Christian Bruns

mass media elections and public policies

Cuvillier Verlag Göttingen
Internationaler wissenschaftlicher Fachverlag

Mass Media, Elections, and Public Policies

Dissertation
zur Erlangung des wirtschaftswissenschaftlichen Doktorgrades
der Wirtschaftswissenschaftlichen Fakultät der
Universität Göttingen

vorgelegt von
Christian Bruns
aus
Bückeburg

Göttingen, 2011

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet über http://dnb.d-nb.de abrufbar.
1. Aufl. - Göttingen: Cuvillier, 2011
Zugl.: (TU) Göttingen, Univ., Diss., 2009

978-3-86955-887-5

Erstgutachter: Prof. Dr. Robert Schwager
Zweitgutachter: Prof. Dr. Andreas Wagener
Tag der mündlichen Prüfung: 8. August 2011

© CUVILLIER VERLAG, Göttingen 2011
Nonnenstieg 8, 37075 Göttingen
Telefon: 0551-54724-0
Telefax: 0551-54724-21
www.cuvillier.de

Alle Rechte vorbehalten. Ohne ausdrückliche Genehmigung des Verlages ist es nicht gestattet, das Buch oder Teile daraus auf fotomechanischem Weg (Fotokopie, Mikrokopie) zu vervielfältigen.
1. Auflage 2011
Gedruckt auf säurefreiem Papier.

978-3-86955-887-5

Contents

List of Tables

List of Abbreviations

ABC	American Broadcasting Company
CBS	CBS Broadcasting Inc. (formerly: Columbia Broadcasting System)
CFFR	Consolidated Federal Funds Report
CPB	Corporation for Public Broadcasting
DMA	Designated Market Area
FCC	Federal Communications Commission
FOX	Fox Broadcasting Company
GS2SLS	Generalized Spatial Two Stage Least Squares
IV	Instrumental Variable
LLA	Landslaget for Lokalaviser
MBL	Mediebedriftenes Landsforening
NBC	National Broadcasting Company
NSD	Norsk Samfunnsvitenskapelig Datatjeneste
OLS	Ordinary Least Squares
PBS	Public Broadcasting Service
TV	Television

1 Mass media and political economics

1.1 Introduction

Mass media have been providing people with information about the men and women who govern them for decades and centuries even. One of the earliest examples was in the 16th century when the printing press served to propagate the ideas of the Reformation. Among other things, the reformers argued against the selfish behavior of bishops, priests, monks and other clerics who often rather strove to live a pleasant life instead of attending to their duties (Barzun 2001, pp. 4-11). Later, the printing press provided the basis for the establishment of a mass medium that has ever since been considered to play a vital role in holding government accountable: the newspaper. Thomas Jefferson, for example, wrote in the late 18th century:

> "...were it left to me to decide whether we should have a government without newspapers, or newspapers without a government, I should not hesitate a moment to prefer the latter."
>
> -Thomas Jefferson to Edward Carrington, 1787. Papers 11:48-49

More recently, political scientist Doris A. Graber (2001, p.1) wrote that it would be "as American as apple pie to regard the press as the chief tool of public political enlightenment" and the results from a survey by the Pew Internet & American Life Project (Horrigan 2004) confirm that the modern press is the most important provider of political information: When American voters were asked about their dominant sources of political news and information, television was on top with 74% of respondents calling it a primary source, followed by radio (54%) and newspapers (51%)[1].

This view of the mass media assumes that media coverage shapes both what people think about their politicians and how the latter act in office. Since politicians control billions of euros and dollars and make decisions that influence everybody's life it appears to be important to understand the role of mass media properly. Economists, however, had little to say about the relation between mass media and government until recently.

Although Downs (1957) already suggested the importance of information provided by the press and later studies analyzed how a given distribution or quality of information influences political and/or economic outcomes (see Persson and Tabellini (2000), Besley (2006) or Mueller (2003) for surveys) there were basically no rigorous theoretical or empirical analyzes

[1]The numbers add to more than 100 due to multiple responses.

that strived to explain the distribution of information in the electorate or its quality and to test the theoretical results empirically. Only within the last decade, there has emerged a literature that studies how the media market influences the informational situation of the electorate. Some of the most influential contributions deal with the topics of informative coverage and policy outcomes (Strömberg 2004a,b), ideological bias of news reports (Mullainathan and Shlaifer 2005) and capture of the press (Besley and Prat 2006) and they provided the bedrock for a now burgeoning literature which is briefly reviewed below.

This literature originates from a political economy perspective on government that regards politicians as self-interested individuals who also pursue their own goals instead of serving a well-defined public interest only, as the traditional public finance approach assumes. A self-interested elected politician may, for example, exploit public funds at his disposal, accept bribes from interest groups and bias policies towards their interest, implement policies according to his own preferences or simply enjoy his status instead of working hard. In principle, elections can serve to limit the extent of selfish behavior of incumbent politicians since they enable voters to kick bad politicians out of office. However, voters usually are not perfectly informed about the characteristics and actions of their politicians. Incumbent behavior then depends on the ability of the electorate to evaluate the incumbent where a more informed electorate should be better able to hold him accountable.

Elections, however, may not only help to ensure accountability but they can also influence incumbent behavior in a different way: When it comes to distributing funds among voters, a politician who seeks reelection has incentives to allocate more funds to informed voters, who consider this benefit when they cast the ballot, in contrast to ignorant voters, who are not aware of their benefactor. So elections may not only discipline politicians but they can also constitute an origin of biased policies which create winners and losers among voters. Then, it is interesting to study how market forces determine which voters are on the winning side because they get informed by the media.

These considerations illustrate that it is worthwhile and interesting to study how the media market determines the quality of information voters rely on or the distribution of information within the electorate and also how these aspects affect incumbent behavior. In the next section we describe how this book contributes to understand the relation between voters, politicians and mass media.

1.2 Contribution of this book

This book contains purely theoretical analyzes as well as empirical results all of which are concerned with the role of modern mass media in political economics. In chapters 2 and 3 we utilize tools from game theory and the economics of information to characterize the informational situation that arises when the electorate wants to obtain information in order to choose between political candidates or policies.

In chapter 2 we use these techniques to investigate a common value election where citizens can obtain additional information from the mass media. The citizens can choose between two alternatives (politicians or policies, for example) whose qualities determine their future well-being and they are uncertain about which alternative is best. When media coverage is interpreted as a public signal to citizens, the impact of newly produced information on an electoral decision is non-rival and non-excludable in nature. Thus, the quality of media coverage is a public good and we show that subscriptions to newspapers can serve as a vehicle to contribute to this public good. Our results suggest that a competitive media market provides information of inefficiently low quality because the citizens have incentives to free-ride and therefore electoral decisions are inefficiently risky. Further, we find that the gap between the optimal and the market quality of information increases in the size of the electorate and that the citizens generally benefit from a more efficient technology to produce information. Besides shedding light on these problems a novel technical feature of the analysis is the use of Bayesian sampling theory to model the production of information.

We extend this framework in chapter 3 to analyze how media activity affects incumbent behavior when the electorate uses elections to select competent politicians. In contrast to the analysis in chapter 2, incumbent politicians are active players who can try to manipulate the electorate's belief about their competence. When voters rely on policy outcomes to infer an incumbent's competence, the latter can influence the policy outcomes by choosing a level of rent-seeking or effort, where less rent-seeking or more effort imply better policy results. From the incumbent's perspective, the effectiveness of this mechanism and thus his choice of rent or effort depends on the voters' ability to monitor policies.

We derive the demand for information of the electorate in this decision problem and characterize the equilibrium level of information the media market provides. This enables us to describe how features of the media market shape the incentives of incumbents. We find that low costs of information production and a technology that provides more precise information for a given amount of resources spent on investigations imply a better informed electorate which results in less rent-seeking. The effect of better prior knowledge about the incumbent on rent-seeking, however, is found to be ambiguous. Further, we find that the electorate does not suffer from reduced information provision when a competitive market is replaced by a monopoly. The monopolist exerts his power to charge a higher price, but it provides the same amount of information as a competitive market. Thus, the incentives of politicians are identical under both market regimes.

After these purely theoretical considerations, we turn to a more applied perspective in the two following chapters. These applied analyzes aim at studying whether popular information services actually in use, such as newspapers and local television programs, influence policy outcomes. The empirical results in these chapters are not meant as a direct test of the theoretical results in the chapters 2 and 3, but they strive to examine in general whether more informed voters receive favorable policies.

In chapter 4, we examine whether the circulation of local newspapers affects public sector efficiency in Norwegian municipalities. In a theoretical model, the share of informed voters, who can perfectly monitor the performance of a politician, originates from two factors: the share of newspaper readers and the amount of news space that the newspapers assign to the politician in a constituency. We take both of these factors as given and show how they shape the incentives of the politician to invest effort on public good production. The model predicts that a larger share of informed voters induces the incumbent to exert more effort.

We use panel data on Norwegian municipalities for the years 2001-2005 to test the prediction of the model. Our main variables are an index of public sector efficiency in Norwegian municipalities as introduced by Borge et al. (2008) and two measures of voter information that are based on data on the circulation of some 150 newspapers in Norwegian municipalities. The results support the prediction of the model that a more informed electorate is associated with higher levels of local government efficiency.

Finally, in chapter 5, we study whether the geography of television markets in the United States affects the allocation of federal funds. Media scholars have long been familiar with the fact that the cost to produce a report from a place increases in the distance of this place to the headquarters of TV stations. We show in a simple model that news programs may be biased towards regions close to media outlets due to this cost effect. As a consequence, citizens who live near media outlets are better informed about the spending decisions by an incumbent politician than citizens living in more distant areas. This distribution of information motivates the incumbent to allocate more funds to regions where media activity is higher because the effect of an extra dollar on expected votes is larger there.

We test this prediction empirically using U.S. data on county-level grant spending, Designated Market Areas (DMAs)[2] and location of licensed television stations. We show that the

[2]DMAs are the current industry standard for defining television markets in the United States.

distance of a county to the next media city in the respective market significantly affects the geographical distribution of grant spending, as predicted by the theoretical model. Counties which are closer to the media cities, where the bulk of television stations is located, receive significantly more funds. These results are robust to correcting for spatial correlation.

Acknowledgement. Part of the content of chapter 4 has been published under the title "Newspaper Circulation and Local Government Efficiency" in the *Scandinavian Journal of Economics* (2011, vol. 113(2), 470-492). We would like to thank the editors of the journal, Blackwell Publishing and Föreningen för utgivande av SJE for allowing to include the study in this book. The results presented in chapter 5 have been published under the title "Media Activity and Public Spending" in the journal *Economics of Governance* (2010, vol. 11, 309-332) published by Springer. Both papers originate from joint work with Oliver Himmler.

1.3 Related literature

This section introduces the reader to the literature on mass media in political economics. We focus especially on some seminal contributions which pioneered this field and on papers which are closely connected to the content of this book.[3] Some special papers which are connected to one of the following chapters only are referred to in the introductory sections of the respective chapters.

The literature that explores the role of mass media in political economics builds on the notion that the informational situation of the electorate has an effect on political outcomes and we can group the studies in this field into the main categories of informative coverage, capture of media and media bias. This book belongs to the first category for it studies how commercial motives of ideologically neutral and free media outlets determine how well voters are informed and why some voters are better informed than others.

The distinction between informed and uniformed voters can already be found in the papers by Baron (1994) and Grossman and Helpman (1996), for example, where informed voters are aware of policies or party positions. In their survey of political economics, Persson and Tabellini (2002) also emphasize the role of the information among voters for political outcomes, be it concerning distributional issues or agency problems, and say that it would be interesting to study how mass media affect the informational situation of voters.

The article by Strömberg (2001) is one of the first attempts to explain how mass media shape the formation of informed and uniformed voters. He shows how the incentives of

[3]For a more detailed and comprehensive overview of the current state of the literature, we refer the interested reader to the recent survey by Strömberg and Prat (2011).

commercial mass media can be introduced into various existing political economy models in order to explain why some voters are informed while others are not. These models deal with issues such as redistribution, the size of the government sector, rents and corruption, the effectiveness of lobby groups and political business cycles. While this study rather provides a first explorative overview, Strömberg (2004a) can be considered the first rigorous theoretical analysis in this area. In this paper, Strömberg adapts the location model by Lindbeck and Weibull (1987) to study both, the competition between media companies and political competition. He derives his results from special characteristics of commercial news production: Strömberg argues that news programs must attract an audience which is relevant for advertisers because advertising receipts constitute the bulk of revenues. Further, the costs to produce a newscast or the first copy of a newspaper are relatively high but the costs of additional copies or viewers are small or even negligible. Thus, the model predicts that mass media bias their programs in favor of large groups due to increasing returns to scale. Moreover, there is more coverage on topics important for groups that are more valuable to advertisers, for groups that attach a high value to information and for groups which are easier to reach in terms of distribution costs. Strömberg argues that when politicians announce their campaign promises before an election via mass media they take the biased allocation of news coverage into account. As a consequence, they promise higher spending levels for groups that receive more media attention because informed voters remember the promises and vote for the party which promises most. So, in a nutshell, voters who are better informed because they are more valuable for commercial mass media receive favorable policies.

In a closely related paper, Strömberg (2004b) analyzes data on the allocation of funds by a major New Deal program in the United States to find out whether politicians provide informed voters with favorable policies. First he shows in a theoretical model how incumbent governors can use radio programs to inform voters about what they have done for them. This influences voting behavior because voters only take benefits into account when they can assign them to a candidate. In line with the theoretical prediction, Strömberg finds in a cross-section analysis that per capita spending is higher in counties with a more informed electorate, where he uses the share of households owning a radio as a proxy for the share of informed voters in a county. But the share of households with a radio is only an indirect measure of voter information. Snyder and Strömberg (2010) acknowledge this problem and provide a very detailed analysis of the relation between media, information and politics. They present a sophisticated empirical analysis of three substantive effects from mass media: on the knowledge of voters, on various dimensions of incumbent behavior and on the allocation of funds. At the core of their analysis is the geographical fit between newspaper markets and congressional districts in the United States. They find that a poor fit implies that voters are less informed about their congressmen because newspapers hardly report about the latter. Another result is that congressmen from districts which are highly congruent with newspaper markets work harder for their constituency as measured by their engagement in congressional hearings, committee work and voting against the party line. Finally, Strömberg and Snyder show that the reduced accountability of politicians results in lower federal spending in poor-fitting districts. In a different context, Revelli (2008) examines how local media influence the behavior of politicians in the jurisdictions within a market and he shows that local media serve as a vehicle for yardstick competition: His findings suggest that yardstick competition affects tax policies of English local governments in districts belonging to the same local media market.

One of the first empirical papers on how media activity enhances the accountability of politicians is the study by Besley and Burgess (2002) who analyze panel data from India. They use newspaper circulation in Indian states to measure the share of informed voters and the results show that state governments provide more public food and calamity relief where newspaper circulation is higher. Reinikka and Svensson (2005) provide empirical evidence from Africa. They show that a Ugandan government program on school improvement is less inflicted with corruption in areas where local newspapers inform people about this program. Svaleryd and Vlachos (2008) add evidence from Swedish municipalities. They find that political parties refrain from (legal) rent-seeking in municipalities where citizens are well-informed due to extensive local newspaper coverage.

So far, all studies have adopted the rather optimistic view that mass media report facts sincerely and biased reporting only consists in certain compilations of true facts when, for instance, news programs favor certain audiences. But there is also the problem that mass media do not transmit all of the information they have collected truthfully: Journalists can omit important facts, emphasize only selected facts or slant stories in various other ways. Media bias can result from various reasons. One possibility is that private interest groups or politicians control or bribe media outlets. Besley and Prat (2006), for example, analyze features of media markets and politics that affect the freedom of the press. In an agency framework, media outlets can inform voters about an incumbent politician, but they suppress bad signals if the politician successfully bribes them. The authors show that the number of independent news outlets and the transaction costs for bribing essentially influence the freedom of the press. Public ownership of media outlets is assumed to imply lower transaction costs which makes media capture more likely and it is shown that captured media reduce the accountability of politicians. In accordance with this theoretical result, Djankov et al. (2003) find that countries with a large share of state ownership of the media exhibit poor government performance. A paper by Petrova (2008) also deals with media capture. Rich voters can bribe the media to manipulate the coverage of redistribution issues. Her formal analysis predicts that lower income inequality and the availability of diverse information sources lead to a lower probability of media capture. The analysis of cross-country panel data supports these predictions. In a similar spirit, Corneo (2005) develops a model where mass media inform voters about the consequences of alternative policies but interest groups can bribe the media in order to manipulate coverage. Corneo shows that media bias is more likely when firm ownership is more concentrated.

Media capture is not the only source of distorted news, for biased reporting may also result from the demand side. Mullainathan and Shleifer (2005) show that media bias can result from a confirmatory cognitive bias of media consumers. Readers who hold beliefs (e.g. about the competence of a politician) which they like to see confirmed in the news can make commercial media slant the information they send to their customers. The authors show that competition may not reduce but even strengthen media bias when readers hold heterogenous believes. Gentzkow and Shapiro (2006) demonstrate in a theoretical model that biased coverage can also result from a the ambition of media firms to build up a reputation of providing accurate news. When their customers perceive information to be more accurate if it is close to their prior beliefs, commercial mass media are willing to distort their reporting towards these beliefs. Furthermore, Baron (2006) introduces the career concerns of journalists as a source of media bias. Journalists may use slanting to make it to the front page because this pushes their careers. Profit-maximizing media companies allow for slanting if journalists accept lower wages. Moreover, Baron shows that even rational media consumers who are fully aware of slanting can be influenced by media bias. Gabszewicz et al. (2001) construct a model that demonstrates how the advertising market affects the political orientation of media outlets. If they largely rely on advertising receipts, commercial media are shown to adopt rather neutral instead of extreme political positions.

Some empirical papers find evidence of biased media. DellaVigna and Kaplan (2007) present empirical evidence that biased media coverage affects voting behavior. They show that the entry of Fox News in U.S. cable markets significantly raises the share of votes for the Republicans in the respective areas. In contrast, Groseclose and Milyo (2005) find a tendency among U.S. media outlets to slant their stories towards leftist political positions. They suspect various factors to cause the leftist bias such as corresponding political preferences of journalists or, as Hamilton (2006) argues, that some groups with more liberal attitudes receive more attention because they are more valuable for advertisers. Gentzkow and Shapiro (forthcoming) provide one of the first studies on media bias in the internet and they find that news consumption in the internet is not heavily segregated ideology-wise.

There is also empirical evidence that mass media affect political participation. Gentzkow (2006) shows that the introduction of television in the United States significantly lowered voter turnout. In addition, he demonstrates that more informative media like radio and newspapers were substituted by television consumption which resulted in less political knowledge of voters. The findings by Gentzkow et al. (forthcoming) support the notion that newspapers play an important role for political participation, since the results indicate that having more newspapers published increases turnout. Oberholzer-Gee and Waldfogel (2009) find that the turnout among Hispanics was increased by the entry of local television programs in Spanish language. Althaus and Trautman (2008) study the impact of television market size on voter turnout. They find that turnout in local elections in the United States is lower in large television markets. Local elections are assumed to be less newsworthy in larger markets so that residents in these markets receive less political information which pushes down voter turnout. Prior (2006) finds that the rise of television in the United States contributed to the incumbency advantage of U.S House representatives in the 1960s. He suggests that television allowed incumbent politicians to have more media appearances than potential challengers to impress voters.

2 Collective decision-making and market provision of information

2.1 Introduction

Voters usually face decisions under uncertainty when it comes to selecting politicians or policy programs because they are not perfectly informed about the ability of candidates, political programs, or the consequences of policies. This implies that there is always the possibility of failing to choose the best alternative. However, voters can obtain additional information in order to reduce the risk of decisions and we have already argued in the introduction of this book that the bulk of political information voters rely on comes from the mass media. In this chapter, we strive to characterize the level of information in the electorate which results from market provision. In line with Downs (1957), we identify a market for information with a situation where voters delegate the task of information production to specialized firms: media companies employ journalists who collect information and then they sell the information to the voters. We show that the media market tends to provide inefficiently low levels of information because voters have incentives for easy-riding behavior and, as a consequence, the risks associated with electoral decisions are inefficiently high.

This result applies to electoral decision problems where the electorate has a collective interest to choose the alternative that is commonly considered the best, i.e. the subject of the election is not primarily concerned with redistributive problems, for example. This includes political offices that are responsible for providing local public goods like roads, parks, public safety, fire departments and schools or also referenda, for example. Even if the voters do not fully agree on the evaluation of an electoral subject in all its dimensions, there are usually dimensions where voters share common preferences. When voters have different party preferences, for example, they should have a common interest to learn about the ability of political candidates affiliated to the parties.

The driving force behind the inefficient provision of information is the public good nature of a better informed electoral decision when it comes to electoral subjects as described in the preceding paragraph. Given common preferences of voters, each piece of information that is used to reduce the risk of a decision benefits the whole electorate. Thus, if the information produced by media companies falls into this category, this information constitutes a public good. A newspaper subscription, for example, then can be interpreted to be a vehicle to contribute to this public good and the voters' decisions to subscribe to newspapers can be described as a game of private public good provision where the newspapers produce the contributions and set prices. For such a setting, our analysis shows that a competitive market fails to provide the efficient level of information because of strategic incentives for easy-riding behavior in the electorate. This result rests on the behavioral assumption that as to contributing to information production the voters act non-cooperatively and that their decisions form a Nash-equilibrium.

In addition to characterizing the potential inefficiency in information provision we also find that the gap between the efficient level of information and the market level tends to increase when the electorate is larger. This tendency is a well-known problem in the literature on the private provision of public goods (see Pecorino (1999), for example) which also applies to the provision of political information. Further, we show that the level of information resulting from market provision is higher if the cost of collecting information is low and if the production technology is of a higher quality such that a given amount spent on investigations yields more precise information. Finally, we find that an increase in the precision of the knowledge prior to information collection can have a comparatively adverse effect on the risk of electoral decisions.

Our analysis is related to several papers that deal with the subject of information acquisition by voters. While we focus on the production of public information, Martinelli (2006), in contrast, presents a model where rational voters condition their decisions to acquire private information on the probability of being decisive in the election. He shows that large electorates need not necessarily be as poorly informed as Downs (1957) suggested. Strömberg (2004) proposes another reason why voters are willing to pay for political information: informed voters are better able to privately adapt to future policies. In his model, voters who do not read newspapers or pay attention to television programs are excluded from the benefits of the information offered by the mass media because information helps to optimize private behavior. Due to this excludability, mass media programs are biased towards large groups. A study by Aldashev (2010), however, suggests that voters also engage in information acquisition because political knowledge is a valuable asset for communication within their social networks.

The idea that citizens cannot be excluded from well-informed electoral decisions also plays a prominent role in the paper by Feddersen and Pesendorfer (1996). They argue that it can be an optimal strategy of uninformed voters to abstain from voting in order to let informed voters decide an election. While the level of information is exogenously given in the model by Feddersen and Pesendorfer, a study by Feddersen and Sandroni (2006) presents a theory that explains why voters acquire costly information. The authors say that pivot probabilities are too small in large elections to explain why voters obtain costly information and therefore they propose a framework where voters are willing to pay for information out of a sense of civic duty.

Finally, since political information constitutes a public good in our model this analysis is also related to the literature on the private provision of public goods (see Cornes and Sandler (1996) for a survey).

2.2 Description of the model

There is an electorate that consists of H identical citizens who can elect either a politician A or a politician B. The elected candidate provides a public good g and we assume that the level of the public good a politician provides depends on some fixed characteristic $\theta_{c,\ c=A,B}$ and thus the politicians enter the model as parameters only. The fixed characteristic can be interpreted as a politician's personal ability, for example.

The citizens are uncertain about the abilities of the politicians but they can obtain additional information to reduce their uncertainty: Two newspapers $n \in \{1, 2\}$ can produce information about the politicians' abilities and sell it to the citizens via subscriptions. A subscription consists of a pair (p_n, x_n) where p_n denotes the price of a subscription and x_n represents an increase in the quality of the information that the newspaper is willing to produce for each sold subscription. The newspapers non-cooperatively announce their offers (p_1, x_1) and (p_2, x_2) to maximize their profits

$$\Pi_1[(p_1, x_1), (p_2, x_2)] = D_1[(p_1, x_1), (p_2, x_2)] \cdot (p_1 - k \cdot x_1) \tag{2.1}$$

and

$$\Pi_2[(p_1, x_1), (p_2, x_2)] = D_2[(p_1, x_1), (p_2, x_2)] \cdot (p_2 - k \cdot x_2), \tag{2.2}$$

where D_n denotes the number of subscribers to a newspaper and k is the cost of producing one unit of x. The details of the production technology are derived below.

The overall quality $x = D_1 \cdot x_1 + D_2 \cdot x_2$ of the information service created by the newspapers determines the risk of the electoral decision between the two politicians and we will show below that each citizen strives to minimize his individual total risk

$$\rho_t = \rho(x) + p_n \tag{2.3}$$

which is equal to the sum of the risk of the electoral decision $\rho(x)$ and the cost to contribute to the overall quality via a subscription to a newspaper. It will become apparent below, that the quality of the information service created by the newspapers constitutes a pure public good because a well-informed electoral decision is of a non-rival nature and no citizen can be excluded from its benefit. Consequently, each citizen benefits from the other citizens' subscriptions and therefore the citizens' decisions to subscribe to a newspaper must form a 'Nash-equilibrium'.

The timing of the game is as follows: First, the newspapers announce their subscription packages and then the citizens choose whether or not to subscribe. Afterwards the newspapers produce their signals, all citizens observe the information produced by the newspapers and vote. Finally, the winning politician implements the public good.

2.2.1 The electoral decision problem

The citizens cast their ballot simultaneously and non-cooperatively. A citizen can elect either politician A or B so that the choice set for his voting decision is $v \in \{A, B\}$. Consider the decision problem of a citizen: The ability of the elected candidate determines the level of the public good and we assume that the value each citizen attaches to the public good is linear in ability such that the value is either $g_A = \theta_A$ or $g_B = \theta_B$. A citizen does not know the value of each candidate's ability but it is common knowledge that each θ_c is drawn from a normal distribution with mean $\bar{\theta}$ and precision $2\tau_\theta$, or written formally $\theta_c \sim N(\bar{\theta}, 2\tau_\theta)$.[4]

We assume that the citizens know that they have identical preferences. Consequently, they know that they all agree on which politician is preferable as long as it is common knowledge that they share the same information about the politicians' abilities. In this case an equilibrium where the citizens vote sincerely for the preferred candidate constitutes a focal point of the electoral game.

Clearly, rational citizens prefer to select the candidate with higher ability and thus they should vote for the candidate with higher expected ability. However, we shall reformulate the electoral decision problem in terms of the *opportunity loss* associated with an electoral decision in order to characterize the quality of this decision. Indeed it will become apparent that this concept yields the same optimal voting decision as relying on expected values but it allows us to analyze the information acquisition of the citizens. From now on we will often simply say *loss* instead of using the term *opportunity loss* to save some space.

[4]The precision is defined to be the reciprocal of the variance (precision=1/variance). The number 2 in the term $2\tau_\theta$ serves only to economize on notation in the remainder of the model.

Each citizen incurs an opportunity loss $\ell_{v,\ v=A,B}$ when the candidate with lower ability is elected accidentally. The loss represents the difference in ability between the best available candidate and the candidate who is actually chosen. In other words, we can interpret the loss to be the cost of a wrong decision. Let $\theta := \theta_A - \theta_B$ denote the difference between the abilities of candidate A and candidate B. Then each citizen's loss associated with an electoral decision is either

$$\ell_A(\theta) = \begin{cases} 0 & \text{for } \theta \geq 0 \\ |\theta| & \text{for } \theta < 0 \end{cases}$$

or

$$\ell_B(\theta) = \begin{cases} \theta & \text{for } \theta \geq 0 \\ 0 & \text{for } \theta < 0. \end{cases}$$

Obviously, the citizens do not suffer a loss if they select the better candidate; otherwise each citizen's loss is $|\theta|$. Given the priors of the candidates' abilities, θ is a normally distributed random variable with mean 0 and precision τ_θ, or written formally $\theta \sim N(0, \tau_\theta)$. Let F denote the distribution function of θ at a given information set and let $E(\theta)$ denote the expected value of θ at the respective information set. Knowing the distribution function F each citizen can compute the *expected loss*

$$\rho_{A^0} := E[\ell_A(\theta)] = \int_{-\infty}^{0} |\theta| \; dF(\theta) \tag{2.4}$$

and

$$\rho_{B^0} := E[\ell_B(\theta)] = \int_{0}^{\infty} \theta \; dF(\theta) \tag{2.5}$$

that is associated with the respective electoral decision. For ease of exposition, we will often use the term *risk* synonymously with expected loss.

Clearly, the citizens shall vote for the alternative with lower risk and it can be shown that:

Result 2.1. *The optimal decision is A if $E(\theta) > 0$ and B if $E(\theta) < 0$.*

Proof. See appendix. □

Thus it is always optimal to elect the candidate with higher expected ability (recall that $\theta = \theta_A - \theta_B$). Since the prior mean of θ is zero the citizens are originally indifferent between the two candidates and $\rho_A^0 = \rho_B^0$. However, this may change once they obtain additional information about the politicians from the newspapers and in the next section, we derive the value each citizen attaches to an information service of a given quality. Then, we will use this result to examine the quality of newspaper coverage that arises in a market equilibrium.

2.2.2 The value of additional public information

When you subscribe to a newspaper, you usually do not know its content in advance but you can only form expectations about the content. Thus, our aim in this section is to determine the value each citizen attaches to an information service before having received a signal from it.

Suppose that additional information about θ is provided by an information service which generates a public signal

$$s = \theta + \varepsilon \tag{2.6}$$

where $\varepsilon \sim N(0, \tau_\varepsilon x)$ and x serves as a measure of the quality of the information service. The signal contains information about θ that the citizens can use to revise their beliefs and possibly change their voting decision.

Obviously, the optimal posterior voting decision is unknown before the signal is observed because the mean of the posterior distribution of $\theta|s$ is not known at this point. However, assuming the citizens to be rational in the sense that they update their beliefs using Bayes' rule we can compute this posterior distribution for every possible realization of the signal given the quality x. It can be shown that the posterior distribution of $\theta|s$ is a normal distribution with mean $\mu = \frac{\tau_\varepsilon x}{\tau_\theta + \tau_\varepsilon x} \cdot s$ and precision $\tau = \tau_\theta + \tau_\varepsilon x$. Thus we can characterize the optimal posterior voting decision and the respective posterior risk for every realization of the signal. From Result 1, it follows directly that the optimal posterior decision is characterized by

$$v^1 = \begin{cases} A & \text{for } E(\theta|s) > 0 \\ B & \text{for } E(\theta|s) < 0 \end{cases}$$

and the respective posterior risks ρ_{A^1} and ρ_{B^1} can be computed analogously to the prior risk. It is shown in the appendix that the posterior risk can be described by either

$$\rho_{A^1} = \tau^{-\frac{1}{2}} \left\{ \varphi(-\mu\tau^{\frac{1}{2}}) - \mu\tau^{\frac{1}{2}} \cdot \Phi(-\mu\tau^{\frac{1}{2}}) \right\} \tag{2.7}$$

or

$$\rho_{B^1} = \tau^{-\frac{1}{2}} \left\{ \varphi(-\mu\tau^{\frac{1}{2}}) + \mu\tau^{\frac{1}{2}} \cdot [1 - \Phi(-\mu\tau^{\frac{1}{2}})] \right\}. \tag{2.8}$$

A brief characterization of the posterior risk is in order: The magnitude of the posterior risk depends on the values of the parameters μ and τ that describe the distribution of $\theta|s$ and both of these parameters depend on the quality parameter x. As to μ, for every realization of the signal a higher quality x implies that more weight is given to the signal as against the prior and this raises the absolute value of the posterior mean which makes the posterior belief more 'decisive'. This effect reduces the posterior risk. Only for $s = 0$, the posterior mean is equal to the prior mean which leaves the citizens indifferent. However, the posterior precision τ strictly increases in x so that the posterior distribution is always less noisy than the prior distribution and this effect reduces the posterior risk. Summarized, we find that:

Result 2.2. *Posterior risk of a citizen.*

(a) The posterior risk is smaller than the prior risk for $x > 0$.

(b) The posterior risk is decreasing in the quality of the information service:
$\partial\rho_{v^1}/\partial x < 0$.

Proof. See appendix. □

Now we have characterized the electoral decision problem for every possible realization of the signal. From the posterior perspective, the value that each citizen attaches to a signal is the difference between the prior risk and the posterior risk: $\rho_{v^0} - \rho_{v^1} = E[\ell_{v^0}(\theta)] - E[\ell_{v^1}(\theta|s)]$. However, prior to observing the signal, a citizen can compute the prior risk, ρ_{v^0}, only while the posterior risk, ρ_{v^1}, depends on an unknown realization of the random variable s.

Although a citizen cannot predict the realization of the signal, he does know that the posterior mean $E(\theta|s)$ determines the electoral decision and he also knows the posterior risk corresponding to each value of $E(\theta|s)$. From the prior perspective, $E(\theta|s)$ is a random variable and the citizen can use his knowledge about the distribution of the signal s to compute the distribution of $E(\theta|s)$ which is a normal distribution with mean zero and precision $r = \frac{\tau_\theta(\tau_\theta+\tau_\varepsilon x)}{\tau_\varepsilon x}$. This allows him to form expectations about the posterior risk: $E\left\{E[\ell_{v^1}(\theta|s)]\right\}$, where the outer expectation operator refers to s and the inner expectation operator refers to θ. Following Raiffa and Schlaifer (1961) we call this prior expectation of the posterior risk the *pre-posterior* risk and define $\rho := E\left\{E[\ell_{v^1}(\theta|s)]\right\}$. Let ω denote the value each citizen expects to receive from the information service. Having established the pre-posterior risk of the electoral decision, we know that this value is

$$\omega = E[\ell_{v^0}(\theta)] - E\left\{E[\ell_{v^1}(\theta|s)]\right\}. \tag{2.9}$$

Given our distributional assumptions the following can be shown:

Result 2.3. *The value every single citizen attaches to a public information service with quality x can be described by the function*

$$\omega(x) = \left(\frac{\tau_\theta(\tau_\theta + \tau_\varepsilon x)}{\tau_\varepsilon x}\right)^{-1/2} \cdot \varphi(0) \tag{2.10}$$

where φ is the density function of the standard normal distribution. The function $\omega(x)$ is strictly increasing in x and concave with infinite slope at the origin.
Each citizen's pre-posterior risk of the electoral decision is

$$\rho(x) = \rho_{v^0} - \omega(x) = \tau_\theta^{-1/2}\varphi(0) - \left(\frac{\tau_\theta(\tau_\theta + \tau_\varepsilon x)}{\tau_\varepsilon x}\right)^{-1/2} \cdot \varphi(0). \tag{2.11}$$

Proof. See appendix. □

This result says that the value of the information service always increases in the quality x but at a decreasing rate. If, for example, journalists already have comprehensively investigated the politicians' abilities further efforts only produce little additional value. It can also be observed that the value of an information service of given quality is lower, the more precise the prior belief is. This is reasonable and closely related to our former statement regarding the quality of the information service: the more information the electorate already has at hand, the less valuable is additional information. Having established the functions $\omega(x)$ and $\rho(x)$ we can now turn our attention to the production of the information service by the newspapers.

2.2.3 Information production

The information service is formed by the behavior of the two newspapers $n = 1, 2$. Each newspaper can produce a noisy signal that provides information about the difference in abilities between the two politicians:

$$s_n = \theta + \varepsilon_n \tag{2.12}$$

where $\varepsilon_n \sim N(0, x_n^+ \cdot \tau_\varepsilon)$. We assume that the noise terms ε_1 and ε_2 are uncorrelated so that the signals of the newspapers are conditionally independent. Both newspapers have identical technologies: A newspaper can manipulate the precision of its signal by choosing a value for the variable x_n^+ where every unit of x_n^+ comes at a cost k.

One possible interpretation of this technology is that k denotes the hourly wage paid to a journalist and that one hour of work produces a signal $s_j = \theta + \varepsilon_j$, where $\varepsilon_j \sim N(0, \tau_\varepsilon)$. The precision of a single signal may be understood to measure how complex the investigated topic is. For instance, the past performances of politicians can be analyzed to gain information about their abilities. If they already occupied offices comparable to the one up to election the observations should be less noisy. Another aspect is the complexity of the office itself. The more dimensions need to be considered, the more complex an evaluation might become. Following this interpretation, x_n^+ denotes the total amount of working hours a newspaper allocates to investigating the politicians' abilities and since we are dealing with sampling from a normal random process the signal s_n we characterized above is simply the mean of the single signals s_j each newspaper produces.

Since we assume that the signals of both newspapers become public the information contained in both signals will shape the electoral decision. We can summarize this information with the sufficient statistic (s, x) where $s := (s_1 + s_2)/2$ denotes the mean of both signals and $x := x_1^+ + x_2^+$ the sum of the quality variables. Once the citizens know the pair (s, x) they cannot learn more about θ from the single signals s_n. Thus we can say that the citizens receive a public signal s with quality x instead of using the single signals. Recalling Result 2.3 we know that each citizen's pre-posterior risk associated with this signal is equal to $\rho(x)$.

The qualities of the signals the newspapers provide are determined by the characteristics of the subscriptions the newspapers offer and the citizens' decisions to accept an offer: Each newspaper announces a subscription package (p_n, x_n) and the resulting quality of a newspaper's signal then is characterized by $x_n^+ = D_n[(p_1, x_1), (p_2, x_2)] \cdot x_n$ where D_n describes the number of subscribers to a newspaper.

In the next section, we put things together and derive the quality of newspaper coverage in a market equilibrium.

2.3 Market equilibrium

The quality x of the information service determines the risk $\rho(x)$ that each citizen attaches to the electoral decision and each citizen decides whether or not to subscribe to a newspaper in order to minimize his total risk

$$\begin{aligned} \rho_t &= \rho(x) + p_n \\ &= \rho_{v^0} - \omega(x) + p_n. \end{aligned}$$

For ease of exposition, however, we prefer to analyze the equivalent problem of maximizing

$$\omega(x) - p_n$$

because this will save us some distracting notation. According to this approach, $\omega(x)$ is the value that each citizen attaches to the overall quality. The level of x is determined by the citizens' contributions that come along with their newspaper subscriptions. It is important to notice that each citizen benefits from the contributions of the other citizens and therefore each citizen considers the others' contributions when he decides about his own contribution. Thus, we can interpret this situation as a game of private public good provision where the newspapers determine the characteristics of the available contributions, i.e. the price of a subscription and the additional quality attached to it. They choose these characteristics in order to maximize their profits

$$\Pi_1[(p_1, x_1), (p_2, x_2)] = D_1[(p_1, x_1), (p_2, x_2)] \cdot (p_1 - k \cdot x_1) \tag{2.13}$$

and

$$\Pi_2[(p_1, x_1), (p_2, x_2)] = D_2[(p_1, x_1), (p_2, x_2)] \cdot (p_2 - k \cdot x_2). \tag{2.14}$$

Our first step to derive the market equilibrium is to examine the citizens' decision to subscribe to a newspaper for given offers by the newspapers. Since all citizens are identical, we can simply analyze a representative citizen to characterize a symmetric equilibrium. A citizen i subscribes to a newspaper if the condition

$$\omega(x_n + x_{-i}) - \omega(x_{-i}) - p_n \geq 0$$

is satisfied: the additional value a citizen derives from contributing x_n via a subscription when all others contribute the amount x_{-i} must not be smaller than the price of the subscription. For given offers (p_1, x_1) and (p_2, x_2) the citizen subscribes to newspaper 1 if and only if

$$\omega(x_1 + x_{-i}) - \omega(x_2 + x_{-i}) > p_1 - p_2$$

and to newspaper 2 if and only if

$$\omega(x_1 + x_{-i}) - \omega(x_2 + x_{-i}) < p_1 - p_2.$$

If the citizen is indifferent between both offers we assume that he subscribes to each newspaper with probability $1/2$. Next, we use this result to study the price setting stage between the newspapers.

Price equilibrium. In this section, we analyze the price-setting behavior of the newspapers for given qualities x_1 and x_2 of the subscriptions. Consider a decision of the citizen for a given pair of offers as derived in the previous section: At least one newspaper has always an incentive to undercut its rival's price in order to attract the citizen. Clearly, in response, the undercut newspaper lowers its price until the citizen favors its offer again and this Bertrand-like price competition pushes prices downwards. We have to distinguish between price-equilibria with equal and unequal given qualities:

- *Unequal qualities.* Assume with no loss of generality that $x_1 > x_2$. If newspaper 1 offers a higher quality, it charges

$$p_1^* = \omega(x_1 + x_{-i}) - \omega(x_2 + x_{-i}) + kx_2$$

  and earns a positive profit. If newspaper 1 charges a higher price, newspaper 2 can undercut this price and capture the citizen. However, to ensure non-negative profits, newspaper 2 cannot announce a price lower than $p_2 = k \cdot x_2$ and thus it cannot undercut if newspaper 1 charges p_1^*. Consequently, p_1^* is the profit-maximizing price for newspaper 1 while the price of newspaper 2 is $p_2^* = k \cdot x_2$.

  For $x_1 < x_2$ it follows that

$$p_2^* = \omega(x_2 + x_{-i}) - \omega(x_1 + x_{-i}) + kx_1 \quad \text{and} \quad p_1^* = k \cdot x_1.$$

- *Equal qualities.* In case of equal qualities the newspapers can undercut each other's prices as long as the prices are above the cost to produce a subscription of quality x_1 or x_2. Thus, the only price equilibrium has both newspapers set equal prices $p_1^* = p_2^*$, which are characterized by

$$p_1^* = k \cdot x_1 \quad \text{and} \quad p_2^* = k \cdot x_2.$$

Having analyzed the price setting stage we can now turn to the first stage where the newspapers choose the qualities of the subscriptions they offer.

Quality equilibrium. The net gains from selling a subscription to the citizen are

$$\Pi_{1i}(x_1, x_2) = \begin{cases} p_1^*(x_1, x_2) - k \cdot x_1 & \text{for } x_1 > x_2 \\ 0 & \text{for } x_1 \leq x_2 \end{cases}$$

and

$$\Pi_{2i}(x_1, x_2) = \begin{cases} 0 & \text{for } x_1 \geq x_2 \\ p_2^*(x_1, x_2) - k \cdot x_2 & \text{for } x_1 < x_2. \end{cases}$$

It is a weakly dominant strategy for each newspaper to choose the quality that maximizes the net gain of a subscription for a given choice of the other newspaper. Due to the symmetry of the newspapers' maximization problem, the optimal quality of a subscription is characterized for both newspapers by the first-order condition

$$\omega'(x_n + x_{-i}) = k. \tag{2.15}$$

This means that the citizen's marginal willingness to pay for higher quality, evaluated at the aggregate amount of contributions to quality equals the cost of producing one more unit of quality. As the function $\omega(x)$ is strictly concave, there is a unique solution to equation (2.15).

Since this condition is identical for both newspapers, they offer identical subscriptions which are characterized by the quality $x_n = x_1 = x_2$ and the price $p_n = k \cdot x_n$ where x_n denotes the contribution of quality that satisfies the first-order condition given x_{-i}. Consequently, both newspapers earn zero profits.

The assumption of identical citizens implies that, in a symmetric equilibrium, the solution to equation (2.15) is characterized by a situation where each citizen contributes an identical amount so that the total quality adds up to $x^* = H \cdot x_n^*$.

Therefore, we can describe the equilibrium of our model as follows:

Proposition 2.1. *The market provides an information service of quality x^* and this quality is implicitly defined by*

$$\omega'(x^*) = k. \tag{2.16}$$

Each citizen contributes the amount $x_n^* = x^*/H$ via a subscription and pays the price $p_n^* = k \cdot (x^*/H)$. Further, the individual total risk is

$$\rho_i(x^*) + k \cdot (x^*/H) \tag{2.17}$$

so that the aggregate total risk of the electorate adds up to

$$H \cdot \rho_i(x^*) + k \cdot x^*. \tag{2.18}$$

In order to examine whether the market provision of information is efficient we compare the market result with a situation where a social planner employs the information technology of the newspapers. He chooses the quality of the information service to minimize the aggregate total risk

$$H \cdot \rho_i(x) + k \cdot x$$

and we assume that he makes every citizen pay an equal share $k \cdot (\hat{x}/H)$ of the cost to produce the optimal overall quality $\hat{x}$. This optimal quality is implicitly defined by the first order condition

$$H \cdot \omega'(\hat{x}) = k. \tag{2.19}$$

This is the Samuelson condition that must be satisfied to ensure an efficient provision of a public good. Comparing the equations (2.16) and (2.19) we can see that the market does not provide the optimal quality. Both equations differ in the factor $H > 1$ only. Since $\omega'(x)$ is a strictly decreasing function the value of $\hat{x}$ must be larger than the value of x^*. Hence, we can say that:

Proposition 2.2. *The market provides an information service whose quality is inefficiently low.*

This result indicates that the market provision of information cannot overcome the free-riding problem among the citizens. The market outcome rather corresponds to a game of private provision of a public good where the citizens would individually choose a contribution to the quality of public information at a cost of k per unit.[5] This follows because the competition between the newspapers prevents them from exploiting their opportunity to determine the characteristics of a subscription by charging higher prices, for example.

Before we briefly discuss some relevant assumptions of our model to interpret this result we further characterize the market provision of information by presenting insights from a comparative statics analysis.

[5] This insight is analogous to Besley and Ghatak (2007) who study the economics of corporate social responsibility.

2.4 Comparative statics analysis

It remains to describe how changes of the parameters in the model alter the equilibrium outcome:

(a) *Size of the electorate (H).* To characterize the comparative statics with respect to H it suffices to note that the unique value of the overall quality x^* that satisfies the first-order condition in Proposition (2.1) is independent of the size of the electorate. Consequently, the overall quality remains the same whereas the individual contributions x^*/H decrease in the size of the electorate. It also follows from Proposition (2.1) that the individual total risk of a citizen decreases and that the aggregate total risk of the electorate increases in H. This result implies further that the gap between the efficient level $\hat{x}$ and the market level x^* is larger, the larger the electorate is.

While the strong result of a constant market level in our setting depends on the quasi-linear structure of the citizens' preferences, the problem that the gap between the efficient level and the market level of a public good increases in the size of the population is of a general nature as Pecorino (1999) shows.

We can further make an interesting comparison to a result by Strömberg (2004a) who finds that mass media allocate more resources to large groups. In his model, voters who would not buy a newspaper are excluded from the benefit of information because the value of information originates from a private action. This is different in our model where non-subscribers cannot be excluded from the benefits of information.

So far, we have simply relied on the concavity of $\omega(x)$ to characterize the market equilibrium. Recalling the functional form of $\omega(x)$ as shown in Result 2.3 the first-order condition that implicitly defines the overall quality in equilibrium reads

$$\frac{1}{2}\varphi(0)\frac{\tau_\theta^2}{\tau_\varepsilon x^2}\left(\frac{\tau_\theta^2}{\tau_\varepsilon x}+\tau_\theta\right)^{-3/2}-k=0 \tag{2.20}$$

and we can use the implicit function rule to derive the remaining comparative statics results (The formal proofs are in the appendix.).

(b) *Production technology (k and τ_ε).* Recall that $\omega'(x)$ is a strictly decreasing function. Thus, lower values of k result in an increase of the overall information production, or $\partial x^*/\partial k < 0$.

The effect of a higher precision of an observation (τ_ε) on x^* is ambiguous:

$$\frac{\partial x^*}{\partial \tau_\varepsilon} \begin{cases} > 0 & \text{for } \tau_\theta > 2\tau_\varepsilon x^* \\ = 0 & \text{for } \tau_\theta = 2\tau_\varepsilon x^* \\ < 0 & \text{for } \tau_\theta < 2\tau_\varepsilon x^*. \end{cases} \tag{2.21}$$

The equilibrium precision of the information service ($\tau_\varepsilon x^*$), however, always increases in τ_ε and it is shown in the appendix that lower values of k as well as larger values of τ_ε lead to a lower total risk of each citizen. So the main message is that a more efficient production technology benefits the electorate.

(c) *Prior degree of uncertainty (τ_θ).* The effect of a more precise prior on information production is ambiguous:

$$\frac{\partial x^*}{\partial \tau_\theta} \begin{cases} > 0 & \text{for } \tau_\theta < \tau_\varepsilon x^*/2 \\ = 0 & \text{for } \tau_\theta = \tau_\varepsilon x^*/2 \\ < 0 & \text{for } \tau_\theta > \tau_\varepsilon x^*/2. \end{cases} \tag{2.22}$$

Though one may intuitively expect that a more informed prior implies less information production this only holds true for relatively 'large' values of τ_θ compared to $\tau_\varepsilon x^*$. Further, a more precise prior can result in an increase of a citizen's total risk (ρ_t), so that we conclude that more prior information can lead to comparatively adverse outcomes.

2.5 Discussion

Our analysis is mainly concerned with describing the free-riding problem that arises when the media market provides information for collective decision-making. The results are derived in a setting where we make rather optimistic assumptions about the market environment while we could say that the assumption of purely non-cooperative behavior is a pessimistic one. The model constitutes a benchmark case that serves to clearly characterize the inefficient information provision due to free-riding in absence of other adverse effects on the level of information in equilibrium. A short discussion of the assumptions is in order.

The assumption of stiff competition between the newspapers serves to achieve a situation where the citizens are not harmed because of market power the newspapers might possess. A monopoly newspaper, for example, would have incentives to charge a price higher than the production costs in order to raise profits. From the citizens' perspective, there is no such waste associated with information acquisition in case of a fully competitive market.

We are optimistic concerning the problem that the newspapers commit to collecting the amount of information the citizens have paid for. Once the newspapers have received their payments they have incentives to provide a signal that is based on fewer observations in order to raise profits. In case of imperfect monitoring, the newspapers can try to build up reputations of being high quality providers which do not cheat. Besley and Ghatak (2007) show that competing firms need to charge prices above marginal costs to make credible promises in such a setting. This result indicates that the media market produces less information when the newspapers cannot credibly commit to sell information at marginal cost.

Further, the informational situation is assumed to be efficient in such a way that every single piece of information that is produced definitely helps to improve the electoral decision. Suppose, for example, that the citizens only receive the information produced by the newspaper they subscribe to but no other information and that there is no communication between citizens. Then, the first-order condition in equilibrium would read $\rho \cdot \omega(x_n^+) = k$, where $0 \leq \rho \leq 1$ denotes the probability that the signal provided by a newspaper n is pivotal for the electoral decision. This follows because the payoff for contributing to the quality of the respective signal is $\omega(x_n^+)$ with probability ρ and zero with probability $1-\rho$. Thus, for $\rho < 1$ all signals are less informative in equilibrium. However, it appears to be more realistic that information once produced will spread out because of communication between journalists or citizens.

Our assumption that the citizens act non-cooperatively yields a pessimistic benchmark. The discussion on public good provision and collective action goes back at least to Olson (1965) and the assumption of non-cooperative behavior appears to be appealing, especially in large groups. However, cooperation can sustain even when groups are large as is shown theoretically by Pecorino (1999), for example. With cooperative behavior, the total level of a public good is higher than in our non-cooperation equilibrium.

2.6 Concluding remarks

We have shown that rational voters have an incentive to pay for information that is not driven by the pivot probability of a vote. This incentive applies when it appears reasonable that the impact of the newly produced information is of a public good nature which is very likely true for information produced and distributed by modern mass media like newspapers, for example. Our analysis suggests, however, that the market for information fails to provide the optimal quality of information in such a scenario precisely because of the public good nature of this kind of information.

Our results could be interpreted as a rationale for public provision of political information. We do not jump to this conclusion, however, but we think that answering this question is beyond the scope of the analysis: Think, on the one hand, of the problem that public institutions may also work inefficiently as it is widely discussed in the literature. Moreover, it is a delicate question to which extent government should be allowed to control modern mass media that do not only act as neutral information providers as in our analysis but may shape our thinking about the world in general (see Prat and Strömberg (2011) for a discussion). Thus, we confine ourselves to believing that our analysis points out an important problem and that it constitutes an interesting benchmark case and starting point for further research.

One promising direction for further research is to leave behind the assumption of identical citizens and to introduce different tastes for public goods or ideologically motivated preferences when it comes to selecting political candidates. It would be interesting to study how these extensions alter our results, especially if the problem of inefficient information provision would become more or less severe in such (more realistic) environments. Further, political candidates enter our model as parameters only but politicians or parties usually have strong incentives to manipulate the voters' beliefs about their talents in order to secure political offices. In the next chapter, we present a model where we extend our framework and introduce an incumbent politician who tries to manipulate the information voters use in an election.

2.7 Appendix

Proof of Result 2.1. The citizens vote for candidate A if

$$E[\ell_A(\theta)] = \int_{-\infty}^{0} |\theta| \; dF(\theta) < E[\ell_B(\theta)] = \int_{0}^{\infty} \theta \; dF(\theta).$$

Moving terms on one side shows that this is equivalent to

$$\int_{-\infty}^{0} \theta \; dF(\theta) + \int_{0}^{\infty} \theta \; dF(\theta) = E(\theta) > 0.$$

Obviously, a vote for candidate B implies that $E[\ell_A(\theta)] > E[\ell_B(\theta)] \Leftrightarrow E(\theta) < 0.$ $\square$

Proof of Result 2.2. **Posterior risk.** The posterior risk is either

$$E[\ell_A(\theta|s)] = \int_{-\infty}^{0} -\theta \; dF(\theta|s)$$

or

$$E[\ell_B(\theta|s)] = \int_{0}^{\infty} \theta \; dF(\theta|s), \tag{2.23}$$

where $F(\theta|s)$ denotes the posterior distribution function of θ which is a normal distribution with mean $\mu = \frac{\tau_\varepsilon x \cdot s}{\tau_\theta + \tau_\varepsilon x}$ and precision $\tau = \tau_\theta + \tau_\varepsilon x$. The integrals on the right-hand side of the equations can be interpreted as *partial expectations* of the random variable θ.

We shall derive the functional form of such partial expectations in general for a normal random variable W with mean μ and precision τ because we can use this general result later. The density function and the distribution function of W are denoted $f(w)$ and $F(w)$, respectively. First, we show that

$$\int_{-\infty}^{0} -wf(w)dw = \tau^{-\frac{1}{2}} \left[\varphi(-\tau^{\frac{1}{2}}\mu) - \tau^{\frac{1}{2}}\mu \cdot \Phi(-\tau^{\frac{1}{2}}\mu)\right].$$

We start with substituting $wf(w)$ in the integral on the left-hand side with $\mu f(w) - \frac{f'(w)}{\tau}$ which follows directly from $f'(w) = -\tau(w-\mu)f(w)$:

$$\begin{aligned}
\int_{-\infty}^{0} -wf(w)dw &= -\int_{-\infty}^{0} wf(w)dw \\
&= -\mu \int_{-\infty}^{0} f(w)dw + \frac{1}{\tau}\int_{-\infty}^{0} f'(w)dw \\
&= -\mu \cdot F(0) + \frac{1}{\tau} f(0) \\
&= \tau^{-\frac{1}{2}} \left[\tau^{-\frac{1}{2}} f(0) - \mu\tau^{\frac{1}{2}} \cdot F(0)\right].
\end{aligned}$$

Transforming to the standard normal form we obtain

$$\int_{-\infty}^{0} -wf(w)dw = \tau^{-\frac{1}{2}} \left\{\varphi[-\mu\tau^{\frac{1}{2}}] - \mu\tau^{\frac{1}{2}} \cdot \Phi[-\mu\tau^{\frac{1}{2}}]\right\}. \tag{2.24}$$

Applying the analog procedure or relying on a symmetry argument confirms that

$$\int_0^\infty wf(w)dw = \tau^{-\frac{1}{2}} \left[\varphi(-\tau^{\frac{1}{2}}\mu) + \tau^{\frac{1}{2}}\mu \cdot (1 - \Phi(-\tau^{\frac{1}{2}}\mu))\right]. \tag{2.25}$$

Proof of part (a). Rearranging the right-hand sides of equation (2.7) and equation (2.8) we have to prove that

$$\tau_\theta^{-\frac{1}{2}} \cdot \varphi(0) > \tau^{-\frac{1}{2}} \cdot \varphi(-\mu\tau^{\frac{1}{2}}) - \underbrace{\mu \cdot \Phi(-\mu\tau^{\frac{1}{2}})}_{>0}$$

and

$$\tau_\theta^{-\frac{1}{2}} \cdot \varphi(0) > \tau^{-\frac{1}{2}} \cdot \varphi(-\mu\tau^{\frac{1}{2}}) - \underbrace{|\mu| \cdot [1 - \Phi(-\mu\tau^{\frac{1}{2}})]}_{>0}.$$

It is straightforward to see that $\tau_\theta < \tau = \tau_\theta + \tau_\varepsilon x$. Thus $\tau_\theta^{-\frac{1}{2}} < \tau^{-\frac{1}{2}}$. Further, the maximum of a standard normal density function $\varphi(z)$ is at $z = 0$ which means that $\varphi(0) \geq \varphi(-\mu\tau^{\frac{1}{2}})$ and consequently $\tau_\theta^{-\frac{1}{2}} \cdot \varphi(0) > \tau^{-\frac{1}{2}} \cdot \varphi(-\mu\tau^{\frac{1}{2}})$. Subtracting a positive term on each right-hand side further reduces the total value of each right-hand side.

Proof of part (b). Straightforward calculations show that

$$\begin{aligned}\frac{\partial\rho_{A^1}(\mu(x), \tau(x))}{\partial x} &= \frac{\partial\rho_{A^1}}{\partial\mu}\frac{\partial\mu}{\partial x} + \frac{\partial\rho_{A^1}}{\partial\tau}\frac{\partial\tau}{\partial x} \\ &= -\Phi(-\mu\tau^{\frac{1}{2}}) \cdot \frac{\tau_\theta\tau_\varepsilon s}{(\tau_\theta + \tau_\varepsilon x)^2} - \frac{1}{2}\tau^{-\frac{3}{2}}\varphi(-\mu\tau^{\frac{1}{2}}) \cdot \tau_\varepsilon < 0\end{aligned}$$

and

$$\begin{aligned}\frac{\partial\rho_{B^1}(\mu(x), \tau(x))}{\partial x} &= \frac{\partial\rho_{B^1}}{\partial\mu}\frac{\partial\mu}{\partial x} + \frac{\partial\rho_{B^1}}{\partial\tau}\frac{\partial\tau}{\partial x} \\ &= -[1 - \Phi(-\mu\tau^{\frac{1}{2}})] \cdot \frac{\tau_\theta\tau_\varepsilon |s|}{(\tau_\theta + \tau_\varepsilon x)^2} - \frac{1}{2}\tau^{-\frac{3}{2}}\varphi(-\mu\tau^{\frac{1}{2}}) \cdot \tau_\varepsilon < 0.\end{aligned}$$

□

Proof of Result 2.3. **The value of the information service.** The value of the information service is:

$$\omega = E[\ell_{v^0}(\theta)] - E\left\{E[\ell_{v^1}(\theta|S)]\right\}$$

and we can rewrite this equation as

$$\begin{aligned}\omega &= E\left\{E[\ell_{v^0}(\theta|S)\right\} - E\left\{E[\ell_{v^1}(\theta|S)]\right\}\\ &= E\left\{E[\ell_{v^0}(\theta|S)] - E[\ell_{v^1}(\theta|S)]\right\}\\ &= E\left\{E[\ell_{v^0}(\theta|S) - \ell_{v^1}(\theta|S)]\right\}\end{aligned}$$

where $E[\ell_{v^0}(\theta)] = E\left\{E[\ell_{v^0}(\theta|S)\right\}$ in the first line can be verified by applying the law of iterated expectation. The next step is to inspect the term in curly brackets in the last line. To do so we need to determine the optimal prior voting decision, v^0, and the optimal posterior voting decision, v^1. Given the prior belief, the citizens are indifferent between the two candidates. We assume with no loss of generality that the citizens elect candidate A whenever they are indifferent. Thus we continue with $v^0 = A$. The optimal posterior decision, v^1, depends on the realization of the signal and is characterized by

$$v^1 = \begin{cases} A & \text{for } E(\theta|s) \geq 0 \\ B & \text{for } E(\theta|s) < 0. \end{cases} \tag{2.26}$$

Substituting the optimal decisions into $\{E[\ell_{v^0}(\theta|S) - \ell_{v^1}(\theta|S)]\}$ we obtain

$$\begin{aligned} E[\ell_A(\theta|s) - \ell_{v^1}(\theta|s)] &= \begin{cases} E[\ell_A(\theta|s) - \ell_A(\theta|s)] & \text{for } E(\theta|s) \geq 0 \\ E[\ell_A(\theta|s) - \ell_B(\theta|s)] & \text{for } E(\theta|s) < 0 \end{cases} \\ &= \begin{cases} 0 & \text{for } E(\theta|s) \geq 0 \\ -E(\theta|s) & \text{for } E(\theta|s) < 0. \end{cases} \end{aligned} \tag{2.27}$$

We have already shown that $E(\theta|s) = \frac{\tau_\varepsilon x \cdot s}{\tau_\theta + \tau_\varepsilon x}$ and we can use this result to compute the pre-posterior distribution of $E(\theta|s)$. From the pre-posterior perspective, $E(\theta|s)$ is the random variable

$$\frac{\tau_\varepsilon x}{\tau_\theta + \tau_\varepsilon x} \cdot (\theta + \varepsilon), \tag{2.28}$$

where $\theta \sim N(0, \tau_\theta)$ and $\varepsilon \sim N(0, \tau_\varepsilon x)$. Recalling the rules for sums of independently distributed normal random variables we have the result that $E(\theta|s)$ is normally distributed with mean $m = 0$ and precision $r = \frac{\tau_\theta(\tau_\theta + x\tau_\varepsilon)}{x\tau_\varepsilon}$. Let G denote the distribution function of $E(\theta|S)$. Then, the value of the information service is

$$\omega = \int_{-\infty}^{0} -w \; dG(w) + \int_{0}^{\infty} 0 \; dG(w) \tag{2.29}$$

$$= \int_{-\infty}^{0} -w \; dG(w). \tag{2.30}$$

This is the partial expectation of a normal random variable. Inserting m and r into the functional form defined by equation (2.24) yields

$$\omega = \int_{-\infty}^{0} -wg(w)dw = r^{-\frac{1}{2}} \cdot \varphi(0).$$

Substituting for r we obtain

$$\omega(x) = \left(\frac{\tau_\theta(\tau_\theta + x\tau_\varepsilon)}{x\tau_\varepsilon}\right)^{-1/2} \cdot \varphi(0).$$

Concavity of $\omega(x)$. The first and second derivatives are

$$\omega'(x) = \frac{1}{2}\left(\frac{\tau_\theta^2}{\tau_\varepsilon x} + \tau_\theta\right)^{-3/2} \varphi(0)\frac{\tau_\theta^2}{\tau_\varepsilon x^2} > 0$$

and

$$\omega''(x) = \frac{1}{2}\varphi(0) \cdot \left[\frac{3}{2}r^{-\frac{5}{2}}\left(\frac{\tau_\theta^2}{\tau_\varepsilon x^2}\right)^2 - 2r^{-\frac{3}{2}}\frac{\tau_\theta^2}{\tau_\varepsilon x^3}\right]$$
$$= \frac{1}{2}\varphi(0)\frac{\tau_\theta^2}{\tau_\varepsilon x^3}r^{-\frac{3}{2}}\left[\frac{-(\tau_\theta + 4x\tau_\varepsilon)}{2(\tau_\theta + \tau_\varepsilon x)}\right] < 0.$$

Thus, $\omega(x)$ is a strictly concave function in x.

The slope of $\omega(x)$ at the origin.
We are interested in the limit value $\lim_{x\to 0} \omega'(x)$. To determine the limit value it is sufficient to inspect $\frac{1}{x^2} \cdot \left(\frac{\tau_\theta^2}{\tau_\varepsilon x} + \tau_\theta\right)^{-3/2}$. First we rearrange:

$$\left(\frac{\tau_\theta^2}{\tau_\varepsilon x}+\tau_\theta\right)^{-3/2}\cdot\frac{1}{x^2}=\left[\left(\frac{\tau_\theta^2}{\tau_\varepsilon x}+\tau_\theta\right)\cdot x^{4/3}\right]^{-3/2}=\left(\frac{\tau_\theta^2}{\tau_\varepsilon}x^{1/3}+\tau_\theta x^{4/3}\right)^{-3/2}.$$

It is easy to see that

$$\lim_{x\to 0}\left(\frac{\tau_\theta^2}{\tau_\varepsilon}x^{1/3}+\tau_\theta x^{4/3}\right)=0$$

and thus

$$\lim_{x\to 0}\left(\frac{\tau_\theta^2}{\tau_\varepsilon}x^{1/3}+\tau_\theta x^{4/3}\right)^{-3/2}=\infty.$$

□

Comparative statics analysis. The optimal overall quality $x^* = x^*(k, \tau_\theta, \tau_\varepsilon)$ is implicitly defined by the equation

$$\frac{1}{2}\left(\frac{\tau_\theta^2}{\tau_\varepsilon x}+\tau_\theta\right)^{-3/2}\varphi(0)\frac{\tau_\theta^2}{\tau_\varepsilon x^2}-k=0.$$

Though we cannot explicitly solve for x^*, comparative statics results can be derived by using the implicit-function rule. We define the function

$$z(x;k,\tau_\theta,\tau_\varepsilon):=\frac{1}{2}\left(\frac{\tau_\theta^2}{\tau_\varepsilon x}+\tau_\theta\right)^{-3/2}\varphi(0)\frac{\tau_\theta^2}{\tau_\varepsilon x^2}-k.$$

Given $z(x; k, \tau_\theta, \tau_\varepsilon) = 0$ the implicit-function rule yields

$$\frac{\partial x}{\partial \alpha}=-\frac{z_\alpha}{z_x}\quad\text{for } \alpha\in\{k,\tau_\theta,\tau_\varepsilon\}.$$

The denominator of $-\frac{z_\alpha}{z_x}$ is the second derivative $\omega''(x) = z_x$ and we have already shown above that $\omega''(x) < 0$. Hence, the sign of $\frac{\partial x}{\partial \alpha} = -\frac{z_\alpha}{z_x}$ is determined by the sign of z_α.

Effect of k. Since $z_k = -1$, we obtain the result that $\partial x^*/\partial k = -z_k/z_x < 0$. It follows for the individual total risk $\rho_t = \rho(x^*) + k \cdot (x^*/H)$ that

$$\frac{\partial \rho_t}{\partial k} = \frac{\partial \rho}{\partial x}\frac{\partial x}{\partial k} + \frac{x}{H} + \frac{k}{H} \cdot \frac{\partial x}{\partial k}$$
$$= \underbrace{\left(\frac{\partial \rho}{\partial x} + \frac{k}{H}\right)}_{<0} \underbrace{\frac{\partial x}{\partial k}}_{<0} + \frac{x}{H} > 0.$$

Effect of τ_ε. Computing z_{τ_ε} yields:

$$z_{\tau_\varepsilon} = \frac{1}{2}\varphi(0) \cdot \frac{\tau_\theta^2}{x^2} \left[\frac{\partial r^{-\frac{3}{2}}}{\partial \tau_\varepsilon} \cdot \frac{1}{\tau_\varepsilon} - r^{-\frac{3}{2}} \cdot \frac{1}{\tau_\varepsilon^2}\right]$$
$$= \frac{1}{2}\varphi(0) \cdot \frac{\tau_\theta^2}{x^2 \tau_\varepsilon^2} r^{-\frac{3}{2}} \left[\frac{\tau_\theta - 2\tau_\varepsilon x}{2(\tau_\theta + \tau_\varepsilon x)}\right],$$

and it follows that

$$\frac{\partial x^*}{\partial \tau_\varepsilon} = \frac{x^*}{\tau_\varepsilon} \cdot \frac{\tau_\theta - 2\tau_\varepsilon x}{\tau_\theta + 4\tau_\varepsilon x} \begin{cases} > 0 & \text{for } \tau_\theta > 2\tau_\varepsilon x^* \\ = 0 & \text{for } \tau_\theta = 2\tau_\varepsilon x^* \\ < 0 & \text{for } \tau_\theta < 2\tau_\varepsilon x^*. \end{cases}$$

Hence, we find that

$$\frac{\partial(\tau_\varepsilon x^*)}{\partial \tau_\varepsilon} = -\frac{z_{\tau_\varepsilon}}{z_x} \cdot \tau_\varepsilon + x = 2x \cdot \frac{(\tau_\theta + \tau_\varepsilon x)}{\tau_\theta + 4\tau_\varepsilon x} > 0.$$

Further, we can show that the individual total risk decreases in τ_ε:

$$\frac{\partial \rho_t}{\partial \tau_\varepsilon} = \frac{\partial \rho}{\partial x}\frac{\partial x}{\partial \tau_\varepsilon} + \frac{k}{H} \cdot \frac{\partial x}{\partial \tau_\varepsilon} + \frac{\partial \rho}{\partial \tau_\varepsilon}$$
$$= \frac{x}{\tau_\varepsilon(\tau_\theta + 4\tau_\varepsilon x)} \cdot \left[-\omega'(x) \cdot 2(\tau_\theta + \tau_\varepsilon x) + \frac{k}{H} \cdot (\tau_\theta - 2\tau_\varepsilon x)\right] < 0.$$

This follows because, in equilibrium, $\omega'(x) > k/H$ and $|2(\tau_\theta + \tau_\varepsilon x)| > |\tau_\theta - 2\tau_\varepsilon x|$.

Effect of τ_θ. Computing z_{τ_θ} yields:

$$\begin{aligned}
z_{\tau_\theta} &= \frac{1}{2}\varphi(0)\frac{1}{x^2\tau_\varepsilon}\left[\frac{\partial r^{-\frac{3}{2}}}{\partial \tau_\theta}\cdot\tau_\theta^2 + r^{-\frac{3}{2}}\cdot 2\tau_\theta\right] \\
&= \frac{1}{2}\varphi(0)\frac{1}{x^2\tau_\varepsilon}\left[-\frac{3}{2}r^{-\frac{5}{2}}\left(\frac{2\tau_\theta+\tau_\varepsilon x}{\tau_\varepsilon x}\right)\cdot\tau_\theta^2 + r^{-\frac{3}{2}}\cdot 2\tau_\theta\right] \\
&= \underbrace{\frac{1}{2}\varphi(0)\frac{\tau_\theta}{x^2\tau_\varepsilon}r^{-\frac{3}{2}}}_{>0}\left[\frac{-2\tau_\theta+\tau_\varepsilon x}{2(\tau_\theta+\tau_\varepsilon x)}\right],
\end{aligned}$$

and it follows that

$$\frac{\partial x^*}{\partial \tau_\theta} = \frac{x^*}{\tau_\theta}\cdot\frac{-2\tau_\theta+\tau_\varepsilon x}{\tau_\theta+4\tau_\varepsilon x}\begin{cases} >0 & \text{for } \tau_\theta < \tau_\varepsilon x^*/2 \\ =0 & \text{for } \tau_\theta = \tau_\varepsilon x^*/2 \\ <0 & \text{for } \tau_\theta > \tau_\varepsilon x^*/2. \end{cases}$$

The effect of an increase of τ_θ on the total individual risk is:

$$\begin{aligned}
\frac{\partial \rho_t}{\partial \tau_\theta} &= \frac{\partial \rho}{\partial x}\frac{\partial x}{\partial \tau_\theta} + \frac{k}{H}\cdot\frac{\partial x}{\partial \tau_\theta} + \frac{\partial \rho}{\partial \tau_\theta} \\
&= -\omega'(x)\cdot r^{\frac{3}{2}}\cdot\frac{\tau_\varepsilon x^2}{\tau_\theta^2}\left(1-\frac{4(\tau_\theta+\tau_\varepsilon x)^{\frac{1}{2}}(\tau_\varepsilon x)^{\frac{1}{2}}}{(\tau_\theta+4\tau_\varepsilon x)}\right) + \frac{k}{H}\cdot\frac{\partial x}{\partial \tau_\theta}.
\end{aligned}$$

We define $\gamma := \left(1-\frac{4(\tau_\theta+\tau_\varepsilon x)^{\frac{1}{2}}(\tau_\varepsilon x)^{\frac{1}{2}}}{(\tau_\theta+4\tau_\varepsilon x)}\right)$ and a straightforward calculation shows that

$$\gamma\begin{cases} >0 & \text{for} \quad \tau_\theta > 8\tau_\varepsilon x^* \\ =0 & \text{for} \quad \tau_\theta = 8\tau_\varepsilon x^* \\ <0 & \text{for} \quad \tau_\theta < 8\tau_\varepsilon x^*. \end{cases}$$

Recalling our results concerning $\frac{\partial x}{\partial \tau_\theta}$ it follows that

$$\frac{\partial \rho_t}{\partial \tau_\theta} < 0 \text{ for } \tau_\theta/\tau_\varepsilon x^* > 8$$

and

$$\frac{\partial \rho_t}{\partial \tau_\theta} > 0 \text{ for } \tau_\theta/\tau_\varepsilon x^* < 1/2.$$

For the domain $1/2 < \tau_\theta/\tau_\varepsilon x^* < 8$, however, the sign of $\frac{\partial \rho_t}{\partial \tau_\theta}$ depends on the respective magnitude of both the change in the electoral risk $(-\omega'(x) \cdot r^{\frac{3}{2}} \cdot \frac{\tau_\varepsilon x^2}{\tau_\theta^2} \cdot \gamma)$ and in the price of a subscription $\frac{k}{H} \cdot \frac{\partial x}{\partial \tau_\theta}$. □

3 Public good provision, electoral selection and information acquisition

3.1 Introduction

In this chapter, we examine how media activity shapes incumbent behavior. As in the last chapter, the electorate uses elections to select competent politicians, but in contrast to the previous analysis, we focus on the incentives of an incumbent politician to behave well in order to get reelected. The intuition of our analysis is simple: When the electorate infers an incumbent's talent from his performance in a policy task, the incumbent can try to boost the policy outcome in order to make the electorate believe that he is very talented. In more technical terms, we are dealing with information manipulation by incumbent politicians.

To which extent a politician's performance can serve to evaluate his competence, however, depends on the ability of the electorate to observe his achievements. If, for example, voters are rather ignorant about politics and have only very fuzzy knowledge about the responsibilities of politicians, then it is not very likely that a good performance of a politician significantly influences the voters' perception of his competence. Think, for instance, of a voter who enjoys driving on a road but who does not know that it was the mayor, for example, who made it possible to build or to maintain the road. This voter will hardly consider the mayor's contribution when he contemplates whether or not to reelect him. Or else, the voter knows that the mayor was involved but, in addition, he knows that the central government also provided funds and thus he wonders how much the mayor contributed.

Our aim is to explain how the media market influences this 'fuzziness' of the voters' perceptions and the resulting behavior of incumbent politicians. We propose a theoretical framework based on a career concern model (see Holmström (1999) and Persson and Tabellini (2002)) where an incumbent politician provides a public good and the electorate can obtain information about his performance by subscribing to a newspaper before an election. The model allows to

- characterize the electorate's willingness to pay for information,
- describe how the informational situation of the electorate at the time of an election depends on the technology of information production and on prior knowledge,
- compare the informational situation of the electorate when either a competitive or a monopoly newspaper market provides information,

- examine how information provision by the newspaper market influences the incumbent's incentives to manipulate public good production.

From a technical perspective, a novel feature of the model is that we integrate Bayesian sampling theory into the career concern framework to model information production. Another feature that distinguishes our approach from other studies on endogenous information provision is that voters do not obtain political information as a by-product of some private activity but they strive to get more informed in order to avoid selecting less competent politicians.

Our results emphasize that media activity is an important factor in explaining incumbent behavior. The central finding is that higher newspaper quality strengthens a politician's incentives to behave well in office. One important determinant of newspaper quality is the technology of information production. We find that a more efficient technology, in terms of lower costs to produce a certain quality or more informative signals resulting from an investigation, improves the informational situation of the electorate which makes incumbent politicians work harder. This means, for example, that cost factors such as the wages of journalists or the geographical distance of a jurisdiction to the next news desk influence policy outcomes. Further, the qualification of journalists or the complexity of policy tasks may affect the amount of information resulting from a certain effort of investigation.

We also find that a monopoly newspaper market creates the same incentives for a politician as a competitive market because both market forms provide the same quality of information. This result appears to be an interesting aspect regarding the current situation where many local newspaper markets become monopolies. Finally, we find that more prior information about a politician's competence can weaken the incentives of politicians when media activity is very low. Hence, having more prior information is not always beneficial from the voters' perspective.

Being closely related to the previous chapter, this analysis shares connections to the political economy literature on endogenous information provision with the previous analysis, in particular to the articles by Strömberg (2004a) and Martinelli (2006). There are further connections to various articles which also use a career concern model to study political economy problems: Alesina and Tabellini (2007, 2008) rely on this framework to compare the incentives of bureaucrats and politicians in order to optimally allocate policy tasks to each type, whereas Ashworth (2005) examines the incumbency advantage of politicians and how the incentives of incumbents change the longer they are in office. In another article, Gehlbach (2007) compares the incentives of national politicians to provide local public goods when there are electoral-college or majoritarian elections. The career concern framework can also be used to study how fiscal equalization affects the accountability of politicians when there is yardstick competition as shown by Kotsogiannis and Schwager (2008).

Within the political economy literature on elections, our analysis belongs to the branch which assumes that elections serve to select good politicians and that the incentives of incumbent politicians arise rather as a by-product of the voters' strategy to select good types. See Ashworth et al. (2010) for a discussion of the question if elections serve to select good politicians or to create incentives.

3.2 Description of the model

There are two time periods. In period 1, there is an incumbent politician (the incumbent) who provides a public good g and there is a homogenous electorate (the citizen) who either reelects the incumbent or selects a challenger at the end of period 1. Before the election takes place, the citizen can obtain information about the incumbent by purchasing an information service that is provided either by two competing newspapers $n = 1, 2$ or by a monopoly newspaper. In period 2, the winner of the election implements the public good.

Next, we will explain the incentives of the players in detail.

3.2.1 Public good production by the incumbent

The level of the public good in each period results from the incumbent's effort $a_t \geq 0$ in the respective period and his competence θ^I according to the additive production function

$$g_t = a_t + \theta^I. \tag{3.1}$$

We assume that the incumbent's competence remains constant over time whereas the level of effort is a period-specific choice. It will become clear below that the citizen has an incentive to retain competent politicians in office and that incumbents can try to appear more competent by investing effort in public good production. Effort, however, is costly and the cost resulting from effort is described by the function $\gamma(a)$ which is assumed to be strictly convex with $\gamma(0) = 0$, $\gamma'(a) > 0$, $\gamma(a)'' > 0$ and $\lim_{a \to 0} \gamma'(a) = 0$.

We build the model in the spirit of the career concern literature originated by Holmström (1999) which was adapted to political agency problems by Persson and Tabellini (2002). A central assumption of this modeling approach is that the incumbent does not know his own competence when he selects his effort level but that competence is a random variable for the incumbent and that he shares a common prior of this random variable with all other players. Here, we assume that all players share the common prior that the competence of a politician is drawn from a normal distribution with mean $\bar{\theta}$ and precision τ_θ, or in formal notation: $\theta^I, \theta^C \sim N(\bar{\theta}, \tau_\theta)$, where θ^C denotes the competence of a randomly selected challenger.

The assumption of symmetric information about competence implies that there are no signaling issues in our model which would arise if we assumed that the incumbent knows his own competence in contrast to the citizen and the newspapers. This assumption is debatable, of course, and one could argue that politicians should be well-informed about their qualifications. This notion, however, would tend to apply a rather narrow definition of competence such as academic degrees, for instance. But when it comes to implementing policies, politicians have to deal with many different actors such as bureaucrats, lobbyists, party leaders, or businessmen and it appears reasonable to assume that a politician is uncertain whether his skills fit well into the political-economic situation. Ashworth (2005) and Besley (2006) argue in a similar way.

So the level of the public good results from the realization of the random variable *competence* and the action *effort level* chosen by the incumbent. This action can be interpreted in various ways: On the one hand, we can take it literally and say that effort denotes how much time an incumbent devotes to activities like attracting grant monies, monitoring bureaucrats or negotiating contracts. According to this interpretation, working hard reduces the time that is left for enjoying the amenities associated with political offices. But the model can also be interpreted as a rent-seeking story. Alesina and Tabellini (2007) or Gehlbach (2007), for example, show how this kind of 'effort story' can easily be transformed into a rent-seeking model. The basic idea in both variants is that politicians can influence policy outcomes via a costly action. In a rent-seeking context, politicians typically would decide how much money they divert from productive purposes to rents which may include opulent amenities and salaries, payments to political parties or inefficient contracts with cronies. Then, less rent-seeking increases public good production while it reduces the immediate benefits of politicians. So whenever we argue in terms of *higher* or **lower** levels of effort there is always a corresponding interpretation in terms of *lower* or **higher** levels of rent-seeking.

The incumbent knows that the citizen will use information about his performance in public good production to infer whether or not to reelect him. Thus he can try to influence the probability of his reelection $p(a)$ by exerting effort and his objective in period 1 is

$$p(a_1) \cdot [R - \gamma(a_2)] - \gamma(a_1), \tag{3.2}$$

where $R > 0$ denotes an exogenous rent from being in office in period 2 for the winner of the election.[6] So the incumbent weighs the cost of effort in period 1 against the expected net rent in period 2. The level of effort he chooses depends on the mapping of effort a into the probability of reelection $p(a)$. We have to analyze the decision problem of the citizen to determine this probability.

3.2.2 The citizen's decision problem

The citizen receives utility $u(g) = g$ from the public good and his only interest is to select a politician who provides a high level of the public good in period 2. He can either reelect the incumbent or replace him by a challenger whose competence is drawn from the same distribution as the incumbent's competence. Formally, we denote the citizen's decision by $v \in \{I, C\}$ where I and C are the decisions to elect the incumbent or the challenger, respectively.

Since g_2^I and g_2^C depend on the realizations of the random variables θ^I and θ^C and the effort of an incumbent in period 2 the citizen can only form expectations and elect the candidate whom he expects to provide more of the public good. Although this is the best rational decision he can make it can be wrong in such a way that the alternative would have provided a higher utility. In this case, the citizen has missed the opportunity to select the best candidate because of the uncertainty in his decision. We denote the difference between the consequence of the optimal decision and the consequence of the decision actually made as the *opportunity loss* of the decision which can be interpreted as a measure of the cost of uncertainty in a decision problem.

It is shown below that the solution to the citizen's decision problem in terms of opportunity loss yields the same optimal decision as relying on expected utility. The opportunity loss concept, however, allows to bring out certain features of the decision problem more clearly. Most of all, it enables us to derive a well-defined willingness to pay for additional information of the citizen.

In our setup, the opportunity loss (the loss) of an electoral decision is the difference between the level of g_2 that is associated with this decision and the level that would result from an optimal decision. Formally, the loss that results from reelecting the incumbent is

$$\ell_I\left(g_2^I, g_2^C\right) = \begin{cases} 0 & \text{for} \quad g_2^I \geq g_2^C \\ g_2^C - g_2^I & \text{for} \quad g_2^I < g_2^C \end{cases}$$

[6] There is no discounting.

and

$$\ell_C\left(g_2^I, g_2^C\right) = \begin{cases} g_2^I - g_2^C & \text{for} \quad g_2^I \geq g_2^C \\ 0 & \text{for} \quad g_2^I < g_2^C \end{cases}$$

is the loss that results from selecting the challenger. The citizen does not suffer a loss if he makes an optimal decision, of course.

At the time of the election, the citizen does not know the exact values of ℓ_I and ℓ_C but he can compute the expected loss (the *risk*) of each decision, $\rho_I := E[\ell_I]$ and $\rho_C := E[\ell_C]$. Before making a decision based on these risks, the citizen can obtain additional information about the incumbent: If he subscribes to a newspaper, he receives a signal s_n that contains information about the incumbent's performance in public good production which he can use to revise his belief about the incumbent's competence.[7] Then he can select a candidate based on the revised belief.

Thus, the citizen contemplates whom to vote for and whether he should subscribe to a newspaper in order to minimize his total risk

$$\rho_t = \rho_v + q_n \tag{3.3}$$

associated with the electoral decision, where ρ_v denotes the risk of the electoral decision and q_n the price of a newspaper subscription.

3.2.3 The newspaper market

We consider both a competitive market, where the two newspapers $n = 1, 2$ offer subscriptions, and the monopoly case with only one newspaper. In both cases, newspapers have the following identical production technology:

[7] We assume that the citizen cannot observe the incumbent's performance without reading a newspaper in order to simplify the analysis. Since the level of the public good affects the citizen's utility, it would be more realistic to assume that the citizen receives a (noisy) signal about public good production and that reading the newspaper provides additional information. For example, total utility could be composed of both the incumbent's contribution and another component independent from the incumbent (think of the economic situation in general or contributions from other politicians). This could be interpreted as a direct noisy signal and information from newspapers then would help the citizen to assess the incumbent's contribution. As long as we assume that the direct signal does not reveal the exact level of g our simplification does not change the results. An alternative interpretation would be that the public good has long-term consequences so that the citizen does not observe his utility before the election as in Besley and Prat (2006).

Each newspaper can produce a noisy signal about the incumbent's performance in public good production:

$$s_n = g + \varepsilon_n \tag{3.4}$$

where $\varepsilon_n \sim N(0, \tau_\varepsilon \cdot x_n)$. A newspaper can manipulate the precision of its signal by choosing a value for the variable x_n where every unit of x_n comes at a cost $k > 0$ that may denote the hourly wage paid to a journalist, for example.

We can interpret this technology as follows: One hour of work produces a signal $s_{ni} = g + \varepsilon_{ni}$, where $\varepsilon_{ni} \sim N(0, \tau_\varepsilon)$. The precision of a single signal (τ_ε) can be understood to measure how complex the investigated topic or how talented a journalist is. Following this interpretation, x_n denotes the total amount of working hours a newspaper allocates to investigating the incumbent's performance.[8] Then, each signal s_n is simply the mean of the single signals s_{ni} a newspaper produces:

$$s_n = \frac{1}{x_n} \sum_{i=1}^{x_n} s_{ni} = g + \frac{1}{x_n} \sum_{i=1}^{x_n} \varepsilon_{ni} = g + \varepsilon_n. \tag{3.5}$$

As we are dealing with sampling from a normal random process, the pair (s_n, x_n) is a sufficient statistic for the whole sample of signals a newspaper produces and we call x_n the quality of a newspaper. The quality of a newspaper and the price q_n together describe a subscription package (q_n, x_n) that each newspaper offers.

The objectives of the newspapers and the timing in this stage are as follows:

Competition. The newspapers non-cooperatively and simultaneously announce their subscription packages (q_1, x_1) and (q_2, x_2) to maximize their profits

$$\Pi_1[(q_1, x_1), (q_2, x_2)] = D_1[(q_1, x_1), (q_2, x_2)] \cdot (q_1 - k \cdot x_1)$$

and

$$\Pi_2[(q_1, x_1), (q_2, x_2)] = D_2[(q_1, x_1), (q_2, x_2)] \cdot (q_2 - k \cdot x_2),$$

where $D_n \in \{0, 1\}$ takes on the value 1 if the citizen subscribes to the respective newspaper and 0 otherwise. After the citizen has subscribed to a newspaper, the respective newspaper investigates the incumbent's performance and then it sends its signal to the citizen.

[8]We disregard integer problems throughout the analysis.

Monopoly. A monopolist offers a subscription package (q_m, x_m) to maximize

$$\Pi_m(q_m, x) = D_m(q_m, x_m) \cdot (q_m - k \cdot x_m).$$

Before we study the equilibrium, we briefly summarize the game:

Period 1:

- Nature selects the competence of the incumbent θ^I which remains unknown to all players.
- The incumbent chooses the effort level a_1 and the level of the public good $g_1 = a_1 + \theta^I$ is realized.
- The newspapers offer their subscription packages and then they investigate the incumbent's performance.
- The citizen reads the newspaper, revises his belief about the competence of the incumbent and then he votes for either the incumbent or the challenger.

Period 2:

- The winner of the election chooses an effort level and the level of public good production is either $g_2^I = a_2^I + \theta^I$ or $g_2^C = a_2^C + \theta^C$.

3.3 Equilibrium

In this section, we analyze the equilibrium behavior of the three players. Please note that, as to the game between the incumbent and the citizen, no out-of-equilibrium action of the incumbent can be observed because the distributions of the random variables θ^I and ε have full support on the real line. From the citizen's perspective, every signal he receives is consistent with his equilibrium belief, because he attributes the difference between a realization of the signal and his belief to realizations of the random variable $\theta^I + \varepsilon$ which can be every value on the real line.

3.3.1 Electoral decision

Consider the public good production in period 2. Neither the incumbent nor the challenger will provide any effort because effort is costly and the game will end after period 2. Thus,

$a_2^I = a_2^C = 0$ and the level of the public good in period 2 will be determined only by the competence of the office holder, i.e. the level is either $g_2 = \theta^I$ or $g_2 = \theta^C$. The citizen understands this result when he has to decide if he should vote for the incumbent or the challenger. As he knows that g_2 solely depends on competence, the losses associated with his decision are

$$\ell_I\left(\theta^I, \theta^C\right) = \begin{cases} \theta^I - \theta^I & \text{for} \quad \theta^I \geq \theta^C \\ \theta^C - \theta^I & \text{for} \quad \theta^I < \theta^C \end{cases}$$

$$\ell_C\left(\theta^I, \theta^C\right) = \begin{cases} \theta^I - \theta^C & \text{for} \quad \theta^I \geq \theta^C \\ \theta^C - \theta^C & \text{for} \quad \theta^I < \theta^C. \end{cases}$$

Let us define the difference between the actual competence values of the incumbent and the challenger as $\theta := \theta^I - \theta^C$. Recalling the distribution of θ^I and θ^C we know that θ is a normal random variable with mean zero and precision $\tau_\theta/2$. With respect to θ the losses from reelecting the incumbent and from electing the challenger are

$$\ell_I(\theta) = \begin{cases} 0 & \text{for } \theta \geq 0 \\ |\theta| & \text{for } \theta < 0 \end{cases}$$

$$\ell_C(\theta) = \begin{cases} \theta & \text{for } \theta \geq 0 \\ 0 & \text{for } \theta < 0. \end{cases}$$

The citizen is uncertain about the value of θ and he strives to minimize the risk ρ_v of his decision. Depending on the voting decision the risk is either

$$\rho_I(\theta) = \int_{-\infty}^{0} |\theta| \; dF(\theta)$$

or

$$\rho_C(\theta) = \int_{0}^{\infty} \theta \; dF(\theta),$$

where F denotes the distribution function of θ. Straightforward calculation (see the appendix) shows that the citizen selects the incumbent if $E(\theta) > 0$ and the challenger if $E(\theta) < 0$. Formally, we have

$$v^0 = \begin{cases} I & \text{for } E(\theta) > 0 \\ C & \text{for } E(\theta) < 0, \end{cases}$$

where v^0 denotes the decision against the prior distribution (F) of competence. Given his prior, the citizen is indifferent between selecting I or C because the prior mean is $E(\theta) = 0$ which means that, without additional information, he expects the incumbent to be as good as the challenger.

Before making a decision, however, the citizen can get additional information about the incumbent's competence from the newspaper market. Recall that subscribing to a newspaper of quality x implies that the citizen receives a signal

$$s = a + \theta^I + \varepsilon \tag{3.6}$$

where $\theta^I \sim N(\bar{\theta}, \tau_\theta)$ and $\varepsilon \sim N(0, \tau_\varepsilon x)$. The signal contains information that the citizen can use to revise his belief about the incumbent's competence and possibly change his voting decision. As to the updating process, we assume that the citizen is rational in the sense that he updates his belief using Bayes' rule, given his belief $\tilde{a}$ about the incumbent's effort.

The citizen does not know the optimal posterior voting decision before having observed the signal, of course, but he can compute the posterior distribution of $\theta|s$ for every possible realization of the signal given a quality x. This posterior distribution of $\theta|s$ is a normal distribution with mean

$$\begin{aligned} \mu &= E\left(\theta^I|s, \tilde{a}\right) - E(\theta^C) \\ &= \frac{\tau_\theta \bar{\theta} + \tau_\varepsilon x \cdot (s - \tilde{a})}{\tau_\theta + \tau_\varepsilon x} - \bar{\theta} \\ &= \frac{\tau_\varepsilon x}{\tau_\theta + \tau_\varepsilon x} \cdot (s - \tilde{a} - \bar{\theta}) \end{aligned}$$

and precision $\tau = \tau_\theta + \tau_\varepsilon x$. In technical terms, s can be considered a noisy signal of θ^I which is biased by the incumbent's action. Therefore, the citizen intends to correct the signal for the bias and the posterior mean is a weighted average of the prior mean $\bar{\theta}$ and the observed value of s corrected by the belief $\tilde{a}$. The weights depend on the precisions of the prior and the signal in an intuitive way: a higher precision increases the weight given to the respective values of the prior belief and of an observation of the signal. Thus, higher newspaper quality (higher x) makes the citizen form higher expectations of the incumbent's competence for a given value of s and belief $\tilde{a}$.[9] Further, we can see that the posterior precision of $\theta|s$ is strictly increasing in x.

[9] A straightforward manipulation shows that $\frac{\tau_\varepsilon x}{\tau_\theta + \tau_\varepsilon x} = \frac{1}{\tau_\theta / \tau_\varepsilon x + 1}$, which facilitates to see this effect from x on updating.

Having determined the posterior distribution of $\theta|s$, we can characterize both the optimal posterior voting decision and the respective posterior risk for every realization of s. The optimal posterior decision is

$$v^1 = \begin{cases} I & \text{for } E(\theta|s) > 0 \\ C & \text{for } E(\theta|s) < 0 \end{cases}$$

and the respective posterior risks ρ_{I^1} and ρ_{C^1} are

$$\rho_{I^1} = \int_{-\infty}^{0} -\theta \; dF(\theta|s)$$

and

$$\rho_{C^1} = \int_{0}^{\infty} \theta \; dF(\theta|s),$$

where $F(\theta|s)$ denotes the posterior distribution function of θ, which is a normal distribution with mean μ and precision τ. From the posterior perspective, the value that the citizen attaches to a signal he has received is the difference between the prior risk and the posterior risk: $\rho_{v^0} - \rho_{v^1} = E[\ell_{v^0}(\theta)] - E[\ell_{v^1}(\theta|s)]$. But the citizen has to compute the value of information provided by a newspaper before he knows the realization of s, i.e. from the prior perspective. Prior to observing s, he can compute the prior risk ρ_{v^0} whereas, for given $\tilde{a}$, the posterior risk ρ_{v^1} depends on an unknown realization of the random variable s.

The citizen can use his knowledge about the distribution of s and about the posterior situation for a given realization s to form expectations: From the prior perspective, the posterior mean, $E(\theta|s)$, that determines the posterior decision is the normal random variable

$$\frac{\tau_\varepsilon x}{\tau_\theta + \tau_\varepsilon x} \cdot (a - \tilde{a} + \theta^I - \bar{\theta} + \varepsilon)$$

with mean zero and precision $r = \frac{\tau_\theta(\tau_\theta + \tau_\varepsilon x)}{\tau_\varepsilon x}$. Both of these parameters do not depend on the action of the incumbent which means that only the parameters of the random variables θ and ε determine the value of information, whereas the incumbent cannot influence it.

Knowing the distribution of $E(\theta|s)$, the citizen can form expectations about the posterior risk: $E\left\{E[\ell_{v^1}(\theta|s)]\right\}$, where the outer expectation operator refers to s and the inner expectation operator refers to θ. Following Raiffa and Schlaifer (1961) we call this prior expectation of the posterior risk the *pre-posterior* risk and define $\rho := E\left\{E[\ell_{v^1}(\theta|s)]\right\}$.

Thus, the value ω that the citizen attaches to a newspaper subscription is the difference between the prior risk and the pre-posterior risk:

$$\begin{aligned}\omega &= E[\ell_{v^0}(\theta)] - E\left\{E[\ell_{v^1}(\theta|s)]\right\} \\ &= \rho_{v^0} - \rho.\end{aligned}$$

Given the parameters of our model, it can be shown that:

Proposition 3.1. *From the pre-posterior perspective, the value of subscribing to a newspaper of quality x can be described by the function*

$$\omega(x) = \left(\frac{\tau_\theta(\tau_\theta + \tau_\varepsilon x)}{\tau_\varepsilon x}\right)^{-1/2} \cdot \varphi(0) \tag{3.7}$$

where φ is the density function of the standard normal distribution. The function $\omega(x)$ is strictly increasing in x and concave with infinite slope at the origin.
Further, depending on x the pre-posterior risk is

$$\rho(x) = \rho_{v^0} - \omega(x) = \tau_\theta^{-1/2}\varphi(0) - \left(\frac{\tau_\theta(\tau_\theta + \tau_\varepsilon x)}{\tau_\varepsilon x}\right)^{-1/2} \cdot \varphi(0). \tag{3.8}$$

Proof. See appendix. □

This result says that the value of a newspaper subscription increases in the quality x but at a decreasing rate. If, for example, journalists already have comprehensively investigated the incumbents' performance further efforts produce little additional value only. Having derived the citizen's willingness to pay for additional information, we examine now which quality of information the newspapers offer in equilibrium.

3.3.2 Information provision

The quality x of the information service determines the pre-posterior risk $\rho(x)$ of the electoral decision and, having determined his optimal voting behavior for every possible information set, the citizen decides whether or not to subscribe to a newspaper in order to minimize the total pre-posterior risk

$$\begin{aligned}\rho_t &= \rho(x) + q_n \\ &= \rho_{v^0} - \omega(x) + q_n.\end{aligned}$$

Again, for ease of exposition, we prefer to analyze the equivalent problem of maximizing the net value of subscribing to a newspaper

$$\omega(x_n) - q_n$$

in order to avoid distracting notation. The citizen subscribes to a newspaper only if the condition

$$\omega(x_n) - q_n \geq 0$$

is satisfied, i.e. the net value of subscribing must not be negative. This is the participation constraint of the citizen.

Competition. Recall that the newspapers choose qualities and prices in order to maximize their profits

$$\Pi_1[(q_1, x_1), (q_2, x_2)] = D_1[(q_1, x_1), (q_2, x_2)] \cdot (q_1 - k \cdot x_1) \tag{3.9}$$

and

$$\Pi_2[(q_1, x_1), (q_2, x_2)] = D_2[(q_1, x_1), (q_2, x_2)] \cdot (q_2 - k \cdot x_2). \tag{3.10}$$

For given offers (q_1, x_1) and (q_2, x_2) the citizen subscribes to newspaper 1 if and only if

$$\omega(x_1) - \omega(x_2) > q_1 - q_2$$

and to newspaper 2 if and only if

$$\omega(x_1) - \omega(x_2) < q_1 - q_2.$$

If the citizen is indifferent between both offers, we assume that he subscribes to each newspaper with probability 1/2. Next, we use this result to study the price setting stage between the newspapers.

Price equilibrium. In this section, we analyze the price-setting behavior of the newspapers for given qualities x_1 and x_2 of the subscriptions. Consider a decision of the citizen for a given pair of offers as derived in the previous section: At least one newspaper has always an incentive to undercut its rival's price in order to attract the citizen. Clearly, in response, the undercut newspaper lowers its price until the citizen favors its offer again and this Bertrand-like price competition pushes prices downwards. We have to distinguish between price-equilibria with equal and unequal given qualities:

- *Unequal qualities.* Assume with no loss of generality that $x_1 > x_2$. If newspaper 1 offers a higher quality, it charges

$$q_1^* = \omega(x_1) - \omega(x_2) + kx_2$$

and earns a positive profit. If newspaper 1 charges a higher price, newspaper 2 can undercut this price and capture the citizen. However, to ensure non-negative profits, newspaper 2 cannot announce a price lower than $q_2 = k \cdot x_2$ and thus it cannot undercut if newspaper 1 charges q_1^*. Consequently, q_1^* is the profit-maximizing price for newspaper 1 while the price of newspaper 2 is $q_2^* = k \cdot x_2$.

  For $x_1 < x_2$ it follows that

$$q_2^* = \omega(x_2) - \omega(x_1) + kx_1 \quad \text{and} \quad q_1^* = k \cdot x_1.$$

- *Equal qualities.* In case of equal qualities the newspapers can undercut each other's prices as long as the prices are above the cost to produce a subscription of quality x_1 or x_2. Thus, the only price equilibrium has both newspapers set equal prices $q_1^* = q_2^*$, which are characterized by

$$q_1^* = k \cdot x_1 \quad \text{and} \quad q_2^* = k \cdot x_2.$$

Having analyzed the price setting stage we can now turn to the first stage where the newspapers choose the qualities of the subscriptions they offer.

Quality equilibrium. The profits of selling a subscription to the citizen are

$$\Pi_1(x_1, x_2) = \begin{cases} q_1^*(x_1, x_2) - k \cdot x_1 & \text{for } x_1 > x_2 \\ 0 & \text{for } x_1 \leq x_2 \end{cases}$$

and

$$\Pi_2(x_1, x_2) = \begin{cases} 0 & \text{for } x_1 \geq x_2 \\ q_2^*(x_1, x_2) - k \cdot x_2 & \text{for } x_1 < x_2. \end{cases}$$

It is a weakly dominant strategy for each newspaper to choose the quality that maximizes the profit of selling a subscription for a given choice of the other newspaper. Due to the symmetry of the newspapers' maximization problem, the optimal quality of a subscription is characterized for both newspapers by the first-order condition

$$\omega'(x_n) = k. \tag{3.11}$$

This condition says that each newspaper chooses a quality by which the citizen's marginal willingness to pay for higher quality equals the cost of producing one more unit of quality. As the function $\omega(x)$ is strictly concave, there is a unique solution to equation (3.11).

Since this condition is identical for both newspapers, they offer identical subscriptions which are characterized by the quality $x_n^* = x_1^* = x_2^*$ and the price $q_n^* = k \cdot x_n^*$ where x_n^* denotes the quality that satisfies the first-order condition. Consequently, both newspapers earn zero profits in equilibrium.

Therefore, we can describe the newspaper quality in a competitive equilibrium as follows:

Proposition 3.2. *The competitive market provides information of a quality x_n^* that is implicitly defined by*

$$\omega'(x_n^*) = k. \tag{3.12}$$

and the citizen pays the price $q_n^* = k \cdot x_n^*$ for a subscription. Further, the citizen's total risk is

$$\rho_t(x_n^*) = \rho(x_n^*) + k \cdot x_n^* \tag{3.13}$$

Having determined the competitive equilibrium, we turn to the case of a monopoly newspaper.

Monopoly. A monopoly newspaper offers a subscription package (q_m, x_m) to maximize

$$\Pi_m(q_m, x_m) = D(q_m, x_m) \cdot (q_m - k \cdot x_m).$$

The monopolist can use his market power to extract the whole surplus the citizen receives from a signal of quality x_m and charges the price

$$q_m^* = \omega(x_m)$$

for a given quality. This is the price where the participation (subscription) constraint of the citizen reads $\omega(x_m) - q_m = 0$, i.e. the citizen is indifferent between subscribing and not subscribing to the monopoly newspaper. Thus, the monopolist chooses the quality that maximizes

$$\Pi_m(x_m) = \omega(x_m) - k \cdot x_m$$

and this quality is characterized by the first-order condition

$$\omega'(x_m^*) = k. \tag{3.14}$$

Hence, information provision with a monopoly newspaper can be characterized as follows:

Proposition 3.3. *The monopoly newspaper provides an information service of quality x_m^* that is implicitly defined by*

$$\omega'(x_m^*) = k. \tag{3.15}$$

and the citizen pays the price $q_m^* = \omega(x_m^*)$ for a subscription. Thus, the citizen's total risk is

$$\begin{aligned}\rho_t(x_m^*) &= \rho(x_m^*) + q_m^* \\ &= \rho_{v^0} - \omega(x_m^*) + \omega(x_m^*) \\ &= \rho_{v^0}\end{aligned}$$

Comparing Proposition 3.2 to Proposition 3.3 we can see that both competitive information provision and monopoly provision result in the same quality of newspapers. This finding originates from the similar incentives of competitive and monopoly newspapers to exploit the citizen's willingness to pay for information as far as their constraints allow: While a monopoly newspaper only has to consider the participation constraint of the citizen, the profit-maximization of competing newspapers is additionally constrained by the strategies of their rivals. For a given strategy of a rival, a competing newspaper can, similar to a monopolist, try to offer a subscription package that is favored by the citizen and to extract the residual surplus of the citizen with respect to the offer by the rival. This incentive just as the incentive of a monopolist makes newspapers offer a quality where the marginal value of additional information is equal to the marginal cost. Proposition 3.2, however, shows that competition prevents newspapers from exploiting the citizen's willingness to pay. This is why the citizen is strictly better off with competitive information provision, where the total pre-posterior risk is smaller than the prior risk $\rho_t(x_n^*) < \rho_{v^0}$ in contrast to a monopoly where $\rho_t(x_m^*) = \rho_{v^0}$ because the monopolist uses his market power to charge the price $q_m^* = \omega(x_m^*)$.

Most important for the remainder of the analysis is the result that the unique quality of information does not depend on the market structure but on the parameters of the model $(\tau_\theta, k, \tau_\varepsilon)$ only.[10] This result is summarized in

Proposition 3.4. *A monopoly newspaper market provides information of the same quality as a competitive market: $x_c^* = x_m^* = x^*$. Thus, recalling the functional form of $\omega(x)$, the quality of information x^* in both scenarios is implicitly defined by*

$$\frac{1}{2}\varphi(0)\frac{\tau_\theta^2}{\tau_\varepsilon x^2}\left(\frac{\tau_\theta^2}{\tau_\varepsilon x}+\tau_\theta\right)^{-3/2} = k. \tag{3.16}$$

Hence, the equilibrium quality of information is a function $x^* = x^*(\tau_\theta, k, \tau_\varepsilon)$ and we can derive comparative statics results[11] in order to characterize the equilibrium quality with respect to the technology of producing information (k and τ_ε) and the precision of the prior knowledge about competence (τ_θ):

(a) *Production technology (k and τ_ε).* Recall that $\omega'(x)$ is a strictly decreasing function. Thus, for lower values of k the quality in equilibrium is higher: $\partial x^*/\partial k < 0$.

We find that the effect of a higher precision of an observation (τ_ε) on x^* is ambiguous:

$$\frac{\partial x^*}{\partial \tau_\varepsilon}\begin{cases} > 0 & \text{for } \tau_\theta > 2\tau_\varepsilon x^* \\ = 0 & \text{for } \tau_\theta = 2\tau_\varepsilon x^* \\ < 0 & \text{for } \tau_\theta < 2\tau_\varepsilon x^*, \end{cases}$$

so that the direction of the effect depends on the relation between the precision of the prior and the precision of newspaper investigations. Intuitively, a higher τ_ε implies that an additional signal s_i contains more information but it also means that a given number of signals has already provided more information which reduces the value of additional information.

As to the total precision of newspaper coverage $(\tau_\varepsilon x^*)$, however, the effect of τ_ε is clear: $(\tau_\varepsilon x^*)$ always increases in τ_ε. Thus, a more efficient technology, i.e. lower k or higher τ_ε, implies higher newspaper quality:

[10] With only one citizen (or identical citizens), the monopolist is a perfectly discriminating monopolist that implements the same allocation as the competitive market. It would be interesting to study information provision with different types of citizens where the types may refer to different valuations of public goods, for example.

[11] The proofs are relegated to the appendix.

$$\frac{\partial(\tau_\varepsilon x^*)}{\partial k} < 0$$
$$\frac{\partial(\tau_\varepsilon x^*)}{\partial \tau_\varepsilon} > 0.$$

(b) *Prior degree of uncertainty (τ_θ).* The effect of a more precise prior on newspaper quality is ambiguous:

$$\frac{\partial x^*}{\partial \tau_\theta} \begin{cases} > 0 & \text{for } \tau_\theta < \tau_\varepsilon x^*/2 \\ = 0 & \text{for } \tau_\theta = \tau_\varepsilon x^*/2 \\ < 0 & \text{for } \tau_\theta > \tau_\varepsilon x^*/2. \end{cases}$$

Hence, in contrast to the previous result, the total precision ($\tau_\varepsilon x^*$) can increase or decrease for larger values of τ_θ. The result depends on the magnitude of two effects from τ_θ on the demand for information. On the one hand, a more precise prior reduces the risk of the voting decision which makes additional information less important. On the other band, a higher τ_θ makes a signal more informative.

Now we have characterized the optimal decisions of the citizen and the newspapers and it remains to explore how these results shape the incentives of the incumbent.

3.3.3 Incumbent behavior

The incumbent understands the citizen's decision problem and the incentives of the newspapers. Thus, he can conclude that (i) he gets reelected if the citizen considers him to be at least as competent as the average and that (ii) newspapers provide information of quality x^*. Further, he can compute the posterior distribution of $\theta|s$ and he knows that the citizen will reelect him if

$$E(\theta|s) = \frac{\tau_\varepsilon x^*}{\tau_\theta + \tau_\varepsilon x^*} \cdot (s - \tilde{a} - \bar{\theta}) \geq 0. \tag{3.17}$$

The incumbent knows that his effort influences this posterior mean via the signal s. He does not know, however, the exact value of $E(\theta|s)$ for a given $\tilde{a}$ when he chooses his effort level because of the random elements of the signal. From the incumbent's prior perspective, the posterior mean $E(\theta|s)$ is the normal random variable

$$\frac{\tau_\varepsilon x^*}{\tau_\theta + \tau_\varepsilon x^*} \cdot (a - \tilde{a} + \theta^I - \bar{\theta} + \varepsilon)$$

with mean

$$m = \frac{\tau_\varepsilon x^*}{\tau_\theta + \tau_\varepsilon x^*} \cdot (a - \tilde{a})$$

and precision

$$r = \frac{\tau_\theta^2}{\tau_\varepsilon x^*} + \tau_\theta.$$

Comparing the incumbent's prior perspective on the distribution of $E(\theta|s)$ to the citizen's perspective, we see that they share the same knowledge about the precision of this distribution. The perspectives on the mean, however, differ in a crucial way: The citizen simply forms expectations, whereas the incumbent can manipulate the posterior mean via his effort, because, for a given belief $\tilde{a}$, an unexpected amount of effort increases the citizen's estimate of θ^I. Hence, one unit of additional effort increases m by the factor $\tau_\varepsilon x^*/\tau_\theta + \tau_\varepsilon x^*$ and we know that more precise newspaper coverage (higher x^*) magnifies this effect of exerting more effort. It follows that the incumbent can shift the distribution that determines the probability of his reelection: Having determined the distribution function of $E(\theta|s)$ the incumbent can compute p, the probability that he gets reelected, where

$$p = prob[E(\theta|s) \geq 0].$$

Thus, the probability of reelection from the incumbent's perspective is

$$p = 1 - \Phi\left[(0 - m(a|\tilde{a}))r^{\frac{1}{2}}\right],$$

where Φ denotes the standard normal distribution function and it can be seen that, since m is increasing in a, the probability of reelection is higher for larger values of effort.

Hence, the incumbent chooses a to maximize the objective function

$$\left\{1 - \Phi\left[(0 - m(a|\tilde{a}))r^{\frac{1}{2}}\right]\right\} \cdot R - \gamma(a), \tag{3.18}$$

taking the citizen's expectation about effort, $\tilde{a}$, as given.[12] The optimal level of effort is characterized by the first-order condition[13]

[12]Recall that the winner of the election exerts effort $a_2 = 0$ in period 2 and thus $\gamma(a_2) = 0$.

[13]The distribution function Φ is both concave and convex on part of its domain. We show in the appendix of chapter 4 that an appropriate assumption on the value of R ensures the concavity of the incumbent's objective function so that the first order condition is sufficient to characterize the optimal level of effort (see also Ashworth (2005)).

$$\left\{\varphi\left[(0-m(a|\tilde{a}))r^{\frac{1}{2}}\right]\sqrt{\frac{\tau_\theta\cdot\tau_\varepsilon x^*}{\tau_\theta+\tau_\varepsilon x^*}}\right\}\cdot R=\gamma'(a), \tag{3.19}$$

where φ is the standard normal density function again and the term in curly brackets is equal to the density of the distribution of $E(\theta|s)$ evaluated at 0. Thus, the optimal effort level is such that the marginal expected rent is equal to the marginal cost of effort.

With rational expectations, the citizen correctly anticipates the incumbent's behavior and thus we have $a=\tilde{a}$ in equilibrium. It follows that $m=0$ and equation 3.19 turns into the equation

$$\varphi(0)\sqrt{\frac{\tau_\theta\cdot\tau_\varepsilon x^*}{\tau_\theta+\tau_\varepsilon x^*}}\cdot R=\gamma'(a), \tag{3.20}$$

which implicitly defines the optimal level of effort in equilibrium. Further, since the citizen is not fooled by the incumbent in equilibrium but correctly anticipates the level of effort, it follows from $m=0$ that the probability of reelection is

$$p^*=1-\Phi(0)=\frac{1}{2}.$$

This result shows that exerting effort does not pay off for the incumbent, since the probability of reelection is not higher than prior to exerting effort, but not exerting effort would even reduce the probability of reelection because the citizen would form a lower estimate of the incumbent's competence in this case.

As $\gamma'(a)$ is invertible we can describe the incumbent's equilibrium behavior as follows:

Proposition 3.5. *In the unique pure-strategy equilibrium, the incumbent selects the effort level*

$$a^*=(\gamma')^{-1}\left(\frac{1}{\sqrt{2\pi}}\cdot\sqrt{\frac{\tau_\theta\cdot\tau_\varepsilon x^*}{\tau_\theta+\tau_\varepsilon x^*}}\cdot R\right). \tag{3.21}$$

Proof. See above. □

Since (γ') is an increasing function, $(\gamma')^{-1}$ is an increasing function, too. Thus, effort is higher for larger values of $\left(\frac{1}{\sqrt{2\pi}}\cdot\sqrt{\frac{\tau_\theta\cdot\tau_\varepsilon x^*}{\tau_\theta+\tau_\varepsilon x^*}}\cdot R\right)$. In this model, we are mainly interested in the question of how the media market influences incumbent behavior and we can directly derive three results from Proposition 3.5:

(1) The term $\frac{\tau_\theta \cdot \tau_\varepsilon x^*}{\tau_\theta + \tau_\varepsilon x^*}$ is the precision of the random variable $\theta^I + \varepsilon$. Transformed into variance notation (precision=1/variance), this precision is equal to $1/(\sigma_\theta^2 + \sigma_\varepsilon^2/x^*)$ and we can see that the precision is higher for larger values of x^*. Thus, more informative coverage always makes the incumbent work harder. This media effect stems from the fact that more precise media coverage increases the effect of an unexpected additional amount of effort. When x^* is higher, a signal is given more weight in the updating process which implies that a better performance of the incumbent due to an unexpected boost of effort has a stronger influence on the citizen's belief about θ^I. This makes the incumbent exert more effort.

(2) According to Proposition 3.4, a monopoly newspaper provides information of the same quality as a competitive newspaper market. It follows that the market form does not influence the incumbent's incentives so that the level of effort/rent-seeking is identical in both market situations. Yet, the citizen is better off with a competitive market because a monopoly newspaper charges a higher price.

(3) The incumbent exerts more effort for larger values of the exogenous rent R, so the prospect of holding a more prestigious office makes incumbents work harder.

Since newspaper quality $x^* = x^*(k, \tau_\theta, \tau_\varepsilon)$ is a function of the cost to produce information (k), the precision of the prior about competence (τ_θ) and the precision of the sampling technology (τ_ε), we can derive further comparative statics results from Proposition 3.5 (The proofs are relegated to the appendix).

(4) The equilibrium level of effort is higher if the cost of sampling is lower (low k) and if the precision of the sampling technology is higher (high τ_ε):

$$\frac{\partial a^*}{\partial k} < 0, \quad \frac{\partial a^*}{\partial \tau_\varepsilon} > 0.$$

In summary, a more efficient technology results in higher levels of effort.

(5) The effect of the precision of the prior of competence on the incumbent's effort is ambiguous and the direction of the effect depends on the ratio of τ_θ and $\tau_\varepsilon x^*$:

$$\frac{\partial a^*}{\partial \tau_\theta} \begin{cases} > 0 & \text{for } \tau_\theta < 2 \cdot \tau_\varepsilon x^* \\ = 0 & \text{for } \tau_\theta = 2 \cdot \tau_\varepsilon x^* \\ < 0 & \text{for } \tau_\theta > 2 \cdot \tau_\varepsilon x^*. \end{cases}$$

This ambiguous result can be explained as follows: For a given x^*, the direct effect of an increase of τ_θ on effort is always positive, while the indirect effect via $x^*(\tau_\theta)$ is ambiguous for the quality of information decreases in τ_θ if $\tau_\theta > \tau_\varepsilon x^*/2$ and it increases if $\tau_\theta < \tau_\varepsilon x^*/2$. In the latter case, both the direct and the indirect effect go in the same direction while they go in opposing directions in the former case. The direct effect dominates an opposing indirect effect for $\tau_\varepsilon x^*/2 < \tau_\theta < 2\tau_\varepsilon x^*$, whereas the indirect effects dominates for $\tau_\theta > 2\tau_\varepsilon x^*$. This explains our result.

3.4 Concluding remarks

Our analysis explains how intensively commercial media investigate the performance of incumbent politicians and the resulting incentives of the latter to engage in public good production. A central finding is that more informative coverage strengthens the incentives of incumbents to behave well. The quality of information is higher when media outlets have a more efficient production technology in terms of low costs of investigations and high informativeness of a given investigation effort. Our empirical result in chapter 5 which says that regions far away from media cities receive less favorable policies seems to support this prediction. This is, on the one hand, because the cost to produce news is larger in more distant regions and, on the other hand, because journalists are often less informed about local institutions in distant regions which can make their investigations less informative.

The model also predicts that monopoly newspapers and competing newspapers provide the same quality of information which implies that the incentives of politicians are identical in both scenarios. This is an interesting aspect concerning the recent development that many local newspaper markets become monopolistic. But the result should be interpreted carefully because, as with most theory, it is only part of the story, of course. Mullainathan and Shlaifer (2005), for example, show that the market form affects the incentives of newspapers to bias information and it is not clear how this aspect would affect our result. Yet, it would be interesting to test it empirically.

The analysis also shows that the incentives of politicians are weakened when media activity is low and the electorate has very precise prior knowledge about the competence of politicians. This may imply that senior incumbents in jurisdictions which receive little coverage invest less effort in public good provision because the electorate is well informed about their competence.

Needless to say, our results shed light on only some components that explain the interaction between media and policy outcomes and the main message of this conclusion is that more research is to be done in order to understand this problem properly. Our analysis achieves to develop a robust framework that can be extended in several directions in order to gain more insights. Some extensions that seem to be promising fall into the following categories:

- *Informational characteristics of policy tasks.* It would be interesting to analyze systematically if some policy tasks are more newsworthy than others. It appears reasonable, for example, that citizens need information in order to privately adapt to certain policies whereas their demand for information originates only from the desire to make an informed vote when it comes to other policies. Think of implementing new projects versus maintaining already existing projects, for example. Thus, it could be studied if the media market creates incentives for politicians to exert more effort in or allocate more funds to certain tasks. Empirically, it would be interesting to examine whether one would find traces of such an effect in the composition of public budgets.

- *Features of the media market.* Having the provision of information via the internet and television in mind, it is important to study the role of advertising in detail because these media platforms often depend on revenues from advertising. Though our 'subscription story' could also be interpreted in terms of advertising (think of the prices in terms of negative utility the citizen suffers from ad intensity), this narrow interpretation of advertising is not sufficient to understand the incentives of providing information on the internet.

- *Heterogenous voters.* We could integrate different preferences of voters for public goods into our framework or let voters have different opinions about politicians in terms of ideology. It would be interesting to study how the heterogeneity of voters affects information provision and the incentives of politicians. Further, the ideological preferences of voters are often considered to be a substitute for valence issues (public goods, for example) that are valued by all voters in the same way. Thus, politicians may try to manipulate the ideological positions of voters in order to reduce the monitoring of the valence policy and earn higher rents.

This brief collection of possible extensions shows that there is much interesting research to be done.

3.5 Appendix

Electoral decision. The citizen elects the incumbent if

$$E[\ell_I(\theta)] = \int_{-\infty}^{0} |\theta| \; dF(\theta) < E[\ell_C(\theta)] = \int_{0}^{\infty} \theta \; dF(\theta).$$

Moving terms on one side shows that this is equivalent to

$$\int_{-\infty}^{0} \theta \; dF(\theta) + \int_{0}^{\infty} \theta \; dF(\theta) = E(\theta) > 0.$$

Obviously, the challenger is elected if $E[\ell_I(\theta)] > E[\ell_C(\theta)] \Leftrightarrow E(\theta) < 0$. □

Proof of Proposition 3.1. **The value of subscribing to a newspaper.** The value ω is equal to the difference between the prior risk and the pre-posterior risk:

$$\omega = E[\ell_{v^0}(\theta)] - E\left\{E[\ell_{v^1}(\theta|s)]\right\}$$

which is equal to

$$\begin{aligned}
\omega &= E\left\{E[\ell_{v^0}(\theta|s)\right\} - E\left\{E[\ell_{v^1}(\theta|s)]\right\} \\
&= E\left\{E[\ell_{v^0}(\theta|s)] - E[\ell_{v^1}(\theta|s)]\right\} \\
&= E\left\{E[\ell_{v^0}(\theta|s) - \ell_{v^1}(\theta|s)]\right\}
\end{aligned}$$

where $E[\ell_{v^0}(\theta)] = E\left\{E[\ell_{v^0}(\theta|s)\right\}$ in the first line can be verified by applying the law of iterated expectation. The next step is to inspect the term in curly brackets in the last line. To do so we need to determine the optimal prior voting decision, v^0, and the optimal posterior voting decision, v^1. Given the prior belief, the citizen is indifferent between the two candidates. We assume with no loss of generality that the citizens elect candidate A whenever they are indifferent. Thus we continue with $v^0 = I$. The optimal posterior decision, v^1, depends on the realization of the signal and is characterized by

$$v^1 = \begin{cases} I & \text{for } E(\theta|s) \geq 0 \\ C & \text{for } E(\theta|s) < 0. \end{cases} \tag{3.22}$$

Substituting the optimal decisions into $\{E[\ell_{v^0}(\theta|s) - \ell_{v^1}(\theta|s)]\}$ we obtain

$$E[\ell_I(\theta|s) - \ell_{v^1}(\theta|s)] = \begin{cases} E[\ell_I(\theta|s) - \ell_I(\theta|s)] & \text{for } E(\theta|s) \geq 0 \\ E[\ell_I(\theta|s) - \ell_C(\theta|s)] & \text{for } E(\theta|s) < 0 \end{cases}$$
$$= \begin{cases} 0 & \text{for } E(\theta|s) \geq 0 \\ -E(\theta|s) & \text{for } E(\theta|s) < 0. \end{cases} \tag{3.23}$$

We know that $E(\theta|s) = \frac{\tau_\varepsilon x}{\tau_\theta + \tau_\varepsilon x} \cdot (s - \tilde{a} - \bar{\theta})$ and we can use this result to compute the pre-posterior distribution of $E(\theta|s)$. From the pre-posterior perspective, $E(\theta|s)$ is the random variable

$$\frac{\tau_\varepsilon x}{\tau_\theta + \tau_\varepsilon x} \cdot (a - \tilde{a} + \theta^I - \bar{\theta} + \varepsilon), \tag{3.24}$$

where $\theta^I \sim N(0, \tau_\theta)$ and $\varepsilon \sim N(0, \tau_\varepsilon x)$. Recalling the rules for sums of independently distributed normal random variables, we have the result that $E(\theta|s)$ is normally distributed with mean 0 and precision $r = \frac{\tau_\theta(\tau_\theta + x\tau_\varepsilon)}{x\tau_\varepsilon}$. Let F denote the distribution function of $W = E(\theta|s)$. Then, the value of a subscription is

$$\omega = \int_{-\infty}^{0} -w \; dF(w) + \int_{0}^{\infty} 0 \; dF(w) \tag{3.25}$$
$$= \int_{-\infty}^{0} -w \; dF(w). \tag{3.26}$$

This is the partial expectation of a normal random variable and we will derive the functional form of such partial expectations in general for a normal random variable W with some mean μ and some precision r before we substitute for the respective values of our analysis, i.e. $\mu = 0$ and $r = \frac{\tau_\theta(\tau_\theta + x\tau_\varepsilon)}{x\tau_\varepsilon}$. The density function and the distribution function of W are denoted $f(w)$ and $F(w)$, respectively. First, we show that

$$\int_{-\infty}^{0} -wf(w)dw = r^{-\frac{1}{2}} \left[\varphi(-r^{\frac{1}{2}}\mu) - r^{\frac{1}{2}}\mu \cdot \Phi(-r^{\frac{1}{2}}\mu)\right].$$

We start with substituting $wf(w)$ in the integral on the left-hand side with $\mu f(w) - \frac{f'(w)}{r}$ which follows directly from $f'(w) = -r(w - \mu)f(w)$:

$$
\begin{aligned}
\int_{-\infty}^{0} -wf(w)dw &= -\int_{-\infty}^{0} wf(w)dw \\
&= -\mu \int_{-\infty}^{0} f(w)dw + \frac{1}{r}\int_{-\infty}^{0} f'(w)dw \\
&= -\mu \cdot F(0) + \frac{1}{r} f(0) \\
&= r^{-\frac{1}{2}} \left[r^{-\frac{1}{2}} f(0) - \mu r^{\frac{1}{2}} \cdot F(0) \right].
\end{aligned}
$$

Transforming to the standard normal form we obtain

$$
\int_{-\infty}^{0} -wf(w)dw = r^{-\frac{1}{2}} \left\{ \varphi[-\mu r^{\frac{1}{2}}] - \mu r^{\frac{1}{2}} \cdot \Phi[-\mu r^{\frac{1}{2}}] \right\}. \tag{3.27}
$$

Inserting $\mu = 0$ and $r = \frac{\tau_\theta(\tau_\theta + x\tau_\varepsilon)}{x\tau_\varepsilon}$ into the functional form defined by equation (3.27) yields

$$
\omega = \int_{-\infty}^{0} -wg(w)dw = \left(\frac{\tau_\theta(\tau_\theta + x\tau_\varepsilon)}{x\tau_\varepsilon} \right)^{-\frac{1}{2}} \cdot \varphi(0).
$$

Thus, we obtain

$$
\omega(x) = \left(\frac{\tau_\theta(\tau_\theta + x\tau_\varepsilon)}{x\tau_\varepsilon} \right)^{-1/2} \cdot \varphi(0).
$$

Concavity of $\omega(x)$. The first and second derivatives are

$$
\omega'(x) = \frac{1}{2} \left(\frac{\tau_\theta^2}{\tau_\varepsilon x} + \tau_\theta \right)^{-3/2} \varphi(0) \frac{\tau_\theta^2}{\tau_\varepsilon x^2} > 0
$$

and

$$
\begin{aligned}
\omega''(x) &= \frac{1}{2}\varphi(0) \cdot \left[\frac{3}{2} r^{-\frac{5}{2}} \left(\frac{\tau_\theta^2}{\tau_\varepsilon x^2} \right)^2 - 2r^{-\frac{3}{2}} \frac{\tau_\theta^2}{\tau_\varepsilon x^3} \right] \\
&= \frac{1}{2}\varphi(0) \frac{\tau_\theta^2}{\tau_\varepsilon x^3} r^{-\frac{3}{2}} \left[\frac{-(\tau_\theta + 4x\tau_\varepsilon)}{2(\tau_\theta + \tau_\varepsilon x)} \right] < 0.
\end{aligned}
$$

Thus, $\omega(x)$ is a strictly concave function in x.

The slope of $\omega(x)$ at the origin.

We are interested in the limit value $\lim_{x\to 0} \omega'(x)$. To determine the limit value it is sufficient to inspect $\frac{1}{x^2} \cdot \left(\frac{\tau_\theta^2}{\tau_\varepsilon x} + \tau_\theta\right)^{-3/2}$. First we rearrange:

$$\left(\frac{\tau_\theta^2}{\tau_\varepsilon x} + \tau_\theta\right)^{-3/2} \cdot \frac{1}{x^2} = \left[\left(\frac{\tau_\theta^2}{\tau_\varepsilon x} + \tau_\theta\right) \cdot x^{4/3}\right]^{-3/2} = \left(\frac{\tau_\theta^2}{\tau_\varepsilon} x^{1/3} + \tau_\theta x^{4/3}\right)^{-3/2}.$$

It is easy to see that

$$\lim_{x\to 0} \left(\frac{\tau_\theta^2}{\tau_\varepsilon} x^{1/3} + \tau_\theta x^{4/3}\right) = 0$$

and thus

$$\lim_{x\to 0} \left(\frac{\tau_\theta^2}{\tau_\varepsilon} x^{1/3} + \tau_\theta x^{4/3}\right)^{-3/2} = \infty.$$

□

Comparative statics - newspaper quality. The optimal quality $x^* = x^*(k, \tau_\theta, \tau_\varepsilon)$ is implicitly defined by the equation

$$\frac{1}{2}\left(\frac{\tau_\theta^2}{\tau_\varepsilon x} + \tau_\theta\right)^{-3/2} \varphi(0)\frac{\tau_\theta^2}{\tau_\varepsilon x^2} - k = 0.$$

Though we cannot explicitly solve for x^*, comparative statics results can be derived by using the implicit-function rule. We define the function

$$z(x; k, \tau_\theta, \tau_\varepsilon) := \frac{1}{2}\left(\frac{\tau_\theta^2}{\tau_\varepsilon x} + \tau_\theta\right)^{-3/2} \varphi(0)\frac{\tau_\theta^2}{\tau_\varepsilon x^2} - k.$$

Given $z(x; k, \tau_\theta, \tau_\varepsilon) = 0$ the implicit-function rule yields

$$\frac{\partial x}{\partial \alpha} = -\frac{z_\alpha}{z_x} \quad \text{for } \alpha \in \{k, \tau_\theta, \tau_\varepsilon\}.$$

The denominator of $-\frac{z_\alpha}{z_x}$ is the second derivative $\omega''(x) = z_x$ and we have already shown above that $\omega''(x) < 0$. Hence, the sign of $\frac{\partial x}{\partial \alpha} = -\frac{z_\alpha}{z_x}$ is determined by the sign of z_α.

Effect of k. Since $z_k = -1$, we obtain the result that $\partial x^*/\partial k = -z_k/z_x < 0$ and thus $\partial(\tau_\varepsilon x^*)/\partial k < 0$.

Effect of τ_ε. Computing z_{τ_ε} yields:

$$
\begin{aligned}
z_{\tau_\varepsilon} &= \frac{1}{2}\varphi(0) \cdot \frac{\tau_\theta^2}{x^2}\left[\frac{\partial r^{-\frac{3}{2}}}{\partial \tau_\varepsilon} \cdot \frac{1}{\tau_\varepsilon} - r^{-\frac{3}{2}} \cdot \frac{1}{\tau_\varepsilon^2}\right] \\
&= \frac{1}{2}\varphi(0) \cdot \frac{\tau_\theta^2}{x^2\tau_\varepsilon^2} r^{-\frac{3}{2}}\left[\frac{\tau_\theta - 2\tau_\varepsilon x}{2(\tau_\theta + \tau_\varepsilon x)}\right],
\end{aligned}
$$

and it follows that

$$
\frac{\partial x^*}{\partial \tau_\varepsilon} = \frac{x^*}{\tau_\varepsilon} \cdot \frac{\tau_\theta - 2\tau_\varepsilon x}{\tau_\theta + 4\tau_\varepsilon x}
\begin{cases}
> 0 & \text{for } \tau_\theta > 2\tau_\varepsilon x^* \\
= 0 & \text{for } \tau_\theta = 2\tau_\varepsilon x^* \\
< 0 & \text{for } \tau_\theta < 2\tau_\varepsilon x^*.
\end{cases}
$$

Hence, we find that

$$
\frac{\partial(\tau_\varepsilon x^*)}{\partial \tau_\varepsilon} = -\frac{z_{\tau_\varepsilon}}{z_x} \cdot \tau_\varepsilon + x = 2x \cdot \frac{(\tau_\theta + \tau_\varepsilon x)}{\tau_\theta + 4\tau_\varepsilon x} > 0.
$$

Effect of τ_θ. Computing z_{τ_θ} yields:

$$
\begin{aligned}
z_{\tau_\theta} &= \frac{1}{2}\varphi(0)\frac{1}{x^2\tau_\varepsilon}\left[\frac{\partial r^{-\frac{3}{2}}}{\partial \tau_\theta} \cdot \tau_\theta^2 + r^{-\frac{3}{2}} \cdot 2\tau_\theta\right] \\
&= \frac{1}{2}\varphi(0)\frac{1}{x^2\tau_\varepsilon}\left[-\frac{3}{2}r^{-\frac{5}{2}}\left(\frac{2\tau_\theta + \tau_\varepsilon x}{\tau_\varepsilon x}\right) \cdot \tau_\theta^2 + r - \frac{3}{2} \cdot 2\tau_\theta\right] \\
&= \underbrace{\frac{1}{2}\varphi(0)\frac{\tau_\theta}{x^2\tau_\varepsilon}}_{>0} r^{-\frac{3}{2}}\left[\frac{-2\tau_\theta + \tau_\varepsilon x}{2(\tau_\theta + \tau_\varepsilon x)}\right],
\end{aligned}
$$

and it follows that

$$
\frac{\partial x^*}{\partial \tau_\theta} = \frac{x^*}{\tau_\theta} \cdot \frac{-2\tau_\theta + \tau_\varepsilon x}{\tau_\theta + 4\tau_\varepsilon x}
\begin{cases}
> 0 & \text{for } \tau_\theta < \tau_\varepsilon x^*/2 \\
= 0 & \text{for } \tau_\theta = \tau_\varepsilon x^*/2 \\
< 0 & \text{for } \tau_\theta > \tau_\varepsilon x^*/2
\end{cases}
$$

and thus

$$\frac{\partial(\tau_\varepsilon x^*)}{\partial \tau_\theta} \begin{cases} > 0 & \text{for } \tau_\theta < \tau_\varepsilon x^*/2 \\ = 0 & \text{for } \tau_\theta = \tau_\varepsilon x^*/2 \\ < 0 & \text{for } \tau_\theta > \tau_\varepsilon x^*/2. \end{cases}$$

□

Comparative statics - effort. Both the function $(\gamma')^{-1}$ and the square root are strictly increasing in their respective arguments. Thus it is sufficient to inspect the term $\frac{\tau_\theta \cdot \tau_\varepsilon x}{\tau_\theta + \tau_\varepsilon x} =: t$ to derive the comparative statics results with respect to the parameters k, τ_ε and τ_θ, because $\operatorname{sgn} \frac{\partial t}{\partial \alpha} = \operatorname{sgn} \frac{\partial a^*}{\partial \alpha}$ for $\alpha \in \{k, \tau_\varepsilon, \tau_\theta\}$. The partial derivative of t with respect to k yields

$$\frac{\partial t}{\partial k} = \frac{\tau_\varepsilon \tau_\theta^2}{(\tau_\theta + \tau_\varepsilon x^*)^2} \cdot \frac{\partial x^*}{\partial k} < 0 \tag{3.28}$$

which follows directly from the comparative statics results regarding newspaper quality. For τ_ε we obtain

$$\frac{\partial t}{\partial \tau_\varepsilon} = \frac{\tau_\theta^2}{(\tau_\theta + \tau_\varepsilon x^*)^2} \cdot \frac{\partial(\tau_\varepsilon x^*)}{\partial \tau_\varepsilon} > 0 \tag{3.29}$$

which also follows directly from the comparative statics results regarding newspaper quality. Finally, for τ_θ we get

$$\frac{\partial t}{\partial \tau_\theta} = \frac{\tau_\varepsilon}{(\tau_\theta + \tau_\varepsilon x^*)^2} \cdot \left(\tau_\varepsilon {x^*}^2 + \tau_\theta^2 \cdot \frac{\partial x^*}{\partial \tau_\theta} \right). \tag{3.30}$$

As the first factor is positive, the sign of $\frac{\partial t}{\partial \tau_\theta}$ depends on the term in brackets only. Substituting for $\frac{\partial x^*}{\partial \tau_\theta}$, we obtain

$$\begin{aligned} \left({x^*}^2 \tau_\varepsilon + \tau_\theta^2 \cdot \frac{\partial x^*}{\partial \tau_\theta} \right) &= \frac{2x^*}{\tau_\theta + 4x^* \tau_\varepsilon} \cdot \left(2{x^*}^2 \tau_\varepsilon^2 + x^* \tau_\theta \tau_\varepsilon - \tau_\theta^2 \right) \\ &= \underbrace{\frac{2x^* \cdot (x^* \tau_\varepsilon)^2}{\tau_\theta + 4x^* \tau_\varepsilon}}_{>0} \cdot \left[-\left(\frac{\tau_\theta}{x^* \tau_\varepsilon} \right)^2 + \frac{\tau_\theta}{x^* \tau_\varepsilon} + 2 \right]. \end{aligned}$$

For the domain of our problem, we have that

$$\left[-\left(\frac{\tau_\theta}{x^*\tau_\varepsilon}\right)^2 + \frac{\tau_\theta}{x^*\tau_\varepsilon} + 2\right] \begin{cases} > 0 & \text{for } \tau_\theta < 2x^*\tau_\varepsilon \\ = 0 & \text{for } \tau_\theta = 2x^*\tau_\varepsilon \\ < 0 & \text{for } \tau_\theta > 2x^*\tau_\varepsilon. \end{cases} \tag{3.31}$$

Thus, we have proved our result because $\operatorname{sgn}\left[-\left(\frac{\tau_\theta}{x^*\tau_\varepsilon}\right)^2 + \frac{\tau_\theta}{x^*\tau_\varepsilon} + 2\right] = \operatorname{sgn} \frac{\partial a^*}{\partial \tau_\theta}$. □

4 Theory and evidence from Norway: Newspaper circulation and local government efficiency

4.1 Introduction

In this chapter, we examine whether the circulation of local newspapers affects public sector efficiency in Norwegian municipalities. Two features make Norway a very suitable object for studying the influence of newspaper reading on policy outcomes: First, the fact that Norwegians are among the most avid newspaper readers worldwide has shaped a very diverse and vibrant newspaper landscape, where most of the newspapers also have a pronounced local focus. The second reason is a highly decentralized structure of government. The local public sector is the main provider of welfare services and about 15% of GDP is spent by local governments on the implementation of these services.

The central government aims at ensuring equal living conditions throughout the country which implies the provision of welfare services like health care, education and social services in high quality across all municipalities. To this end, the central government allocates a budget to each jurisdiction and each municipality can allocate autonomously funds to the production of different services in order to match local preferences. Within each municipality, a locally elected government is responsible for the provision of public services. As service production is the very task of local governments, voters should mainly consider public service performance when they decide on the reelection of local officials. Thus, elections provide incentives for the latter to perform well because a more efficient use of disposable funds - implying a higher level of services, given the fixed budget - should increase the chances of reelection.

At the municipal level, individuals usually are quite well informed about who is in charge of local services and they can observe service provision in their day-to-day life. It is, however, likely that they are not perfectly informed about all factors that determine public service provision. In particular, when it comes to evaluating politicians it is important to know whether factors that are not in control of local government have influenced service provision. If such local or country-wide shocks affect public service production, voters who are aware of the magnitude of the shocks can judge government performance more accurately.

As local newspapers play an important role in informing voters in Norway, this analysis strives to examine whether certain features of the newspaper market can explain differences across municipalities in public service efficiency. We build a formal model to illustrate how the informational situation of the electorate shapes an incumbent's incentives to behave well. The central prediction of the model is that a larger share of informed voters in the electorate makes the incumbent work harder because external factors are not confounded with an incumbent's effort. This prediction is tested using panel data on Norwegian municipalities for the years 2001-2005. Our main variables are an index of public sector efficiency in Norwegian municipalities as introduced by Borge et al. (2008) and two measures of voter information that are based on newspaper circulation at the municipality level. We define political accountability in terms of the efficient provision of public services or, put differently, reduced budgetary slack and we find that higher circulation of local newspapers in a municipality is associated with higher levels of efficiency in public service provision, i.e. increased accountability. This effect is larger when circulation is weighted by the amount of community-specific information contained in the various newspapers, and magnified even more in small and de-central municipalities.

The efficiency measure we apply was introduced by Borge et al. (2008) to analyze how political and budgetary institutions affect efficiency in public service provision. Their main finding is that higher revenues are associated with increased budgetary slack. Revelli and Tovmo (2007) also use the index and find that local government efficiency in Norway shows a spatial pattern due to yardstick competition.

Empirical evidence that the mass media affect accountability is provided by Besley and Burgess (2002), Reinikka and Svensson (2005), Shi and Svensson (2006), and Svaleryd and Vlachos (2009). Strömberg (2004a,b) theoretically and empirically shows that the mass media channel more political information to certain population subgroups, which then receive favorable redistributive policies. Most closely related to our analysis, Snyder and Strömberg (2010) find empirical evidence that newspaper coverage raises voters' information levels, thus increasing electoral accountability. Although our paper is mainly concerned with the role information plays in establishing accountability, it is also related to the empirical literature on local government efficiency surveyed in De Borger and Kerstens (2000).

4.2 Theory: Newspapers and efficiency

The link between newspapers and efficiency in the public sector can be analyzed theoretically in a principal-agent framework. The electorate (the principal) delegates service provision to a politician (the agent) and wants the politician to produce a maximum of services on a given budget. The politician's incentives regarding production are shaped by the electorate's ability to monitor his performance and this ability depends on information provided by newspapers. In contrast to the previous chapters, we do not explain why voters buy newspapers but we simply show how exogenously given shares of newspaper readers and given intensity of coverage affects the politician's incentives. Thus, our model also applies to scenarios where many if not most local newspapers are bought for other reasons than to obtain information about local politics. Even though such readers may care more about sports, weddings, obituaries, clubs and the like, they will still obtain political information as a by-product when scanning the paper (Popkin, 1991).[14]

Further, newspapers are not the only source of information, of course: social networks also play a crucial role in distributing information. However, Mondak (1995) finds evidence that media fill information into these networks rather than being substituted by them. This corresponds to evidence presented by Snyder and Strömberg (2010) that newspapers enhance the political knowledge of voters, and it underscores that local newspapers have an effect on information levels independent of other important arenas of debate.

We build a political economy model of a single constituency with a continuum of voters normalized to unity. There are two periods of time.

Production of a public good

In the first period, an incumbent government provides the public good y according to the production technology

$$y_1 = a_1 + \theta^I + \varepsilon_1. \tag{4.1}$$

The level of the public good in period 1 results from the incumbent's effort ($a_1 \in [0, \infty)$), her competence (θ^I) and a temporary shock (ε_1). Subscripts denote the time period and superscript I refers to the incumbent. We assume that competence, θ^I, is a permanent feature of the incumbent.
With regard to Norway, one can think of y as the services that a municipality provides to its citizens. To provide these services, municipal governments are endowed with a fixed budget by the central government. Hence, incumbent politicians cannot raise y by increasing revenues, but effort they exert may affect the level of public services in the following ways:

- *Bargaining between politicians and local public administration.* Administration competes with service production for money from the municipal budget. Local government does have the final say on the allocation of funds, yet the public administration heavily takes part in budgeting (Kalseth & Rattso 1998). Thus, the politicians' bargaining power affects the allocation of funds. If, for example, politicians invested effort to

[14]For these readers, political information works much like an advertisement. While no one buys a newspaper for the advertisements, they certainly do not go unnoticed.

raise their expertise this could increase their bargaining power and, finally, lead to a higher level of services due to resource allocation.

- *Monitoring the administration.* The administration not only takes part in budgeting but also implements political decisions. Once funds are allocated to public service provision, the level of services is higher if these funds are used in an efficient manner. If politicians try harder to monitor the implementation this will raise efficiency and, consequently, the level of services.

The effect of the incumbent's competence θ^I on public good production resembles the effect of effort. θ denotes the exogenously given talent of a politician to influence public good provision. At a given effort level, a more competent politician will provide a higher level of public services. Finally, the shock component ε can be either a local shock or a country-wide shock which affects the level of y.

Information

θ^I and ε_1 are random variables. At the time when the incumbent decides about her effort level, neither she nor the electorate know the realizations of θ^I and ε_1. Common knowledge are the distributions $\theta \sim N(\bar{\theta}, \sigma_\theta^2)$ and $\varepsilon \sim N(0, \sigma_\varepsilon^2)$.

After the incumbent has chosen effort and the values of θ^I and ε_1 have been realized, all voters observe the level of the public good, y_1. An exogenously defined share of voters λ reads local newspapers for reasons that are independent of politics, e.g., because they care about news on sports, weddings, obituaries, clubs and the like. We assume that newspapers have perfect information about ε_1 and report it.[15] The probability q that a voter i finds the information about ε_1 in the newspaper depends on how much news space s the editors assign to local politics. Thus, $q = q(s)$ and we presume $q(0) = 0$, $q' > 0$ and $q'' < 0$. So the probability that a voter i is informed about ε_1 is $\lambda \cdot q(s)$. Given a continuum of voters, the share of informed voters is $\alpha := \lambda \cdot q(s)$.

Let us briefly illustrate this argument about informed voters by an example. Suppose that there is no newspaper published in a municipality but citizens read newspapers that come from contiguous municipalities. If these newspapers bias local coverage in favor of their home municipalities only little space will be allocated to news about our municipality of interest. Still, there can be many readers in this municipality but it is unlikely that they are well informed about local politics. On the other hand, we would expect a more informed electorate if a newspaper market and a municipality coincide geographically.

[15] The extreme assumption that newspapers perfectly observe ε_1 serves to simplify the analysis. Alternatively, we could assume that newspapers only receive a less noisy signal than voters. This would not substantially change our results.

So we have two groups of voters in the electorate. A share of α voters are perfectly informed about the shock component ε_1. Observing y_1, these voters can clearly distinguish between the shock and the sum of the two components attributed to the incumbent. A share of $1-\alpha$ voters only know the distribution of ε_1 and may confound the effect of the shock on y_1 with the impact of a_1 and θ^I.

Timing

Now we have determined what voters know when casting the ballot. Before we examine the incumbent's incentives and the optimal voting behavior, we briefly summarize the game between the incumbent and the electorate. The timing is as follows:

Period 1:

- The incumbent politician chooses effort a without knowing his own competence.
- The values of both the incumbent's competence, θ^I, and the shock, ε_1, are realized.
- All voters observe y_1. Additionally, a share α of voters learn the value of ε_1 from the newspaper.
- Elections are held. The incumbent faces a challenger whose competence is drawn from a normal distribution with mean $\bar{\theta}$ and variance σ_θ^2.

Period 2:

- The winner of the election chooses effort.
- y_2^I is realized if the incumbent of period 1 still is in office or y_2^C is realized if the challenger has won the election.

The incumbent's incentives

The incumbent knows that the level of y_1 will affect her chances of reelection. She can influence the level of y_1 by choosing a_1. Effort brings along cost $C(a)$ with $C' > 0$ and $C'' > 0$. At the end of period 1, there is an election where the incumbent faces a randomly drawn challenger. If the incumbent wins, she will receive an exogenous rent $R > 0$ from staying in office. Thus, the incumbent chooses a_1 to maximize

$$p^I(a_1) \cdot [R - C(a_2)] - C(a_1), \tag{4.2}$$

where p^I denotes the probability that the incumbent is reelected. There is no discounting. The first-order condition is

$$\frac{\partial p^I}{\partial a_1} \cdot [R - C(a_2)] = C'(a_1). \tag{4.3}$$

The first-order condition shows that the incumbent weighs the expected net rent against present cost when she chooses a_1. She will increase effort as long as the marginal effect on the expected payoff in period 2 is larger than the marginal cost of effort in period 1. In order to analyze how effort affects the probability of reelection, p^I, we have to examine the voting decisions in the electorate.

Optimal voting behavior

All voters receive utility $u = u(y) = y$ from the public good. Each voter cares about y_2, the level of y in period 2, and about her own ideological position in relation to the incumbent, β_i. β_i is drawn from a uniform distribution with support $[-\beta_0, \beta_0]$. Negative values of β_i imply an ideological bias of voter i in favor of the incumbent, whereas positive values mean a bias in favor of the challenger. Voter i prefers the incumbent if

$$\tilde{y}^I_{2i} \geq \tilde{y}^C_{2i} + \beta_i, \tag{4.4}$$

where $\tilde{y}^I_{2i}$ and $\tilde{y}^C_{2i}$ denote the expected level of y in period 2 under the incumbent and under the challenger respectively. Voters are rational and expect the competence of the randomly drawn challenger to be $\bar{\theta}$. Furthermore, as there is no incentive to invest effort in period 2 for any politician in office, all voters correctly expect that $a^I_2 = a^C_2 = 0$. Hence, the expected level of y_2 depends only on the competence of the incumbent politician in period 2 so that

$$\tilde{y}^I_{2i} = \tilde{\theta}^I_i \quad \text{and} \quad \tilde{y}^C_{2i} = \tilde{\theta}^C_i = \bar{\theta}.$$

Consequently, (4.4) reduces to

$$\tilde{\theta}^I_i \geq \bar{\theta} + \beta_i. \tag{4.5}$$

The expected competence of the incumbent depends on the information a voter possesses. Voters are rational and update their prior beliefs about competence using Bayes' rule. We have to distinguish informed voters from uninformed voters. In each group, $\tilde{\theta}^I_i$ is the same for every single voter. From now on, $\tilde{\theta}^I_m$ denotes the competence as estimated by informed voters and $\tilde{\theta}^I_n$ labels the competence as estimated by uninformed voters.

As informed voters observe y_1 and, additionally, ε_1, they expect the incumbent's competence to be

$$\tilde{\theta}_m^I = y_1 - \tilde{a}_1 - \varepsilon_1 = \theta^I + a_1 - \tilde{a}_1, \tag{4.6}$$

where $\tilde{a}_1$ denotes effort in period 1 as expected by the voters. Uninformed voters only observe the value of y_1 and form the expectation[16]

$$\tilde{\theta}_n^I = \frac{\sigma_\varepsilon^2 \bar{\theta} + \sigma_\theta^2 (y_1 - \tilde{a}_1)}{\sigma_\varepsilon^2 + \sigma_\theta^2}. \tag{4.7}$$

The incumbent wins the election if he gets more than half of all votes cast which we can write as

$$\alpha \cdot \frac{\tilde{\theta}_m^I - \bar{\theta} + \beta_0}{2\beta_0} + (1-\alpha) \cdot \frac{\tilde{\theta}_n^I - \bar{\theta} + \beta_0}{2\beta_0} \geq \frac{1}{2}. \tag{4.8}$$

Inserting (4.6), (4.7) and (4.1) and rearranging, we obtain

$$\left[\alpha + (1-\alpha)\frac{\sigma_\theta^2}{\sigma_\varepsilon^2 + \sigma_\theta^2}\right] \cdot (a_1 - \tilde{a}_1 + \theta^I - \bar{\theta}) + (1-\alpha)\frac{\sigma_\theta^2}{\sigma_\varepsilon^2 + \sigma_\theta^2} \cdot \varepsilon_1 \geq 0. \tag{4.9}$$

The probability of reelection, p^I, is given by the probability that (4.9) is met.[17] The left-hand side of (4.9) is a normal random variable with mean

$$\mu = \left[\frac{\alpha \cdot \sigma_\varepsilon^2 + \sigma_\theta^2}{\sigma_\varepsilon^2 + \sigma_\theta^2}\right] \cdot (a_1 - \tilde{a}_1) \tag{4.10}$$

and variance

$$\sigma^2 = \frac{\alpha^2 \cdot \sigma_\varepsilon^2 \sigma_\theta^2 + \sigma_\theta^4}{\sigma_\varepsilon^2 + \sigma_\theta^2}. \tag{4.11}$$

Now we can compute the probability of reelection as $p^I = 1 - F(0; \mu, \sigma^2)$, where F is the distribution function of the left-hand side of (4.9).

Equilibrium

The incumbent maximizes her objective function, (4.2), taking the voters' expectations about effort, $\tilde{a}_1$, as given. Taking $p^I = 1 - F(0; \mu, \sigma^2)$ and $a_2 = 0$ into account, the first-order condition turns into

[16] In Appendix A1 it is shown how uninformed voters update their expectations.

[17] To be precise, this condition only holds true for $\bar{\theta} - \beta_0 < \tilde{\theta}_m^I, \tilde{\theta}_n^I < \bar{\theta} + \beta_0$. However, for large values of β_0 the approximation error is very small and, in equilibrium, p^I is the probability of reelection for all $\tilde{\theta}_m^I, \tilde{\theta}_n^I$.

$$-\frac{\partial F(0;\mu,\sigma^2)}{\partial \mu}\frac{\partial \mu}{\partial a_1}\cdot R = f(0;\mu,\sigma^2)\frac{\partial \mu}{\partial a_1}\cdot R = C'(a_1). \tag{4.12}$$

An equilibrium with rational expectations requires $a_1 = \tilde{a}_1$. Thus, in equilibrium, $\mu = 0$ and the first-order condition is

$$\frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\left[\frac{\alpha\cdot\sigma_\varepsilon^2+\sigma_\theta^2}{\sigma_\varepsilon^2+\sigma_\theta^2}\right]\cdot R - C'(a_1) = 0, \tag{4.13}$$

with $1/(\sqrt{2\pi}\cdot\sigma) = f(0;0,\sigma)$. From (4.13) one arrives promptly at the following result:

Proposition 4.1. *In equilibrium, the incumbent's effort level in period 1, a_1^*, is uniquely defined by*

$$a_1^* = (C')^{-1}\left(\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sqrt{\sigma_\varepsilon^2+\sigma_\theta^2}}\cdot\frac{\alpha\cdot\sigma_\varepsilon^2+\sigma_\theta^2}{\sqrt{\alpha^2\cdot\sigma_\varepsilon^2\sigma_\theta^2+\sigma_\theta^4}}\cdot R\right). \tag{4.14}$$

Proof. See Appendix. □

Proposition 1 shows that equilibrium effort hinges on the variance of the shock, σ_ε^2, the variance of competence, σ_θ^2, the rent, R, and the share of informed voters, $\alpha = \lambda\cdot q(s)$. As we strive to study how the information in the electorate affects the incentives of the incumbent, we are mainly interested in the impact of λ and s on the effort level. Proposition 1 leads directly to the main message of the model.

Proposition 4.2. *The incumbent's effort increases in the share of informed voters.*

Proof. See Appendix. □

So the central prediction of the model is that a larger share of informed voters makes the incumbent work harder. Consequently, for given θ^I and ε_1, the level of the public good, y_1, is higher when many voters read newspapers and when newspapers devote more news space, s, to information about the shock, ε_1.

4.3 Empirical analysis

4.3.1 The Norwegian institutions

In the following, we use panel data from Norway in trying to find empirical support for the mechanisms described above. The most important reason for picking Norway is its very diverse newspaper landscape. Local newspapers play a primary role in providing local information (Høst, 1999) and circulation data at the municipality level is available for more than 150 newspapers. This is a large number considering that Norway had a population of only roughly 4.8 million in 2008, and it also implies that many of these newspapers have a rather local focus. This is not the only feature of the Norwegian newspaper sector that makes it particular. Until 2004 Norway had the highest per (adult) capita newspaper reach worldwide. As of today it is only surpassed by Japan in that respect (according to World Press Trends, approximately 600 newspapers per 1000 adults were sold in Norway in 2007 on a daily basis). Seven in ten Norwegian households had one or more newspaper subscriptions in 2007. In the large cities, 64 of every 100 households had a newspaper subscription, while the percentage was 75 in more scarcely populated areas.[18] In fact, the reach of newspapers in Norway is higher than that of television, as can be seen from the top portion of Table 1.

Table 1 also demonstrates that newspaper use does not depend on educational levels as much as one may expect. Differences in readership (defined as the percentage of residents that *read* a newspaper on a given day) between the university educated and lower secondary school educated are rather small and have actually decreased in the years 2001-2005, the time period we will be using in the estimations. This is particularly important because it means that the possible confounding of newspaper readership with general interest in politics (as measured by education) is not as big an issue as it may be in other countries. While the gap between educational levels shrank, the total reach of newspapers is slowly declining, a trend that Norway has in common with other western countries.

Another feature that makes Norway an attractive object of research is the fact that there are more than 400 municipalities. This provides an excellent opportunity for studying the effects of newspapers on relatively comparable small government units. The municipalities are responsible for a vast array of services and can decide autonomously on how they spend their budget. Even though local government revenues are mostly fixed, this leaves a lot of leeway for officials in shaping the public service structure in their municipality.[19] Services

[18]Figures from the Norwegian Media Barometer 2007 at `http://www.ssb.no/medie_en/arkiv/`.

[19]There is also another tier between the local and the national level. These *fylke* districts are sizewise somewhat akin to the US counties. Their main responsibilities include higher secondary education, dental services and public transportation. Since municipalities provide a broader range of services and people tend to identify with their municipality rather than the *fylke*, we only consider the municipality level.

Table 1: MEDIA USE IN NORWAY.

TV, daily reach$^{(a)}$	**2001**	**2002**	**2003**	**2004**	**2005**
among population aged 12+	71	71	71	71	71
Newspaper readers, average day$^{(b)}$	**2001**	**2002**	**2003**	**2004**	**2005**
Lower secondary school	70	77	78	74	76
Upper secondary school	83	78	80	78	77
University/college ≤4 yrs	85	81	81	81	79
University/college >4 yrs	86	86	84	87	82
All (9-79 years)	78	77	77	75	74

Note: (a) share of the population aged 12+ that watches television on an average day; (b) share of the population aged 9-79 years that reads at least one newspaper on an average day, by levels of education. Sources: (a) and (b) from Statistics Norway, accessible at the Medianorway website: `http://www.medienorge.uib.no/`.

provided by the municipality include primary and lower secondary education, daycare, care for the elderly, welfare benefits, primary health care and child custody. These services enter into the calculation of the efficiency index (Borge et al., 2008), which we use to evaluate the performance of local governments.

In what follows, we describe in detail the newspaper data as well as the efficiency index and its components. We also briefly go through the controls used in the estimations.

4.3.2 Data sources and media measures

We use data from various sources. While the efficiency indicator and its components were provided by Lars-Erik Borge, municipality level controls were obtained from Statistics Norway (SSB) and the Norwegian Social Science Data Services (NSD). Our main explanatory newspaper variables are derived from data provided by the Norwegian Media Businesses' Association (Mediebedriftenes Landsforening, MBL), whose members include most of the Norwegian newspapers. We add data from the Local Newspaper Association (Landslaget for lokalaviser, LLA) for one specification, as some newspapers are organized in LLA only.[20] Thus, the data can be broken down into three categories: newspaper data, efficiency data and controls.

Newspaper data. The data on newspaper circulation is taken from *Aviskatalogen*, a database maintained by MBL. It contains annual information on the circulation of Norwegian newspaper publications at the community level, where circulation figures are given as average circulation per edition. Following Høst (2006), we categorize the newspapers as follows: (a) tabloids, (b) national newspapers, (c) big-city local newspapers, (d) local newspapers and (e) specialty newspapers.[21]

[20]While some LLA members are organized in MBL as well, most of the LLA newspapers are weeklies that our primary dataset does not cover.

[21]Specialty newspapers are mostly weekly newspapers that cater to special interests, such as sports

In general, we define newspaper reach as: $reach_i = \sum_n circulation_{ni}/households_i$, where n denotes newspapers and i indexes municipalities.[22] Reach can be calculated for all newspapers or subgroups of newspapers. The most important definition of newspaper reach for our purposes is local reach, which includes the above categories (c) and (d). This variable measures the reach of all newspapers that mainly convey local or regional information, because their journalistic focus is on a well-defined local or regional market (Høst, 1999). Table 2 shows that average local reach is very stable over the years at about one newspaper per household. The lower portion of the table establishes that the municipalities for which the efficiency index is available, i.e., the ones used in the estimations, are comparable to all Norwegian communities when it comes to the reach of local newspapers.

Table 2: LOCAL NEWSPAPER REACH IN NORWAY AND ESTIMATION SAMPLE.

local newspaper reach per household	**2001**	**2002**	**2003**	**2004**	**2005**
local papers, all municipalities	1.037	0.996	0.985	0.978	0.969
N	431	431	431	431	431
local papers, efficiency available	1.042	0.996	0.982	0.987	0.968
N	359	379	370	361	374

Note: Newspaper reach per household according to the data employed in this paper. Reach is defined as the total newspaper circulation in a municipality per household, where circulation figures are given as average circulation per edition summed over all newspapers. Average reach in the sample for which the efficiency measure is available is very similar to reach in the full sample. Source: Mediebedriftenes Landsforening (MBL).

Since newspapers in the other categories are very limited in the amount of local information they provide, we expect only local newspapers to enhance local efficiency. As a way of testing this we use the above newspaper classification to construct the additional measures tabloid reach, national reach and specialty reach – none of which should exhibit significant effects on local government performance.

A high local newspaper reach by itself doesn't necessarily mean that a lot of truly local information reaches the community level. It seems reasonable that a newspaper will devote more attention content-wise to those municipalities where it sells most of its circulation (Snyder and Strömberg (2010) provide compelling evidence for this argument). To illustrate, picture a small community i where a high percentage of residents read the newspaper n originating in the nearest city, which is located in municipality j. If the share of newspaper n 's total readers living in municipality i is small, n 's reporting from i will be slim. Reach in i thus may not appropriately capture the information level in i. To account for newspapers' differing local focus, the reach of every newspaper n sold in i is discounted by the $readershare_{ni} = circulation_{ni}/\sum_i circulation_{ni}$, resulting in the variable

or computer weeklies. Even though these are nationally distributed newspapers, they are included in a category of their own, due to their highly specialized focus. A full list of newspapers in the dataset and their categorization is provided at `http://wwwuser.gwdg.de/~uwvw3/norway_webappendix.htm`.

[22]This definition of *reach* is the same as in the systematic description of the Norwegian newspaper system provided by Høst (1999).

$content_i = \sum_n reach_{ni} \cdot readershare_{ni}$. Content states how many equivalents of newspapers that are exclusively concerned with the municipality under consideration are sold per household.[23]

Even though Norway consistently ranks among the top countries worldwide in terms of press freedom, we also calculate a Herfindahl index of market shares within a given municipality, to allow for the possibility that monopoly newspapers can be bribed into reporting favorably more easily. In that case, high levels of concentration may counteract any efficiency-enhancing effects (Besley and Prat, 2006).

Efficiency data. Public sector efficiency is measured via an index developed by Borge et al. (2008), which is available for the period 2001-2005. It relates public service production to disposable revenues. Thus, efficiency increases when, e.g., higher levels of production are extracted from given revenues.

Production is quantified by the production index, an aggregate output measure that was developed by Borge et al. (2001) for the Norwegian authorities. Output comprises 17 indicators of production in the six main municipal service sectors: care for the elderly, primary and lower secondary education through 10th grade, day care, welfare benefits, child custody, and primary health care. The production index accounts for both quantity and quality aspects of public good provision. Quality of public services is obviously very hard to capture, yet the production index includes indicators such as the share of single rooms in nursing homes and teachers per student in an attempt to do so.[24]

Dividing output by local government revenues per capita measures the efficiency of a municipality.[25] Revenues comprise own tax revenues and block grants from the central government. As the costs of service production vary across municipalities, the revenues are adjusted by a cost index, which accounts for factors such as population size, settlement pattern, the age composition of the population and social factors.[26] Finally, the efficiency *index* is calculated by relating the efficiency of a municipality to the country-wide mean efficiency. Since the efficiency index is normalized such that mean efficiency equals 100, a municipality's

[23] An alternative measure – congruence – was proposed by Snyder and Strömberg (2010). It is defined as the sum over the readershares of newspapers that sell in municipality i, weighted by their market shares within that municipality. Intuitively, in our setting congruence describes how well the municipalities coincide with the newspaper markets. The correlation between content and congruence in our sample is very high and results when using the congruence variable are similar to those obtained with the content variable.

[24] The 17 sub-indices and their exact weighting are explained in detail in Borge et al. (2008).

[25] Both output and revenues are normalized such that the respective population-weighted mean equals 100.

[26] The deflating of revenues is necessary because unfavorable cost conditions lead to more per capita revenues from the block grant system; e.g. small municipalities receive more funds because they fail to exploit economies of scale. Without deflating, these municipalities would be evaluated as less efficient simply because of the unfavorable cost conditions.

index value can be interpreted as the deviation from mean efficiency in percentage terms. Thus, an efficiency index value of, say, 110 indicates that the municipality's efficiency is 10% above average. We also propose a crude way of translating the abstract efficiency ratings into monetary terms. Consider a benchmark municipality with efficiency 100 that has population-weighted mean per capita revenues of 26,170 Norwegian kroner (NOK), or 4,067 US-$ at its disposal.[27] The mechanics of the index imply that one way for this municipality to attain an efficiency rating of 110 is by producing the same output on per capita revenues of 3,697 US-$. We will return to this monetization of efficiency in the empirical analysis in order to illustrate the magnitude of possible efficiency gains.

The efficiency index is global in the sense that it is meant to relate overall service provision to revenues. As we believe newspaper circulation to affect government efficiency in general, such a global efficiency measure seems very appropriate. Indeed, considering the well-known problem of measuring public output, Borge et al. (2008) argue that the index has an edge over other efficiency measures, because it relies on a fairly large number of production indicators. However, an ideal efficiency measure for our purpose would relate output in terms of 'things of primary interest to the citizen-consumer' (Bradford et al., 1969) to revenues, i.e., we would like to measure citizens' *utility* resulting from public sector production rather than production per se. This would better fit the analysis because we argue that newspapers affect efficiency through electoral incentives and voters will much more care about their utility than about pure service provision. This perspective renders the efficiency index a somewhat flawed measure, which cannot account for differing preferences across municipalities. If a group of voters in a municipality successfully lobbies for a service not included in the production index, the efficiency index will decrease, while utility-based efficiency might increase. However, the services included in the production index account for about 75% of local budgets while spending on administration represents another 10% (Borge et al., 2008). Thus, lobbying for services outside the index should not be too relevant. On the other hand, when voters weight included services different from what is implied in the index and engage in successful lobbying, the direction of bias in the efficiency index depends on the resulting allocation of funds to services.

Following Borge et al. (2008), we use the efficiency measure as our outcome variable. Because division bias may occur when local government revenue appears both as a control variable and as the denominator of the efficiency index (Borjas, 1980), we also estimate an alternative specification which uses local output as the dependent variable while at the same time controlling for local government revenue. Local output is again defined as the production index at the community level and a positive coefficient on newspaper circulation

[27] These are the mean revenues over the sample years 2001-2005. All calculations are based on 2005 prices and NOK are converted to US-$ using the average 2005 exchange rate.

while holding local government revenue constant indicates an efficiency enhancing effect of increases in newspaper reach.

Controls. Variables that account for heterogeneity at the local level are taken from Statistics Norway and the NSD. These community characteristics include the percentage of the population classified as urban, total population, average household size, income level, religious share of the population, immigrant share, commuter share and average educational level as well the shares of five age groups in the total population. While the efficiency index already corrects for effects of most of these variables on the local cost structure, we include them as controls because they may have an impact on efficiency beyond what is captured in the official index.[28] Political controls are voter turnout as a proxy for unobserved interest in local affairs, the number of municipality council seats per 1000 inhabitants, the seat share of local lists in the municipal council, the share of votes received by the strongest party in the council, and a Herfindahl index of party fragmentation in the local council. Finally, the effect of having an election year is captured by year dummies because local elections take place on the same date in all Norwegian municipalities. Table 3 shows summary statistics for all variables.

Table 3: SUMMARY STATISTICS (NORWAY).

	N	Mean	Std. Dev.	Min	Max
newspaper reach (all newspapers)	1843	1.89	0.54	0.34	6.56
local newspaper reach	1843	0.99	0.27	0.04	1.83
local newspaper reach excluding big-city local newspapers	1843	0.86	0.31	0.03	1.82
local LLA newspaper reach	1840	0.06	0.20	0.00	1.69
tabloid newspaper reach	1843	0.64	0.36	0.03	4.04
national newspaper reach	1843	0.78	0.48	0.03	4.90
specialty newspaper reach	1843	0.12	0.12	0.03	1.02
non-local newspaper reach	1843	0.90	0.50	0.09	5.50
local newspaper content	1843	0.17	0.19	$1.6e{-05}$	0.86
herfindahl local newspapers	1843	54.70	18.09	4.98	92.20
local public sector efficiency	1843	103.73	10.83	44.92	137.00
local public sector production	1843	109.98	15.73	78.92	188.12
local government revenue	1843	107.46	22.90	84.68	330.22
secondary school pct	1843	57.99	3.95	39.84	67.19
university educated pct	1843	16.50	4.98	7.93	41.97
average gross income (1000 NOK)	1843	211.88	23.47	146.00	299.70
religious population pct	1840	90.65	4.89	2.21	100.00
immigrant population pct	1840	4.08	2.42	0.23	23.01
commuter pct	1840	26.82	13.50	5.40	67.90
age 0 to 5 pct	1835	7.48	1.05	3.49	11.01
age 5 to 15 pct	1835	13.87	1.45	10.05	19.28
age 16 to 19 pct	1835	5.17	0.67	2.72	9.74
age 20 to 66 pct	1835	58.47	2.46	52.22	66.63
age 67 plus pct	1835	15.00	3.26	6.48	25.66
voter turnout pct	1843	60.83	5.55	45.12	81.83
local list seats in council pct	1843	5.88	11.52	0.00	100.00
local council seats per 1000 inh.	1840	6.12	4.32	0.10	37.57
share largest party	1843	36.10	9.96	18.52	100.00
herfindahl council	1843	24.89	7.85	13.95	100.00
average household size	1840	2.40	0.18	1.91	3.26
urban population pct	1840	51.42	26.86	0.00	100.00
population (1000)	1840	11.42	30.41	0.35	538.41

Note: National and *specialty* newspapers add up to *non-local newspaper penetration. Tabloid* newspapers are also included in *national* newspapers. The extremely high values of 6.56 and 4.90 for overall reach and national reach are for the municipality that harbors Oslo airport. Variable descriptions are given in the appendix.

[28]The correction procedure in the official index is documented in Ministry of Local Government (2006).

4.3.3 Empirical strategy

The general estimation strategy is OLS with time and municipality fixed effects. Thus, the estimation equation is $E_{it} = \delta \cdot info_{it} + \mathbf{x_{it}}\gamma + c_i + u_{it}$, where E_{it} denotes local public sector efficiency (or production) in municipality i in time period t, $info_{it}$ denotes voter information as measured by the newspaper variables discussed above, $\mathbf{x}_{it}$ is a vector of municipality level controls, and the unobserved effect c_i is allowed to be correlated with $\mathbf{x}_{it}$ and $info_{it}$. While the fixed effects take care of time-constant unobserved heterogeneity, if there are omitted time-varying variables that influence both the newspaper reach and efficiency, we have that $E(u_{it}|\mathbf{x}_{it}, info_{it}, c_i) \neq 0$. Because we cannot provide a variable that induces enough exogenous variation in reach to generate precise results in an instrumental variable approach, in the following we discuss how endogeneity of reach might affect the results obtained.

Unobserved political interest may induce voters to read more newspapers and at the same time have an independent effect on politicians' efforts. Because political interest should be a rather slow changing personal trait (see the below example of reverse causality for one exception), it can be argued that at least in a very short panel such as the one at hand, it will be taken care of by the municipality fixed effects in combination with additional controls. Another variable that may drive both readership and local public sector efficiency is newspaper quality. We do not have a newspaper quality measure and because there are so many aspects to quality, it is in fact unclear how it should be measured. Most quality aspects such as layout do not pose a problem as they may be correlated with reach, yet they should not have an effect on efficiency other than through newspaper reach. However, there is also a quality side to the provision of local information. Higher quality of reporting on local politics may increase reach because more people buy the improved paper – this is the reach effect. At the same time those who read the paper might receive more information than before the quality increase – this is the quality effect. The reach coefficient will pick up both the quality effect as well as the pure reach effect and therefore may lead us to overestimate the effect of pure readership on efficiency, i.e., the effect if we could hold political coverage quality fixed. In the end this doesn't pose a problem, because whether it is newspaper quality or newspaper reach that drives efficiency, both affect efficiency via higher information levels in the population. The reach coefficient then translates to the aggregate effect of changes in information levels on efficiency.

Reverse causality may be an issue, too. One can easily imagine a slack local government or public administration whose actions lead to very low efficiency and this in turn may lead to a spike in the public's interest in obtaining information via the local newspapers.

Whenever readership increases with mismanagement or corruption this implies observing low levels of government efficiency simultaneously with high newspaper readership, thus underestimating the effect of newspaper reach on government efficiency. However, voters may also enjoy reading about successful political leadership and the ensuing achievements, therefore causing an overestimation of the effect. As stated above, we believe that a local newspaper's reach is mainly driven by its entertainment value determined by sports news and the like, and information about politics is received as a by-product. Consequently, applying the fixed effects approach and controlling for municipality characteristics should greatly reduce the scope for the endogeneity issues considered above.[29] This may be especially true in the short run and in a country like Norway, where reading habits across socioeconomic groups do not differ by much anyway.

With both the efficiency as well as the newspaper circulation data it is quite natural to assume that they are error-ridden – the dependent index variables for their complexity and all the general issues involved in measuring efficiency or public sector production described above, newspaper circulation because of its aggregate nature. The presence of measurement error in the efficiency index will lead to imprecise estimates, while measurement error in reach biases our estimates towards zero, causing an underestimation of the efficiency enhancing effects. The limited time series variation of the newspaper variables exacerbates this problem, as the implied high noise-to-signal ratio might entail a large bias towards zero in the coefficient of the newspaper variables.[30]

4.3.4 Results

This section presents the empirical results on the association between local newspapers and efficiency. First, we show results for the full sample of municipalities as a benchmark. Then, we check the robustness of the baseline results by using alternative measures of newspaper activity and restricting the sample to subgroups of municipalities.

Baseline results, full sample of municipalities

The baseline results are presented in Table 4. In the top panel, media influence is measured as the household reach of local newspapers, while local newspaper content is the main explanatory variable in the bottom part. For both panels, column (1) displays results when only local household reach and local newspaper content are included, respectively. Column

[29] Svaleryd and Vlachos (2009) use a similar control approach.

[30] The within variation of reach is 15% the size of its total variation. This number is 30% for content, 40% for efficiency and 25% for production.

(2) adds population and urbanization controls, and column (3) adds variables that account for characteristics of the local political sphere. A variety of other municipality level controls are included in column (4), and column (5) includes local government revenues as a control in order to check whether greater fiscal capacity leads to higher levels of budgetary slack, as suggested by Borge et al. (2008) and Revelli and Tovmo (2007).

The coefficients indicate that increasing the reach of local newspapers by one percentage point increases local public sector efficiency by around .031 points. To put these numbers into perspective, a one standard deviation increase in reach (.27 points) raises efficiency by .8 points or roughly 8% of a standard deviation. The results from the content estimations also suggest that local politicians produce services more efficiently in municipalities where newspapers provide voters with larger amounts of local information. An increase in content by one standard deviation (.19) increases efficiency by 1 point or 10% of a standard deviation. That the content effect is larger than the reach effect is not surprising, because unlike raw circulation, the content variable also accounts for the amount of truly local information that the newspapers convey. In monetized terms, depending on the media variable used, a one standard deviation increase in the media variables implies total annual efficiency gains between $368,000 and $460,000 for a municipality of average population 10,600 and efficiency 100.

The controls suggest that an increase in urban population share leads to higher levels of efficiency, a result that may point to large communities being able to better exploit economies of scale. Higher shares of immigrant and religious population, on the other hand, are associated with lower levels of efficiency, possibly pointing to these groups being less interested in local politics. Having said that, it is not very surprising that many community level controls are of low or no significance at all, since the efficiency index already partly accounts for their effects. When it comes to the political variables, the result that a larger share of seats in the municipal parliament being taken by local lists is associated with higher efficiency is in line with the notion that these parties are not tied to national party politics, that is, they are assumed to exclusively have local issues on their agenda. The share of votes for the strongest party also bears a positive coefficient, possibly due to the fact that a stronger party can accelerate the decision-making processes as there is not as much need for negotiations with other parties.[31] This is consistent with the finding from Borge et al. (2008) that a higher fragmentation of the local council lowers efficiency.[32] The single largest predictor

[31]At the same time, a negative coefficient wouldn't have been too surprising, either, as a larger share may make it easier to extract rents. In this respect, the positive coefficient on seats per capita is also a bit surprising as it may lead to more need for negotiation. On the other hand a larger number of seats may make it harder to form a rent-extracting cartel.

[32]The negative sign on the additional control for party fragmentation doesn't suggest differently, as it only picks up the effect of fragmentation of the opposition parties in the parliament, given a certain size of the largest party.

of efficiency is local revenue. Just as Borge et al. (2008) and Revelli and Tovmo (2007) have stated, higher revenues diminish efficiency, presumably via budgetary slack, which may materialize in exaggerated levels of public employment, low effort, or increased salaries.

The inclusion of the local government revenue variable on both sides of the equation – as a control variable and as the denominator of the efficiency index – leads to unbiased estimates as long as there is no measurement error. In the presence of measurement error in the revenue variable, however, the coefficient of revenue will be biased towards negative unity (Borjas, 1980). As stated before, a variable such as the revenue index will almost certainly be mismeasured and we therefore also estimate the model with local government production as the dependent variable. These estimates are displayed in Table 5. Results with local reach as the main explanatory variable are shown in column (1), whereas local newspaper content is used in column (2). Again, the estimates imply that increases in reach and content both lead to higher levels of public sector production. A one standard deviation increase in content increases production by 1.1 points or 8% of a standard deviation, whereas the effect is 1 point or 6.5% of a standard deviation for reach. Overall the results are very similar to those presented in Table 4.[33]

Robustness checks and placebo tests

As a first check of the validity of our results, in Table 6 we use two other measures of local information. The first one is local newspaper reach including small newspapers registered with the LLA.[34] This variable also shows a significant positive connection with efficiency and production. The second alternative measure is local newspaper reach excluding seven big-city newspapers. The excluded newspapers have a significantly larger radius of operation than the smaller local papers, i.e., in addition to covering local stories they are also regional newspapers. The magnitude and significance of the effect is now very similar to the results obtained earlier, which suggests that the local reach coefficient isn't driven by these high-circulation regional newspapers alone.

In other robustness checks (not reported) we find that the results hold when we exclude all observations where reach or content changed by more than 10 percentage points from

[33]Throughout the paper, the coefficients in the 'production' specifications are larger than in the 'efficiency' specifications, indicating measurement error in the revenue variable.

[34]As we could not obtain data on the geographical distribution of the circulation of these newspapers, all sold newspapers are assigned to the municipality where the newspaper's headquarters is located. As the assignment of the LLA circulation to only one community may bias estimates, we refrain from using it in other estimations.

Table 4: DEPENDENT VARIABLE: PUBLIC SECTOR EFFICIENCY, FULL SAMPLE.

	(1)		(2)		(3)		(4)		(5)	
local newspaper reach	2.9713*	(1.659)	3.1854*	(1.646)	3.1716**	(1.591)	3.9842**	(1.803)	3.1312**	(1.299)
herfindahl local papers							−0.0264	(0.019)	−0.0218	(0.015)
local government revenue									−0.5264***	(0.074)
secondary school pct							−0.9228***	(0.348)	−0.2358	(0.292)
university educated pct							−0.7262	(0.599)	−0.1135	(0.538)
average gross income							0.1684***	(0.056)	0.0931**	(0.046)
religious population pct							−0.0536*	(0.029)	−0.0419**	(0.019)
immigrant population pct							−1.0830***	(0.346)	−0.4739*	(0.282)
commuter pct							−0.2821*	(0.171)	−0.0734	(0.151)
age 0 to 5 pct							0.7527	(0.591)	−0.0235	(0.489)
age 5 to 15 pct							0.8351	(0.606)	0.4953	(0.532)
age 16 to 19 pct							0.5455	(0.653)	0.0108	(0.524)
age 67 plus pct							−0.0933	(0.468)	0.1445	(0.393)
voter turnout pct					0.0599	(0.075)	0.0390	(0.071)	−0.0224	(0.062)
local list seats in council pct					0.0596	(0.040)	0.0563	(0.038)	0.0595**	(0.028)
local council seats per 1000 inh.					0.6522**	(0.301)	0.5270*	(0.287)	0.3580	(0.260)
share largest party					0.1403*	(0.074)	0.1431**	(0.070)	0.1175**	(0.053)
herfindahl council					−0.1178	(0.111)	−0.1345	(0.108)	−0.1091	(0.084)
average household size			0.0895	(3.403)	0.3756	(3.328)	−1.6845	(3.389)	−3.6717	(2.925)
urban population pct			0.2014***	(0.065)	0.1917***	(0.063)	0.1568**	(0.061)	0.1169**	(0.054)
population (1000)			0.5071**	(0.231)	0.5082**	(0.232)	0.3440**	(0.156)	0.0853	(0.129)
year fixed effects	Yes		Yes		Yes		Yes		Yes	
N	1843		1840		1839		1835		1835	

	(1)		(2)		(3)		(4)		(5)	
local newspaper content	7.6835***	(2.887)	7.9884***	(2.905)	7.7621***	(2.737)	6.9039***	(2.666)	5.4136**	(2.614)
herfindahl local papers							−0.0054	(0.015)	−0.0052	(0.014)
local government revenue									−0.5272***	(0.074)
secondary school pct							−0.8893**	(0.350)	−0.2083	(0.292)
university educated pct							−0.7629	(0.605)	−0.1413	(0.541)
average gross income							0.1731***	(0.056)	0.0967**	(0.046)
religious population pct							−0.0486*	(0.028)	−0.0379**	(0.019)
immigrant population pct							−1.0798***	(0.345)	−0.4703*	(0.282)
commuter pct							−0.2819	(0.172)	−0.0729	(0.152)
age 0 to 5 pct							0.7445	(0.594)	−0.0313	(0.491)
age 5 to 15 pct							0.7809	(0.613)	0.4521	(0.533)
age 16 to 19 pct							0.5379	(0.661)	0.0039	(0.530)
age 67 plus pct							−0.1200	(0.466)	0.1239	(0.392)
voter turnout pct					0.0540	(0.075)	0.0364	(0.071)	−0.0245	(0.062)
local list seats in council pct					0.0593	(0.040)	0.0542	(0.038)	0.0578**	(0.028)
local council seats per 1000 inh.					0.6463**	(0.302)	0.5348*	(0.288)	0.3638	(0.261)
share largest party					0.1427*	(0.073)	0.1444**	(0.070)	0.1185**	(0.053)
herfindahl council					−0.1232	(0.110)	−0.1366	(0.107)	−0.1107	(0.083)
average household size			0.8999	(3.315)	1.2070	(3.238)	−0.5125	(3.284)	−2.7538	(2.892)
urban population pct			0.1982***	(0.065)	0.1886***	(0.063)	0.1540**	(0.062)	0.1146**	(0.055)
population (1000)			0.5025**	(0.224)	0.5042**	(0.225)	0.3379**	(0.148)	0.0801	(0.124)
year fixed effects	Yes		Yes		Yes		Yes		Yes	
N	1843		1840		1839		1835		1835	

Note: All estimations include municipality fixed effects. Standard errors in parentheses allow for clustering on the panel variable (municipality code). Dependent variable is local public sector efficiency. The top panel of the table uses 'local newspaper reach' as the main explanatory variable, whereas the bottom panel uses 'local newspaper content'.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Dependent variable: public sector production, full sample.

	(1)		(2)	
local newspaper reach	3.8212**	(1.610)		
local newspaper content			5.7729**	(2.905)
herfindahl local papers	−0.0296	(0.019)	−0.0096	(0.017)
local government revenue	0.0874**	(0.036)	0.0863**	(0.037)
secondary school pct	0.0861	(0.333)	0.1178	(0.333)
university educated pct	0.3420	(0.597)	0.3057	(0.602)
average gross income	0.0552	(0.059)	0.0602	(0.059)
religious population pct	−0.0422**	(0.016)	−0.0370**	(0.016)
immigrant population pct	−0.5098	(0.311)	−0.5067	(0.311)
commuter pct	−0.1290	(0.162)	−0.1280	(0.163)
age 0 to 5 pct	0.0012	(0.573)	−0.0144	(0.575)
age 5 to 15 pct	0.7746	(0.636)	0.7189	(0.640)
age 16 to 19 pct	−0.5427	(0.774)	−0.5491	(0.781)
age 67 plus pct	−0.2838	(0.502)	−0.3068	(0.502)
voter turnout pct	−0.0567	(0.069)	−0.0596	(0.069)
local list seats in council pct	0.0790**	(0.031)	0.0768**	(0.031)
local council seats per 1000 inh.	0.3382	(0.358)	0.3469	(0.361)
share largest party	0.0861	(0.060)	0.0873	(0.059)
herfindahl council	−0.0685	(0.100)	−0.0705	(0.099)
average household size	−6.3259*	(3.230)	−5.1677	(3.172)
urban population pct	0.0682	(0.059)	0.0655	(0.060)
population (1000)	−0.0259	(0.151)	−0.0340	(0.144)
year fixed effects	Yes		Yes	
N	1835		1835	

Note: All estimations include municipality fixed effects. Standard errors in parentheses allow for clustering on the panel variable (municipality code). The dependent variable is local public sector production. Column (1) uses 'local reach' as the main explanatory variable, column (2) employs 'local content'. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

one year to the next. The same is true if we follow Borge et al. (2008) and exclude all communities with efficiency levels below 80 and above 120.

Five additional placebo tests are reported in the bottom part of Table 6. The first four of these variables are the household reach of tabloid newspapers, national newspapers, specialty newspapers as well as the reach of all newspapers excluding those that make up local reach. None of the above newspaper groups convey much local content, and the results in Table 6 show that none of these media measures significantly affects public sector performance. This is in line with the idea that only media that actually carries information on municipal affairs is suited to pressure local politicians into using funds more efficiently. The final measure is the average reach of local papers in the other municipalities that are part of municipality i's county (*fylke*). Changes in neighboring communities' local reach should not have an effect on efficiency in municipality i, which is reflected in the insignificant coefficient.[35] All of these measures combined strongly suggest that the media effect is closely related to the local nature of the newspapers that voters read.

Municipality size, centrality and media effects

The impact of newspapers on government efficiency may depend on municipality population or level of urbanization for a couple of reasons. First, many smaller or non-urban commu-

[35]Such effects may arise through yardstick competition. When more informed voters in the neighboring municipality j receive more efficient politics, yardstick competition may induce higher levels of efficiency in municipality i. Yardstick effects may also be picked up by the commuter share control variable. Revelli (2008) finds such yardstick effects in tax policies of English local governments that share a local television market.

Table 6: Robustness checks and placebo tests.

	(1) production		(2) efficiency	
other measures of local information				
local newspaper (+LLA) reach	3.8386**	(1.596)	3.2383**	(1.297)
local newspaper reach excluding big-city local newspapers	3.6561**	(1.590)	3.1204**	(1.287)
variables unrelated to local information				
tabloid newspaper reach	−1.9953	(1.960)	0.7786	(1.204)
national newspaper reach	−1.1074	(1.461)	1.0043	(0.947)
specialty newspaper reach	−7.2473	(8.818)	−.3235	(7.158)
non-local newspaper reach	−1.1877	(1.428)	0.9333	(0.938)
neighbors' local reach (fylke)	2.4447	(3.648)	0.9931	(3.068)

Note: All specifications are as in column (5) of the base regressions in Table 4. All estimations include municipality fixed effects. Standard errors in parentheses allow for clustering on the panel variable (municipality code). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

nities are rather sparsely populated and newspapers might therefore play a more important role in distributing information than they do in larger communities. Whenever voters in less populous or less urban places rely more on newspapers for information on the community, this implies a larger media effect on efficiency. Second, in smaller communities a larger part of the news may be taken up by local politics, as there is much less going on that the newspapers can actually pick up on, i.e., there is less news competition.[36] If a larger share of the news is made up of local politics in the less populated or non-urban communities, the effect of newspapers on efficiency and production should be larger there. A third argument comes from Kalseth and Rattsø (1995, 1998) who find that it is the smaller jurisdictions that overspend the most on public services. In a similar vein, Sørensen (1984) finds that financial stress, measured as the ratio of expenditure growth compared to the growth of tax revenues is positively correlated with centrality. This can be interpreted as the smaller or non-central municipalities having the largest potential for efficiency gains whenever public information levels and pressure rise.

Municipalities which have been classified as 'central' by *Statistics Norway* are excluded in the top panel of Table 7, leaving us with about 75% of the original sample. Communities with more than 10,000 inhabitants are excluded in the middle panel. The cutoff point is chosen as it marks the 75th percentile of municipality population and 10,000 is at the same time about the average population of a Norwegian municipality. The coefficients in these samples are much larger than in the full sample. In the non-central municipalities the reach effect is more than 50% larger than in the full sample and the content effect almost triples.

When considering municipalities that have a population of less than 10,000, the coefficients are highly significant and again much larger than in the full sample. Finally, the bottom panel of Table 7 excludes all municipalities that are classified as central and at the same

[36]This also means that it takes a much larger scandal to make the news in Oslo or Bergen than it does in some small municipality up north.

Table 7: SAMPLES EXCLUDING LARGE/CENTRAL MUNICICIPALITIES.

	(1) production		(2) efficiency	
central municipalities excluded$^{(a)}$				
local newspaper reach	5.945**	(2.547)	5.001**	(2.041)
local newspaper content	15.739***	(5.558)	13.528***	(5.116)
municipalities above 10,000 in population excluded$^{(b)}$				
local newspaper reach	5.539***	(1.899)	4.736***	(1.524)
local newspaper content	11.622***	(3.514)	9.285***	(2.498)
central municipalities above 10,000 in population excluded.$^{(c)}$				
local newspaper reach	5.466***	(1.790)	4.681***	(1.419)
local newspaper content	11.635***	(3.193)	10.310***	(2.477)

Note: All estimations include municipality fixed effects. Standard errors in parentheses allow for clustering on the panel variable (municipality code). Only the coefficients of the main explanatory variables are reported. All specifications are as in column (5) of the base regressions in Table 4. Column (1) and (2) display the effect on local production and local efficiency, respectively. (a) Municipalities classified as *central* by *Statistics Norway* excluded (N=1356 remaining). (b) Municipalities with population above 10,000 excluded (N=1354 remaining). (c) Municipalities with population above 10,000 and at the same time classified as 'central' are excluded (N=1588 remaining).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

time have a population above 10,000. This leaves more than 85% of the observations that are in the full sample, and again all coefficients are statistically significant and of a larger magnitude than when all municipalities are considered.

The subsample coefficients imply efficiency gains of 1.2 to 2.5 points when reach or content increases by one standard deviation. Similarly, production is up between 1.5 and 2.9 points when reach or content is increased by a standard deviation. Put differently, a one standard deviation increase in reach or content increases the outcome variable by between 10% and 22% of a standard deviation. Consequently, the per capita monetary efficiency gains of newspaper circulation are also larger in these subsamples, and taken together the results strongly suggest that newspapers are especially important for political accountability in small or rural communities.[37]

4.4 Conclusion

We find robust evidence that having an active and locally focused newspaper landscape can be a cornerstone in establishing political accountability. The basic idea is that a more informed electorate provides incentives for incumbent politicians to behave well. Because newspapers serve as a primary source of information for voters, their circulation in a jurisdiction should have an impact on policy outcomes. Using panel data on public sector efficiency and various measures of electorate information based on newspaper circulation, we find empirical support for this argument. The results suggest that a larger share of informed voters goes with larger efficiency. This connection is stronger in small and non-central municipalities. Summarizing all results from the full and restricted samples, we can

[37]As a robustness check, we excluded all municipalities with a population below 1,000. This leads to slightly higher precision in the estimations.

state that a one standard deviation increase in the media variable is associated with an increase in efficiency or production of .8 to 2.9 points – 6.5% to 22% of a standard deviation – depending on media variable and sample.

The effects are moderate, in that newspaper circulation explains only a fraction of the total variation in the outcome variable. This is consistent with results reported in other research on the impact of mass media on political economy variables (Gentzkow et al. (2009), Snyder and Strömberg (2010), Svaleryd and Vlachos (2009)) and should not come as a surprise, considering that there are many determinants of efficiency aside from information levels in the electorate. We believe another important reason to be that while voters care about efficiency in general, the weighting in the production index and its sub-indices cannot mirror varying local preferences across municipalities. An additional factor may be that newspaper circulation in Norway is unusually high. If information returns from newspaper circulation are diminishing, effects on accountability in countries with less evolved media markets may be larger.[38] The fact that studies focusing on such countries (Reinikka and Svensson (2005), Besley and Burgess (2002)) tend to find larger media effects than studies concerned with developed countries may be interpreted as tentative evidence for such decreasing returns to information.

4.5 Appendix

Updating. For reasons of clarity, we drop all super- and subscripts. The competence variable θ refers to the incumbent and a, y and ε denote effort, public good production and the shock respectively in period 1. All voters are rational Bayesians who use all disposable information to form their beliefs. First off, an uninformed voter knows the distributions $\theta \sim N(\bar{\theta}, \sigma_\theta^2)$ and $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. Thus, his prior belief about the incumbent's competence is $\tilde{\theta} = \bar{\theta}$. Then, he receives the signal y which is informative about θ. The uninformed voter knows that $y = a + \theta + \varepsilon$ but he cannot distinguish between the three terms of the sum on the right-hand side. Expecting the level of effort in period 1 to be $\tilde{a}$, uninformed voters face a signal extraction problem yielding that the density function of $\theta|y$ is:

[38]We would like to thank an anonymous referee for pointing this out to us.

$$f(\theta|y) = \frac{f(y|\theta)\cdot f(\theta)}{f(y)} = \frac{\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\cdot \exp^{-\frac{1}{2\sigma_\varepsilon^2}(y-\theta-\tilde{a})^2}\cdot\frac{1}{\sqrt{2\pi\sigma_\theta^2}}\cdot\exp^{-\frac{1}{2\sigma_\theta^2}(\theta-\bar{\theta})^2}}{\frac{1}{\sqrt{2\pi(\sigma_\theta^2+\sigma_\varepsilon^2)}}\cdot\exp^{-\frac{1}{2(\sigma_\theta^2+\sigma_\varepsilon^2)}(y-\tilde{a}-\bar{\theta})^2}}$$

$$= \frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sqrt{\frac{\sigma_\theta^2\cdot\sigma_\varepsilon^2}{\sigma_\theta^2+\sigma_\varepsilon^2}}}\cdot\exp^{-\frac{1}{2\cdot\left(\frac{\sigma_\theta^2\cdot\sigma_\varepsilon^2}{\sigma_\theta^2+\sigma_\varepsilon^2}\right)}\left(\theta-\frac{\sigma_\varepsilon^2\bar{\theta}+\sigma_\theta^2(y-\tilde{a})}{\sigma_\varepsilon^2+\sigma_\theta^2}\right)^2}.$$

So, for a given level of public good production, the incumbent's competence is drawn from a normal distribution with mean $\frac{\sigma_\varepsilon^2\bar{\theta}+\sigma_\theta^2(y-\tilde{a})}{\sigma_\varepsilon^2+\sigma_\theta^2}$ and variance $\frac{\sigma_\theta^2\cdot\sigma_\varepsilon^2}{\sigma_\theta^2+\sigma_\varepsilon^2}$. Consequently, after having received the signal y, uninformed voters expect the incumbent's competence to be $\tilde{\theta}|y = \frac{\sigma_\varepsilon^2\bar{\theta}+\sigma_\theta^2(y-\tilde{a})}{\sigma_\varepsilon^2+\sigma_\theta^2}$. The posterior belief, $\tilde{\theta}|y$, is a weighted average of the prior and the information contained in y. It is intuitive that if the prior is relatively noisy (high σ_θ^2), more weight is given to the signal. In contrast, if the signal is not very precise (high σ_ε^2), a greater weight is given to the prior. □

Proof of Proposition 4.1. The incumbent chooses a to maximize her objective function

$$p^I(a)\cdot R - C(a),$$

taking as given the voters' expectation about effort $\tilde{a}$. This objective function must be concave for the first-order condition to be sufficient for a maximum. To prove concavity, we have to pay attention particularly to p^I, the probability of reelection for the incumbent. p^I is defined by $1 - F(0;\mu,\sigma^2)$, where F is the distribution function of a normal distribution with mean μ and variance σ^2. F is both convex and concave on part of its domain. We will show, however, that there exists an upper bound for R that ensures the incumbent's objective function to be concave (see also Ashworth (2005)).

First off, to ease the analysis we standardize F so that $p^I = \Phi(-\frac{\mu}{\sigma})$. Next, we define $x := -\frac{\mu}{\sigma}$ so that the objective function turns into

$$[1-\Phi(x)]\cdot R - C(a).$$

Thus, the first derivative is

$$-\phi(x)\cdot\frac{\partial x}{\partial a}\cdot R - C'(a),$$

yielding the second derivative

$$-x\phi(x) \cdot (-1) \left(\frac{\partial x}{\partial a}\right)^2 \cdot R - C''(a). \tag{4.15}$$

The term $-x\phi(x)$ is the slope of the density function $\phi(x)$. At the inflection points $x = -1$ and $x = 1$, the slope has its largest absolute values. Thus, $-x\phi(x)$ is bounded between the minimal value $-1/\sqrt{2\pi e}$ and the maximal value $1/\sqrt{2\pi e}$. In addition, straightforward calculation shows that

$$\frac{\partial x}{\partial a} = -\frac{1}{\sigma} \cdot \frac{\partial \mu}{\partial a} = -\frac{1}{\sigma} \cdot \left[\frac{\alpha \cdot \sigma_\varepsilon^2 + \sigma_\theta^2}{\sigma_\varepsilon^2 + \sigma_\theta^2}\right] = -\frac{1}{\sigma} \cdot \mu'.$$

Now, with the upper bound

$$R \leq \frac{\sigma^2}{(\mu')^2} \cdot \sqrt{2\pi e} \cdot C''(0)$$

the maximal value of the left-hand term in (4.15) is given by

$$x\phi(x) \cdot \frac{(\mu')^2}{\sigma^2} \cdot R \leq \frac{\sigma^2}{(\mu')^2} \cdot \sqrt{2\pi e} \cdot C''(0) \cdot \frac{(\mu')^2}{\sigma^2} \cdot \frac{1}{\sqrt{2\pi e}} = C''(0).$$

So, finally, with $C''' \geq 0$, for the second derivative we have that

$$x\phi(x) \cdot \left(\frac{\partial x}{\partial a}\right)^2 \cdot R - C''(a) \leq 0.$$

Consequently, the objective function is concave for $a \geq 0$ and the solution to the first-order condition, a^*, constitutes a maximum.

In equilibrium, the voters correctly expect $\tilde{a} = a^*$ resulting in $\mu = 0$ and $p^I = 1 - F(0; 0, \sigma^2)$. Then, the first-order condition turns into

$$\frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \left[\frac{\alpha \cdot \sigma_\varepsilon^2 + \sigma_\theta^2}{\sigma_\varepsilon^2 + \sigma_\theta^2}\right] \cdot R = C'(a).$$

As $C'(a)$ is invertible, the incumbent's effort level in equilibrium is uniquely defined by

$$a^* = (C')^{-1} \left(\frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{\sigma_\varepsilon^2 + \sigma_\theta^2}} \cdot \frac{\alpha \cdot \sigma_\varepsilon^2 + \sigma_\theta^2}{\sqrt{\alpha^2 \cdot \sigma_\varepsilon^2 \sigma_\theta^2 + \sigma_\theta^4}} \cdot R \right).$$

□

Proof of Proposition 4.2. Proposition 4.1 says that

$$a^* = (C')^{-1}\left(\frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{\sigma_\varepsilon^2 + \sigma_\theta^2}} \cdot \frac{\alpha \cdot \sigma_\varepsilon^2 + \sigma_\theta^2}{\sqrt{\alpha^2 \cdot \sigma_\varepsilon^2 \sigma_\theta^2 + \sigma_\theta^4}} \cdot R\right).$$

To proof that $\frac{\partial a^*}{\partial \alpha} > 0$, we define

$$\upsilon := \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{\sigma_\varepsilon^2 + \sigma_\theta^2}} \cdot \frac{\alpha \cdot \sigma_\varepsilon^2 + \sigma_\theta^2}{\sqrt{\alpha^2 \cdot \sigma_\varepsilon^2 \sigma_\theta^2 + \sigma_\theta^4}} \cdot R.$$

Recalling that $C(a)$ is a strictly increasing function, C^{-1} is strictly increasing, too. Thus, the effort level, a, is higher for larger values of υ. So we have to show that $\frac{\partial \upsilon}{\partial \alpha} > 0$, where α denotes the share of informed voters. The partial derivative yields:

$$\frac{\partial \upsilon}{\partial \alpha} = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{\sigma_\varepsilon^2 + \sigma_\theta^2}} \cdot \frac{1}{\sqrt{\sigma_\theta^2}} \cdot \frac{\sigma_\theta^2 \sigma_\varepsilon^2 (1-\alpha)}{(\alpha^2 \sigma_\varepsilon^2 + \sigma_\theta^2)^{\frac{3}{2}}} > 0$$

for $0 \leq \alpha < 1$. □

Table 8: DESCRIPTION OF VARIABLES (NORWAY).

Variable	Description
Aviskatalogen data[a]	
local newspapers	Includes big-city daily newspapers, local dailies and non-daily local newspapers that are published at least twice a week. Of the weekly newspapers, those organized in *LLA* are included in the variable 'local newspaper (+*LLA*) reach' only.
local newspapers excl. big-city	Local newspapers excluding big-city dailies.
tabloids	Newspapers *Dagbladet* and *Verdens Gang*.
national newspapers	Tabloids plus *Aftenposten* (nationwide focus).
specialty newspapers	Newspapers that cater to special interest audiences.
newspaper penetration	Share of households reading a newspaper. Values > 1 denote more than one paper on average.
newspaper content	The newspaper penetration of newspaper n in municipality i, weighted by the share of newspaper n's total sales occuring in that municipality. Summed over all newspapers sold in i.
herfindahl local papers	Herfindahl index of local papers sold within a municipality ($\sum {s_n}^2$, where s_n are the market shares of newspapers n in the municipality).
LLA data	
LLA newspapers	Local newspapers not included in the *MBL Aviskatalogen*. Regionally disaggregated circulation data not available, the full circulation is assigned to the paper's home municipality.
Local gov't data	
local public sector efficiency	Efficiency index, developed by Borge et al. (2008).
local public sector production	Production index, developed by Borge et al. (2001).
local government revenue	Revenue index, developed by Borge et al. (2008).
NSD variables	
secondary school pct	% of population over 16 years whose highest degree is secondary.
university educated pct	% of population over 16 years whose highest degree is tertiary.
KOSTRA variables (ssb.no)	
average gross income (1000 NOK)	Gross income per taxpayer.
religious population pct	Percent of residents registered with the state church.
immigrant population pct	None
commuter pct	% of population aged 20 to 66 commuting to other municipalities.
0 to 5 years pct	% of population aged 0 to 5 years.
5 to 15 years pct	% of population aged 5 to 15 years.
16 to 19 years pct	% of population aged 16 to 19 years.
20 to 66 years pct	% of population aged 20 to 66 years.
67 plus pct	% of population aged 67 and older.
voter turnout pct	Turnout in municipal elections.
local list seats in council pct	Percent of seats in the municipal council taken by non-national parties (only one election in the period under consideration: 2003).
local council seats per 1000 inh.	The number of seats in the municipal council is at the municipalities' discretion, as long as it doesn't fall below a lower limit.
share largest party	Vote share of the strongest party in the municipal council.
herfindahl council	Herfindahl index of party fragmentation: ($\sum {sh_p}^2$, where sh_p are the seat shares of parties p in the local council).
average household size	None
urban population pct	None
population (1000)	None

Note: (a) A complete list of newspapers included in the estimations, their classification as local, big-city, tabloid, national and specialty newspapers along with some additional material can be accessed at `http://wwwuser.gwdg.de/~uwvw3/norway_webappendix.htm`.

5 The geography of local television markets and the allocation of federal funds in the United States

5.1 Introduction

In this chapter, we examine whether local television news, the most popular source of political information in the United States (Napoli and Yan 2007), affects the allocation of federal grants. More precisely, we are interested in whether a geographical pattern of grant allocation exists, which reflects a well-documented bias in local TV news programming: local news content focuses on the core areas of TV markets, where the majority of TV stations is located (Adams 1980; Althaus and Trautman 2008; Heider 2000; Kaniss 1997). Thus, political actions in the core area of a TV market are more likely to receive coverage which suggests that voters in that area are better informed about politics than voters in remote regions. Due to this publicity effect, the amount of funds channeled to a location should systematically decline in the distance to the core area of a TV market.

We analyze the relation between TV markets and public spending both theoretically and empirically. At first, we present a simple model based on Strömberg (2004a,b) which serves to establish the link between local TV news and grant allocation. The model shows that television stations bias their news programs towards the markets' core areas because news production is less expensive there. As a consequence, there are more voters who are informed about politics in these regions. Then, the model explains how this information structure in the electorate affects a politician's incentives concerning grant allocation. The main prediction of the model is that politicians direct more funds to areas close to media centers.

We test this prediction empirically using U.S. data on the county-level. The dependent variable is the amount of federal grants per capita allocated to a county. Using a county's distance to the nearest media city in the market as our main explanatory variable, we find a substantial effect of distance on grant spending. A county which is located twice as far away as the mean distance receives 3.4% less in grants per capita which makes for an average total loss of about $2.8 million per county per annum.

Our analysis builds on the pioneering work by Strömberg (2004a,b) who was the first to explicitly analyze the influence of mass media on fiscal policy. In a formal model, Strömberg (2004a) shows that certain groups of voters receive favorable policies when the mass media report more heavily on campaign promises concerning these groups. Analyzing data from the United States, Strömberg (2004b) tests empirically whether better informed groups in the electorate receive favorable policies. He argues that the diffusion of radio affected the level of political information and finds that counties where many households owned a radio received more funds from a major New Deal program in the 1930s. In contrast to Strömberg's empirical paper, our analysis examines how policy is affected by the geographic distribution of information, which results from biased news programming in a market where everyone has access to the medium. Other research that deals with the media's role for accountability includes Snyder and Strömberg (2010), who find empirical evidence that newspaper coverage affects voters' information about their U.S. House representatives. A better fit of newspaper markets with districts is found to increase information levels, resulting in more accountable Congressmen. Besley and Burgess (2002) also deal with the media's role for political accountability. They analyze panel data from India and find that state governments provide more public food and calamity relief in hard times when newspaper circulation in the respective state is higher.

5.2 Theory

The model builds on the framework by Strömberg (2004a,b). A politician distributes a given budget within his constituency. From now on, we refer to the politician as a U.S. governor who receives grants from the federal government and allocates the money across counties in his state. The governor distributes grants strategically in order to win the next election. He knows that funds generate more votes in counties where many voters attribute the benefits derived from grant spending to him. He further knows that the share of these informed voters is larger in counties which are extensively covered by local television news. The model shows that economic forces make local television stations broadcast more news from counties located in the core area of a TV market. This geographical bias in local news programming yields lower shares of informed voters in remote counties and larger shares of informed voters in core areas. Thus, incentives arise for the governor to spend more in these core areas.[39]

The timing of the game, which is described in the next section, goes as follows: the governor distributes grants across counties. Then, the TV stations decide on their reporting strategy and voters choose a station to watch. Finally, an election takes place.

[39] At first glance, the choice of telling the story from a governor's perspective may appear arbitrary, since there is a large number of other players involved in the allocation procedure (such as senators, House members, mayors, or non-governmental organizations). However, a large share of federal funds is distributed to local recipients via the state administration. As states often enjoy quite some discretion in the allocation decisions (Lee 2003), we believe governors as chief executives of their states to be very important players in the allocation process. However, it is important to notice that local television affects the incentives of all players involved in the allocation process in the same way: a higher probability to receive coverage implies more political benefits for voters.

The governor's allocation problem. The governor aims at winning votes by allocating total grants G across the counties $c = 1, 2, ..., C$ in his state such that

$$\sum_c p_c \cdot g_c = G, \tag{5.1}$$

where g_c denotes grant spending per capita and p_c the population in county c. A voter v in county c enjoys utility $\omega_c = \ell_c \cdot \omega(g_c)$ from grant spending. We assume $\omega' > 0$, $\omega'' < 0$ and $\omega'(0) = \infty$. The parameter ℓ_c implies that voters may value grant monies differently across counties.

Voter v votes for the incumbent if his total utility under the incumbent's regime exceeds some reservation utility $\overline{\omega_v}$:

$$\tau_v \cdot \omega_c(g_c) - \sigma_v - \delta \geq \overline{\omega_v}, \tag{5.2}$$

and for a challenger otherwise. Besides grant spending, voters evaluate the incumbent in terms of ideology. Voter v's ideological position is described by $\sigma_v + \delta$, where σ_v denotes a personal component and δ is the incumbent's general popularity in the electorate; both are random variables and may take on positive as well as negative values (see Persson and Tabellini 2002).

The indicator variable τ_v introduces the media effect into the model. τ_v equals one if voter v attributes grant spending to the governor and zero otherwise. Thus, only informed voters consider grant spending in their voting decision. We assume that television news constitutes the only source of information about grant monies.[40] In the next section, we show that television stations bias their news programs towards counties which are located in the TV market's core area. Consequently, the share of informed voters, ϕ_c, for whom $\tau_v = 1$ is larger in those counties and smaller in remote counties.

Local TV news. We assume that the television market is congruent with the state territory such that the electorate and the TV audience coincide. Two TV stations A and B compete for audience by broadcasting local news. Each station allocates total air time N across counties such that $\sum_c n_c^s = N$, where n_c^s is the news time devoted to county c by station $s = A, B$. As the probability of a single vote being decisive is zero, voters consume news for its entertainment value only. They like to know what is going on in their community and their expected utility of watching a station's news is $u_c(n_c^s) = \phi(n_c^s) \cdot \bar{u}_c$. ϕ denotes the probability that a voter actually catches sight of a newscast. We assume $\phi'(n_c) > 0$,

[40] We make this assumption for ease of exposition. Obviously, there are alternative sources like newspapers, radio or social contacts.

$\phi''(n_c) < 0$ and $\phi'(0) = \infty$. $\bar{u}_c$ denotes the exogenous utility a voter in county c derives from seeing the news from his resident county. In other words, a voter knows that his expected utility from seeing an interesting newscast increases in the news time a station devotes to his home county. A voter v chooses to watch station A if

$$u_c(n_c^A) - u_c(n_c^B) \geq \xi_v \tag{5.3}$$

and station B otherwise. ξ_v denotes how voter v evaluates fixed characteristics of station A relative to station B, e.g. sympathy for anchormen, the style of presenting news or the ideological bias of a station. A positive value of ξ_v implies that voter v favors station B whereas negative values indicate that station A is favored, leaving news levels out of consideration. F denotes the distribution function of ξ_v with density f. Consequently, a voter watches station A's newscasts with probability $F[u_c(n_c^A) - u_c(n_c^B)]$.

Having described the demand for news, we now turn to the supply side. When television stations decide on which events in their market to cover, they have to consider the costs of news production. Kaniss (1997) and Heider (2000) argue that it is more expensive to report from distant counties for a number of reasons. First, the set-up costs of sending a news team to remote areas are higher. Long travel time often means overtime payments and higher expenditures for gasoline, meals and hotels. Further, it may be necessary to purchase satellite time to immediately transfer a report to the newsroom. The alternative of employing freelancers is also expensive, because freelance fees are often relatively high (Heider 2000). Second, as journalists usually reside in the core area of a media market, they find gathering information to be easier there than in remote counties. This is because they have better access to contact persons and a more comprehensive knowledge of local institutions in the core area (Kaniss 1997). With budgets for the production of news being rather low (Kaniss 1997), geographical cost differentials should critically affect the decision of assignment editors. Being situated at the same place, both stations A and B face the same marginal cost of reporting from a county, k_c. These costs are high in distant counties whereas they are low in counties near the stations.

Both stations maximize expected profits with revenues coming from advertising. Advertisers on local television usually are less concerned with targeting specific socio-demographic groups than advertisers in newspapers (Kaniss, 1997). Consequently, both stations simply maximize their audience regardless of composition and we normalize advertising revenues per viewer to one. Station A maximizes expected profit

$$E[\pi^A] = \sum_c \left[p_c \cdot F[u_c(n_c^A) - u_c(n_c^B)] - k_c \cdot n_c^A\right] \tag{5.4}$$

subject to the air time constraint, and stations B's optimization problem is identical. Both stations decide simultaneously and non-cooperatively about allocating news time across counties. As the two stations face exactly the same optimization problem, the unique Nash-Equilibrium[41] has both stations broadcasting the same amount of news on each county in the market. Thus, equilibrium news allocation is given by a pair of strategies $(\mathbf{n^A}, \mathbf{n^B})$ satisfying $n_c^A = n_c^B = n_c^*$, the air time constraint and the first-order condition

$$p_c f \phi_c'(n_c^*)\bar{u}_c - k_c = \lambda, \quad \lambda > 0, \tag{5.5}$$

which implies that in equilibrium the marginal effect of a news unit on expected profit must be equal across all counties. Equation (5.5) summarizes the message of the model regarding news time allocation. Assuming for $n_c^A = n_c^B = n_c^*$ the density f to be $f(0) = 1$, we obtain the equilibrium news $n_c^* = n^*(k_c, p_c, \bar{u}_c)$. The stations broadcast more news about counties where producing news is less expensive (low k_c), where the intrinsic utility of watching the news is higher (high $\bar{u}_c$) and where the population is large (high p_c).

The amount of local news from a county determines the share of informed voters, $\phi_c = \phi(n_c^*(k_c, p_c, \bar{u}_c))$. Our main focus is on the result that more voters obtain information in counties where the reporting costs, k_c, are low. Recalling our argument on the geography of reporting costs, the model predicts that TV newscasts contain less news from remote counties. Consequently, the share of informed voters in a county decreases in distance to the center of the TV market.

Strategic Allocation of Grants. Having determined the shares of informed voters in the counties, we can now compute the expected number of votes for the governor. For simplicity, we assume that $\sigma_v + \overline{\omega_v}$ in (5.2) is distributed uniformly with mean m_c and density ψ_c. Let γ_c denote voter turnout in county c, then the expected number of votes for the governor is given by the left hand side of

$$\sum_c \gamma_c p_c \left(\frac{1}{2} + \psi_c(\phi_c \cdot \omega_c - m_c - \delta)\right) \geq \frac{1}{2}\sum_c \gamma_c p_c. \tag{5.6}$$

Inequality (5.6) states that the governor wins the election if he accumulates more than half of all votes cast. Apparently, for any allocation of grants it depends on the realization of the

[41]Basically, the model of competition between the two television stations is analogous to models of redistributive politics, as introduced by Lindbeck and Weibull (1987) and extended by Dixit and Londregan (1996). As the basic model has already attained textbook status (see, e.g., Persson and Tabellini, 2002) we abstain from extensively proving uniqueness and existence of the Nash-Equilibrium in this simple setting. Lindbeck and Weibull (1987) or Strömberg (2004a) clearly characterize equilibrium strategies and give proof. We assume F to be of a form which satisfies the concavity condition for the objective functions of both stations for existence of equilibrium.

general popularity shock, δ, whether (5.6) is satisfied and the incumbent wins the election. Contingent on grant allocation the probability of reelection, P, is given by

$$P = \Omega \left[\frac{1}{\sum_c \gamma_c p_c \psi_c} \sum_c \gamma_c p_c \psi_c (\phi_c \cdot \omega_c - m_c) \right],$$

where Ω denotes the distribution function of δ. In equilibrium, the optimal allocation of grants $\mathbf{g}^*$ which maximizes the probability of reelection satisfies the first-order condition

$$\gamma_c \cdot \psi_c \cdot \phi_c \cdot \ell_c \cdot \omega'(g_c^*) = \mu, \quad \mu > 0 \tag{5.7}$$

and the budget constraint.[42] Equation (5.7) implicitly defines the equilibrium spending levels $g_c^* = g^*(\gamma_c, \psi_c, \phi_c, \ell_c)$. Recalling the concavity of $\omega(g_c)$, it is easy to see that counties with high electoral clout ($\gamma_c \cdot \psi_c \cdot \phi_c \cdot \ell_c$) receive more grants. Specifically, the governor directs more grants to counties where voter turnout (γ_c) is high, where many swing voters live (high ψ_c), where an extra dollar generates more utility (high ℓ_c) and where the share of informed voters is large (high ϕ_c). This is intuitively reasonable since larger values of each of these factors imply that an extra dollar is converted more effectively into votes. Thus, more funds are channeled to counties with high electoral clout while counties with low electoral clout receive less money. This allocation of funds finally equalizes the marginal payoff of funds across counties.

The central message of the model for the empirical analysis is that the governor allocates more money to counties which are located near or in the core area of a TV market. This follows directly from the result that counties with larger shares of informed voters (ϕ_c) receive more grants. We know that $\phi_c = \phi(n_c^*(k_c, \bar{u}_c, p_c))$ and also that the TV stations broadcast more news from distant counties due to low reporting costs (low k_c). Thus, the share of informed voters is larger there (high ϕ_c), which implies larger amounts of grants. It also follows from the results in the media section that more grants are distributed to counties where the intrinsic utility of watching the news is high (high $\bar{u}_c$) and where the population is large (high p_c).

The next section outlines how these theoretical predictions can be empirically tested using data on United States television markets and the allocation of federal grant money across counties.

[42]Due to the concavity of $\omega(g)$, the Hessian of the incumbent's payoff function is a diagonal matrix with all elements on the diagonal being negative. Thus, the Hessian is negative definite and the incumbent's payoff function is concave. Finally, as the constraint set is convex, $\mathbf{g}^*$ constitutes a unique global constrained maximum.

5.3 Data and estimation approach

With the theoretical idea that a politician will direct larger amounts of funds towards counties where they generate more media attention comes the need to identify such places. The main argument of our analysis is that television news programs are biased in favor of counties in the core area of a television market whereas distant counties receive only little attention. Thus, in order to measure this geographic determinant of voter information we need to identify the core areas of TV markets. To this end, we denote places that appear in the name of a media market defined by Nielsen Media Research as 'media centers' or 'media counties' and use as a main explanatory variable the distance of a county to the nearest media center. The United States is split up into 210 of these media markets – called Designated Market Area (DMAs) – whose inhabitants tend to watch the same television stations. DMAs can cross state borders and are named after the city or cities where most TV stations are located (Figure 1 shows DMA boundaries).[43] Guided by the theoretical predictions, we expect public expenditure to generate more media coverage in counties that are media centers or are close to media centers, due to lower reporting cost in the proximity of the stations' headquarters. This in turn leads to higher shares of informed voters and consequently to higher levels of funds. As the distances vary a great deal across DMAs, we use a relative distance measure, calculated as county c's distance to the nearest media center divided by the average distance to a media center within c's DMA. The relative measure captures that TV stations divide their local news time mainly among events that occur within their own DMA.

We would ideally like to have as a dependent variable funds that politicians can freely distribute across their constituency. Since such monies are hardly ever available, the strategic use of resources is measured by the per capita amount of US federal grants allocated to counties in the 48 contiguous states in 2000. These grants are awarded in hundreds of programs ranging in focus from agriculture to health and the total amount awarded in 2000 was $309 billion, or $1050 per capita. The allocation of funds across counties results from a complex political bargaining process. Many different actors participate in this process, among them governors, senators, House members, mayors or private organizations. As we have already outlined in the theory section above, we tell the story paradigmatically from a governor's perspective. However, it is important to note that the all politicians face incentives created by television market geography that affect political outcomes in the same way: regions where media activity is high receive favorable policies.[44]

[43]Whenever multiple cities of importance to the media market are present, they enter the DMA name (e.g. the San Francisco DMA is called San Francisco-Oakland-San Jose). When this is the case, our distance variable measures distance to the nearest city appearing in the DMA name.

[44]Studies by Albouy (2009a and 2009b), Knight (2004), Levitt and Snyder (1997) as well as Snyder and

The following baseline specification empirically represents this main concept of grant spending contingent on media coverage, which was already derived in the theoretical model. All variables except for those measured in percentages are entered as logs:

$$g_c \;=\; t_d \cdot \phi_c^D \;+\; t_m' \cdot \phi_c^M \;+\; \eta' x_c \;+\; \varepsilon_c. \tag{5.8}$$

Here, g_c is the dollar amount of federal grants per capita awarded to county c. The share of informed voters ϕ_c from the theoretical model is proxied for by ϕ_c^D, the county's distance to the nearest media center within its DMA. While this is our main predictor of grant spending, further media related variables that may also have an effect on the level of information in the population make up vector ϕ_c^M. The control vector x_c consists of socio-demographic county characteristics, political economics variables as well as a number of population and geography controls.

Additional TV market variables. The first variable in ϕ_c^M is the number of TV stations licensed in county c. Aside from being tantamount to zero distance to the nearest media outlet whenever the variable takes on values greater than zero, it accounts for possibly having multiple outlets at one's disposal.[45] These stations are made up in large part of affiliates of the four big networks ABC, CBS, NBC, FOX plus the PBS stations. To these full-service stations we add stations that have been classified as 'Class-A' by the Federal Communications Commission (FCC).[46] Taken together, there are over 850 counties which contain at least one station, compared to about 340 media cities. The number of stations may be a predictor of federal grant spending, because having at least one media outlet in the immediate vicinity will increase chances of news being picked up on, even without being a media city. In addition, the variable captures possible competition effects in the local media market. We expect the variable's coefficient to bear a positive sign because the share of informed voters could be larger in counties with more TV stations. Obviously, all media center counties are host to at least one TV station and so the distance and number of TV stations variables are to some extent collinear.

As quite a few DMAs cross state borders, we add to ϕ_c^M an indicator for whether a county is *'out-of-state'*, i.e. all media cities in the DMA are located in another state than county c. It accounts for counties possibly being marginalized within their own DMA information-wise, because it may be more costly for TV stations from the media center to cover them. This

Strömberg (2010) also use federal grants as an outcome in research on the strategic distribution of resources.

[45] The variable is calculated as log(number of TV stations +1). The estimated coefficient t_m^{no} will therefore not represent an elasticity. t_m^{no} can be transformed into an elasticity as follows: $t_m^{no} \cdot [j/(j+1)]$, where j is the number of TV stations.

[46] 'Class-A' stations are given protected status by the FCC because they convey local information. They are typically unaffiliated with networks and sometimes have lower reach than network TV stations.

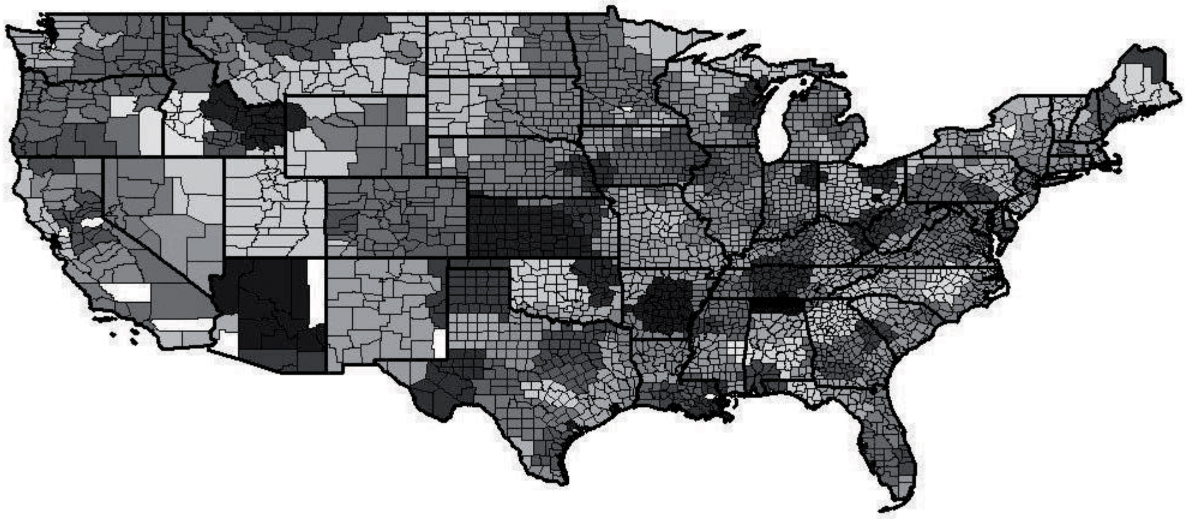

Figure 1: DMAs and state borders

could happen whenever networks for gathering information are state-specific, e.g. because they require connections to the state administration. This would effectively create a home-state-bias in reporting.

We also add the share of the DMA population living in the same state as county c. This variable may also measure marginalization, such that a smaller share means less media attention because stories from that state are of relevance to a smaller part of the total DMA audience. However, if county c is *not* out-of-state and TV stations have a home-state-bias, a smaller share of DMA population living in the same state as c might lead to higher grants, as the TV station caters to a smaller within-state audience among which the news time is divided. In an attempt to disentangle these effects, we add an interaction of the two variables.

Finally, we include the distance to the nearest media city outside the own DMA in order to check whether the 'distance to own DMA media center' variable merely measures distance to a metro area rather than a true media effect. We expect this variable to exhibit no significant relation to grant allocation.

Political economics controls. Theory advises to include variables measuring voter turnout (γ_c) and voter mobility (the number of swing voters, ψ_c). The model predicts that high voter turnout as well as high voter mobility lead to more grant awards. Relative voter turnout in county c is calculated as the share of the population that voted in the 1996 presidential election, divided by the average turnout in the state county c is located in.

Voter mobility is accounted for by including the number of times the majority in presidential elections in county c has shifted from 1980 to 1996.[47]

In addition, we include the distance to the state capital relative to all other counties in the same state. This measure is believed to be negatively related to federal grants due to the higher lobbying costs remote counties face (Borck and Owings, 2003). A high percentage of residents employed by the federal or local government supposedly is associated with greater political influence and thus higher grants per capita. Because we do not have data on the number of politicians hailing from each county, we assume that the percentage of residents employed by federal, state and local governments in county c is highly correlated with the number of political agents operating on behalf of county c. The government employment variables then also measure 'politician density' and thus account for differences in how well counties are represented in the political process. A Herfindahl index of ethnic fragmentation is included, because Alesina et al. (1999) find that fragmentation leads to higher intergovernmental spending. An index value of unity indicates a completely homogenous population. We add the white percentage of the population, because different ethnic compositions can result in the same Herfindahl index.

Sociodemographic and urbanity controls. The sociodemographic controls measure the financial needs of a county and therefore the relative importance of federal grants to different groups in the population (ℓ_c in the model). In addition to the percentage of residents with at least a bachelor's degree, which is essentially another control for how informed voters are, we include per capita income, poverty rate, unemployment rate, share of females, percentage of high school dropouts, county expenditure per capita, percentage of population under 24, bank deposits per capita and the ratio of mean to median income.

We must take care not to confound the effect of media outlets with general urban/rural characteristics that may lead to different levels of grant allocation. Thus, we include in x_c the county population, land area, population growth in the 1990s, the county longitude and latitude, an indicator for counties of more than 200,000 inhabitants, an indicator for whether the county is classified as a metropolitan/micropolitan area, and population density. A nice side effect of adding these controls is that they should at the same time account for the attractiveness of a county for setting up a television station. As Greenberg (1969) states: 'station profitability depends primarily on audience size'. In the same vein, Gentzkow (2006) finds that it was counties with high population and population density that were historically most profitable for TV station owners. Consequently, these counties were the first to be granted television licenses and to this day the number of TV stations in a county is largely predicted by these variables. Adding the controls listed above thus reduces the potential for the coefficient on 'number of television stations' to be confounded with unobserved market

[47]The log of mobility is calculated as log(majority shifts +1).

size effects that may drive both the number of stations and grant allocation. All estimations allow for clustering of standard errors by state and include state fixed effects. Summary statistics are shown in Table 15 and variable descriptions in Table 16.

5.4 Empirical results

This part of the analysis presents the baseline OLS estimations in section 5.4.1. Because our two key variables hardly change over time, panel analysis in order to net out fixed county characteristics is not an option.[48] Rather, we use a wide array of controls in order to isolate the association between the media and grant spending. Robustness checks and corrections for spatial dependency are shown in 5.4.2 and 5.4.3, respectively. In all estimations, the 48 counties containing the state capitals have been excluded from the analysis. As can be seen from Table 15, the average county in the sample excluding state capitals is awarded around $1,000 per capita in grants. However, because some grants that cannot be attributed to a single county are assigned to the state capital, counties that contain the state capital receive about four times that amount. Since many of these counties are also media centers or at least contain a television station, including them would lead us to overestimate the effect of the media on grant spending. The sample used for the full specifications includes 2934 counties because some controls are missing for a number of counties.

5.4.1 The link between media activity and spending

On a descriptive level, the connection between media proximity and federal grant distribution is far from clear. While the theoretical model predicts higher spending in the vicinity of media centers, the raw correlation between 'distance to media center' and grant spending is actually 0.15, i.e. remote counties receive more rather than fewer grants. The same is true for the number of television stations and grants: the correlation is close to zero and not of the expected sign at −0.03.

Column (1) of Table 9 shows that by simply adding the mean per capita income and state fixed effects to a regression, the negative correlation between grants and distance is reversed. The distance coefficient implies that being twice as far from a media center as the average county is associated with receiving 5% fewer grants. This is consistent with the theoretical

[48] The number of major TV stations licensed in a given county shows only very limited, variation over time, the distance a county is located from the nearest media city is fixed. The only possible reasons for this distance to actually change would be for a county to change DMA affiliation. This rarely happens and thus the variation in the data will be insufficient to identify any effects. The stationarity is acknowledged by Gentzkow (2006), who assumes current DMA borders to be a valid approximation to those in the 1960s.

Table 9: Dependent variable: grants per capita, OLS full sample.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Media variables							
distance to media center ratio	−.05***	−.0516***	−.0509***	−.0386***	−.0385***		−.0344***
	(0.017)	(0.014)	(0.012)	(0.010)	(0.009)		(0.009)
tv stations						.0412***	.0276*
						(0.015)	(0.015)
distance to media city outside DMA					−.0082	$-7.5e{-}04$	−.0095
					(0.028)	(0.030)	(0.028)
out of DMA-state county					−.1016*	−.1174**	−.1025*
					(0.057)	(0.058)	(0.057)
DMA share in home state					−.0017**	−.0018**	−.0017**
					(0.001)	(0.001)	(0.001)
DMA share X out of DMA-state					.003**	.0031**	.003**
					(0.001)	(0.001)	(0.001)
Socio-demographic county characteristics							
mean per capita income (1000$)	−1.643***	−1.41***	−.4899***	−.5417***	−.5547***	−.5596***	−.556***
	(0.117)	(0.088)	(0.163)	(0.135)	(0.137)	(0.141)	(0.136)
bachelor or higher pct			.0209***	.0151***	.0146***	.0148***	.0146***
			(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
native american pct			.0064**	.0036*	.0036*	.0031	.0036*
			(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
poverty pct			.035***	.0329***	.0331***	.0326***	.0329***
			(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
high school dropout pct			.019***	.0177***	.0176***	.0172***	.0176***
			(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
under 24 yrs pct			−.0155***	−.0198***	−.0197***	−.0194***	−.0197***
			(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
unemployed pct			$3.7e{-}04$	.0044	.004	.0037	.0039
			(0.008)	(0.006)	(0.006)	(0.006)	(0.006)
bank deposits per cap (1000$)			.0203	.0333	.0367	.0312	.0349
			(0.032)	(0.032)	(0.032)	(0.033)	(0.032)
female pct			.0011	.0095	.0094	.0102	.0092
			(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
expenditure per cap (1000$)			.0039	−.0421	−.0424	−.0455	−.0415
			(0.045)	(0.036)	(0.035)	(0.036)	(0.035)
mean to median income			.1264	.0164	.0197	.0213	.0204
			(0.079)	(0.040)	(0.040)	(0.040)	(0.040)
Political economics variables							
white pct				−.0023	−.0021	−.0025*	−.0022
				(0.001)	(0.001)	(0.001)	(0.001)
ethnic fragmentation				−.374***	−.3838***	−.3772***	−.3782***
				(0.129)	(0.126)	(0.130)	(0.127)
distance to capital ratio				−.0225	−.0404	−.0451*	−.042
				(0.023)	(0.025)	(0.026)	(0.026)
fed gov employed pct				.0172**	.0165**	.0169**	.0167**
				(0.008)	(0.008)	(0.008)	(0.008)
other gov employed pct				.0197***	.0206***	.0208***	.0207***
				(0.006)	(0.006)	(0.006)	(0.006)
turnout ratio 1996				.1454	.1583	.1596	.1587
				(0.136)	(0.136)	(0.140)	(0.137)
voter mobility				.0778***	.0711***	.0716***	.0708***
				(0.024)	(0.023)	(0.023)	(0.023)
population/geography controls							
population growth		−.0089***	−.007***	−.0055***	−.0053***	−.0053***	−.0053***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
county latitude		−.6079	.7621	.8961	.867	.9157	.9059
		(0.892)	(0.700)	(0.704)	(0.669)	(0.679)	(0.674)
county longitude		1.43**	1.26*	.8053	.8077	.7799	.8236
		(0.613)	(0.643)	(0.573)	(0.520)	(0.529)	(0.520)
big city indicator		.0242	−.0407	−.0371	−.0325	−.0177	−.0403
		(0.042)	(0.039)	(0.032)	(0.031)	(0.032)	(0.033)
metro/micro indicator		−.1143***	−.0261	−.0108	−.009	$7.7e{-}05$	−.0123
		(0.026)	(0.028)	(0.026)	(0.025)	(0.026)	(0.026)
population density (1000/sqm)		.0669***	.0588***	.0371**	.0364**	.0392**	.0364**
		(0.017)	(0.014)	(0.015)	(0.015)	(0.015)	(0.015)
land area (1000 sqm)		.0059	.01	.0062	.0087	.0048	.0074
		(0.011)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
population (100,000s)		.0052	−.0085*	$-3.7e{-}04$	$1.2e{-}04$	$4.6e{-}05$	$-8.2e{-}04$
		(0.005)	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)
constant	22.81	23.2	11.76	11.19	11.52	11.49	11.56
	(1.126)	(1.581)	(2.014)	(1.733)	(1.744)	(1.789)	(1.740)
state effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
observations	3041	3041	2980	2934	2934	2934	2934
adjusted R^2	0.351	0.430	0.548	0.587	0.590	0.589	0.591

Note: Standard errors in parentheses allow for clustering at the state level. All state capitals excluded from the estimation.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

model. It predicts that politicians will spend less money in these remote counties due to the lower media coverage, which in turn is due to the high cost of reporting from these places. Column (2) adds population and geography controls in order to correct the distance coefficient for the fact that most counties that are far away from the media center are also rural. While some of these urbanity controls are highly significant, the distance coefficient remains very stable. Even adding a wide range of sociodemographic controls in column (3) does not change the distance coefficient. It seems that, while many of these controls are highly significant, they are mainly related to income rather than the distance variable. An interesting result in this specification is that higher education levels in the population are positively related to received grants. Just as the negative coefficient on distance, this is in line with the idea of better informed voters receive higher levels of grants. Seven political economics variables are introduced in column (4). All of the political variables have the expected sign and they are also the first controls to actually affect the distance coefficient. It is scaled down to three fourths its original size but remains highly significant.

Even though we focus on the distance variable ϕ_c^D, we are also interested in the other media variables, which make up ϕ_c^M. Column (5) adds the out of state indicator, the DMA home share and an interaction of the two. The sign on out-of-state counties is as expected (-10%), consistent with the idea of higher cost of reporting from these counties and a resulting home state bias of TV stations. Keep in mind, though, that this is evaluated at a DMA home share of zero. For out-of-state counties, an increase in the home share is associated with a gain in grants, whereas in-state counties lose money when their DMA home share increases. This means that a county without a media city in its home state can make up for this disadvantage when its state's DMA home share is larger, that is reporting from this state becomes relatively more profitable for TV station operators. For in-state counties the negative coefficient on home share implies that, given a home-state biased media city in your state, it is best shared with as few people as possible. The mechanism is the same as with the distance variable: the increased media attention leads to a higher probability of receiving information on the events concerning your county, which in turn will provide politicians with incentives to direct spending there.

As can be seen in column (6), having TV stations in the county also leads to higher grant awards, yet the effect decreases in the number of stations. As mentioned above, a pseudo-elasticity can be calculated as: TV station coefficient times $[j/(j+1)]$, where j indicates the number of TV stations. Accordingly, the loss of the only TV station in a county incurs a drop in grants per capita of approximately 2% whereas an increase from 10 to 11 stations only generates $[10 \cdot 10/11] \cdot 0.04 = 0.35\%$ more grants per capita. Finally, column (7) includes both the TV stations and the distance variable at the same time. As expected because of the collinearity, both coefficients are somewhat diminished. Most importantly, though, counties far from the nearest media center still receive significantly lower amounts of grants.

Table 10: DEPENDENT VARIABLE: GRANTS PER CAPITA, METRO/MICRO SAMPLE.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
distance to media center ratio	$-.0902^{***}$	$-.058^{***}$	$-.0467^{***}$	$-.0346^{***}$	$-.0355^{***}$		$-.03^{***}$
	(0.015)	(0.014)	(0.012)	(0.011)	(0.010)		(0.010)
tv stations						$.048^{***}$	$.0364^{***}$
						(0.014)	(0.013)
Observations	1691	1691	1662	1652	1652	1652	1652
Adjusted R^2	0.310	0.437	0.607	0.633	0.637	0.636	0.638

Note: Standard errors in parentheses allow for clustering at the state level. All state capitals excluded from the estimation. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In sum, we find the hypothesized effect of our key variables. This leads to the conclusion that vote-maximizing politicians seem to indeed favor counties where media activity is high and counties less exposed to media coverage receive less attention when grants are distributed.

5.4.2 Robustness checks

Having established a positive link between media activity and grant spending, this section is concerned with how robust the results are to changes in specification and sample.

As a first robustness check, we employ a subset containing only the 1652 micropolitan and metropolitan counties in the dataset.[49] We are confident that the population and geography controls already take care of things such as scale effects, minimum grants per county leading to higher per capita grants in less populous counties, overrepresentation in the political process or flat rate grants per county. However, in the event that there are other unobserved differences in urban and rural counties that affect both the media and grant spending, we try to account for this by estimating the same specifications as in Table 9 on the non-rural counties only. These estimates – shown in Table 10 – are very similar to the ones obtained from the full sample. While the coefficient on distance is again highly significant and virtually identical to the full sample, the association between television stations and grant spending seems to be a bit stronger and more precisely estimated. The coefficients on out of state counties, the DMA home share and their interaction are almost identical to the full sample in terms of size and significance and are not reported here. The urban sample estimates therefore underscore the earlier results.

In a second variation, we exclude Medicaid spending from the grants variable. Medicaid accounts for roughly half the federal grant money in our data and is also considered to be rather "fixed" (Levitt and Snyder 1997). Excluding these payments tests whether the

[49]Metro areas are defined by the Bureau of the Census as areas containing a core urban area of 50,000 or more population. Micropolitan areas contain an urban core of at least 10,000 population. Metro or micro areas include one or more counties, specifically the core urban area, as well as many adjacent counties. Around 80% of the US population resides in metropolitan areas.

Table 11: MEDICAID EXCLUDED.

	(1)	(2)	(3)
Full sample			
distance to media city ratio	−.0443***		−.0424***
	(0.011)		(0.011)
full service tv stations		.0295	.0127
		(0.020)	(0.020)
observations	2932	2932	2932
Metro/Micro sample			
distance to media center ratio	−.039***		−.0327**
	(0.013)		(0.013)
tv stations		.0543***	.0417**
		(0.017)	(0.017)
observations	1650	1650	1650

Note: The dependent variable is 'grants per capita, excluding Medicaid'. Only media variable coefficicents shown. The specifications in columns (1)-(3) correspond to columns (5)-(7) of tables 9 (full sample) and 10 (metro/micro sample). All state capital counties (n=48) excluded from the estimation. Standard errors in parentheses allow for clustering by state. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

media variables pick up health status in the population or some other characteristic we have not controlled for. The specifications in columns (1)-(3) of Table 11 correspond to columns (5)-(7) in tables 9 and 10. In four of the six specifications, the estimated media coefficients are now larger than before. While the coefficients on distance are larger in each of the estimations, the coefficient of television stations is larger in the urban sample and smaller in the full sample. The full sample coefficients are also less precisely estimated. The larger effects are in line with what we would expect if Medicaid cannot be influenced by political agents as much as other grant schemes. The fact that the TV coefficient actually decreases in the full sample weakens this argument somewhat, yet the hypothesized effects put forward in the earlier sections of the chapter are still very much present and in most cases even strengthened.

The baseline estimates obtained so far may still be biased due to omitted variables, such as the presence of interest groups and lobbying activity. One could think of the agricultural sector or heavy industry as branches that have traditionally been successful in acquiring grants. Sheer economic size in terms of having many private businesses may also be advantageous to acquiring grants, since some of them go directly to business owners. To account for these factors, we add as controls the number of privately owned nonfarm businesses as well as indicator variables for the economic type of a region. The indicators are based on a county's primary source of employment and classify them as farming-dependent, mining-dependent, manufacturing-dependent, federal/state government-dependent, services-dependent, or non-specialized. The F-tests in columns (1) and (3) of Table 12 show that a county's specialization is significantly related to grants received. The number of businesses does not seem to be associated with grant spending (coefficient not reported). Most importantly, the coefficients of the media variables remain unchanged in magnitude and significance level.

Measurement error and reverse causality may be an issue, too. TV stations in our data are assigned to the county where the station is licensed or where the main transmitter is located. In most cases this will be identical with the county where the actual TV studio is located, but sometimes the two locations do not coincide, causing the TV stations variable to be measured with error. Because measurement error biases the estimates towards zero while we mainly want to avoid overestimating the media effect, we are not too concerned about it. Somewhat more of an issue is possible reverse causality. The location of TV stations may be endogenous to government spending, as they include PBS affiliates, which are funded by the Corporation for Public Broadcasting (CPB), federal and state governments. As a first measure to account for this, CPB grants to local jurisdictions are already excluded from our dependent variable. Around 2000, these accounted for about 80% of federal money that went to public broadcasting.[50]

We also already control the main factors for placing television stations: population and population density. If stations emerge or are licensed, where grant spending is higher due to reasons other than our population and geography controls, we may still overestimate the effect of having a TV station in the county. Estimates from an IV approach to this issue are shown in columns (2) and (4) of Table 12. We instrument the number of TV stations with the number of low power (LP) TV stations that are not classified as Class-A (CA) and have low broadcasting power of less than 10kw. Most of these stations are non-profit with a niche focus, e.g. educational, religious or other stations low in reach and information content that is of interest to a broad audience. Consequently these stations cater to very small audiences. Therefore they should not be targeted by politicians and their location should not be contingent on public spending or any kind of political clout a county may have. After controlling market size, the number of low power and full-service TV stations should be primarily correlated due to technical issues such as whether the geography of a county favors broadcasting. Despite finding a strong correlation in the first stage, the second stage confidence intervals of the TV coefficient include both the OLS coefficient and zero. While not very informative, the IV estimates shown in columns (2) and (4) of Table 12 thus do not preclude the baseline estimates.

The robustness checks support the notion that federal grant spending is related to differences in the amount of media attention a county gets due to its location. The very stable and highly significant coefficient on the distance variable contrasts with the less stable TV stations coefficient. However, F-tests show that the distance and TV stations variable are jointly significant at the .1% level in all specifications. Taken together, these results corroborate the results from the baseline specification.

[50] See `http://www.cpb.org` or `http://www.newenglandfilm.com/news/archives/00december/pbs.htm`.

Table 12: Additional controls and IV estimates.

	full sample		metro/micro	
	(1)	(2)	(3)	(4)
distance to media city ratio	−.032***	−.029*	−.031***	−.032**
	(0.009)	(0.016)	(0.009)	(0.014)
full service tv stations	0.028*	0.064	.038**	.021
	(0.015)	(0.100)	(0.014)	(0.092)
F-test county classifications, F(5,47)	6.58		5.39	
F-test instrument, F(1,47)		27.32		20.27
observations	2934	2933	1652	1651

Note: The dependent variable is 'grants per capita'. Only media variable coefficients shown. Columns (1) and (3) add further controls to the baseline specification in column (7) of Tables 9 (full sample) and 10. Columns (2) and (4) show IV estimates of column (7) from Tables 9 (full sample) and 10 (metro/micro sample). The instrument is the number of low power TV stations. All state capital counties (n=48) excluded from the estimation. Standard errors in parentheses allow for clustering by state. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.4.3 Spatial dependencies in grant spending

Another concern is that strategic interaction could be responsible for possible similarities between grants awarded to neighboring counties. Not correcting for such interdependencies may bias the media coefficients. One mechanism that can induce a positive correlation between neighboring counties' spending is yardstick competition. Because it requires two governments that strategically interact, yardstick competition may not be so relevant in the case of a governor distributing grants across his state.[51] However, other political agents such as congressmen make strategic interaction seem more realistic. In addition, strategic interaction can always occur in counties bordering another state.[52] In addition, locally correlated shocks may lead to spatially correlated error terms, and thus to inefficient estimates. For these reasons, we estimate the three step spatial-IV estimator (GS2SLS) proposed by Kelejian and Prucha (1998).[53] It takes into account the endogeneity of the spatially lagged dependent variable and allows for simultaneous presence of spatial lag and error:

$$g \;=\; \rho\, W\; g + X\beta + \lambda\; W\; \varepsilon + u. \tag{5.9}$$

Again, g is a vector representing grant spending, W is a spatial weight matrix and Wg gives the measure of grant spending in neighboring counties. The interaction between own and neighbors' spending is captured in ρ, which we would expect to have a positive sign in the presence of yardstick competition. Another reason for including the spatial lag could be spillovers in the provision of public goods, which we may not be able to capture in

[51] If people are envious of the amount of grants their neighboring counties receive, an interdependence would be introduced, such that the governor may have to abide to equity concerns and follow up a grant award to county A with an award to county B, creating positive correlation.

[52] To some extent the variable *DMA home share* may be viewed as a measure of cross-state yardstick competition in our estimations, because a higher share of the DMA population living outside the home state implies getting more information about what is going on in the neighboring state. This may create yardstick competition among governors.

[53] As suggested by Kelejian and Prucha (1998), we use the full set of spatially-first-lagged exogenous variables as instruments to account for the endogeneity of the spatial lag.

the baseline specification. In this case the spatial correlation expressed in ρ may point in either direction. Spatial correlation in the errors is measured by λ, ε is a vector of spatially autocorrelated error terms, and u is a vector of i.i.d. error terms.[54]

Table 13: SPATIAL DEPENDENCY – GS2SLS ESTIMATIONS.

	(1)		(2)		(3)	
distance to media center	$-.039^{***}$	(0.009)			-0.035^{***}	(0.009)
tv stations			0.038^{**}	(0.015)	0.025	(0.015)
spatially lagged grants	0.165^{***}	(0.034)	0.165^{***}	(0.034)	0.165^{***}	(0.034)
N	2934		2934		2934	

Note: Only coefficients on main media variables and spatial lag of grants reported. Full specifications are as in column (7) of Table 9. All state capital counties (n=48) excluded from the estimation. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Estimations are carried out on the full sample only. Table 13 displays the coefficients on the media variables and the spatial lag of federal grant spending. We find highly significant positive spatial dependency, yet both our media variables remain virtually unaltered in comparison to the OLS results. We have already provided some intuition on what factors may be underlying the spatial effects, yet this is not the focus of this research. We are content with being able to state that the significance in both media activity variables cannot be an artefact of spatial dependencies.

5.5 Conclusion

We analyzed the question of whether politicians allocate more funds to regions where they receive more coverage from local television. We developed a simple model which explains the incentives of TV stations to bias coverage towards the core area of a market, where the majority of these stations is located. This geographic bias was shown to create a specific distribution of political knowledge in the electorate: in the core areas, there are more voters who attribute funds to the respective politician than in remote areas. The resulting main prediction of the model is that politicians allocate more funds to the core regions because of higher payoffs in terms of votes.

Next, we carried out an empirical analysis to test this main theoretical prediction. Using data on the allocation of federal grants to U.S counties, we find evidence that public spending decreases in the geographical distance to the core areas of TV markets. This supports our theoretical contention that the well-known geographical bias in local television news programming affects public policy. In addition, the relation between TV markets and state

[54] The weighting matrix W is row standardized based on rook contiguity, i.e. counties sharing a common border are treated as neighbors.

borders seems to affect spending levels, too. The larger an in-state county's home state share of total market population, the less grants the county receives. Interestingly, if a county is out-of-state this effect is reversed, that is if all media cities in the market are located in a neighbor state, a larger home state share is associated with higher funding levels.

Taken together, the empirical results provide evidence to support the theoretical argument that the geography of local television markets influences public policy. In particular, the main hypothesis on the relation between the distance to media cities and grant spending is very reliably confirmed. These findings contribute to the field of media and political economics the important aspect that the geography of media markets plays an important role for the allocation of public funds. A further implication is that future research should consider the geography of media markets. Relevant fields include yardstick competition and the (de)centralization of policies.

5.6 Appendix

Table 14: Counties by number of TV stations (incl. Class A).

	Full sample		Metro micro sample	
number of TV stations	Frequency	pct	Frequency	pct
0	2,149	73.24	1,019	61.68
1	433	14.76	306	18.52
2	129	4.40	113	6.84
3	71	2.42	64	3.87
4	50	1.70	48	2.91
5	35	1.19	35	2.12
6	15	0.51	15	0.91
7	13	0.44	13	0.79
8	5	0.17	5	0.30
9	9	0.31	9	0.54
10	8	0.27	8	0.48
11	4	0.14	4	0.24
12	2	0.07	2	0.12
13	4	0.14	4	0.24
14	2	0.07	2	0.12
15	1	0.03	1	0.06
16	2	0.07	2	0.12
18	1	0.03	1	0.06
24	1	0.03	1	0.06
Total	2,934	100.00	1652	100.00

Table 15: Summary statistics (TV).

	N	Mean	Std. Dev.	Min	Max
grants per capita ($)	2934	1053.03	729.10	96.28	10937.40
grants per cap w/o Medicaid($)	2933	529.65	582.08	0	9986.81
distance to media center ratio	2934	1.01	0.57	0	3.36
tv stations	2934	0.65	1.71	0	24
distance to media city outside DMA	2934	117.78	57.18	4.28	400.39
out of state county	2934	0.15	0.35	0	1
DMA share in home state	2934	78.26	30.39	0.08	100
Per capita income 1999 ($)	2934	17403	3822.02	7069	44962
bachelor or higher pct	2934	13.19	6.28	3.70	53.40
native american pct	2934	1.49	5.43	0	86
poverty pct	2934	14.11	6.40	2.11	50.88
high school dropout pct	2934	30.770	10.23	4.50	68.40
under 24 yrs pct	2934	34.34	4.29	20.20	66
unemployed pct	2934	4.76	2.53	0.70	27.60
bank deposits per cap (1000$)	2934	2.33	0.46	−0.17	4.69
female pct	2934	50.47	1.86	32.75	57.44
expenditure per cap (1000$)	2934	2.24	0.86	0.02	10.48
mean to median income	2934	1.68	0.27	0.80	10.13
white pct	2934	85.16	15.59	12.60	99.5
ethnic fragmentation	2934	0.76	0.18	0.27	0.99
distance to capital ratio	2934	101.18	53.76	3.75	362.08
fed gov employed pct	2934	0.69	1.22	0.04	37.25
other gov employed pct	2934	6.57	2.83	2.25	52.53
turnout ratio 1996	2934	99.65	13.52	16.45	185.33
voter mobility	2934	0.72	0.83	0	3
nonfarm businesses	2934	2048	7004	13	219933
farming-dependent cty	2933	0.14	0.34	0	1
mining-dependent cty	2933	0.04	0.19	0	1
manufacturing-dep. cty	2933	0.30	0.46	0	1
fed/state gov.-dep. cty	2933	0.11	0.31	0	1
services-dep. cty	2933	0.10	0.30	0	1
nonspecialized cty	2933	0.31	0.46	0	1
population growth	2934	11.20	15.84	−26.31	191.04
county latitude (rad)	2934	0.67	0.08	0.85	0.43
county longitude (rad)	2934	−1.60	0.20	−1.18	−2.17
big city indicator	2934	0.16	0.37	0	1
metro/micro indicator	2934	0.56	0.50	0	1
population density (1000/sqm)	2934	0.17	0.64	0	16.40
land area (1000 sqm)	2934	0.96	1.29	0.02	20.05
population (100,000s)	2934	0.84	2.82	0	95.19

Table 16: DESCRIPTION OF VARIABLES (TV).

Variable	Description
Dependent variables[(a)]	
grants per capita ($)	Federal grants per capita, taken from the Consolidated Federal Funds Report (CFFR).
grants per cap w/o Medicaid ($)	Federal grants per capita (CFFR), without Medical Assistance Program (Medicaid).
Media related variables[(b)]	
tv stations	From the Federal Bureau of Communications *Wireless Telecommunications Bureau Database* 2006. Earlier versions of the database could not be obtained, yet the number of TV stations does not change much over time.
distance to media center ratio	distance to nearest media city in county c's DMA, divided by the average distance to the nearest media city across all counties in the same DMA as county c.
distance to media city outside DMA	distance to nearest media city outside county c's DMA.
DMA share in home state	the share of the total population in county c's DMA that lives in county c's state.
out of DMA-state county	a county c that belongs to a DMA with no media city in county c's state.
Sociodemographic controls[(c)]	
per capita income 1999 (1000$)	per capita income 1999 in 1000$.
bachelor or higher pct	share of the population that holds at least a bachelor's degree.
native american pct	share of the population that is native American.
poverty pct	share of the population in poverty.
high school dropout pct	share of the population that does not possess a high school diploma.
under 24 yrs pct	share of the population that is under 24 years of age.
unemployed pct	share of the population that is unemployed.
bank deposits per cap (1000$)	average amount of bank deposits per capita in county i.
female pct	share of the population that is female.
expenditure per cap (1000$)	total local expenditures per capita in 1000$.
mean to median income	mean per capita income divided by median per capita income.
Politics controls[(d)]	
white pct	share of the population that is white.
ethnic fragmentation	herfindahl index of ethnic fragmentation, 1 if homogenous.
distance to capital ratio	distance to county c's state capital, divided by avg. distance to capital across all counties in the same state as county i.
fed gov employed pct	share of population employed by federal government (non-military).
other gov employed pct	share of population employed by state and local governments.
turnout ratio 1996	number of democratic and republican votes cast in the 1996 presidential election divided by population of that year, relative to the average turnout in the state county c is located in.
voter mobility	number of times the majority in presidential elections in county c has shifted from 1980 to 1996.
county classification	main source of employment. Provided by U.S. Department of Agriculture Economic Research Service.
private businesses	number od privately owned nonfarm businesses.
Geography and urbanity controls[(e)]	
population growth	county population change 1990-2000 (%).
latitude	county centroid latitude (rad).
longitude	county centroid longitude (rad).
big city indicator	takes on value 1 if county population $> 100,000$.
metro/micro indicator	takes on value 1 if county is part of a statistical metropolitan or micropolitan area.
population density (1000/sqm)	population per square mile.
land area (1000 sqm)	in square miles.
population (100,000s)	county population.

(a) Washington, D.C. is excluded as are counties that cannot be unambiguously assigned to a single media market. (b)-(e) taken from *County Databook 2000* published by the US Census Bureau, *County Profiles* published by the US Department of Agriculture (available online at `http://maps.ers.usda.gov/profiles/webcensusdownload.aspx`) and *County Quickfacts* provided by the US Census Bureau.

6 Concluding remarks

Mark Fowler, a former chairman of the Federal Communications Commission (FCC), once declared that television would simply be 'a toaster with pictures'. His statement reflects an opinion on media markets that rejects the importance of media content in democratic societies.[55] We don't agree.

In contrast, our analysis is guided by the idea that media markets deserve special attention from researchers because market forces determine how well-informed voters are about politics and this, finally, shapes the incentives of politicians. This is the tenor of the recent political economy literature on mass media. This book contributes new theoretical insights and also empirical evidence which support the political economy perspective, and the main message of the book thus reads: The media matter! As the results and their implications have already been discussed separately in each chapter, we restrict us to summarizing the main insights and offer a synthesis.

Information must be produced, i.e. even when there is an abundance of data available it takes journalists who gather data, derive meaning from it and sell their knowledge to the citizens. We have learned in chapter 2 that the production of information can be afflicted with free-riding behavior on the demand side which results in an inefficiently low quality of information and high risks of electoral decisions. As this finding can be regarded as a typical example of market failure, one could argue for some kind of public provision of information. But we do not jump to this conclusion, because public control of media may also imply that politicians try to manipulate the coverage of political issues, not to speak of propaganda. Thus, in crude analogy to central banks, the independence of a public provider of information would be a prerequisite to ensure objective coverage. These considerations may appear to be interesting regarding the current debate about the role of public broadcasting organizations in Germany. As the production of information will remain a fundamental aspect of information provision in the internet era, proponents of public media could refer to the free-riding problem in order to justify their position. Anyway, the economics of public media are an unexplored field and a promising direction for further research.

We have also shown in a formal analysis in chapter 3 that more informative coverage of the performance of politicians strengthens the incentives of the latter to behave well and how the quality of news reports depends on fundamental features of media markets such as the technology of producing information, the prior knowledge of citizens or the market form. The model is a first step that allows to study the interaction of media, voters and politicians

[55] See Hamilton (2006, p.1) for a brief characterization of this perspective on media markets.

in a coherent theoretical framework and it should be extended in various directions in order to check the robustness of our results. It would be interesting, for example, to take a closer look at the role of advertising or to introduce heterogeneity among voters. Having gained more knowledge, the predictions should be tested empirically. Many local newspaper markets, for example, have become monopolies and we could use data on these markets and local policies to test the prediction that the transition to a monopoly does not affect the incentives of politicians.

The empirical results presented in chapter 4 and chapter 5 both can be interpreted as evidence for the theoretical prediction that a more efficient technology of information production strengthens the incentives of politicians to provide voters with favorable policies. The result that the geography of local television markets in the United States affects the allocation of federal funds fits well to this argument: Large distances from media cities to remote places usually imply both higher costs to produce reports and less informative investigations given a certain effort because journalists are not familiar with local institutions. Thus, voters in more distant regions are less informed which weakens the incentives of politicians. This geographical effect could also explain our finding that newspaper circulation influences the efficiency of local governments in Norway, because distance also matters for Norwegian print journalists.

Moreover, the empirical results concerning local television markets and newspaper circulation definitely suggest that voters who are well informed have a larger political clout and receive favorable policies. This result is unproblematic as long as voters are homogenous, but when it comes to distributional issues, it implies that informed voters benefit at the expense of uninformed voters. Thus, with respect to our analysis one could argue for a regulation of media coverage that aims at leveling informational disparities between regions. Strömberg (2004a) and Hamilton (2006), however, show that there are many different groups of people (poor and rich, old and young, male and female, etc.) which are favored or ignored by mass media and thus a consistent regulation would be difficult to implement. Thus, we are cautious about policy implications of the empirical results and we prefer to say that the results provide new and interesting insights which together with other studies help us to understand the political economics of mass media.

In summary, the findings in this book support the political economy perspective on the role of mass media and we think that there are many more interesting questions that need to be answered.

7 References

Adams, W.C. (1980), Local Television News Coverage and the Central City, *Journal of Broadcasting* 24(2), 253-265.

Albouy, D. (2009a), The Unequal Geographic Burden of Federal Taxation, *Journal of Political Economy* 117(4), 635-667.

Albouy, D. (2009b), Partisan Representation in Congress and the Geographic Distribution of Federal Funds, *NBER Working Papers*, No. 15224.

Aldashev, G. (2010), Political Information Acquisition for Social Exchange, *Quarterly Journal of Political Science* 5(1), 1-25.

Alesina, A., Baqir, R. and Easterly, W. (1999), Public Goods And Ethnic Divisions, *Quarterly Journal of Economics* 114(4), 1243-1284.

Alesina, A. and Tabellini, G. (2007), Bureaucrats or Politicians? Part I: A Single Policy Task, *American Economic Review* 97(1), 169-179.

Alesina, A. and Tabellini, G. (2008), Bureaucrats or Politicians? Part II: Multiple Policy Tasks, *Journal of Public Economics* 92, 426-447.

Althaus, S.L. and Trautman, T.C. (2008), The Impact of Television Market Size on Voter Turnout in American Elections, *American Politics Research*, 36(6), 824-856.

Ashworth, S. (2005), Reputational Dynamics and Political Careers, *Journal of Law, Economics, and Organization* 21(2), 441-466.

Ashworth, S., Bueno de Mesquita, E. and Friedenberg, A. (2010), Creating Incentives and Selecting Good Types Revisited.
`http://home.uchicago.edu/~bdm/PDF/incentives-selection.pdf`.
Accessed 6-27-2011.

Baron, D. (1994), Electoral competition with informed and uninformed voters, *American Political Science Review* 88(1), 33-47.

Baron, D. (2006), Persistent media bias, *Journal of Public Economics* 90, 1-36.

Barzun, J. (2001), *From Dawn to Decadence: 500 Years of Western Cultural Life 1500 to the Present*, 1st Edition, Harper Perennial.

Besley, T. (2006), *Principled Agents? The Political Economy of Good Government*, Oxford: Oxford University Press.

Besley, T. and Burgess, R. (2002), The Political Economy of Government Responsiveness: Theory and Evidence from India, *Quarterly Journal of Economics* 117(4), 1415-1452.

Besley, T. and Ghatak, M. (2007), Retailing public goods: The economics of corporate social responsibility, *Journal of Public Economics* 91(9), 1645-1663.

Besley, T. and Prat, A. (2006), Handcuffs for the Grabbing Hand? Media Capture and Government Accountability, *American Economic Review 96(3)*, 720-736.

Borck, R. and Owings, S. (2003), The political economy of intergovernmental grants, *Regional Science and Urban Economics* 33(2), 139-156.

Borge, L.-E., Falch, T. and Tovmo, P. (2001), *Produksjonsindeks for kommunale tjenester* (An Aggregate Output Measure for Local Government Services), Report, ALLFORSK, Norwegian University of Science and Technology.

Borge, L.-E., Falch, T. and Tovmo, P. (2008), Public Sector Efficiency: The Roles of Political and Budgetary Institutions, Fiscal Capacity, and Democratic Participation, *Public Choice 136(3)*, 475-495.

Borjas, G. (1980), The Relationship Between Wages and Weekly Hours of Work: The Role of Division Bias, *Journal of Human Resources 15(3)*, 409-423.

Bradford, D.F., Malt, R.A., and Oates, W.E. (1969), The Rising Cost of Local Public Services: Some Evidence and Reflections, *National Tax Journal 22(2)*, 185-202.

Corneo, G. (2006), Media capture in a democracy: the role of wealth concentration, *Journal of Public Economics* 90, 37-58.

Cornes, R. and Sandler T. (1996), *The Theory of Externalities, Public Goods, and Club Goods*, 2nd Edition, Cambridge University Press, Cambridge.

De Borger, B. and Kerstens, K. (2000), What is Known About Municipal Efficiency? The Belgian Case and Beyond. In J.L.T. Blank (Ed.), *Public Provision and Performance*, 299-330, North-Holland, Amsterdam.

DellaVigna, S. and Kaplan, E. (2007), The Fox News effect: media bias and voting, *Quarterly Journal of Economics* 122(3), 1187-1234.

Dixit, A. and Londregan, J. (1996), The Determinants of Success of Special Interests in Redistributive Politics, *Journal of Politics* 58, 1132-1155.

Djankov, S., McLeish, C., Nenova, T. and Shleifer, A. (2003), Who owns the media?, *Journal of Law & Economics* 46(2), 341-81.

Downs, A. (1957), *An Economic Theory of Democracy*, HarperCollins Publishers, New York.

Feddersen, T. and Pesendorfer, W. (1996), The Swing Voter's Curse, *American Economic Review* 86, 408-424.

Feddersen, T. and Sandroni, A. (2006), Ethical Voters and Costly Information Acquisition, *Quarterly Journal of Political Science* 1(3), 287-311.

Gabszewicz, J., Laussel, D. and Sonnac,N. (2001), Press advertising and the ascent of the 'Pensee Unique', *European Economic Review* 45(4-6), 641-651.

Gehlbach, S. (2007), Electoral Institutions and the National Provision of Local Public Goods, *Quarterly Journal of Political Science* 2(1), 5-25.

Gentzkow, M. (2006), Television and Voter Turnout, *Quarterly Journal of Economics* 121(3), 931-972.

Gentzkow, M. and Shapiro. J.M. (2006), Media Bias and Reputation, *Journal of Political Economy* 114(2), 280-316.

Gentzkow, M., Shapiro, J. and Sinkinson, M. (forthcoming), The Effect of Newspaper Entry and Exit on Electoral Politics, *American Economic Review*.

Gentzkow, M. and Shapiro, J. (forthcoming), Ideological Segregation Online and Offline, *Quarterly Journal of Economics*.

Graber, D.A. (2001), *Processing Politics: Learning from Television in the Internet Age*, University of Chicago Press.

Greenberg, E. (1969), Television Station Profitability and FCC Regulatory Policy," *Journal of Industrial Economics* 17(3), 210-238.

Groseclose, T. and Milyo, J. (2005), A measure of media bias, *Quarterly Journal of Economics* 122(3), 1191-1237.

Grossman, G.M. and Helpman, E. (1996), Electoral competition and special interest politics, *Review of Economic Studies* 63(2), 265-286.

Hamilton, J.T. (2006), *All the news that's fit to sell*, Princeton: Princeton University Press.

Heider, D.B. (2000), *White News: Why Local News Programs Don't Cover People of Color*, Mahwah: Lawrence Earlbaum Associates.

Holmström, B. (1999), Managerial Incentive Problems - A Dynamic Perspective, *Review of Economic Studies*, 66(1), 169-182.

Horrigan, J., Garrett, K. and Resnick, P. (2004), *The Internet and Democratic Debate*, Pew Internet & American Life Project.

Høst, S. (1999), Newspaper Growth in the Television Era - The Norwegian Experience, *Nordicom Review 20(1)*.

Høst, S. (2006), Avisåret 2005 (The Newspaper Year 2005), *Institutt for Journalistikk Rapport* 1.

Kalseth, J. and Rattsø, J. (1995), Spending and Overspending in Local Government Administration: A Minimum Requirement Approach Applied to Norway, *European Journal of Political Economy 11(2)*, 239-251.

Kalseth, J. and Rattsø, J. (1998), Political Control of Administrative Spending: The Case of Local Governments in Norway, *Economics and Politics 10(1)*, 63-83.

Kaniss, P.C. (1997), *Making Local News*, Chicago: University of Chicago Press.

Kelejian, H. H. and Prucha, I. (1998), A Generalized Spatial Two Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Errors, *Journal of Real Estate Finance and Economics*, 17 (1), 99-121.

Knight, B. (2004), Legislative Representation, Bargaining Power, and the Distribution of Federal Funds: Evidence from the U.S. Senate, *NBER Working Papers*, No. 10385.

Kotsogiannis, C. and Schwager, R. (2008), Accountability and fiscal equalization, *Journal of Public Economics* 92(12), 2336-2349.

Lee, F.E. (2003), Geographic Politics in the U.S. House of Representatives: Coalition Building and Distribution of Benefits, *American Journal of Political Science* 47(4), 714-728.

Levitt, S. and Snyder, J. (1997), The Impact of Federal Spending on House Election Outcomes, *Journal of Political Economy* 105(1), 30-53.

Lindbeck, A. and Weibull, J. (1987), Balanced-Budget Redistribution as the Outcome of Political Competition, *Public Choice* 52, 273-297.

Martinelli, C. (2006), Would rational voters acquire costly information?, *Journal of Economic Theory* 129(1), 225-251.

Ministry of Local Government (2006), Beregningsteknisk dokumentasjon til St.prp. nr. 1 (Technical Appendix to the Block Grant System) (Report H-2193).

Mondak, J. (1995), Media Exposure and Political Discussion in U.S. Elections, *Journal of Politics 57(1)*, 62-85.

Mueller, D.C. (2003), *Public Choice III*, Cambridge: Cambridge University Press.

Mullainathan, S. and Shleifer, A. (2005). The Market for News, *American Economic Review* 95(4), 1031-1053.

Napoli, P.M. and Yan, M.Z. (2007), Media Ownership Regulations and Local News Programming on Broadcast Television: An Empirical Analysis, *Journal of Broadcasting & Electronic Media* 51(1), 39-57.

Oberholzer-Gee, F. and Waldfogel, J. (2009), Media Markets and Localism: Does Local News en Espanol Boost Hispanic Voter Turnout?, *American Economic Review* 99(5), 2120-28.

Olson, M. (1965), *The Logic of Collective Action*, Harvard University Press, Cambridge, MA.

Pecorino, P. (1999), The effect of group size on public good provision in a repeated game setting, *Journal of Public Economics* 72(1), 121-134.

Persson, T. and Tabellini, G. (2002), *Political Economics: Explaining Economic Policy*, 1st Edition, The MIT Press.

Petrova, M. (2008), Inequality and media capture, *Journal of Public Economics* 92, 183-212.

Popkin, S. L. (1991), *The Reasoning Voter, Communication and Persuasion in Presidential Campaigns*, Chicago: University of Chicago Press.

Prat, A. and Strömberg, D. (2011), The political economy of mass media, *CEPR Discussion Paper No. DP8246.*

Prior, M. (2006), The incumbent in the living room: the rise of television and the incumbency advantage in U.S. House elections, *Journal of Politics* 68(3), 657-673.

Raiffa, H. and Schlaifer, R. (1961), *Applied Statistical Decision Theory*, 1st Edition, Harvard University Press.

Reinikka, R. and Svensson, J. (2005), Fighting Corruption to Improve Schooling: Evidence from a Newspaper Campaign in Uganda, *Journal of the European Economic Association 3(2-3)*, 259-267.

Revelli, F. (2008), Performance Competition in Local Media Markets, *Journal of Public Economics 92(7)*, 1585-1594.

Revelli, F. and Tovmo, P. (2007), Revealed Yardstick Competition: Local Government Efficiency Patterns in Norway, *Journal of Urban Economics 62(1)*, 121-134.

Shi, M. and Svensson, J. (2006), Political Budget Cycles: Do They Differ Across Countries and Why?, *Journal of Public Economics 90(8-9)*, 1367-1389.

Snyder, J. and Strömberg, D. (2010), Press Coverage and Political Accountability, *Journal of Political Economy* 118(2), 355-408.

Strömberg, D. (2001), Mass Media and Public Policy, *European Economic Review* 45, 652-663.

Strömberg, D. (2004a), Mass Media Competition, Political Competition, and Public Policy, *Review of Economic Studies 71(1)*, 265-284.

Strömberg, D. (2004b), Radio's Impact on Public Spending, *Quarterly Journal of Economics 119(1)*, 189-221.

Svaleryd, H. and Vlachos, J. (2009), Political Rents in a Non-Corrupt Democracy, *Journal of Public Economics 93(3-4)*, 355-372.

Sørensen, R. J. (1984), The Urban Causes of Financial Stress: An Analysis of Norwegian Municipalities and the Economic Resource-Squeeze, *Housing, Theory and Society 1(3)*, 165 - 178.