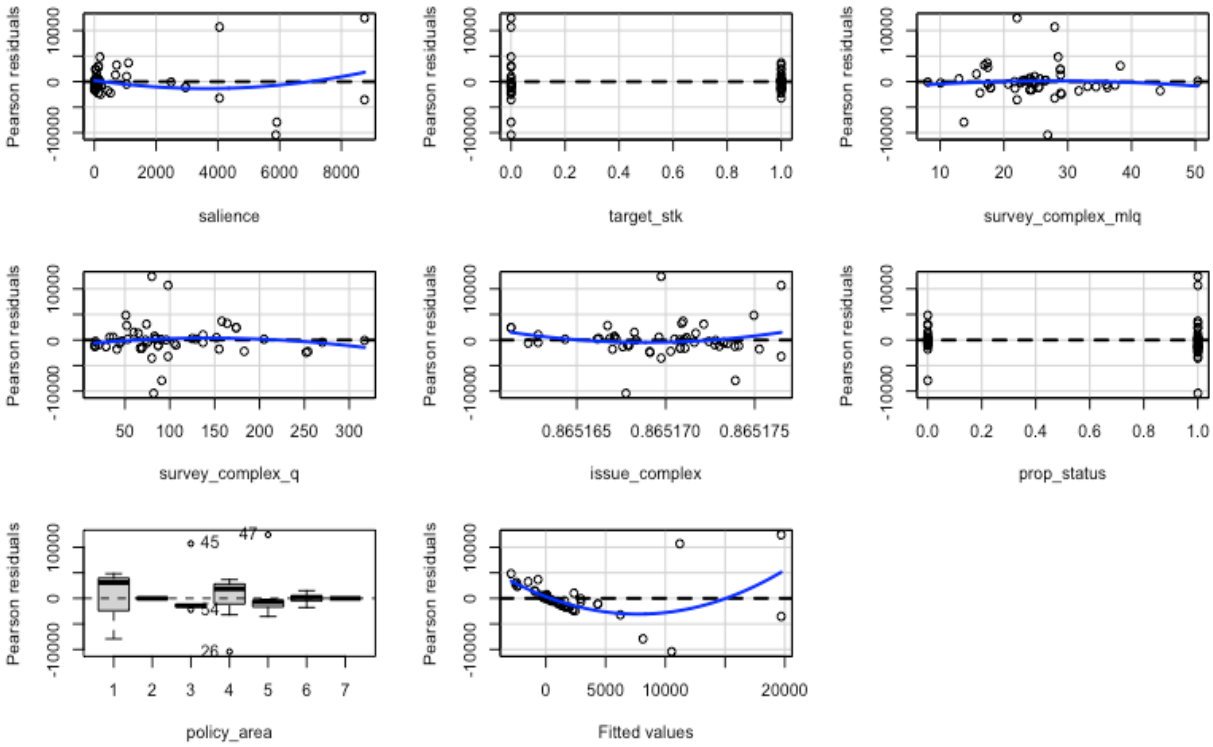
5.3.3.1 Assumptions for the Linear Regression Model

When assessing whether the usage of linear modeling for multivariate regression is appropriate for the data, the relationship between Y and X needs to be linear. As seen in Figure 5.6, linearity can be seen among most explanatory variables and the dependent variable. The residuals for the variable policy area are plotted in the form of box plots because this variable is a factor

variable. Additionally, linearity can not be seen for targeted stakeholders and proposal status as they are dichotomous. Assessing the fit for the remaining explanatory variables, most can be seen to have a sufficiently linear form based on the fitted line (blue).

The residuals plot in Figure 5.6 shows the residuals plot for the data. Ideally, unequal error variance should be avoided as it shows a better fit for some instances than others within our observations. Getting unequal error variances, also known as homoscedasticity or heteroscedasticity, can cause problems in estimating confidence intervals (Kellstedt and Whitten 2018, 209). When addressing the variance within Figure 5.6, it shows an overall less favourable distribution. Salience indicates the worst fit, as the observed variance almost forms like a nest, indicating tendencies for heteroscedasticity. This gives reasons for being causes when addressing the confidence interval.

Figure 5.6 *Residuals Plot*



Note: The residuals plot, from the car package, shows the residuals for all variables in the dataset.

For the assumption of variance in Y is held constant across X, the ANOVA test have been conducted. By applying the ANOVA function to the data it assumes normality among thedistribution of variance across the variables. The variable salience gained statistically

significant predictions with a p = 6.378e-10. This indicates a potential for quadratic relationship with the dependent variable (Finch, Bolin, and Kelley 2014, 19). As salience is the only variable significant, a Kruskal-Wallis test has been performed to check for at least similarities in the shape of salience and the weighted distribution of respondents in (k). The Kruskal-Wallis test is a nonparametric version of ANOVA. The poor performance of *salience* in the ANOVA test is as well not surprising as it's the inputs received, which is not presumed to me normally distributed. Instead of assuming normality in mean, the Kruskal Wallis test assumes same-shaped distribution (Long and Teetor 2019). Salience reaches a p-value of 0.2731, indicating the same shape as the weighted distribution of respondents in (k), and can thus be assumed included without problems for the distribution for the regression.

The general distribution among the data within the model is shown in Figure 5.7. Ideally this should appear normally distributed with most observations in the middle, and fewer observations on both the left side and right side of the distribution. The distribution of the data is shown to be sufficiently symmetric as there are signs of decrease at both ends from the middle. There are more observations in the distribution on the left side but not enough to determine the entire distribution as negatively-skewed: entailing most distribution on the left side of the mean for the distribution.

Figure 5.7 *Distribution Plot*

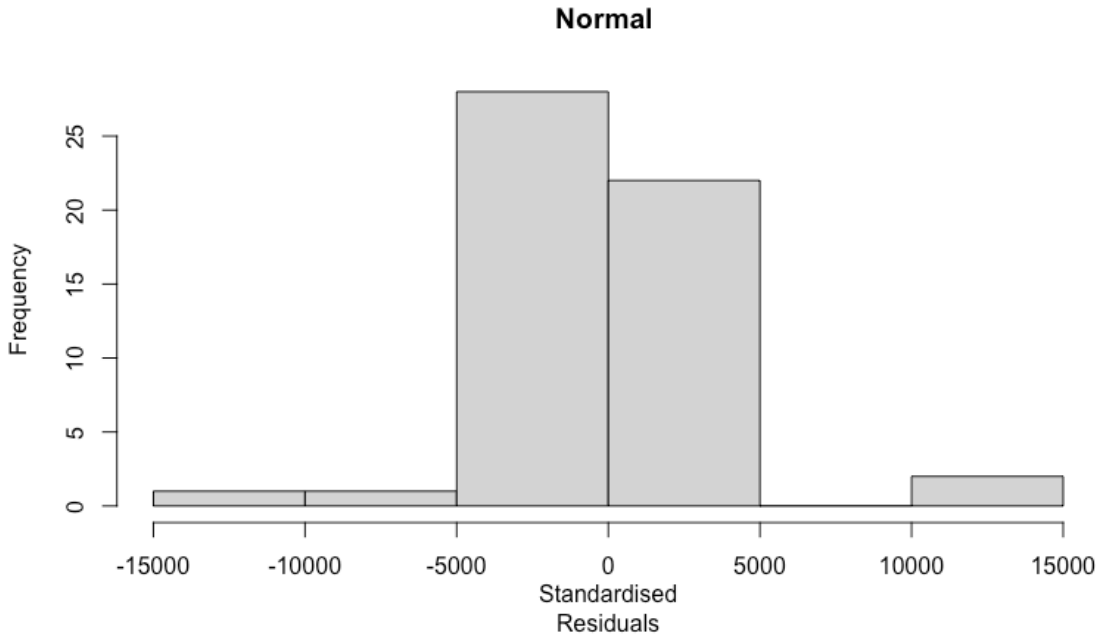
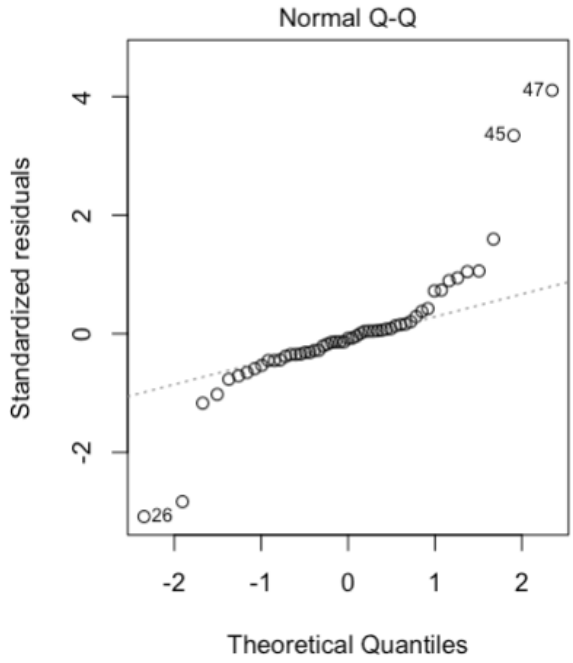


Figure 5.8 *Quantile-Quantile-Plot*



In addition to checking the distribution, the data has been assessed through a quantile-quantile plot. Shown in Figure 5.8, most observations are distributed along the line, indicating the same distribution as Figure 5.8. Indications from the distribution in the QQ plot allows for assuming fairly close normality in the data, as the line indicates a 95% confidence interval allowing for confirming normal distribution for observations on the line (Finch, Bolin, and Kelley 2014, 20).

Finally, for assessing the independence of the observations a Durbin Watson test has been applied on the model. The Durbin Watson test indicates whether X are not autocorrelated with the other variables, indicating that they are independent (Toth 2010, 141). The model for the dataset gets a p-value on 0.7047, indicating that the null hypothesis cannot be rejected. Additionally, the ACF test have been conducted. The ACF test shows no sign for autocorrelation[11]. This gives sufficient evidence for stating the independence assumptions for the model are met.

Overall, presented information can provide a good enough basis for using multiple regression on the data. Despite potential for quadratic relationship between salience and weighted

[11] See Appendix D. for ACF test on full dataset.

proportion of responses in (k), salience is still included in the analysis. This is because salience is an important prediction for potential effect on the outcome of the allocation of responses to the clusters.

5.4 Considerations of Other Methodological Approaches

This thesis has selected the methodological approach based on the deductive research design adopted for exploring diversity in stakeholder preferences within OPC surveys. Deductive research characterizes research applying a "top-bottom" approach to a study (Moses and Knutsen 2012, 22). The hypothesis and effects are drawn from previous research when conducting deductive research, and theory and previous findings guide the empirical evidence used for analysis. The purpose will then not be to develop new theories but rather to test theories or to test for previous findings on empirical data.

For this thesis to best explore preference diversity and observe the effects of survey-level elements, a thorough assessment of the methodological approach has been done. Clustering has been assessed as the best approach to exploring diversity empirically, and different elements within clustering can provide different results. Based on the results gained from clustering, both negative binomial and multiple linear regression have been conducted to explore the potential effects of salience, targeted stakeholders, survey complexity, issue complexity, proposal status, and policy area on preference diversity. Thus, other approaches have been considered, or at least encountered, when exploring the effects of survey-level elements on preference diversity.

5.4.1 Clustering

For the clustering, other cluster algorithms could have been used, but PAM and CLARA are considered the best based on the data used. K-Means and Hierarchical clustering, to name a few, would have required more re-coding of the variables, without assumptions for better results (Reynolds, Richards, and Rayward-Smith 2004, 177). Appliance of another clustering algorithm would as well require another distance measure. As for the algorithm, appliance of another distance measure, such as the Euclidean distance for example, would as well require more data re-coding.

During the clustering, the use of the elbow method is can be questioned. This is because in several of the cases, as illustrated in Figure 5.1, it is seen that it is somewhat difficult to decide

which number will be able to give the best optimal k. Here, other techniques for determining the optimal number of clusters could have been more precise, including the Silhouette method. Silhouettes calculates the average proximities for average group linkage within observations. Though silhouettes might provide a clearer and compact separation between clusters, it is highlighted for usage when scaling ratio variables (Rousseeuw 1987, 55). In the case for the dataset within this thesis, the inputs have either a nominal or an ordinal scale, which would not have fully benefitted from application of the silhouette's calculation. By using silhouette plots in combination with the elbow method, the calculation of the specified number of optimal clusters could have been more precise. Despite this, the elbow method was prioritized as the silhouette additionally required more time to produce an optimal number of clusters calculation.

Overall, it would be difficult to obtain preferences from the response given across all collected OPC surveys without the use of clustering. The use of the Clustering makes it possible to compare the groupings of preferences observed in individual surveys at a general survey level.

5.4.2 Negative Binomial Regression

Negative binomial regression has been selected to test for effects on the dependent variable measuring the optimal number of clusters (k) for each survey. The optimal number of (k) is a count variable, which determines what regression model to run. In the case of count variables, the most frequently used regression is Poisson. Due to the variance exceeding the mean for the dataset, different regression models that can handle overdispersion have been evaluated. In total, four different count models have been explored, including the Poisson. Based on the test result, negative binomial regression (NB2), with a quadratic calculation of the mean for variance, turned out to be the most appropriate for modeling the dataset.

There are several other regression models for count data. Amongst them, zero-inflated models are often used. Poisson regression with zero-inflated counts is used to model a data set with an excess of zero counts. The optimal number of k extends between 1 and 15, which gives no reason for encountering the zero-inflated model. The chosen and tested regression models included have been addressed as sufficient for modeling the explored effects on optimal numbers of k.

5.4.3 Linear Regression

Multiple linear regression has been conducted to test survey-level elements' effects on the weighted average of responses within k. In contrast to the optimal number of k, the weighted proportion of responses is not conditioned as the variable is a cluster size calculation, indicating a continuous variable. Several different regression models could suit modeling the effects on a continuous variable in terms of both linear or non-linear regression models. From assessing the distribution and linearity within the dataset, conducting a non-linear regression would not be suitable. From observations, neither the distribution nor the linearity is deviant enough to assume other than linearity within the dataset. When evaluating additional assumptions, neither multicollinearity nor autocorrelation is observed. In sum, this allows for assuming the multiple linear regression as well suited to estimating potential effects on weighted proportion size by fitting a regular multiple linear regression using OLS.

6 Results from Analyses

Negative binomial regression and linear regression have been used to test the effects of survey-level elements on diversity in stakeholder preferences received through open public consultation surveys. The two dependent variables are composed of different segments related to preference diversity, which can account for the following: *1. how many different grouped opinions are received in OPC surveys,* and *2. an account for the size of the numbers of groups observed in OPC survey responses*. In comparison between the different goals, and the relationship the explanatory variables have on either the number of clusters or the size of the clusters, it will be possible to give an overview of how surveys influence the observed diversity of stakeholder preferences.

In this chapter, I present the results gained from clustering, which will answer how diverse stakeholder preferences are received through OPC surveys. In the second part, I test for predictions presented in the variable chapter and assess how they may affect diversity in preferences measured as the number of k's. For doing so, I have conducted a negative binomial regression. In the third part, I assess how the same predicted variables affect respondents' distribution within the clusters (weighted proportioned responses). Eight regressions have been run: four negative binomial and four linear regressions. For comparative reasons, the variables are gradually included in the regression models to properly assess the explanatory variables' different effects on both dependent variables. All continues explanatory variables have been z-standardized for comparison.

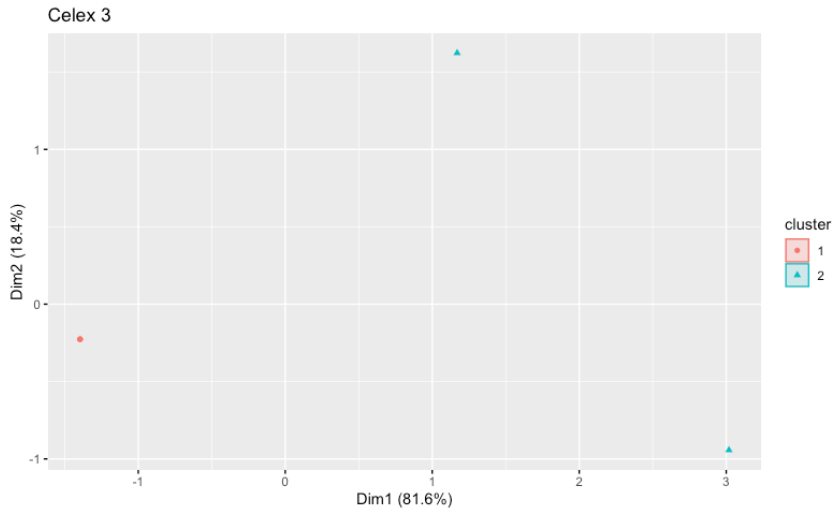
6.1 Clustering: Diversity in Stated Preferences

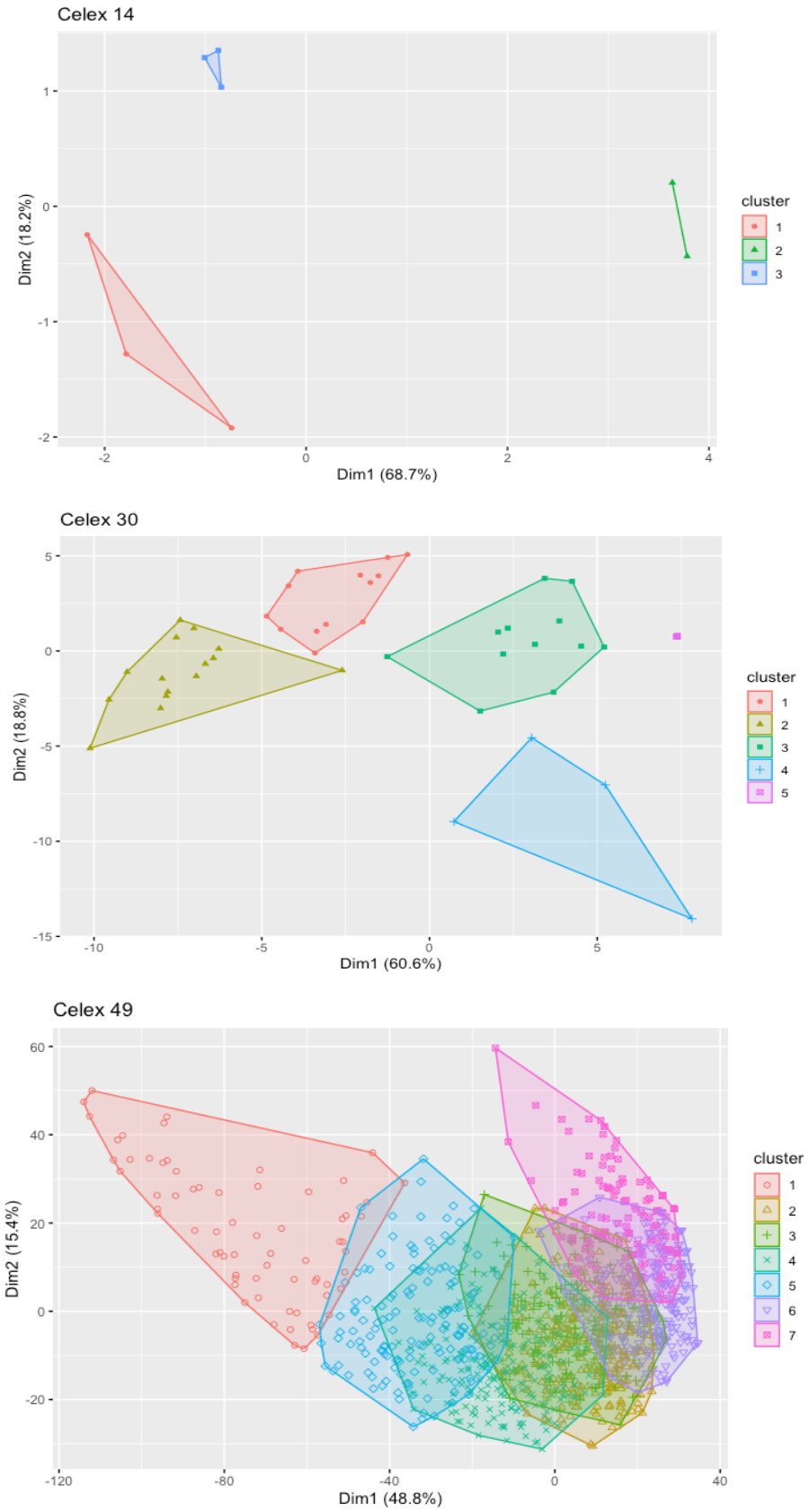
For addressing diversity, the measurements for preference diversity are a combined assessment of the optimal number of clusters and the weighted proportion of assigned responses within each cluster. Through the appliance of PAM and CLARA clustering algorithms, insight for the optimal number of k has been assessed. The first sign indicating how preferences were distributed within the surveys came during the data-wrangling. For each survey, questions asking for preferences within the survey have been crucial for estimating the optimal k. Additionally, the number of responses has influenced how many k's are deemed optimal. By creating the data, additional information gave insight into how distributed the preferences were.

As shown in Figure 5.1, six surveys did not entail more than one cluster to which stakeholders were assigned. Those six included the surveys entailing the fewest responses in the dataset. In contrast to these surveys which are characterized by a minimum level of preference diversity, the two surveys having an approximate optimal number of clusters equal to 15 were the two surveys entailing among the most responses. The majority of surveys turned out to have an optimal number of clusters between 2 to 7 for representing the preferences expressed by stakeholders. As PAM is considered to best predict and process data with between 2 to 7 predicted cluster groupings, the surveys entailing clusters within that scope are considered more accurately optimal than those consisting of 1 or 15 clusters. With better accountability for the majority of the units used, this implies stating that there is a general tendency of at least some diversity among stated preferences across consultations.

Properly assessing diversity within this thesis implies diversity in the sense of two measures. The weighted proportion has been calculated for the size of responses within each cluster. When creating the clusters, information regarding each distribution size has been abstracted for analysis. The size count varied due to the units of responses within each cluster and the predicted optimal number of k for each survey.

Figure 6.1 *Visualization of Clusters Using fviz_cluster for four Examples*





Note: Illustration of clusters created in four surveys: same surveys as illustrated in Figure 5.2. The medoids are organized in different shapes and colors to indicate both size and number of clusters within the represented surveys.

The *fviz_cluster* function from *the factoextra* package allows for visualizing the clusters 2 dimensionally with reduced distance using principal components analysis (PCA). PCA is used within machine learning as it is an unsupervised statistical technique that finds a low-dimensional representation of the data while keeping variation as original as possible (Boehmke and Greenwell 2019). Fviz_cluster plots the size through points for observations and an additional shape surrounding the sphere for the assigned observations within one cluster grouping. In Figure 6.1, different outcomes from the clustering are shown for four surveys[12]. The four surveys were selected due to their representativeness for all surveys included in the dataset. In Celex 3, only two clusters have been set as an optimal number of clusters. The observations assigned to the two medoids are differentiated through color and shape, one red dot and two blue triangles. Medoid one is centered quite the opposite from medoid two, which is scattered on the positive side of the 1-dimension. The entailed units in Celex 14 are estimated to suit three clusters. The first two medoids, also referred to as cluster groups, are characterized the same way as in Celex 3, but the third is shown as blue squares. For the medoids within Celex 14, clear, distinct results are shown as medoids one and two differ on the 1-dimension, whereas medoid three differs from the medoid one on the 2-dimension. The units within Celex 3 and Celex 14 are differentiated, indicating diversity in the included observations. The dimension of responses is thus assessed as smaller indicating less distance between observations.

Both Celex 30 and Celex 49 entail more observations distributed to larger amounts of medoids. Celex 30 shows a total of five medoids, consisting of larger numbers of observations compared to Celex 3 and Celex 14. Each of the five medoids encountered in Celex 30 includes diverse responses without overlapping. This indicates diversity; hence the medoids encountered vary in both dimensions. Medoid five, characterized by a pink square, shows distinction from medoids 3 and 4, but is considerably smaller than the other medoids. This indicates fewer observations for medoid 5. For Celex 49, overlapping between the medoids is observed. Medoids three, four, five, and six indicate resemblance between provided responses. Although observed overlapping, the dimension is considerably larger than Celex 3, Celex 14 and Celex 30. The observations in Celex 49 are anging from -120 to 40 on the 1-dimension and from below -20 to 60 on the 2-dimension. Knowing that the PCA reduces the data for observation

[12] For visualization of clusters for all surveys, see Appendix E. Note that surveys 24, 25, and 26 were too large to appropriately adapt the fviz_clusters function, only showing a limited dimension.

implies large distances, explaining why such overlapping appears when visualizing the medoids. For interpretation, the observations within Celex 49 are considerably more than the other Celex's within Figure 6.1.

These samples, which illustrate different outcomes based on potentially related to the estimation of optimal k within clusters, indicate greater diversity among medoids when entailing more observations. At the same time, a lower number of observations creates more distinct medoids within a smaller dimension. This implies less but still observed diversity when receiving few responses to an OPC survey. Despite the overlap, the nesting of the observations and dimension scale still implies differentiation of preferences present among more significant responses received for OPC surveys. For instants, the first medoid within Celex 49 clearly indicates differentiated preferences compared to the units encountered in medoids two, three, four, and five. This is also seen in Celex 30 for the observations in medoids two and five, compared to one, three, and four.

6.2 Negative Binomial Regression: Survey Elements on Diversity

The results presented from the negative binomial regression are reported with coefficients estimates and standard errors. Using the "raw data", without application of estimates, allows for interpretation with reliance on the standard errors for the coefficient. As the coefficient of the data are predicted using a logarithmic log estimate, interrelated through the NB2, the coefficient provided for each explanatory variable must me multiplied with e for each unit increase in the explanatory variable (Hilbe 2011, 20,187). A change in the predictor variable of one unit is expected to result in a change in the difference in the logs of expected counts of the response variable given the other explanatory variables within the model are held constant. Often the coefficients presented in negative binomial regression are interpreted as IRR, Incidence Rate Ratio. The IRR coefficients estimate the rate for each events occurrence over time, reducing credibility for the coefficients given standard errors.

Model 1 includes only explanatory variables focusing on reach for the OPC surveys: salience and Targeted Stakeholders. In order to assess the effects of the different variables and address the hypothesis adequately, its deemed appropriate. *Salience* shows a significant positive relationship to numbers of clusters with a significance level of 0.1%. This indicates that the higher salience a survey experience results in a slight increase for assumingly more clusters of

preference. Despite this relation, the significance is not on a 95% level, enabling to state a generally significant positive relationship. In contrast, targeted stakeholders are not proven significant in this Model. Although not significant, targeted stakeholder still indicates a negative effect on the numbers of diverse preference clusters.

In Model 2, the survey complexity-related variables have been included. Model 2 then consists of *salience, targeted stakeholders, mean length of questions within a survey*, and *total numbers of questions*. Survey-related items are included due to their assumed impact on stakeholders to engage and complete surveys. In Model 2, *salience* is still shown to significantly affect the numbers of (k) present in the total responses. The significance is still on a 0.1% level, indicating that salience increases the possibility for more diverse expressed preferences with a 0.134 multiple by e, which is considered a marginal effect. In the first Model, targeted stakeholders were non-significant with a negative association. In Model 2, the impact has substantially changed. Targeted stakeholders now significantly negatively impact how many distinct clusters are seen in the response in OPC surveys. The significance is at 0.05, implying that targeting stakeholders in when organizing an OPC survey decreases the possibility for receiving diverse inputs by -0.427 multiple by e in 95% of the cases. This is a remarkable growth in significance, implying it might be related to the other variables' presence. Both survey complexity variables show a negative association without significance. The mean length of questions shows a barely higher negative association than the number of questions. Both are not significant, indicating that the association cannot be assumed.

Model 3 entails all explanatory variables, including the remaining *issue complexity*. As in the two previous models, Salience still shows a significant effect of 10%, indicating a positive relation for optimal numbers of clusters when all explanatory variables are present. *Targeting stakeholders* shows a marginal decrease, implying a -0.428 multiple by e, a significant adverse effect on the number of clusters (k). The significance for targeted stakeholders remains at 0.05, which gives reasons for stating a negative impact on the dependent variable. Both survey complexity measures – *mean length of questions* and *number of questions*, remain non-significant with low negative values. The now added *issue complexity* shows a negative impact on the dependent variable with a value of -0.027 multiple by e. This implies a reduction in observed preference diversity as the more technical and complex the wording within the survey is. Despite the association, issue complexity is not significant, which leaves one unable to state a general relation between the implied association.

Lastly, model 4 includes all explanatory variables and the control variables: *proposal status* and *policy area*. Comparing all explanatory variables in model 4, only targeted stakeholders show significance. The association is now significant on a 99% level with an increase in impact and significance. Targeting stakeholders is stated to negatively impact the potential for diversity among received inputs in OPC surveys. *Salience* has had a significant positive estimate at 0.1 through all three previous models. However, it decreases both in significance and size in Model 4, having no effect on optimal numbers of (k) when control variables are included. *Issue complexity* increases drastically, showing over the double negative impact as in model 3. This increase implies a relation with one of the control variables, but as it remains non-significant, the effect would, however, not be assumed. *Mean length of survey questions* shows an increase in negative association, from -0.004 to -0.010 multiple by e, but remains non-significant—the same holds for *numbers of questions*, which grow from -0.001 to -0.002 and remains non-significant. The control variable *proposal status* has a positive sign, which indicates that new proposals have higher probabilities of receiving more diverse groupings in their response. However, the effect is not significant, stating that the associationis unreliable. For *policy areas*, the different DGs experience mixed results. OPC surveys run by the DGs CLIMA, EMPL, and MOVE have a higher chance of receiving more diverse responses than those surveys composed by the DGs GROW, HOME and ENER. As the policy area is non-significant, this effect cannot be assumed. Overall, the inclusion of control variables indicates reduction in salience effect on observed preference clusters within OPC survey responses, but improvements for all the other explanatory variables.

Assessing the measure of the model fit, the AIC indicates the lowest value for model 1. This indicates that Model 1 is has the best model fit for the data at hand. Model 1 gets the lowest BIC value, indicating Model 1 as the best fit among the compared models in table 6.1. Both BIC and AIC prefer smaller models, which leaves the Log-Likelihood estimation to be the best for predicting which model captures best the effects for the dependent variable. Model 4, entailing both explanatory and control variables, shows the best explanatory power with the highest log-likelihood.

Table 6.1 *Negative Binomial Regression Results for Optimal Number of (k)*

		DP: *Optimal Number (k)*			
		(1)	(2)	(3)	(4)
Salience		0.143*	0.134*	0.137*	0.081
		(0.073)	(0.075)	(0.075)	(0.080)
Targeted Stakeholder		-0.402	-0.427**	-0.428**	-0.542***
		(0.173)	(0.179)	(0.179)	(0.182)
Survey Complexity: Mean Length Q.			-0.004	-0.004	-0.010
			(0.010)	(0.010)	(0.010)
Survey Complexity: Numbers of Q.			-0.001	-0.001	-0.002
			(0.001)	(0.001)	(0.001)
Issue Complexity				-0.027	0.072
				(0.081)	(0.090)
Control Variables					
Proposal Status					0.303
					(0.208)
Policy Area	*CLIMA*				0.519
					(0.472)
	GROW				-0.502
					(0.419)
	EMPL				0.288
					(0.375)
	HOME				-0.339
					(0.355)
	MOVE				0.149
					(0.355)
	ENER				-0.127
					(0.707)
Constant		1.567***	1.726***	1.736***	1.871***
		(0.121)	(0.338)	(0.339)	(0.408)
Loglikelihood		-113.779	-113.628	-133.575	-106.879
AIC		233.558	237.256	239.149	239.757
BIC		241.513	249.190	253.072	267.603
N		54	54	54	54

Significance values: '*'p< 0.1, '**'p< 0.05, '***'p<0.01

6.3 Linear Regression: Survey Elements on Proportion of Diversity

Clear effects have been found for the dependent variable *weighted proportion responses in (k)*. In model 1, the two variables indicating the reach for the OPC survey have been included. *Salience* shows a significant positive relationship, and this implies that the weighted proportion of responses for clusters within surveys will grow when a survey experiences a higher degree of salience. Whereas salience has a significant positive effect on the dependent variable, targeting stakeholders indicates negative implications for the proportion of observed responses. With a coefficient estimate of -237.502, this implies drastic reductions for both number of preferences and the size of diverse preferences as the survey deliberately mentions specific stakeholder groups as associated. Despite the negative result, as the targeted stakeholder coefficient is not significant, it cannot be assumed either.

For model 2, survey complexity measures have been included. *Salience* remains with a significant positive relationship with a 99% certainty. With a small growth, salience shows an increased effect for receiving more and larger sets of diverse preferences within OPC surveys. The variable *targeted stakeholder* is similar to its values seen in model 1 but experiences some reduction of negative impact. Despite the reduction, targeted stakeholder is non-significant and cannot be stated to have an effect. For the variable of survey complexity, the *mean length of questions* shows positive results, whereas the *number of questions* entails negative results. This indicates more responses within (k) when more words are used for fewer questions. Still, as neither the mean length of questions nor the number of questions is significant, survey complexity cannot be stated to effect the proportions of preferences received in OPC surveys.

In Model 3, all explanatory variables are included. *Salience* remains significant with 99% certainty, and *targeted stakeholders* reduce its negative impact but remain non-significant. The coefficients increase for *mean length of questions* while remaining positive and non-significant. *Number of questions* experiences a reduction in its negative value but remains non-significant. Collectively neither targeted stakeholders, the length of questions, and the number of questions within a survey show any credible effect on the proportion of respondents within a cluster. In model 3 *issue complexity* have been included. Issue complexity implies a high positive coefficient related to the weighted proportion of responses assigned within clusters, but it cannot be guaranteed as it is not significant.

Model 4 includes all variables, both explanatory and control variables. Compared with the three previous regression models, *salience* maintains a significant positive relation on 0.01. The effect has grown, increasing from 1.993 to 2.087. Thus, the more salience surveys experience, the more received responses for an OPC survey. For *targeted stakeholders*, the negative effect has increased but remains non-significant. The increase for targeted stakeholders is notably more extensive than previously observed between models 1, 2, and 3, but it is not counted as it remains non-significant. The survey complexity variables additionally experience drastic increases. Mean length of questions goes from a positive coefficient of 41.069 to 62.019. Numbers of questions goes from being negative at -2.212 to positive with 0.844. Though appearing more prominent and positive, they remain non-significant and cannot guarantee any effects. Issue complexity reduces its negative estimate but remains non-significant. When addressing the now included control variables, both consist of positive estimates. Proposal status has a positive coefficient, implying higher proportion of preferences when OPC surveys are drafted for new legislative proposals. Although the positive coefficient, proposal status is not shown significant, making it unreliable. Two of the DGs included show significance in their effect when addressing the variable policy area. All DGs included for policy area are positively estimated relative to the reference category, DG ENV. The DGs CLIMA and HOME can be stated to have a increasing effect for the total proportion of preferences received. As DGs CLIMA and HOME do not have significance on a 0.05 level, but at a 0.1 level, reducing possibility to state a significant relationship. Instead, they indicate assumptions for positive relations between the two DGs and the received preferences through OPC surveys.

When comparing the models, both model 1 and 4 entails an adjusted R^2 on 0.540. The adjusted R^2 for the model represents the variance explained of the multiple linear models, taking into account all variables. Looking at both AIC and BIC for the four models, model 1 receives the lowest value, indicating the best fit and most explanatory power. Despite the best results, it is also the least inclusive model, only encountering effects of two variables. Because AIC and BIC favor smaller models, model 4 is considered the best model for explaining the weighted proportion of responses in clusters among preference clusters from surveys in OPC.

Table 6.2 *Multiple Linear Regression for Weighted Proportion of Responses in (k)*

		DP: *Weighted Proportion Responses in (k)*			
		(1)	(2)	(3)	(4)
Salience		1.978***	2.004***	1.993***	2.087***
		(0.275)	(0.285)	(0. 289)	(0.308)
Targeted Stakeholder		-237.502	-211.139	-182.776	-784.302
		(1,152.923)	(1,212.333)	(1,224.657)	(1,323.162)
Survey Complexity: Mean Length Q.			40.121	41.069	62.019
			(68.547)	(69.173)	(70.637)
Survey Complexity: Numbers of Q.			-2.946	-2.212	0.844
			(7.985)	(8.247)	(8.760)
Issue Complexity				227.354	92.924
				(550.809)	(642.038)
Control Variables					
Proposal Status					446.962
					(1,391.236)
Policy Area:	*CLIMA*				2,121.074
					(3,784.909)
	GROW				5,551.222*
					(3,072.938)
	EMPL				1,361.054
					(2,988.596)
	HOME				4,788.391*
					(2,673.858)
	MOVE				3,923.300
					(2,738.702)
	ENER				4,219.865
					(4,798.595)
Constant		43.680	-700.761	-803.338	-5,231.243
		(981.758)	(2,410.475)	(2,443.812)	(3,289.074)
Adj. R^2		0.540	0.526	0.518	0.540
Std. Error		3,791.464	3,847.388	3,880.378	3,788.615
F-test		32.073***	15.706***	12.386***	6.193***
		(df = 2; 51)	(df = 4; 49)	(df = 5; 48)	(df = 12; 41)
AIC		1048.134	1051.555	1053.363	1056.267
BIC					
N		54	54	54	54
Significance values: '*'p< 0.1, '**'p< 0.05, '***'p<0.01					

6.4 Summary of the Regressions Results

Through the negative binomial regression, model 4 emerges as the most suitable model. Model 4 has the highest loglikelihood, which considers the number of variables included, and is therefore considered the best model for explaining effects on optimal numbers of clusters. For the regression, only targeted stakeholders are significant, indicating that targeted stakeholders are the only variable with an actual effect on the number of optimal clusters. Hence, open public consultation surveys mentioning desired stakeholder groups to receive feedback from experience a -0.542 effect on the total number of observed preference groupings within the responses.

For accounting effects on the weighted proportion of responses within a cluster, linear regression has been run. Through the multiple linear regression, only the variable salience shows an apparent constant positive significant effect on responses within clusters in all four models. Salience maintains a significance at 99%, stating that the more publicly salient a survey is, the more diverse responses are expected to be received. When encountering which model to see best for explaining the effect, both model 1 and model 4 score the highest adjusted r-square. Both entail significant F-testes, but model 1 entails the best AIC and BIC. Despite entailing lower AIC and BIC values, model 4 is assessed as the best model fit. With inclusion of contro variables, model 4 present model results with a reduced risk of omitted variable bias. When encountering the total entity of variables included in model 4, additional comments must be made regarding the control variable policy area. Surveys organized by the DGs CLIMA and HOME have a significance of p=0.1, assuming their effect is positively associated with the weighted proportion of responses in clusters.

7 Discussion

The discussion chapter presents four aspects of discussion for the analyses and assessments done in this thesis. The first section of this chapter evaluates the four presented hypotheses from chapter 3. For this assessment, considerations of both dependent variables will be made. Secondly, the empirical results are assessed in light of the presented theoretical explanations. Accordingly, this will be done in subsection four but with focusing on the thesis's explanatory findings. Lastly, I present assessments of this analysis's main implications and limitations. The thesis has been seeking to answer the two following research questions:

"1) How diverse are the opinions stakeholders express in EU open public consultations?
2) How can we explain differences in the diversity of expressed opinions across consultations?"

Assessing the diversity of stated preferences in EU open public consultations, clustering received inputs from 54 OPC surveys have shown to imply a general diversity among stated preferences. As the size varies, the mean number of clusters observed within the whole dataset implies a broad diversity of ≈ 4 (3,94) clusters representing diverse preferences. This indicates diversity in terms of different opinions as quite diverse. In addition, measuring the weighted average proportion size for responses included in each survey represents a mean of 124.6 (cluster size measurement), implying a relatively large size distribution for each cluster. The analyses imply a rather great diversity in opinions stated in OPC surveys. It is hard to interpret whether these results imply a high degree of diversity, or a low degree, due to the non-existence for comparison. Usage of clustering stakeholder preferences has never, or as far as I have been able to find, been assessed using clustering algorithms.

To explain differences impacting diversity in preferences received through OPC surveys, the analysis shows primary findings. These two, salience and targeted stakeholders, have been shown to explain diversity in each way. Drawing from the findings from the analyses, the hypothesis presented in table 7.1 has proved significant indications, allowing for stating two out of the four hypotheses to be partially supported by empirical evidence in the case of this thesis.

7.1 Evaluation of Hypotheses

Based on the research question, this thesis has formulated four hypotheses to assess the effect of survey-level elements on diversity in received preferences in OPC surveys. The presented literature shows that the research field is primarily focused on stakeholder characteristics for explaining preference and preference attainment. The survey level was selected to assess preference diversity and expand understanding of expressed preferences. As survey elements have been little researched, understanding of participation from different stakeholder groups has given indications of what can be expected of diversity in stated preferences and under which conditions to expect diversity in stated preferences.

Concretely, the hypotheses shown in table 7.1 are based on indications that will reduce the participation of various stakeholder actors. Hence, indications have caused expectations for reducing participation when dealing with complex thematical issues and creating complex surveys. Additionally, assumptions have been made regarding elements generally shown by research to affect participation, such as targeting stakeholders and publicity.

Table 3.1 *Evaluation of Hypothesis*

Hypothesis	**Assessment**
H1: Open public consultations on more salient issues are associated with more diversity in expressed stakeholder preferences than open public consultations on less salient issues.	Partially proven. *Note: stated to effect diversity assessed as Weighted Proportion responses in (k)*
H2: Open public consultations that target specific stakeholder types are associated with less diversity in stated stakeholder preferences than open public consultations that refer to "all" stakeholder types.	Partially proven. *Note: stated to effect diversity assessed as Optimal N of (k).*
H3: Open public consultations on more complex issues are associated with less diversity in expressed stakeholder preferences than open public consultations on less complex issues.	Rejected.
H4: Open public consultation that use more complex surveys are associated with less diversity expressed stakeholder preferences than open public consultations that use less complex surveys.	Rejected.

H1: Effect of publicly known legislative proposals for received preferences

Through regression, salience, the degree of responses received for each survey, indicates an apparent effect on diversity in stated preferences. Salience in general, based on both regressions, has a positive association with preference diversity but can be stated to affect the proportion of received responses in clusters. This is indicated throughout in all four multiple linear regression models. The positive relation between salience and the proportion of responses received makes much sense as the number of respondents for OPC surveys provides a larger pool for preferences, hence a higher probability for more clusters and larger size of clusters. Salience is also stated in three out of four models in the negative binomial regression, indicating a positive association for number of clusters appearing within the survey responses as well.

As noted from previous research, salience has also been the most prominent factor in being able to look at participation and stakeholder preferences, both in a positive and negative sense for preference attainment. Whereas the degree of publicity for a legislative proposal is often associated with reducing the potential for stakeholders to obtain preference attainment as competition grows, it provides more preferences for the EC to consider. Hence, salience allows for a more diverse representation of preferences, creating more evident indications when considering raw preferences received through OPC surveys.

H2: Effect of specifying stakeholders of interests on received preferences

Targeting stakeholders implies devotion to specific stakeholders within the roadmap or in the OPC survey itself. A reasonably thorough review of how the European Commission organizes its consultations signals that DGs in management often have intentions about what kind of information is desired. The actual OPC surveys facilitate the broadest possible representation across different stakeholder groupings. Despite this, some of the included surveys in the dataset have been consistent with specifications for interested stakeholder groups. Indications for specification, or devotion, for specific stakeholder groups have been encountered through the regression.

By appliance of a dichotomous variable, where 1 indicates specification, targeting stakeholders has proven to affect perceiving diversity in clusters significantly in three out of four models using negative binomial regression. In the same way as for salience, the stated effect of targeting stakeholders on observed clusters of diverse preferences is neither shocking nor unexpected. Insight from the presented literature on characteristics for preferences represented by different

stakeholder groups, it is assumed to encounter a reduction in diverse preferences when targeting specific stakeholder groupings. As the heavenly choicer have been documented to dominate consultations, they often prefer remaining status quo. In contrast, public and societal stakeholders have been shown to prefer generalization, hence the compliance with new regulations. When OPC surveys specify which stakeholder groups entail the desired responses, it is natural for those groups to dominate the represented preferences. As stated from the regression, this leaves OPC surveys to reduce the response from diverge groupings of preference, mainly reducing the potential for encountering diversity among the received responses.

H3: Effect of issue complexity for received preferences

While addressing survey level elements' effect on diversity in preferences received through open public consultation surveys, the complexity of the issue being addressed has in previous research been shown to impact expectations for received inputs. Stakeholders compose different forms of knowledge and competence. Stakeholder groups such as interest groups, financial corporations, and public authorities, are often considered to possess broad expertise and a high degree of resources. Compared to less advantaged stakeholders, such as individual citizens and NGOs, the complexity of the issue being consulted can create an uneven distribution of expectations regarding the representation of stakeholders. The demographic representation among different stakeholders is expected to be limited through composing complex and advanced wordings for the issue being consulted.

Issue complexity has been tested on the dataset within this thesis. The variable issue complexity is an internal assessment of words used within each survey, giving an overall score of how frequent and less frequent words appear. The thought has been to map for more complex wording through the rarity of words. Based on the dataset, it is not possible to determine whether complex wording within surveys affects the diversity of received preferences.

Additionally have the policy area been controlled in hopes of encountering other issue-related effects. Policy areas show positive associations for the representation of preferences received within open public consultations. Whereas the issue complexity neither generates nor reduces diversity in preferences, the policy areas can initially positively impact expectations for observing diverse preferences. The drafted legislative proposals organized by MOVE and HOME show positive associations for attracting more diverse preferences expressed by

stakeholders. In sum, the issue's complexity does not obtain associated effects, so the hypothesis is rejected. Notions can, though, be made regarding the policy area's positive association with cumulating diverse representation of responses.

H4: Effect of survey complexity for received preferences

Previous research, as mentioned, signals complexity to be an essential factor influencing the conditions for participation. In the study of survey complexity, two different measures have been used to address the scope of survey complexity. The first variable accounts for the average question length, as of how many words were used, for the surveys. The second variable addresses the total number of questions included in the survey. Similar to issue complexity, previous research considers survey complexity to reduce overall participation in consultations. In line with this assumption, it has been expected to find adverse effects of survey complexity on diversity in stated preferences. Fewer stakeholder types will submit responses to surveys composing several long questions.

Despite assumptions of less participation, hence less stated preferences, for surveys entailing several and more extended questions, the regressions provide no stated effect. Through no significant accounted effects, the mean length of questions and the number of questions provide no insight. Hypothesis 4, like hypothesis 3, is therefore rejected.

7.2 Empirically: How Divers are Stated Preferences in OPC Surveys

Defining how diverse stakeholder preferences are calls for subjective interpretations. Diversity can be understood as how many groupings, addressed as clusters, are seen within the responses to the survey but can also be based on interpretation of how distribution within surveys is segmented. In Appendix E, the visualization of all responses collected for the surveys is presented. The approach applied through the thesis allows combining both measures to address diversity in stated preferences. The analysis shows that the number of submissions is a significant predictor for understanding the diversity in stated preferences, as respondents determine how diverse the preferences can be.

Drawing on the findings from Bunea (2014b) indicate that diversity is more present when experiencing the broad inclusion of different stakeholder groups. This can be seen within surveys entailing a higher salience: receiving more submissions. These surveys, such as survey

49 from figure 7.1, entail seven groupings interpreting different preferences. In contrast, surveys receive less participation, which is the case for survey 3; the overall domain for observed diversity is reduced to encountering only preferences from the few respondents.

Although fewer preference groupings appear within surveys entailing low numbers of respondents, apparent differences are still indicated. Among other things, it has been noted from previous research that one should not assume equal preferences from equal stakeholder groups (Chalmers 2018, 391). Not having the prerequisites for broad representation with low support for the response can indicate differences among observed responses. Even in surveys receiving among the fewest responses accounted for, seen entailing one cluster in figure 5.1, indications for diversity to some degree among the stated preferences are observed. When locating the preferences included in the survey, they appear at different locations, and the space between the few accounted responses still indicates the presence of diversity among preferences.

In sum, the overall tendencies indicated through clustering responses from the selected open public consultation surveys show diversity to be stated within surveys. When addressing how diverse, this is conditioned to the number of respondents for each survey. Considering the dataset, diversity appears relatively frequent, indicating a high degree of diversity in the included consultations.

7.3 Explanatory: How do Survey Elements Affect Diversity in Stated Preferences

The main emphasis of explanatory variables has also been explained through the hypotheses' evaluation. Overall, there are clear indications that salience and targeting stakeholders are decisive for which expectations one can have for diversity in stated preferences. Salience has, through previous literature, shown to be necessary for obtaining more extensive amounts of preferences. With significant results throughout, salience shows to be positively influential on the degree of observed preferences in surveys. Whether salience again affects the preference attainment has not been taken into account, and one can thus not say anything concrete about the consequences salience has for the preferences that will be to get in through OPC surveys.

It can be assumed that the more salience the survey experiences, the more variation in preferences it will get. Based on previous research, these may be indications of responses from

several different stakeholder groups, which in turn have been found to have a "(...) 70% higher chance of verbally formulating similar responses" (Bunea 2014a). In addition, salience is generally vital for meeting the expected conditions for OPC surveys. The OPC surveys are constructed and used to facilitate the broadest possible representation of interests. Through a high degree of salience, a greater degree of coincident representation of preferences can be ascertained.

Again, salience indicates no substantial effect on how many diversified preferences are present in responses. Targeted stakeholders, by contrast, do this. Targeted stakeholders indicate a negative effect, which is in line with indications given in previous literature. Dür and De Bièvre (2007, 4) refers to targeting stakeholders as an effective way to ensure concretized desired competence, which states the adverse effects of receiving diverse preferences within the response.

Open public consultation surveys are constantly used to ensure transparency and create opportunities for the entire public to participate with input on presented EU policies. Based on the analysis, it can be pointed out that specifying the stakeholders preferred to receive responses is less appropriate, recalling the fundamental prerequisite for using OPC surveys. On the other hand, targeting stakeholders can be necessary to reduce the potential for redundant responses received. For the EC, the reduced preferences received can also make some of the work more straightforward by referring to the correspondence between the desired outcome and public preferences. By only receiving a limited preferred view, fewer preferences will have to be taken into account, making the process more favorable for the EC.

Overall, the findings on the relationship between salience and weighted proportion of responses within (k), targeted stakeholder and the optimal number of (k), correspond with the assumptions made based on previous research for stakeholder preference and participation.

7.4 Implications and Limitations

As this thesis uses a quantitative method, some considerations must be considered when addressing the results. First and foremost, the thesis has a deductive research design, allowing the usage of quantitative methods to test hypotheses drawn from previous research. No documentation has been found to apply the same approach to investigate diversity in

stakeholder preferences within OPC surveys. The thesis provides an explanatory approach to previous findings related to stakeholder participation and stakeholder preferences within the EU. Drawing on the presented literature, the thesis applies hypotheses that assumingly can explain or impact diversity in stated preferences. Although embedding previously addressed explanatory variables, the quantitative approach limits testing only to effects that are accounted for within the thesis. Hence, the thesis does not approach testing other than the selected survey-level elements. The literature on stakeholder participation in the EU has primarily focused on stakeholder characteristics when addressing stakeholder participation and preferences, and neither stakeholder characteristics nor stakeholder groups have been encountered. The limited scope of identifying all possible explanatory elements which might impact diversity in stakeholder preferences within OPC surveys can therefore not be stated as exclusively applicable.

The variables used for the study have also been single-humanly coded, leading to somewhat weaker reliability and validity. Most variables are based on either available information or calculations, which are considered highly likely to recreate. The most challenged reliability concerns recreate issue complexity as this variable has been implemented with a code not available digitally for use in R. Also, it is doubtful that the clustering can be recreated identically as it involves judgment calls. As specifically highlighted in Chapter 4.4.2, the resilience of the clustered foundations is difficult to recreate. This is because the questions are subjectively assessed based on given categorical labels. However, the dependent variables' validity can be considered relatively high as an external third party has validated the labeling. This means that the outcome of the clustering and the basis for the response used can be said to address preferences stated in surveys accurately.

For the actual implementation of the clustering, the choice of method for identifying the optimal number of (k) may be somewhat weak as the function fviz_nbclust depends on large enough data to give a good interpretation of which “elbow” to rely on. With this noted, surveys entailing smaller sizes can thus not be sure of giving the most correct and precise prediction for numbers of clusters.

When addressing the exploratory part of the thesis, the dataset created has experienced challenges when satisfying all formal associations related to the regressions that have been used.

The regressions have initially been run based on the type of variable the two dependent variables are and assumptions for the best and most precise fit.

The multiple linear regression has been assessed as sufficient for testing the effects of explanatory variables on the proportion of responses observed from the surveys. Despite not being perfectly linear, the normal distribution and the VIF test have made it possible to rely on the results. The negative binomial regression generally is unfavorable for use on the dataset. The dataset consists of 54 units, which can be interpreted as a small selection, and negative binomial regressions are not in favor of small datasets. The most remarkable change can also be observed from models 3 to 4 within the negative binomial regression. Specifically, the variable issue complexity is experiencing an exceedingly big increase. Since issue complexity is not significant, neither in model 3 nor model 4, the increase indicates less reliability for the variable's coefficient for the optimal numbers of clusters.

The regression results interpretation must generally be considered as potential relationships rather than substantial effects. There have only been investigated seven possible effects on stakeholder diversity, whereas most show no effect, which implies the potential for other effects that have not been taken into account. What can be conclusive is that the analysis accounted for within this thesis allows for further investigation of potential effects on diversity in stated preferences.

8 Concluding Remarks

Through various methodological approaches, this thesis has extracted preferences across 54 diverse open public consultation services in the European Commission and tested for various effects that may affect diversity in stated preferences. Through clustering, preferences have been extracted from responses to OPC surveys and grouped based on response similarity. Generally, between two and seven preference clusters were identified in a majority of responses, indicating diversity in preferences. In addition, the units within surveys entailing only one cluster to be distanced, indicating additional diversity.

Differences in preferences received through OPC surveys can, to some extent, be explained by salience and targeted stakeholder. Different survey elements have been tested through negative binomial and multiple linear regression to account for what can explain diversity in stated preferences. To assess the potential effect on diversity, understood as the number of preference groupings within responses, surveys that specify targeted stakeholders experience a decline in diversity. Simultaneously, salience positively impacts the weighted proportion of responses in observed preference groupings. In line with previous research on stakeholder participation, this provides reliable estimations for salience and targeted stakeholders as important components when encountering diversity in stated preferences received within an OPC survey.

8.1 Contribution

The empirical analysis, I would argue, provides then a very resourceful insight for the field of research on relationships between stakeholders and their participation in consultations arranged by EC. The image illustrates a remarkable diversity, indicating that OPC surveys broadly impact the stakeholder population. To what extent one has succeeded with representation will be another discussion. What can be stated is that diversity is shown here, which provides suitable conditions for further research on stakeholder participation and understanding of how stakeholder participates. In addition, the empirical research provides insight into how public preferences behave in consultations and how this will relate to the EC.

I arguably would highlight the negative relationship between targeted stakeholders and the number of grouped preferences, as well as the stated increased salience for proposing preferences in the observed groups, as the most insightful findings attained from the regressions. The two findings help to provide a clear picture of how OPC surveys facilitate and

potentially prevent broad preference representation. The findings also positively contribute to the school of thought for stakeholder preference and stakeholder participation in the EU, as this finding shows significance for preference diversity.

It can also be mentioned that several of the non-significant findings also give reasonable indications. The assumption that complexity will have a more significant role than is shown indicates that the surveys are better adapted to a diverse stakeholder audience. Alternatively, the lack of significance for survey complexity and issue complexity may also indicate that respondents to selected surveys are better placed to state preference than otherwise. The assumption is related to previous literature that refers to motivation to participate and the intention of policy alignment. This may be the motive, but based on the diversity in preferences seen in the more prominent response masses, one can assume a wide range of stakeholders, making the assumptions related to competence and resources less critical for participation in OPC surveys. OPC surveys will facilitate the opportunity for everyone to participate, something that could potentially have been successful with non-significant effects in the regressions.

8.2 Notions for Further Research

First, using clustering to assess preference diversity within consultations is stated as both practical and insightful. Through mapping, clustering allows navigating through large amounts of data and outlining distribution internally in survey responses based on similarities. As the dataset in this thesis can be perceived as somewhat limited as it only entails 54 surveys, the same procedure can be tested on larger amounts of surveys. For this, it will require a more remarkable time aspect to assess questions that provide the correct form of information and address all the surveys as this is done individually.

Several other segments will also be relevant when applying clustering on a larger unit. Among other things, it will be insightful to address a similar approach to collecting OPC surveys organized by all DGs and collect a similar number of surveys from each DG.

If regression is to be done to address effects on diversity, the inclusion of other variables from the stakeholder level will be very enriching. As this analysis focuses exclusively on survey-level items, the range of variables has been somewhat limited. Assessing the presented results,

only two out of seven accounted effects are shown to have an active effect on stakeholder diversity, indicating there might be more fruitful effects at other levels.

This thesis has been focused on clustering closed answered questions where responses are based on given indications to address preferences diversity within OPC surveys. To gain more depth when addressing preference diversity quantitatively, the appliance of a text-analysis tool for "open" response can provide helpful insight. As this has not been encountered, nor possible with the given time frame of the thesis, text mining approaches are assumed to provide additional fruitful insight for preference diversity.

9 References

Agresti, Alan. 2013. *Categorical data analysis*. Hoboken, New Jersey: Wiley-Interscience.

Arras, Sarah, and Caelesta Braun. 2017. "Stakeholders wanted! Why and how European Union agencies involve non-state stakeholders." *Journal of European Public Policy* 25 (9): 1257-1275. https://doi.org/10.1080/13501763.2017.1307438.

Austgulen, M. H. 2020. "Understanding National Preferences in EU Consumer Policy: A Regime Approach." *Journal of consumer policy* 43 (4): 767-798. https://doi.org/10.1007/s10603-020-09468-8.

Bernauer, Thomas, and Robert Gampfer. 2013. "Effects of civil society involvement on popular legitimacy of global environmental governance." *Global Environmental Change* 23 (2): 439-449. https://doi.org/10.1016/j.gloenvcha.2013.01.001.

Bernhagen, Patrick, Andreas Dür, and David Marshall. 2015. "Information or context: what accounts for positional proximity between the European Commission and lobbyists?" *Journal of European Public Policy* 22 (4): 570-587. https://doi.org/10.1080/13501763.2015.1008556.

Beyers, Jan. 2016. "Voice and Access." *European Union Politics* 5 (2): 211-240. https://doi.org/10.1177/1465116504042442.

Beyers, Jan, and Sarah Arras. 2019. "Who feeds information to regulators? Stakeholder diversity in European Union regulatory agency consultations." *Journal of Public Policy* 40 (4): 573-598. https://doi.org/10.1017/s0143814x19000126.

Beyers, Jan, and Caelesta Braun. 2013. "Ties that count: explaining interest group access to policymakers." *Journal of Public Policy* 34 (1): 93-121. https://doi.org/10.1017/s0143814x13000263.

Binderkrantz, Anne Skorkjær, Jens Blom-Hansen, Martin Baekgaard, and Søren Serritzlew. 2022. "Stakeholder consultations in the EU Commission: instruments of involvement or legitimacy?" *Journal of European Public Policy*: 1-21. https://doi.org/10.1080/13501763.2022.2058066. https://doi.org/10.1080/13501763.2022.2058066.

Boehmke, Brad , and Brandon Greenwell. 2019. *Hands-On Machine Learning with R*. New York: Chapman and Hall/CRC.

Bouwen, Pieter. 2004. "Exchanging access goods for access: A comparative study of business lobbying in the European Union institutions." *European Journal of Political Research*

43 (3): 337-369. https://doi.org/https://doi.org/10.1111/j.1475-6765.2004.00157.x. https://ejpr.onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-6765.2004.00157.x.

Bunea, Adriana. 2013. "Issues, preferences and ties: determinants of interest groups' preference attainment in the EU environmental policy." *Journal of European Public Policy* 20 (4): 552-570. https://doi.org/10.1080/13501763.2012.726467.

---. 2014a. "Explaining Interest Groups´ Articulation of Policy Preferences in the Europen Commission´s Open Consultations: An Analysis of the Environmental Policy Area." *Journal of Common Market Studies* 52 (6): 1124-1241. http://dx.doi.org/10.1111/jcms.12151.

---. 2014b. "Sharing ties and preferences: Stakeholders' position alignments in the European Commission's open consultations." *European Union Politics* 16 (2): 281-299. https://doi.org/10.1177/1465116514558338.

---. 2015. "Evaluating Pluralism: Interest Groups' Policy Demands and Lobbying Success in the European Commission's Open Consultations. A Case Study in Environmental Policy." In *Interest Groups in the European Union: Changing Times, Changing Minds*, edited by Paul A. Shotton and Paul G. Nixon, 49-68. London: Routledge.

---. 2017. "Designing stakeholder consultations: Reinforcing or alleviating bias in the European Union system of governance?" *European journal of political research* 56 (1): 46-69. https://doi.org/10.1111/1475-6765.12165.

Bunea, Adriana, Raimondas Ibenskas, and Anne S. Binderkrantz. 2017. "estimating interest groups? policy positions through content analysis: a discussion of automated and human-coding text analysis techniques applied to studies of EU lobbying." *European political science* 16 (3): 337-353. https://doi.org/10.1057/eps.2016.15.

Cameron, Adrian Colin, and Pravin K. Trivedi. 2005. *Microeconometrics : methods and applications*. Cambridge: Cambridge University Press.

Chalmers, Adam William. 2014. "In over their heads: Public consultation, administrative capacity and legislative duration in the European Union." *European Union Politics* 15 (4): 595-613. https://doi.org/10.1177/1465116514529849.

---. 2018. "Unity and conflict: Explaining financial industry lobbying success in European Union public consultations." *Regulation & Governance* 14 (3): 391-408. https://doi.org/10.1111/rego.12231.

Clarke, Kevin A. 2009. "Return of the Phantom Menace:Omitted Variable Bias in Political Research." *Conflict Management and Peace Science* 26 (1): 46-66.

https://doi.org/10.1177/0738894208097666. https://journals.sagepub.com/doi/abs/10.1177/0738894208097666.

Dudley, Geoffrey, and Jeremy Richardson. 1999. "Competing advocacy coalitions and the process of 'frame reflection': a longitudinal analysis of EU steel policy." *Journal of European Public Policy* 6 (2): 225-248. https://doi.org/10.1080/135017699343694.

Dür, Andreas. 2008a. "Interest Groups in the European Union: How Powerful Are They?" *West European Politics* 31 (6): 1212-1230. https://doi.org/10.1080/01402380802372662.

---. 2008b. "Measuring Interest Group Influence in the EU." *European Union Politics* 9 (4): 559-576. https://doi.org/10.1177/1465116508095151.

Dür, Andreas, Patrick Bernhagen, and David Marshall. 2015. "Interest Group Success in the European Union." *Comparative Political Studies* 48 (8): 951-983. https://doi.org/10.1177/0010414014565890.

Dür, Andreas, and Dirk De Bièvre. 2007. "The Question of Interest Group Influence." *Journal of Public Policy* 27 (1): 1-12. https://doi.org/10.1017/s0143814x07000591.

European Commission, 25. July, 2001, "European Governance A white Paper," White Paper, https://ec.europa.eu/commission/presscorner/detail/en/DOC_01_10.

European Commission 29. April, 2021a, "Better Regulation: Joining forces to make better EU laws and to prepare for the future."

---, 2021b, "Better Regulation: Joining forces to make better laws," https://ec.europa.eu/info/files/better-regulation-joining-forces-make-better-laws_en.

---. 2022. "Better Regulation: why and how." Accessed 1. February. https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulation-why-and-how_en#whatthecommissionisdoing.

European Union. 2022. "Transparency Register - Statistics." Accessed 19.02. https://ec.europa.eu/transparencyregister/public/consultation/statistics.do?locale=en&action=prepareView.

Falleti, Tulia G., and Julia F. Lynch. 2009. "Context and Causal Mechanisms in Political Analysis." *Comparative political studies* 42 (9): 1143-1166. https://doi.org/10.1177/0010414009331724.

Finch, W. Holmes, Jocelyn E. Bolin, and Ken Kelley. 2014. *Multilevel Modeling Using R*. Vol. 16.*Chapman & Hall/CRC statistics in the social and behavioral sciences series*. Baton Rouge: Baton Rouge: CRC Press.

Fink, Simon, Eva Ruffing, Tobias Burst, and Sara Katharina Chinnow. 2021. "Less complex language, more participation: how consultation documents shape participatory

patterns." *Interest Groups & Advocacy* 10 (3): 199-220. https://doi.org/10.1057/s41309-021-00123-2.

Fischer, Corinna, Loet Leydesdorff, and Malte Schophaus. 2004. "Science shops in Europe: the public as stakeholders. ." *Science and Public Policy* 31 (3): 199-211.

Fraussen, Bert, Adrià Albareda, and Caelesta Braun. 2020. "Conceptualizing consultation approaches: identifying combinations of consultation tools and analyzing their implications for stakeholder diversity." *Policy Sciences* 53 (3): 473-493. https://doi.org/10.1007/s11077-020-09382-3.

Gerring, John. 2012. *Social science methodology : a unified framework*. 2nd ed.*Strategies for social inquiry*. Cambridge: Cambridge University Press.

Golden, Marissa Martino. 1998. "Interest Groups in the Rule-Making Process: Who Participates? Whose Voices Get Heard?" *Journal of Public Administration Research and Theory: J-PART* 8 (2): 245-270. http://www.jstor.org/stable/1181558.

Gornitzka, Åse, and Ulf Sverdrup. 2010. "Access of Experts: Information and EU Decision-making." *West European Politics* 34 (1): 48-70. https://doi.org/10.1080/01402382.2011.523544.

Gower, J. C. 1971. "A General Coefficient of Similarity and Some of Its Properties." *Biometrics* 27 (4): 857-871. https://doi.org/10.2307/2528823.

Grant, Wyn. 2015. David B. Truman, The Governmental Process: Political Interests and Public Opinion. Oxford University Press.

Grimaud, Jean Micallef 2018. "Defining Key Terms in Relation to Decision-Shaping and Decision-Taking in EU Decision-Making Processes." In *Small States and EU Governance*, 57-100.

Hermansson, Henrik. 2016. "The European Commission's environmental stakeholder consultations: Is lobbying success based on what you know, what you own or who you know?" *Interest Groups & Advocacy* 5 (3): 177-199. https://doi.org/10.1057/iga.2016.6.

Hilbe, Joseph M. 2011. *Negative binomial regression*. Cambridge: Cambridge University Press.

Holzer, Boris. 2008. "Turning Stakeseekers into stakeholders, A political coalition perspective on the politics of stakeholder influence." *Sage Publications* 47 (1): 50-67. https://doi.org/I10.1177/0007650307306341.

Kellstedt, Paul M., and Guy D. Whitten. 2018. *The fundamentals of political science research*. Third edition. ed. Cambridge: Cambridge University Press.

Kettenring, Jon R. 2006. "The Practice of Cluster Analysis." *Journal of Classification* 23 (1): 3-30. https://doi.org/10.1007/s00357-006-0002-6. https://doi.org/10.1007/s00357-006-0002-6.

Klüver, Heike. 2013a. "Lobbying as a collective enterprise: winners and losers of policy formulation in the European Union." *Journal of European Public Policy* 20 (1): 59-76. https://doi.org/10.1080/13501763.2012.699661.

---. 2013b. *Lobbying in the European Union: interest groups, lobbying coalitions, and policy change*. First Edition ed.

Klüver, Heike, and Elisabeth Zeidler. 2019. "Explaining Interest Group Density Across Economic Sectors: Evidence from Germany." *Political Studies* 67 (2): 459-478. https://doi.org/10.1177/0032321718774685.

Kohler-Koch, Beate. 2010. "Civil society and EU democracy: 'astroturf' representation?" *Journal of European Public Policy* 17 (1): 100-116. https://doi.org/10.1080/13501760903464986.

Larsson, Bengt. 2015. "Trade Union Channels for Influencing European Union Policies." *Nordic Journal of Working Life Studies* 5 (3). https://doi.org/10.19154/njwls.v5i3.4809.

Long, James, and Paul Teetor. 2019. *R Cookbook*. In *Proven Recipes for Data Analysis, Statistics & Graphics*https://rc2e.com/index.html.

Lowery, David, Frank R. Baumgartner, Joost Berkhout, Jeffrey M. Berry, Darren Halpin, Marie Hojnacki, Heike Klüver, Beate Kohler-Koch, Jeremy Richardson, and Kay Lehman Schlozman. 2015. "Images of an Unbiased Interest System." *Journal of European Public Policy* 22 (8): 1212-1231. https://doi.org/10.1080/13501763.2015.1049197.

Löfgren, Karl, and Kennet Lynggaard. 2015. "Assessing the EU's Transatlantic Regulatory Powers Using the Choice of Policy Instruments as Measurement of Preference Attainment." *International Journal of Public Administration* 38 (12): 915-925. https://doi.org/10.1080/01900692.2014.982291.

Majone, Giandomenico. 2002. "Delegation of Regulatory Powers in a Mixed Polity." *European Law Journal* 8 (3): 319-339. https://doi.org/ https://doi.org/10.1111/1468-0386.00156.

Messer, Anne, Joost Berkhout, and David Lowery. 2010. "The Density of the EU Interest System: A Test of the ESA Model." *British Journal of Political Science* 41 (1): 161-190. https://doi.org/10.1017/s0007123409990329.

Mingers, John, and John Brocklesby. 1997. "Multimethodology: Towards a framework for mixing methodologies." *Omega (Oxford)* 25 (5): 489-509. https://doi.org/10.1016/S0305-0483(97)00018-2.

Moses, Jonathon Wayne, and Torbj□rn L. Knutsen. 2012. *Ways of knowing : competing methodologies in social and political research*. New York, New York: Palgrave Macmillan.

Pagliari, Stefano, and Kevin Young. 2016. "The interest ecology of financial regulation: interest group plurality in the design of financial regulatory policies." *Socio-Economic Review* 14 (2): 309-337. https://doi.org/10.1093/ser/mwv024.

Pagliari, Stefano, and Kevin L. Young. 2013. "Leveraged interests: Financial industry power and the role of private sector coalitions." *Review of International Political Economy* 21 (3): 575-610. https://doi.org/10.1080/09692290.2013.819811.

Pakull, Dominic, David Marshall, and Patrick Bernhagen. 2020. "Shop till you drop? Venue choices of business and non-business interests in the European Union." *Interest Groups & Advocacy* 9 (4): 520-540. https://doi.org/10.1057/s41309-020-00092-y.

Persson, Thomas. 2007. "Democratizing European Chemicals Policy: Do Consultations Favour Civil Society Participation?" *Journal of Civil Society* 3 (3): 223-238. https://doi.org/10.1080/17448680701775648.

Pijnenburg, Bert. 1998. "EU lobbying by ad hoc coalitions: an exploratory case study." *Journal of European Public Policy* 5 (2): 303-321. https://doi.org/10.1080/135017698344009.

Pisano, Umberto, Lisa K. Lange, Katrin Lepuschitz, and Gerald Berger. 2015. *The role of stakeholder participation in European sustainable development policies and strategies.* https://www.esdn.eu/fileadmin/ESDN_Reports/2015-December-The_role_of_stakeholder_participation_in_European_sustainable_development_policies_and_strategies.pdf.

Quittkat, Christine. 2011. "The European Commission's Online Consultations: A Success Story?: THE EUROPEAN COMMISSION'S ONLINE CONSULTATIONS." *Journal of common market studies* 49 (3): 653-674. https://doi.org/10.1111/j.1468-5965.2010.02147.x.

Ragin, Charles C. 1989. *Comparative Method : Moving Beyond Qualitative and Quantitative Strategies*. Berkeley: University of California Press.

Rasmussen, Anne, and Brendan J. Carroll. 2013. "Determinants of Upper-Class Dominance in the Heavenly Chorus: Lessons from European Union Online Consultations." *British Journal of Political Science* 44 (2): 445-459. https://doi.org/10.1017/s0007123412000750.

Rauh, Christian. 2021. "One agenda-setter or many? The varying success of policy initiatives by individual Directorates-General of the European Commission 1994–2016."

European Union Politics 22 (1): 3-24. https://doi.org/10.1177/1465116520961467. https://journals.sagepub.com/doi/abs/10.1177/1465116520961467.

Reynolds, Alan P., Graeme Richards, and Vic J. Rayward-Smith. 2004. "The Application of K-Medoids and PAM to the Clustering of Rules." Berlin, Heidelberg.

Rodrigo, Delia, and Pedro Andrés Amo, 2006, "Background document on public consultation," https://www.oecd.org/mena/governance/36785341.pdf.

Rousseeuw, Peter J. 1987. "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis." *Journal of Computational and Applied Mathematics* 20: 53-65. https://doi.org/https://doi.org/10.1016/0377-0427(87)90125-7.

Scharpf, Fritz W. 2006. "The Joint-Decision Trap Revisited." *Journal of Common Market Studies* 44 (4): 845-864. https://doi.org/https://doi.org/10.1111/j.1468-5965.2006.00665.x.

Stirling, A. 2006. "Analysis, participation and power: justification and closure in participatory multi-criteria analysis." *Land Use Policy* 23 (1): 95-107. https://doi.org/10.1016/j.landusepol.2004.08.010.

---. 2008. ""Opening up" and "Closing down": Power, Participation, and Pluralism in the Social Appraisal of Technology." *Science, Technology, & Human Values* 33 (2): 262-294. http://www.jstor.org/stable/29734034.

Thomson, Robert, and Madeleine Hosli. 2006. "Who Has Power in the EU? The Commission, Council and Parliament in Legislative Decision-making." *Journal of common market studies* 44 (2): 391-417. https://doi.org/10.1111/j.1468-5965.2006.00628.x.

Todorova, Tatyana. 2020. "Quantitative analysis on interest group influence using data from EU open consultations." *VUZF Review* 5 (2): 48-54. https://doi.org/10.38188/2534-9228.20.2.05.

Toth, Gergely. 2010. "The replacement of the Neumann trend test and the Durbin-Watson test on residuals by one-way ANOVA with resampling and an extension of the tests to different time lags." *J. Chemometrics* 24 (3-4): 140-148. https://doi.org/10.1002/cem.1293.

Van Ballaert, Bart. 2017. "The European Commission's use of consultation during policy formulation: The effects of policy characteristics." *European Union Politics* 18 (3): 406-423. https://doi.org/10.1177/1465116517702004.

Van Hecke, Matti, Peter Bursens, and Jan Beyers. 2016. "You'll Never Lobby Alone. Explaining the Participation of Sub-national Authorities in the European Commission's

Open Consultations." *JCMS: Journal of Common Market Studies* 54 (6): 1433-1448. https://doi.org/10.1111/jcms.12425.

Xu, Rui, and Don Wunsch. 2008. *Clustering*. 1. Aufl. ed. Vol. 10.*IEEE Press Series on Computational Intelligence*. Hoboken: Hoboken: Wiley-IEEE Press.

Yackee, Susan Webb. 2015. "Participant Voice in the Bureaucratic Policymaking Process." *Journal of Public Administration Research and Theory: J-PART* 25 (2): 427-449. https://doi.org/10.1093/jopart/muu007.

Yang, Shengping , and Gilbert Berdine. 2015. "The Negative Binomial regression." *The Southwest respiratory and critical care chronicles* 3 (10): 50-54. https://doi.org/10.12746/swrccc2015.0310.135.

Appendix A.

Detailed overview of policy issue drafts encountered.
Note: Both included and excluded.

Proposal ID (CELEX number)	Number of Response Files	DG responsible	Included in the Dataset	ID Name in Thesis
52017PC0253	1	EMPL	Yes	Celex 1
52017PC0257	6	GROW	Yes	Celex 2-7
52017PC0275	1	MOVE	No	-
52017PC0276	1	MOVE	Yes	Celex 8
52017PC0277	2	MOVE	Yes	Celex 9
52017PC0278	2	MOVE	Yes	Celex 10-11
52017PC0281	1	MOVE	Yes	Celex 12
52017PC0282	1	MOVE	Yes	Celex 13
52017PC0289	1	MOVE	Yes	-
52017PC0489	2	HOME	No	Celex 14-15
52017PC0548	2	MOVE	Yes	Celex 16-17
52017PC0647	1	MOVE	Yes	Celex 18
52017PC0753	1	ENV	Yes	Celex 19
52017PC0793	1	HOME	Yes	Celex 20
52017PC0794	1	HOME	Yes	Celex 21
52017PC0797	1	EMPL	Yes	Celex 22
52018PC0033	1	MOVE	Yes	Celex 23
52018PC0131	3	EMPL	Yes	Celex 24-26
52018PC0209	1	HOME	No	-
52018PC0213	3	HOME	Yes	Celex 27-29
52018PC0252	1	HOME	No	-
52018PC0274	1	MOVE	Yes	Celex 30
52018PC0277	2	MOVE	Yes	Celex 31-32
52018PC0278	1	MOVE	Yes	Celex 33
52018PC0279	1	MOVE	Yes	Celex 34
52018PC0284	1	CLIMA	Yes	Celex 35
52018PC0302	2	HOME	Yes	Celex 36-37
52018PC0317	2	GROW	No	Celex 38-39
52018PC0381	1	ENV	Yes	Celex 40
52018PC0382	6	EMPL	Yes	Celex 41-45
52018PC0438	1	MOVE	Yes	Celex 46
52018PC0441	1	GROW	Yes	Celex 47
52018PC0471	2	HOME	No	-
52018PC0472	2	HOME	No	-

52018PC0473	2	HOME	No	-
52018PC0639	117	MOVE	No	-
52018PC0640	3	HOME	Yes	Celex 48-51
52020PC0080	1	CLIMA	Yes	Celex 52
52020PC0798	1	ENV	Yes	Celex 53
52020PC0824	1	ENER	Yes	Celex 54
52020PC0829	1	HOME	No	-
52021PC0096	1	HOME	Yes	Celex 55
52020PC0202	1	GROW	Yes	Celex 56

Appendix B.

Overview of Question Labels.

Label	Description
info	Items collecting background information on the responding stakeholder
know	Items collecting information on the responding stakeholder's knowledge of or familiarity with a particular issue or instrument
eval	Items collecting information on the responding stakeholder's assessment or evaluation of existing policies
eval_info	Items collecting information on the status quo, indirectly providing an evaluation of existing policies
alter	Items collecting information on the responding stakeholder's opinion on or preferences over new policy alternatives
alter_eval	Items collecting information on the responding stakeholder's evaluation of the impacts of new policy alternatives or on future implementations
other	Other items

Appendix C.

VIF test results.

Note: Including all variables for both the negative binomial regression and the multiple linear regression.

C.1 Results from explanatory variables used in negative binomial regression (NB2)

Variables	VIF values	Df	GVIF^(1/(2*Df))
Salience	1.717888	1	1.310682
Targeted stk.	1.623313	1	1.274093
Survey C: mean length of words used in questions	1.193650	1	1.092543
Survey C: number of questions	1.371865	1	1.171267
Issue complexity	1.485440	1	1.218786
Proposal status	1.998812	1	1.413793
Policy area	3.821639	6	1.118203

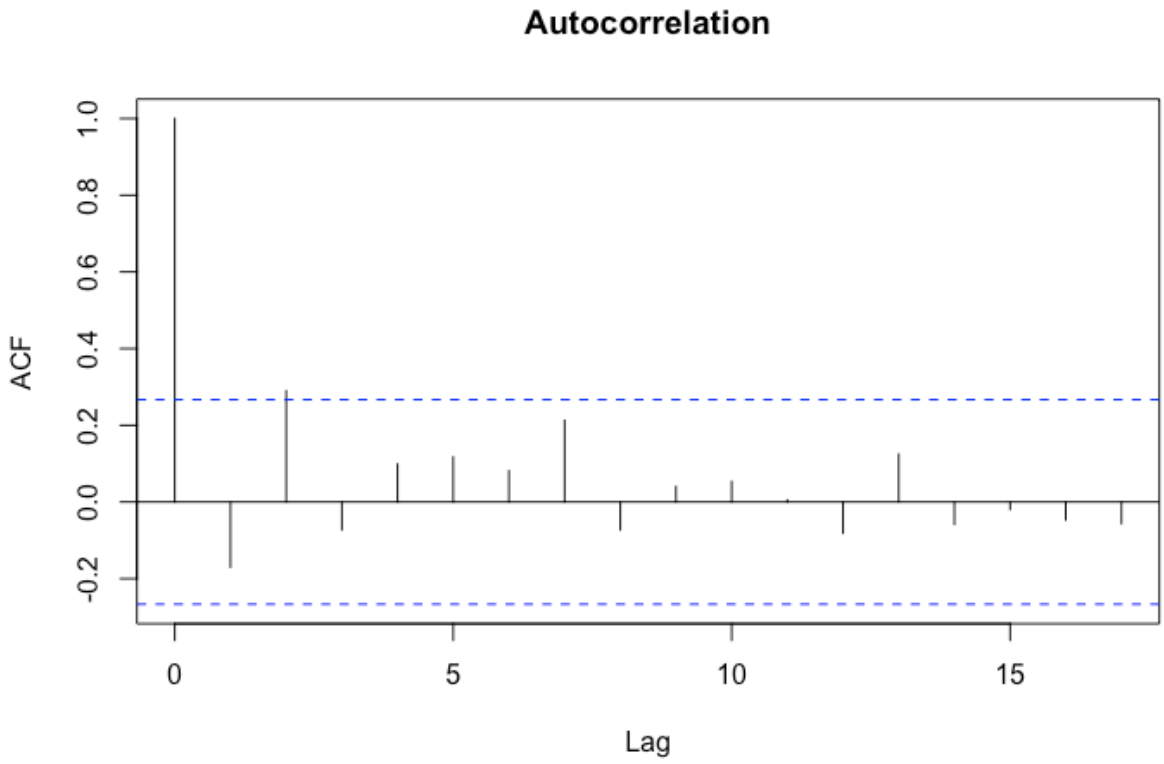
C.2 Results from explanatory variables used in multiple linear regression

Variables	VIF values	Df	GVIF^(1/(2*Df))
Salience	1.519601	1	1.232721
Targeted stk.	1.590172	1	1.261020
Survey C: mean length of words used in questions	1.159249	1	1.076684
Survey C: number of questions	1.304343	1	1.142079
Issue complexity	1.522077	1	1.233725
Proposal status	1.730535	1	1.315498
Policy area	3.428716	6	1.108139

Appendix D.

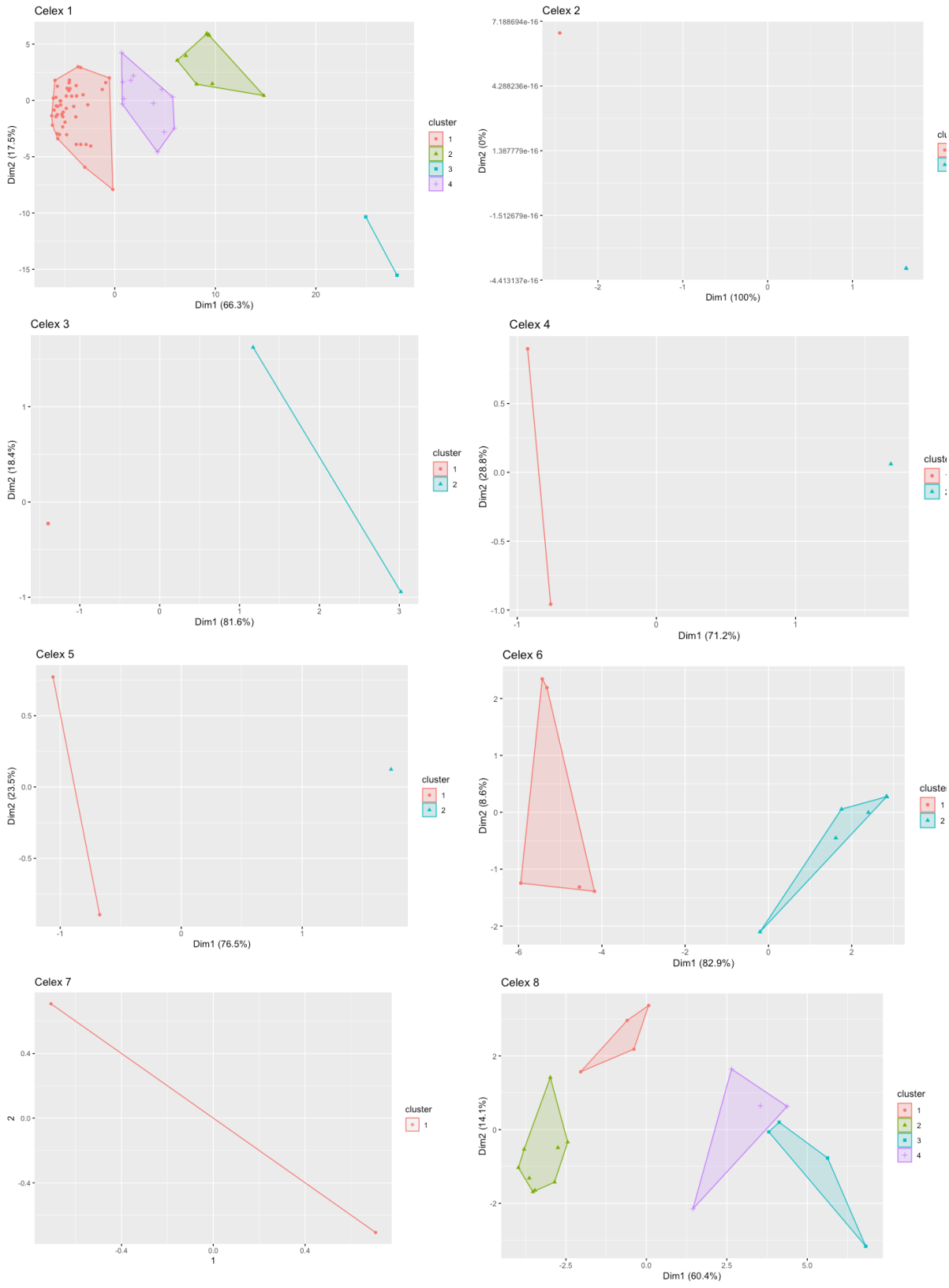
ACF test for multiple linear regression.

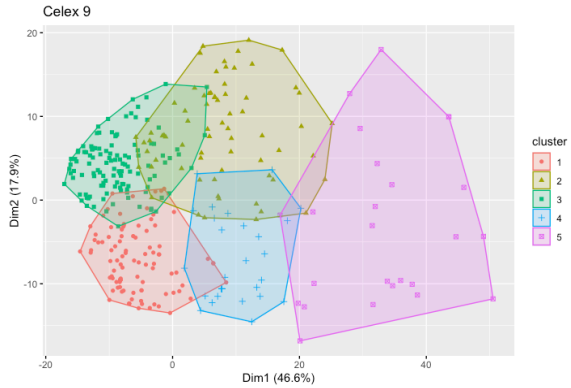
Note: ACF results include all variables.

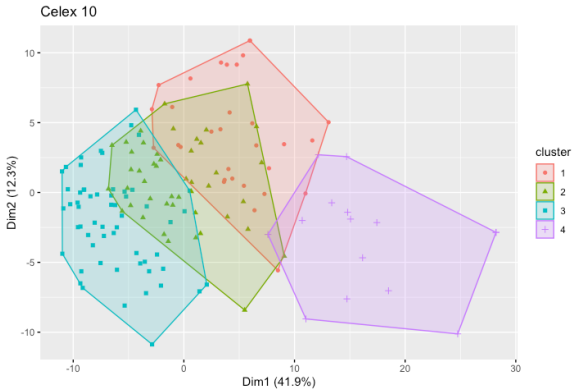


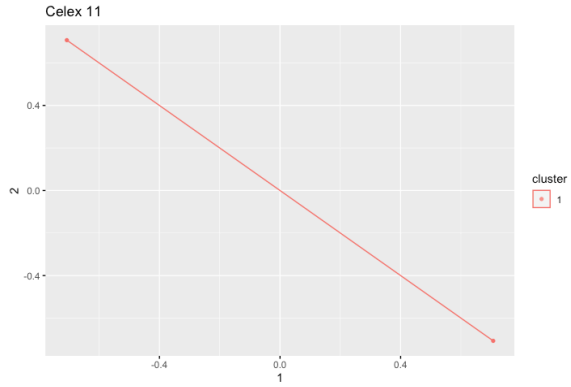
Appendix E.

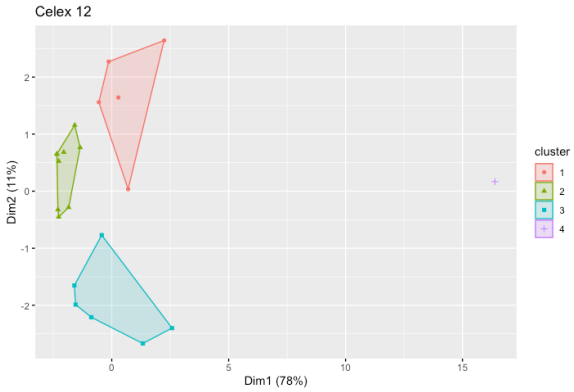
Visualization of clusters within all Surveys.

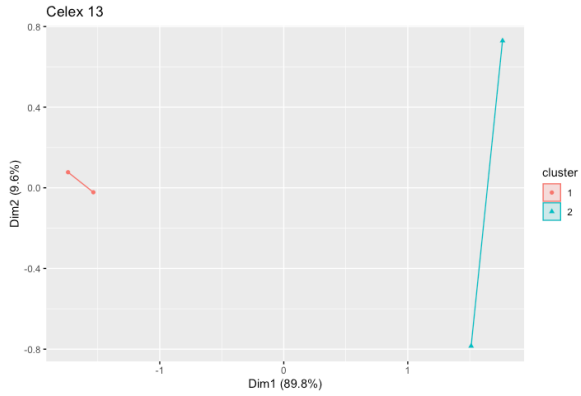


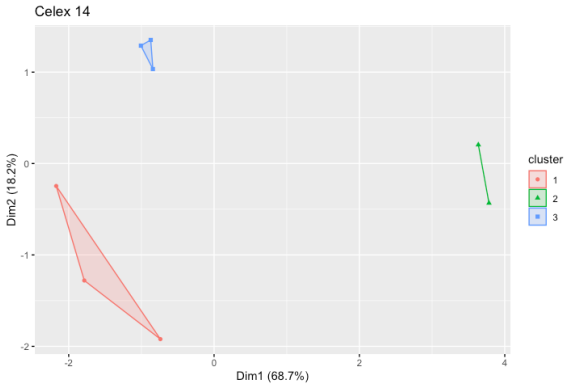

Celex 9
Dim2 (17.9%)
Dim1 (46.6%)
cluster
1
2
3
4
5

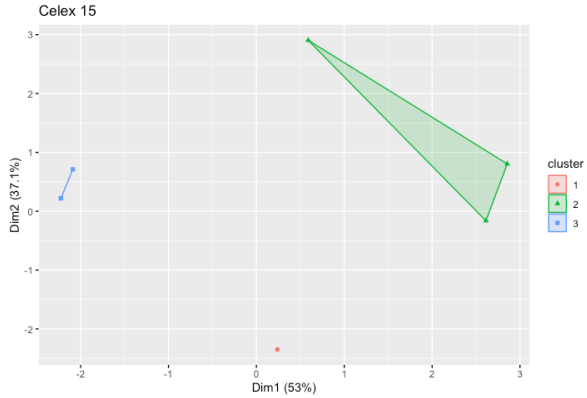

Celex 10
Dim2 (12.3%)
Dim1 (41.9%)
cluster
1
2
3
4

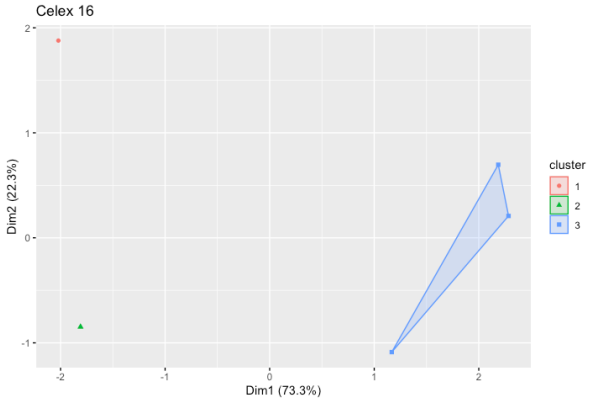

Celex 11
2
1
cluster
1

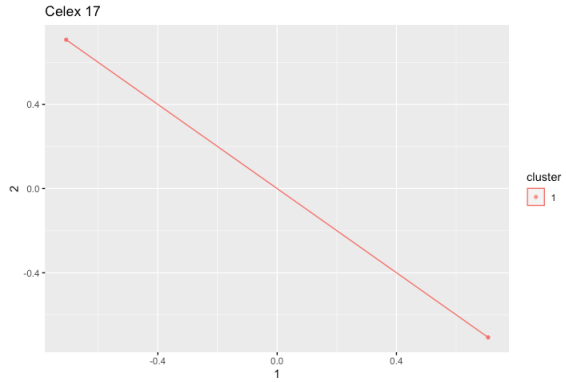

Celex 12
Dim2 (11%)
Dim1 (78%)
cluster
1
2
3
4

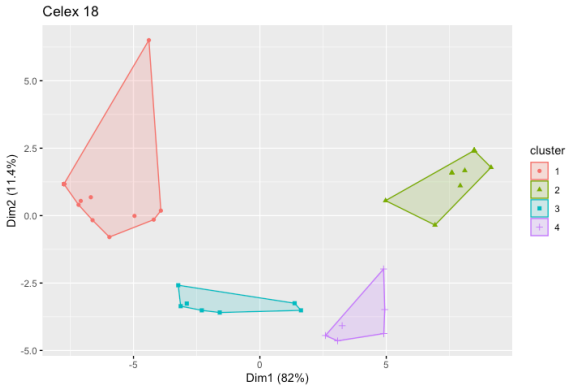

Celex 13
Dim2 (9.6%)
Dim1 (89.8%)
cluster
1
2

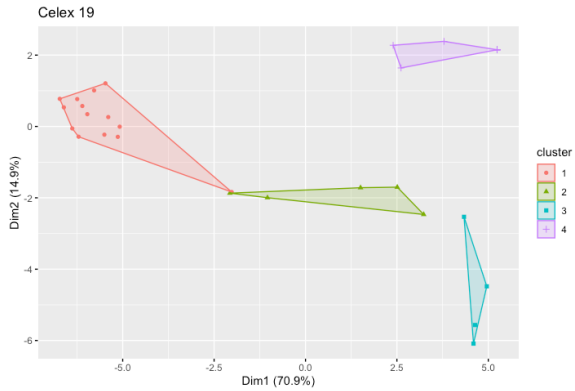

Celex 14
Dim2 (18.2%)
Dim1 (68.7%)
cluster
1
2
3

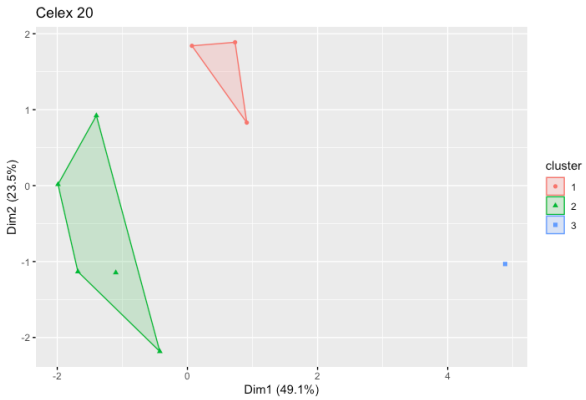

Celex 15
Dim2 (37.1%)
Dim1 (53%)
cluster
1
2
3

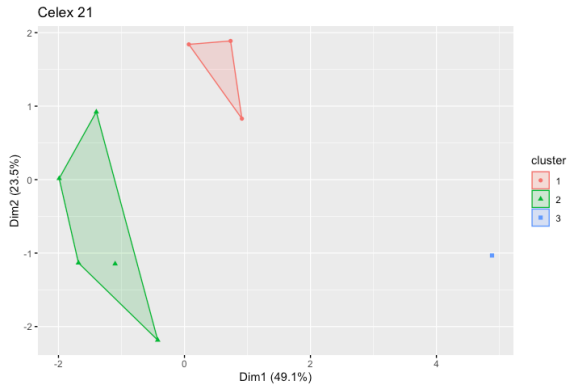

Celex 16
Dim2 (22.3%)
Dim1 (73.3%)
cluster
1
2
3

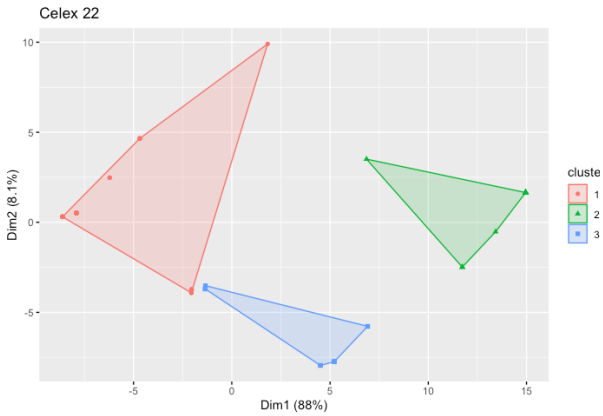

Celex 17
cluster
1
2
1

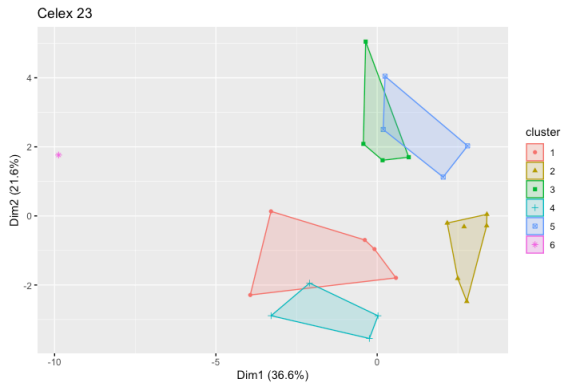

Celex 18
cluster
1
2
3
4
Dim2 (11.4%)
Dim1 (82%)

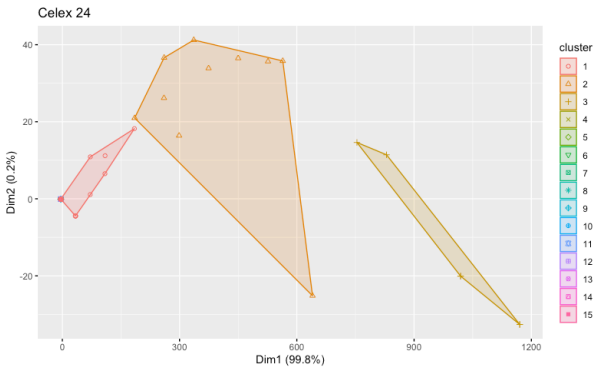

Celex 19
cluster
1
2
3
4
Dim2 (14.9%)
Dim1 (70.9%)

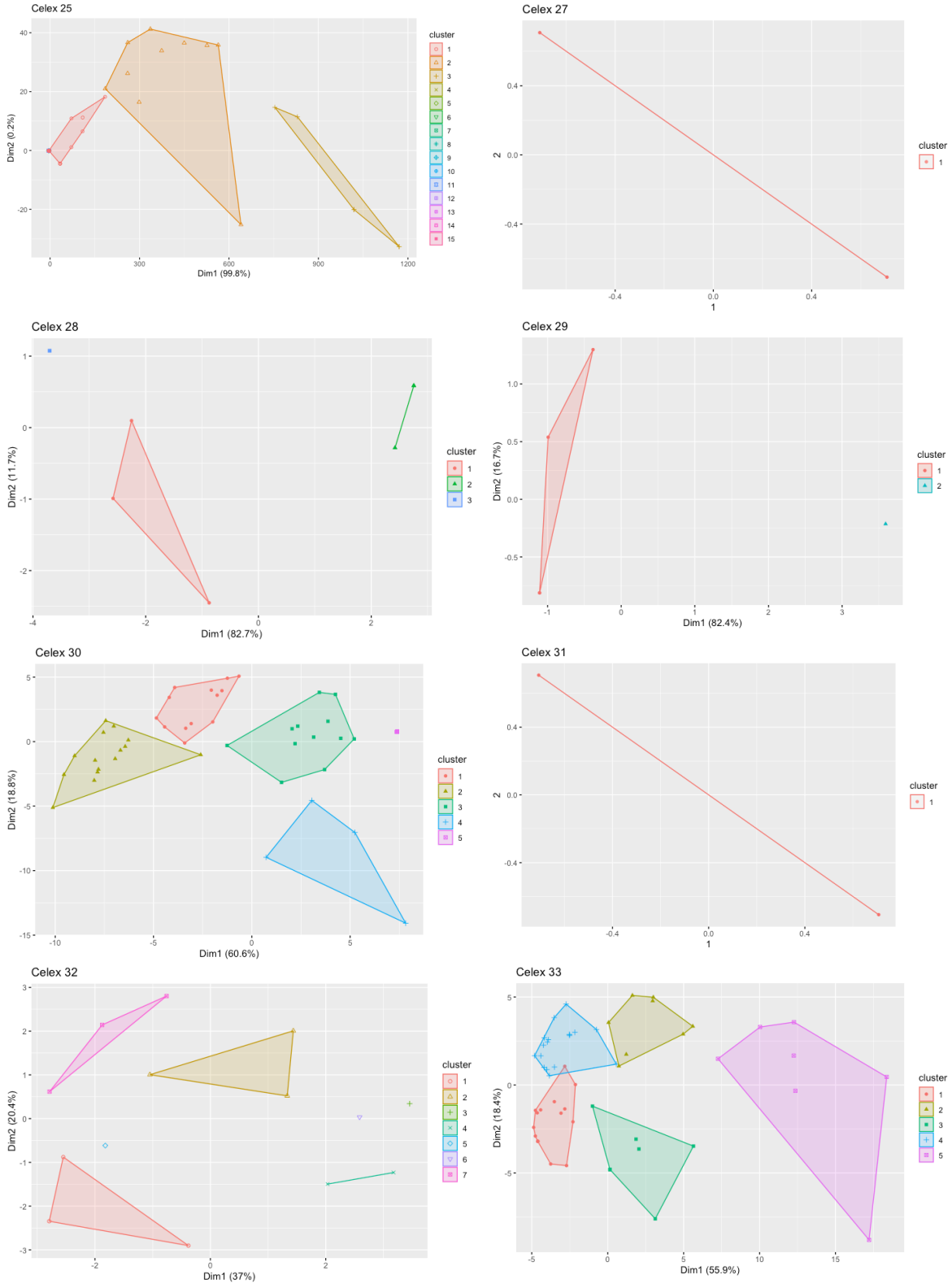

Celex 20
cluster
1
2
3
Dim2 (23.5%)
Dim1 (49.1%)

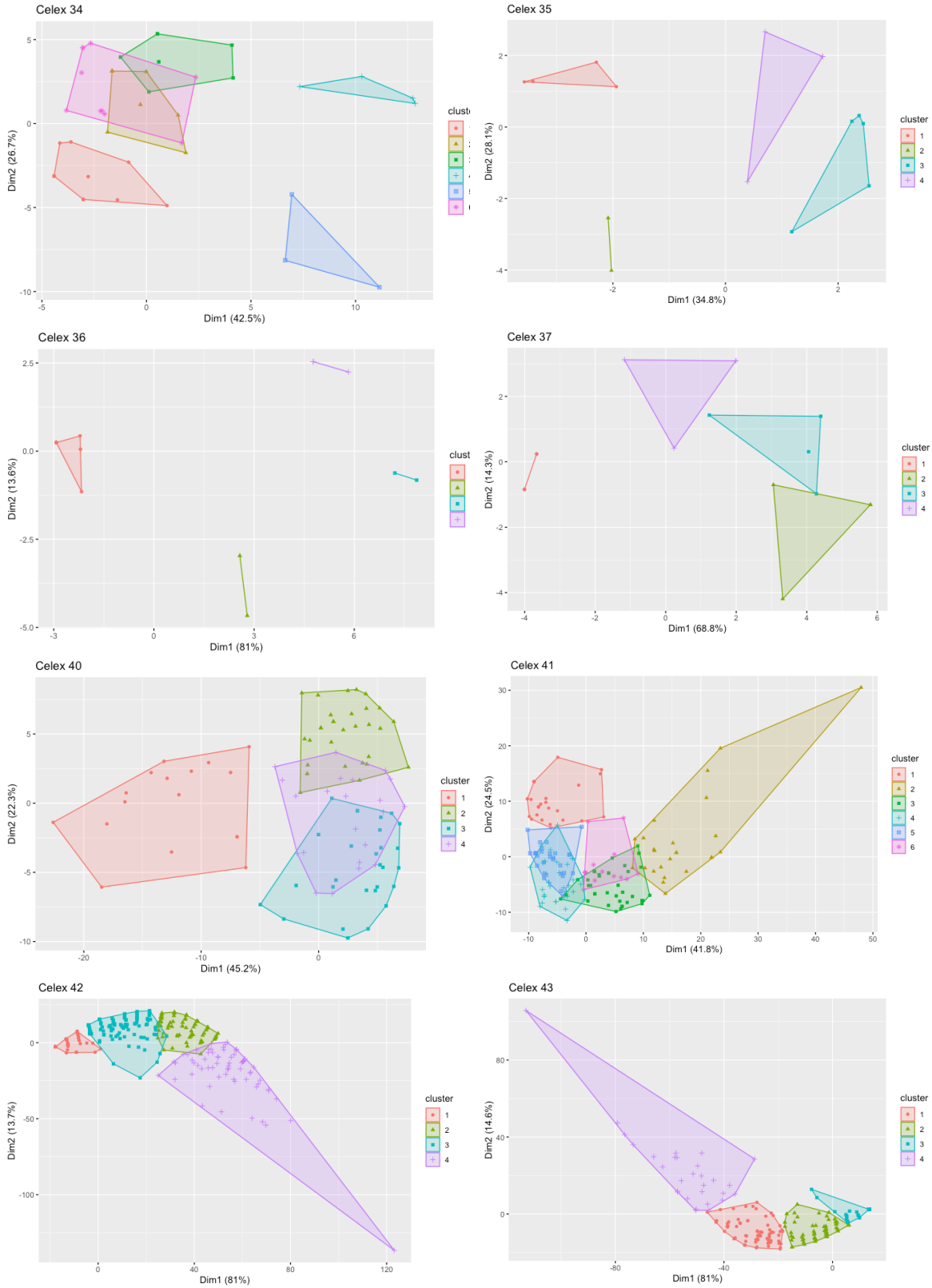

Celex 21
cluster
1
2
3
Dim2 (23.5%)
Dim1 (49.1%)

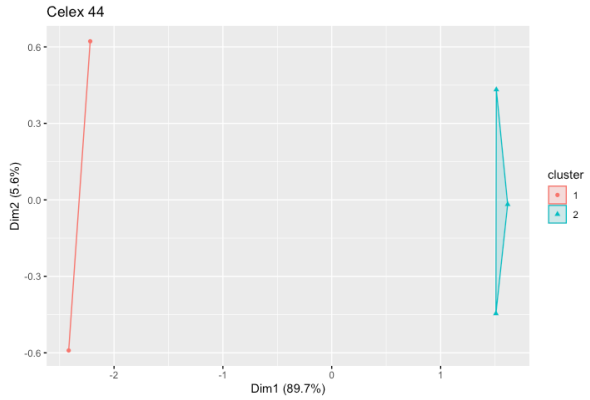

Celex 22
cluste
1
2
3
Dim2 (8.1%)
Dim1 (88%)

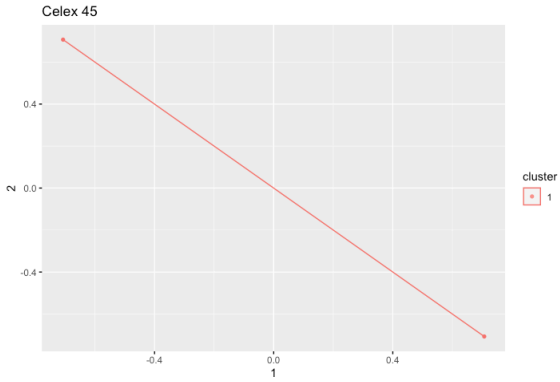

Celex 23
cluster
1
2
3
4
5
6
Dim2 (21.6%)
Dim1 (36.6%)

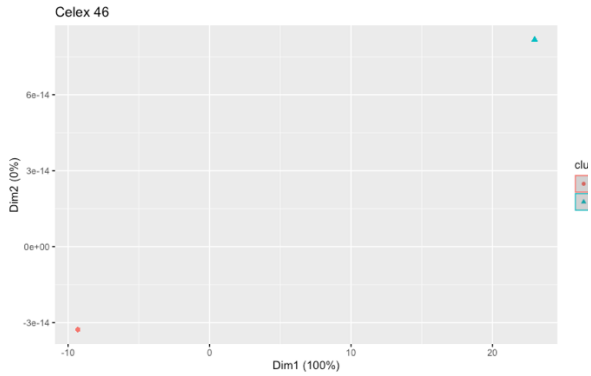

Celex 24
cluster
Dim2 (0.2%)
Dim1 (99.8%)

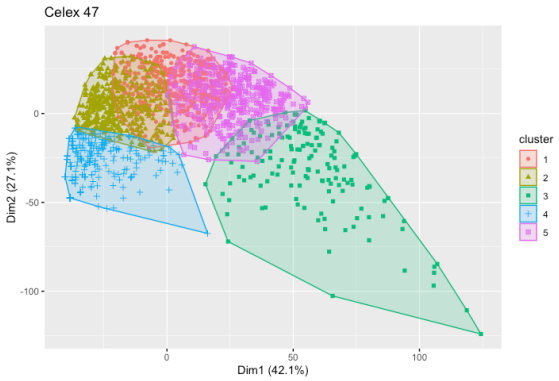

Celex 25
Dim2 (0.2%)
Dim1 (99.8%)
cluster
Celex 27
2
1
cluster
Celex 28
Dim2 (11.7%)
Dim1 (82.7%)
cluster
Celex 29
Dim2 (16.7%)
Dim1 (82.4%)
cluster
Celex 30
Dim2 (18.8%)
Dim1 (60.6%)
cluster
Celex 31
2
1
cluster
Celex 32
Dim2 (20.4%)
Dim1 (37%)
cluster
Celex 33
Dim2 (18.4%)
Dim1 (55.9%)
cluster

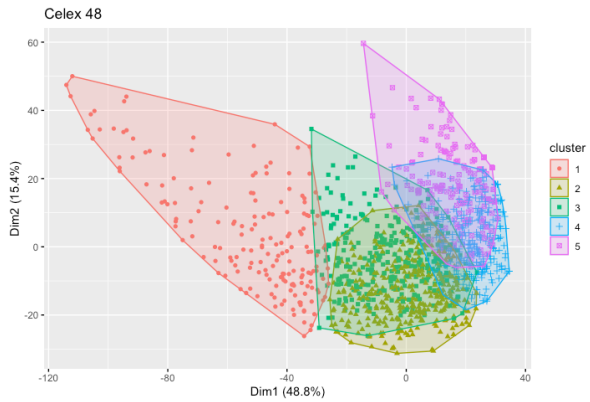

Celex 34
Dim1 (42.5%)
Dim2 (26.7%)
Celex 35
Dim1 (34.8%)
Dim2 (28.1%)
cluster
Celex 36
Dim1 (81%)
Dim2 (13.6%)
Celex 37
Dim1 (68.8%)
Dim2 (14.3%)
Celex 40
Dim1 (45.2%)
Dim2 (22.3%)
Celex 41
Dim1 (41.8%)
Dim2 (24.5%)
Celex 42
Dim1 (81%)
Dim2 (13.7%)
Celex 43
Dim1 (81%)
Dim2 (14.6%)

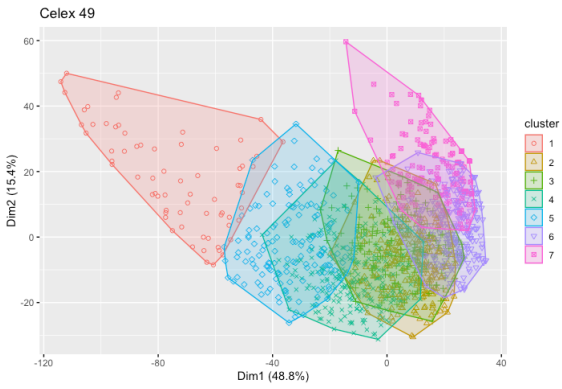

Celex 44
Dim2 (5.6%)
Dim1 (89.7%)
cluster
1
2

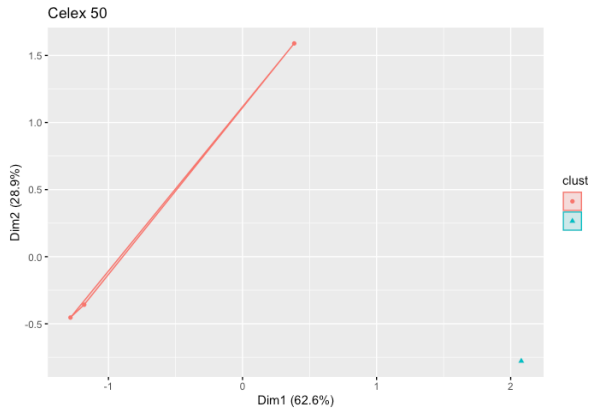

Celex 45
cluster
1

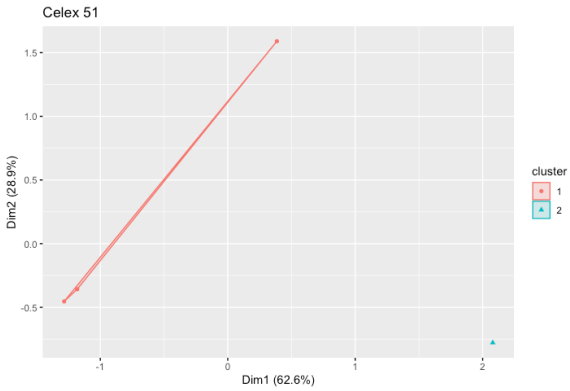

Celex 46
Dim2 (0%)
Dim1 (100%)

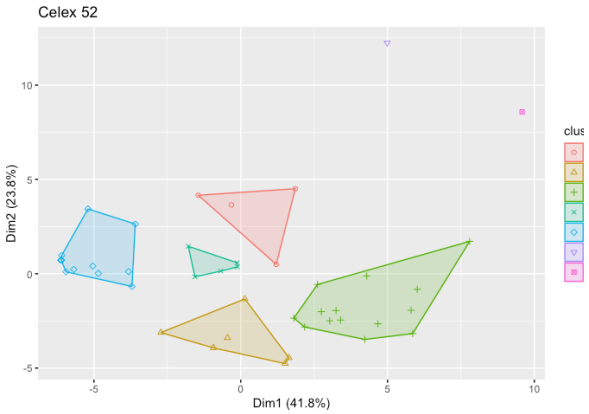

Celex 47
Dim2 (27.1%)
Dim1 (42.1%)
cluster
1
2
3
4
5

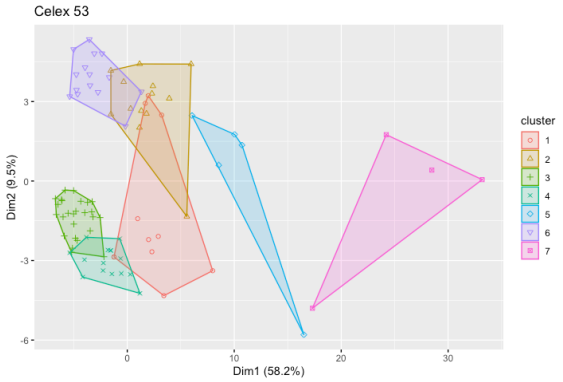

Celex 48
Dim2 (15.4%)
Dim1 (48.8%)
cluster
1
2
3
4
5

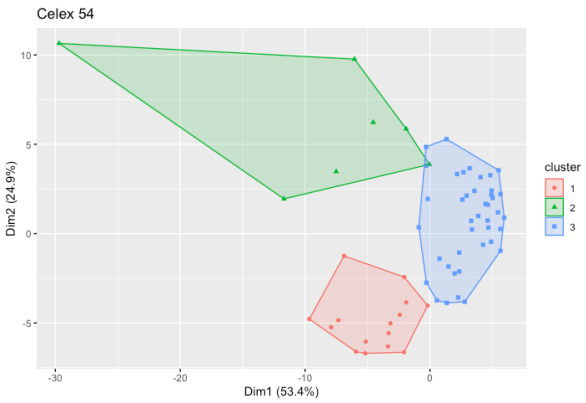

Celex 49
Dim2 (15.4%)
Dim1 (48.8%)
cluster
1
2
3
4
5
6
7


Celex 50
Dim2 (28.9%)
Dim1 (62.6%)

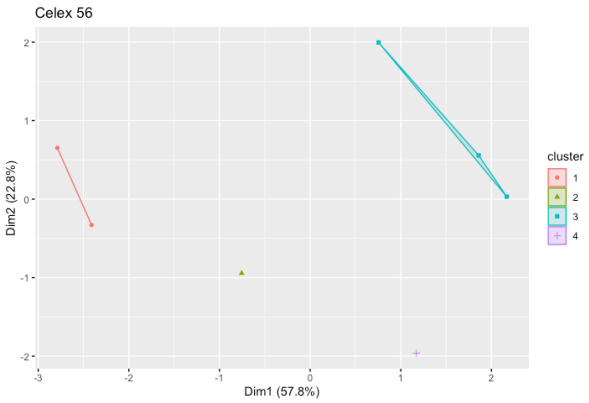

Celex 51
Dim2 (28.9%)
Dim1 (62.6%)
cluster
1
2


Celex 52
Dim2 (23.8%)
Dim1 (41.8%)
clus


Celex 53
Dim2 (9.5%)
Dim1 (58.2%)
cluster
1
2
3
4
5
6
7


Celex 54
Dim2 (24.9%)
Dim1 (53.4%)
cluster
1
2
3


Celex 55
Dim2 (15.1%)
Dim1 (66.8%)
cluster
1
2
3
4
5


Celex 56
Dim2 (22.8%)
Dim1 (57.8%)
cluster
1
2
3
4