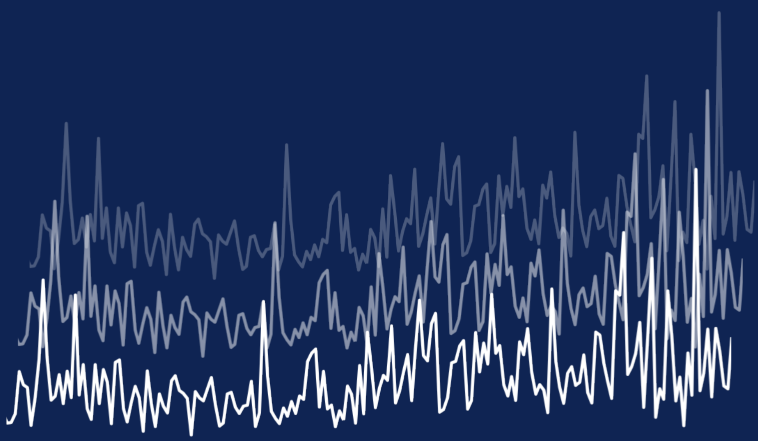
SoDa Laboratories
Monash Business School
Monash University Melbourne, Australia



Socio-economic determinants of Public-Interest Journalism in Australia

Prepared for the Public Interest Journalism Initiative

Simon D. Angus, Klaus Ackermann, David H. Kreitmeir, Paul A. Raschky

December 15, 2023

About this Report

SoDa Laboratories at Monash Business School

Monash Soda Laboratories ("SoDa Labs"), founded in October 2018 by members of the Departments of Economics and Econometrics & Business Statistics, is an empirical research laboratory in the Impact Labs of Monash Business School whose members combine new tools from data science, machine learning and beyond with powerful techniques in the quantitative social sciences to tackle major research challenges in the social sciences. The vision of the lab is, "to be the top social science from alternative data group in the country, and in the top 10 world-wide."

SoDa is the only lab of its kind in Australia, working at the intersection of quantitative and causal social sciences on the one hand, and machine learning and artificial intelligence on the other. Whilst primarily an academic research group, SoDa is about real-world impact, and has partnered with multiple organisations in research partnerships to achieve this. Perhaps most well known is SoDa's global internet measurement platform, the Monash IP Observatory, which conducts over 3 billion active measurements of the internet daily and is the key internet data provider to the United Nations Office of the High Commissioner for Human Rights. Less well known is the lab's partnerships with a range of government, philanthropic and impact organisations in Australia such as the Defence Science and Technology Group, the Paul Ramsay Foundation, the Scanlon Research Institute, and the Judith Neilson Institute.

SoDa has also partnered with high profile international news media organisations to support their public interest journalism, including *The Washington Post*, *The Economist*, *Reuters*, the *ABC*, and *WIRED*, to name a few.

About the Authors

Simon D. Angus Associate Professor; Department of Economics and Monash SoDa Laboratories, Monash University, Melbourne, Australia.

Klaus Ackermann Senuior Lecturer; Dept. of Econometrics & Business Statistics and Monash SoDa Laboratories, Monash University, Melbourne, Australia.

David H. Kreitmeir Research Fellow; Monash SoDa Laboratories, Monash University, Melbourne, Australia.

Paul A. Raschky Professor; Department of Economics and Monash SoDa Laboratories, Monash University, Melbourne, Australia.

Contents

Executive Summary

In the last two decades, local journalism in Australia, particularly in regional and rural areas, has experienced a significant decline. The availability of local news is influenced by various characteristics of Local Government Areas (LGAs). This report embarks on an initial empirical investigation into the socio-economic factors associated with the presence of print and digital local news publishers across Australia. Preliminary findings suggest a developing narrative. Firstly, the size of the local market, in terms of both readership and consumer base, is crucial. Local news publishers become unviable when readership numbers are too low or the market size is insufficiently attractive for advertising. Secondly, there appears to be a correlation between the presence of large companies in primary (mining) or secondary (manufacturing) sectors and a robust local news publisher market.

Among the 540 LGAs studied, 29 (5.4%) lack any local print or digital news publishers. This absence is pronounced in remote and sparsely populated areas, with 8 LGAs in the Northern Territory and 12 in Queensland devoid of such services. An LGA's total population is a key demographic factor in determining the presence of local news publishers. Larger populations generally support the existence and sustainability of these outlets. Moreover, the structure of local industries influences this, with the number of businesses in a sector showing a stronger correlation with news availability than the number of sector employees. This indicates that a diverse business environment may better support local journalism.

In smaller LGAs, with populations below 10,000, the total population is a consistent predictor for the presence of local news publishers. Here, a higher proportion of First Nations people and more businesses in mining, health care, and social assistance positively correlate with news availability. In larger LGAs, with populations over 10,000, a similar trend is observed, with the presence of more manufacturing, professional, scientific, and technical businesses correlating with a competitive local news market. These patterns imply that local economies with larger companies might foster a more diverse range of local news publishers.

However, it is important to acknowledge the limitations of this data and the lack of exogenous variation in the key variables, which precludes making causal claims about the impact of individual socio-economic factors on the availability of local news publishers or the competitiveness of local news markets. Therefore, the findings in this report should be interpreted as correlational at this stage. The report concludes by highlighting the current data constraints that limit the empirical analysis on this topic and offers recommendations for addressing these constraints to enable future, potentially causal studies on the determinants of local news outlet sustainability and market structure.

Key Findings

- 29 LGAs (of 540, 5.4%) in Australia do not have any local print or digital news outlet. These are often remote, sparsely populated LGAs in the Northern territory (8 LGAs) or Queensland (12 LGAs).
- Among demographic factors, *LGA total population* in an LGA is the variable most consistently correlated with the availability of a local news outlet and a healthy local news market (= 2 or more local news publishers). Smaller LGAs simply have a smaller buyer base and are also less attractive markets from an advertising point of view.
- For local industry factors, the number of businesses in a given sector rather than the number of employees in a sector seem to be more systematically correlated with the availability of a local news outlet and a healthy local news market.
- For the group of 242 small LGAs (pop. $<$ 10,000):
  - *LGA population* is the variable that is most consistently correlated the availability of a local news publishers.
  - LGAs with a higher fraction of first nations people are less likely to have a local news outlet.
  - LGAs with more business operating in the mining sector as well as more health care & social assistance businesses, are more likely to have at least one local news publishers.
  - A higher fraction of education & training, financial services, information and telecommunication services and wholesale trade businesses is negatively related with he availability of a local news publishers.
- For the group of 298 large LGAs (pop. $>$ 10,000):
  - *LGA population* is the variable that is most consistently correlated with a more competitive local news market.
  - LGAs with more business operating in the manufacturing sector as well as professional, scientific & technical services businesses, have a more competitive local news market.
  - A higher fraction of administrative & support services and wholesale trade businesses is correlated with a less competitive local news market.

Part 1

Introduction

The objective of this project is to conduct a preliminary empirical analysis of the socio-economic determinants affecting the availability of print and digital local news publishers in Australia. In collaboration with the Public Interest Journalism Initiative (PIJI), our team has compiled a comprehensive cross-sectional dataset encompassing 540 Local Government Area (LGA) level regions across all Australian States and Territories. This dataset includes supply-side information on the number of local news publishers at the LGA level, alongside a broad spectrum of socio-economic variables that could influence the demand for these outlets.

The raw data, observational in nature, is derived from PIJI's collection efforts in 2022/23 (Dickson, 2023) and publicly available data from the Australian Bureau of Statistics (ABS) for 2021. Due to inherent limitations such as the cross-sectional data format (i.e., no within-variation in local news coverage and socio-economic characteristics) and the absence of exogenous variation in key variables, this report cannot infer any *causal* relationships between individual socio-economic factors and the availability or competitiveness of local news publishers. Instead, our findings, based on descriptive statistics, visualizations, and simple regression analyses, draw attention to geographic patterns in local news coverage and identify factors consistently correlated with the structure of local news markets in Australia.

Over the past two decades, local journalism, especially in Australia's regional and rural areas, has witnessed a marked downturn. This decline is primarily attributed to the faltering of traditional news publisher business models, heavily reliant on commercial and classified advertising revenues. Since the early 2000s, the rise of online advertising platforms has redirected advertiser spending away from local print news publishers. From 2001 to 2016, inflation-adjusted classified advertising revenue plummeted from $2b to $200m ACCC (2018). This phenomenon is not unique to Australia and is a global issue affecting local news markets Bhuller et al. (2023); Djourelova and Martin (2023).[1]

Further exacerbating the situation, local news publishers face challenges due to demographic shifts like population aging, evolving industry trends, and the necessity to cover extensive areas with limited resources. Sometimes, the void is partially filled by community-run social media groups commenting on council activities or local councils providing news via their websites and social media channels Simons and Dickson (2019). However, these alternatives cannot fully substitute the objective, consistent reporting offered by professional journalists in established news publishers.

The shutdown of a local news outlet or reduced competition in the market can negatively impact politics and society. Magasic (2023)'s case study of Lightning Ridge, a small mining town in central New South Wales, following the closure of *The Ridge News*, underscores this. The community became less informed about local political matters, losing a vital platform for opinion expression and political advocacy. This led to a weakened council-citizen relationship

[1] Djourelova and Martin (2023) explored the staggered introduction of *Craigslist* across US counties between 1995 and 2009 to isolate the impact of competition for classified advertising from other Internet-related changes. Post *Craigslist*'s entry, local newspapers saw a notable decrease in newsroom and management staff, particularly affecting political editors. These changes led to diminished political news coverage, a decline in newspaper readership, and were not offset by increased online news consumption or other media.

and a sense of political disengagement. Socially, the absence of the newspaper eroded the community fabric, particularly affecting older residents less inclined to use digital alternatives. In terms of civic culture, the study by Magasic revealed a decline in civic discourse and participation. Local news publishers like *The Ridge News* are pivotal in fostering civic engagement by keeping residents informed about community events and volunteer opportunities. Their closure posed significant challenges for civic institutions in promoting and engaging with the community.

These case study observations are corroborated by various quantitative studies from the US and India, demonstrating that local news media coverage leads to a more informed electorate, increased electoral participation, and subsequently, greater political responsiveness from elected officials Besley and Burgess (2002); Snyder and Strömberg (2010).

This report aims to pave the way for a comprehensive, nationwide quantitative study to deepen our understanding of the factors contributing to the decline of local news publishers and their implications for politics and society in Australia. The remainder of the report is structured as follows: Section 2 introduces the data sources and the empirical methods; Section 2.2 presents the results for news outlet coverage in LGAs below 10,000 inhabitants and "competitive" news outlet markets for LGAs with a population larger than 10,000; Section 2.3 concludes and presents suggestions for future research directions.

Part 2

Socio-economic Determinants of Public Interest Journalism in Australia

2.1 Data Used in the Study

The empirical analysis is conducted at the Local Government Area (LGA) level. The final data set is a cross-sectional dataset with information about the number of print and digital local news publishers and a set of demographic and economic variables (*Socioeconomic variables)* for 540 LGAs covering all Australian states and territories.[1] The dataset combines variables on public interest journalism news publishers with socioeconomic characteristics as follows.

Public Interest Journalism Print and Digital Local News Publishers

News outlet and entity data is assembled by the Public Interest Journalism Initiative from public sources including existing and historical adjacent databases, through stakeholder engagement with industry and government, and through its own investigative efforts.[2] Analysis of news publishers against PIJI's eligibility standards for inclusion in the project is undertaken independently and relies on public access to relevant policies and content.

For inclusion in the ANDP, a news outlet should primarily and regularly produce: (i) original public interest journalism for (ii) a local, metropolitan, state or national (iii) public audience, and (iv) adhere to identifiable professional and ethical standards. The definition of public interest journalism which we have adopted for these projects is:

original content that records, reports or investigates issues of public significance for Australians; issues relevant to engaging Australians in public debate and in informing democratic decision making, or content which relates to community and local events.

PIJI's assessment criteria for news publishers does not currently take into account the way that news production in Indigenous communities may be guided by specific cultural protocols and history, and outside of a Western and commercial model that is common across other markets. There may be Indigenous journalists, outlets and news media entities producing news in local government areas that do not fit within PIJI's data framework but are of no less value to their communities. Broadcast (radio and television) news publishers are not included in this study, as PIJI does not have reliable, independent access to transcripts of news content

[1] The following seven mainly unincorporated LGAs are excluded from the analysis due to missing data on news publishers (2023 ABS LGA code in parentheses): Christmas Island (51710), Cocos Islands (51860), Unincorporated NSW (19399), Unincorporated NT (79399), Unincorporated SA (49399), Unincorporated VIC (29399), Unincorporated Other Territories (99399).

[2] Data on news publishers corresponds to coverage at June 30, 2023 (Dickson, 2023).

and coverage data is therefore less developed. This exclusion is not to suggest that broadcast news publishers are not important providers of public interest journalism in their communities.

For the remainder of the analysis, we focus on *print* and *digital* news publishers and treat them as one entity (news publishers, hereafter). We do so to account for the observed shift from print to digital versions of news publishers. We construct two mutually exclusive outcome variables of interest. First, we classify LGAs with a population count of *less than* 10,000 into "with coverage" and "without coverage".[3] We focus on sparsely populated LGAs to ensure *comparability* of news outlet markets. The binary outcome equals one if at least one print or digital news outlet is present in the LGA, and zero otherwise. Second, LGAs with a population count of *at least* 10,000 are classified into two groups: *(i)* LGAs with a "monopolistic" news outlet market (binary indicator equals zero) and *(ii)* LGAs with a "competitive" news outlet landscape (binary indicator equals one). A "monopolistic" market is defined as all print and digital news publishers in an LGA being owned by *one* news publisher company, while a "competitive" news outlet market requires at least two news publisher corporations operating in an LGA. "Competitive" news outlet markets are viewed as an indicator for more balanced reporting.

Socioeconomic variables

We retrieve information on socioeconomic variables from various government sources. First, we obtain data on demographic and labor market characteristics of LGAs from the Australian Bureau of Statistics' (ABS) *2021 Census of Population and Housing*. As a convention, variables expressed as the share relative to the LGA's total population, respectively, labor force are denoted "% Population" and "% Labor force". We complement our Census data with LGA-level information on *(i)* the industry composition from the ABS's *Counts of Australian Businesses, including Entries and Exits* (CABEE) in June 2020, *(ii)* on unemployment rates from *Jobs and Skills Australia* for the period from March 2022 to March 2023, and *(iii)* on remoteness from the *Australian Statistical Geography Standard Edition 3 Remoteness Areas*. More detailed information on each socioeconomic variable used in the analysis is provided in Table 4.1 in the Appendix.

Empirical Methods

Our analysis splits the data into two subsets, based on the size of the LGAs: The first subset contains 242 LGAs with a population of less than 10,000 inhabitants.

In the first phase, we provide a simple comparison of means to investigate how socioeconomic factors differ across LGAs *with* and *without* coverage, as well as between LGAs with a *monopolistic* and a *competitive* news outlet market. We concentrate on 30 different regional, demographic, economic, and labour market variables.

In a subsequent stage, this basic analysis is supplemented by a linear regression, which allows us to isolate the link between each independent variable and the dependent variable. In other words, the regression coefficient can be interpreted as the mean change in the dependent variable for each 1 unit change in an independent variable *while maintaining all other predictors constant*. However, using linear regression models necessitates pre-selection of variables to meet the OLS assumption that none of the predictors are highly correlated. Intuitively, if two or more predictors are closely related, the *independent* effect of each predictor on the outcome variable can no longer be reliably assigned, as independent variables tend to change in unison. This concern is addressed in two ways: First, based on economic intuition, we identify a group of key predictors. Second, we employ a data-driven strategy to identify the most powerful predictors and assess the robustness of our regression results.[4] In addition, we

[3]Population counts are retrieved from the *2021 Census of Population and Housing*.

[4]The interested reader is referred to Appendix Section 3.1 for a more detailed discussion on the automated feature selection process.

present in the main regression tables the *Shapley* value for each selected variable. Intuitively, the Shapley value provides a reference number on how much more of the variance the addition of a particular variable can explain.[5]

[5]For details on the Shapley value, please refer to Section 3.2.

2.2 Findings

In this section we present the findings for the set of 242 "small" LGAs with a population of less than 10,000 inhabitants followed by the subset of 298 "large" LGAs with a population of more than 10,000 inhabitants. In each section, we first present a map illustrating the spatial distribution of local news outlet markets across Australia. This is followed by balance tables that show the differences in means of socio-demographic and local economic variables between different types of LGAs, divided by the type of local news media market. Finally, we present results of multivariate regression models that show correlations between various subsets of socio-demographic as well as local economic variables and our outcome measures of local news publisher outlets.

News Outlet Coverage in LGAs with population $<$ 10,000

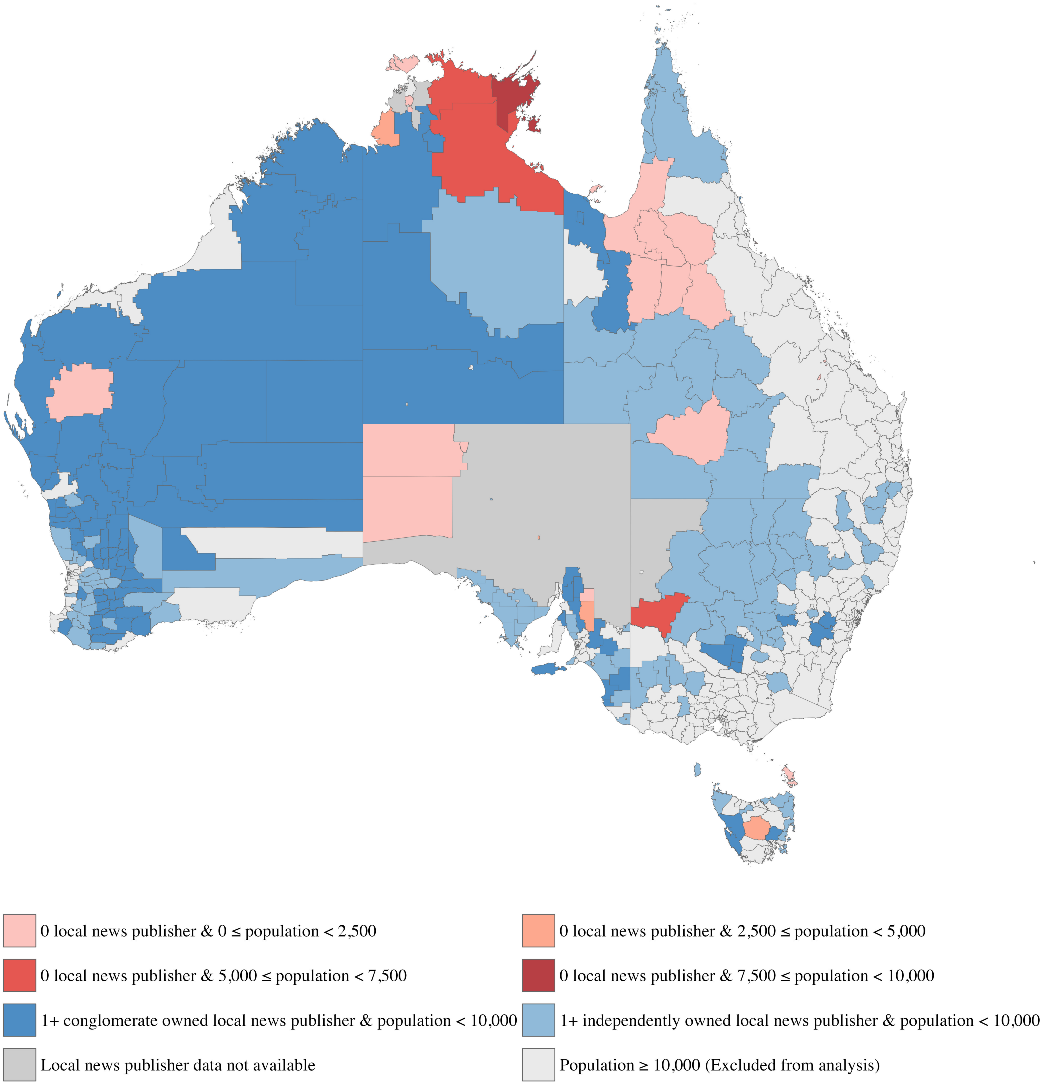
We commence our descriptive analysis of the data for differences in the presence local news publisher outlets in LGAs with population $<$ 10,000 by mapping the different types of local news markets across Australia. Figure 2.1 presents all the LGAs in Australia. The light grey ones, are LGAs with a population above 10,000, which are part of the sub-sample used in the next section. LGAs in light and dark blue and red represent the 242 LGAs with less than 10,000 residents. We observe that the majority of LGAs has access to at least one news outlet (blue) but coverage is not all-encompassing; 29 LGAs are without coverage (red shadings). The latter are located in primarily remote areas and characterised by low population density.[6] For the former, we highlight 91 LGAs that are exclusively serviced by news media "conglomerates" (dark blue) which we define as news media corporations with print and/or digital news publishers in more than one state.[7] Light blue areas denote LGAs that have at least one "independent" news outlet.

In the next step, we compare differences in socio-demographic and local economic variables between LGAs without any local news publisher outlets and those with one or more outlets. We present the results in form of a simple balance test. For each of the variables, we calculate the mean and standard deviation for each of the two groups (LGAs with 0 vs. LGAs with 1+ outlets) and then calculate the difference between the means. We then apply a standard t-test to check if the averages in each value are statistically significant. Table 2.1 highlights significant differences (grey) in socioeconomic characteristics of LGAs with and without coverage. The mean differences confirm the observation that public interest journalism covered areas are on average less remote and more populous. Notable demographic differences are also a lower share in Indigenous population and a larger share of first and second generation immigrants. Interestingly, no significant difference in the mean of the median total household income is observed, while the share of population with tertiary education is higher for LGAs with coverage.

[6] Darker red scales indicate higher population counts in LGAs without coverage and less than 10,000 residents, with population counts categorized into four equally spaced bins. Figure 5.1 in the Appendix presents the number of LGAs in each of the 2,500 population count bins in more detail.

[7] Note that we classify "Seven West Media" as a conglomerate to account for its size, public perception, and nation wide television interests despite its print and digital news publishers being restricted to Western Australia.

Figure 2.1: News Outlet Coverage in sparsely populated LGAs



Notes: The Australian mainland LGAs are displayed. Outlet information is not available for unincorporated and external territories. LGAs with a population above 10,000 are not considered in the analysis and depicted in (light) grey.

Table 2.1: Difference in means for LGAs with and without news outlet coverage (Part I)

	0 outlets (N = 29)		1+ outlets (N = 213)			
	Mean	SD	Mean	SD	Diff. in Mean	p-value
Average household size	3.0	0.8	2.4	0.4	-0.6	0.002
Birthplace outside AUS (% Population)	5.7	4.3	10.8	6.2	5.1	0.000
Dependency ratio	54.0	15.7	65.2	15.9	11.2	0.001
Foreign language used at home	28.5	34.5	9.3	16.4	-19.2	0.006
Indigenous (% Population)	47.5	38.5	14.5	22.5	-32.9	0.000
Major City (%)	0.0	0.0	2.3	15.2	2.3	0.025
Median total household income ($/weekly)	1214.7	481.8	1351.8	428.1	137.1	0.154
Parent born overseas (% Population)	2.5	2.5	5.2	2.5	2.7	0.000
Regional (%)	10.3	31.0	46.5	50.0	36.1	0.000
Remote (%)	89.7	31.0	51.2	50.1	-38.5	0.000
Tertiary education (% Population)	18.0	8.7	23.8	6.0	5.8	0.001
Total population	2219.3	2330.6	3476.3	2833.9	1257	0.011

Notes: Grey colored rows indicate variables that are significantly different between LGAs with and without coverage at the 5% level. The "dependency ratio" is defined as the share of the population that is younger than 15 years and older than 64 relative to the working age population of 15 to 64 year old. "Parent born overseas" comprises all individuals who have *at least* one parent born overseas and "tertiary education" comprises all individuals who hold *at least* a certificate or a higher level degree. LGAs are classified as "remote", "regional", or "major city" if the majority of their area lies within remoteness area class {"Remote Australia";"Very Remote Australia"}, {"Inner Regional Australia";"Outer Regional Australia"}, or {"Major Cities of Australia"}, respectively. Variables are ordered alphabetically.

Turning to labor market characteristics in Table 2.2 reveals that average employment is higher in areas with coverage and a larger share of the labor force is employed in the service industry and primary sector as opposed to the public sector.

Table 2.2: Difference in means for LGAs with and without news outlet coverage (Part II)

	0 outlets (N = 29)		1+ outlets (N = 213)			
	Mean	SD	Mean	SD	Diff. in Mean	p-value
Avg. unemployment rate (Mar 2022 to Mar 2023)	9.4	8.6	5.6	7.8	-3.8	0.031
Working in health care & social assistance (% Labor force)	9.9	5.8	9.3	4.2	-0.6	0.624
Working in retail trade (% Labor force)	5.1	3.2	5.9	2.4	0.8	0.198
Working in professional, scientific & technical services (% Labor force)	1.4	1.6	2.1	2.2	0.8	0.024
Working in construction (% Labor force)	4.5	3.0	5.7	2.5	1.3	0.031
Working in accommodation & food services (% Labor force)	3.7	3.2	5.2	3.2	1.5	0.026
Working in administrative & support services (% Labor force)	2.0	1.7	2.1	1.2	0.1	0.922
Working in manufacturing (% Labor force)	1.4	2.3	3.4	3.4	2.0	0.000
Working in transport, postal & warehousing (% Labor force)	2.4	2.3	3.6	1.8	1.2	0.009
Working in wholesale trade (% Labor force)	1.0	1.1	2.1	1.7	1.1	0.000
Working in financial & insurance services (% Labor force)	0.2	0.4	0.7	0.9	0.5	0.000
Working in education & training (% Labor force)	12.9	9.4	8.3	3.1	-4.6	0.014
Working in rental, hiring & real estate services (% Labor force)	0.3	0.6	0.6	0.7	0.3	0.03
Working in agriculture, forestry & fishing (% Labor force)	13.8	15.2	22.5	15.2	8.7	0.006
Working in mining (% Labor force)	3.2	9.1	5.9	10.7	2.7	0.156
Working in information, media & telecommunications (% Labor force)	0.1	0.3	0.3	0.3	0.2	0.017
Working in electricity, gas, water & waste services (% Labor force)	1.1	1.2	0.9	0.8	-0.2	0.411
Working in public administration & safety (% Labor force)	15.6	9.9	8.3	6.6	-7.3	0.001

Notes: Grey colored rows indicate variables that are significantly different between LGAs with and without coverage at the 5% level. Industry divisions are in descending order of their share of total employment in Australia in June 2021.

As discussed above, differences in mean ignore the correlation of a particular variable with other socioeconomic characteristics. In Table 2.3 we provide a more nuanced analysis in which we isolate the correlation between news outlet coverage and a particular variable while holding all other factors fixed (*ceteris paribus*). To isolate the correlation coefficient of a particular variable, we have to restrict the set of socioeconomic characteristics to variables that are not highly correlated with each other. Figure 5.2 and 5.3 present the pairwise correlation for a pre-selected set of variables of interest.[8] These inform our final selection of variables in Table 2.3. Note that we additionally present pairwise correlation coefficients for the top-5 predictors of coverage selected by *Lasso* (in bold). We see that either predictors are already among the selected set of characteristics or highly correlated with selected variables. We can, therefore, confidently assume that the final regression models cover the main socioeconomic predictors in our data set.[9]

We gradually introduce characteristics in Table 2.3. Column 1 restricts the set of variables to demographic characteristics. The three variables alone explain about 17.4% of the variation in news outlier coverage (R^2). The *Shapley* value presented in brackets below the regression coefficient captures how much each variable explains of the this variation.[10] In column 2 and 3, we gradually introduce variables capturing an LGA's labor market structure with the employment share in the primary and secondary, respectively tertiary sector. In column 4 and 5, we focus on the labor supply in lieu of the demand side. We observe that focusing on the business structure in LGAs in lieu of the labor force composition improves the predictive power of the model.

[8] For illustrative purposes, pairwise correlation coefficients below an absolute value of 0.4 are blanked out.
[9] See Table 4.1 for a detailed list of all socioeconomic variables in our data set.
[10] Note that the individual Shapley values add up to R^2.

Table 2.3: Conditional correlation between socioeconomic characteristics and news outlet coverage

		Labor force		Businesses	
	Demographic Variables (1)	Primary & Secondary Sector (2)	Tertiary Sector (3)	Primary & Secondary Sector (4)	Tertiary Sector (5)
Indigenous	-0.004***	-0.003***	-0.004	-0.004**	-0.001
	[0.140]	[0.098]	[0.055]	[0.091]	[0.048]
Median total household income	0.057	0.037	0.068	0.063	0.045
	[0.013]	[0.010]	[0.007]	[0.008]	[0.005]
Total population	0.036**	0.055**	0.066*	0.041*	0.041*
	[0.021]	[0.026]	[0.018]	[0.015]	[0.015]
Agriculture, forestry & fishing		0.003	0.002	0.001	-0.002
		[0.027]	[0.015]	[0.014]	[0.009]
Mining		0.003	0.002	0.013**	0.010*
		[0.009]	[0.006]	[0.010]	[0.009]
Manufacturing		0.000	-0.002	0.007	0.003
		[0.015]	[0.007]	[0.011]	[0.008]
Accommodation & food services			0.007		0.002
			[0.007]		[0.008]
Administrative & support services			0.005		-0.029***
			[0.001]		[0.026]
Construction			-0.005		-0.006
			[0.006]		[0.009]
Education & training			-0.011		-0.024***
			[0.038]		[0.055]
Electricity, gas, water & waste services			-0.071**		-0.004
			[0.024]		[0.001]
Health care & social assistance			0.009		0.028***
			[0.006]		[0.034]
Professional, scientific & technical services			-0.011		-0.016***
			[0.004]		[0.026]
Public administration & safety			0.002		-0.004
			[0.025]		[0.007]
Retail trade			-0.005		0.000
			[0.003]		[0.008]
Transport, postal & warehousing			-0.003		0.000
			[0.009]		[0.003]
Wholesale trade			0.019		-0.012***
			[0.015]		[0.015]
Financial & insurance services					-0.006***
					[0.009]
Information, media & telecommunications					-0.013***
					[0.031]
Rental, hiring & real estate services					0.003
					[0.015]
R^2	0.174	0.186	0.245	0.151	0.340

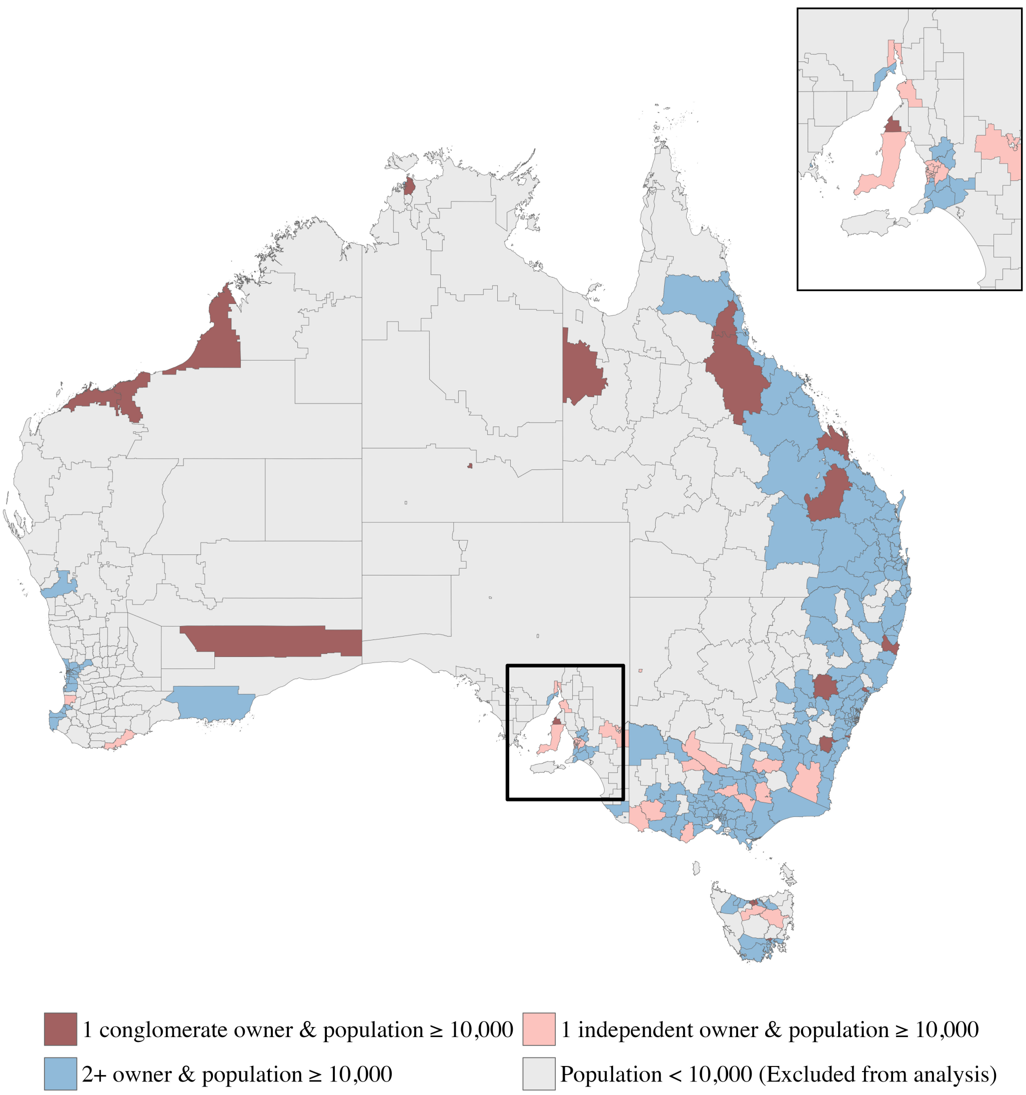
Notes: No. of obs. 242. The outcome variable is a binary indicator that takes the value of one if the LGA has at least one local news publisher outlet and zero if the LGA does not have a local news publisher outlet. Shapley values for R^2 decomposition are presented in brackets. Robust standard errors: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

"Competitive" news outlet markets in LGAs with population $\geq$ 10,000

In this section, we present our findings for the sub-sample of 298 LGAs with 10,000 or more residents. As all of these LGAs have at least one local news publisher outlet, we focus on the market structure instead of a comparison between coverage and no coverage. In particular, we distinguish between LGAs with a "monopolistic" local news publisher market (only 1 news outlet or multiple news publishers with one conglomerate owner) and LGAs with "competitive" market structure (2 or more local news publisher outlets with independent owners).

Figure 2.2 presents a fairly homogeneous geographic distribution of "monopolistic" (red shades) and "competitive" (blue) news outlet markets in populous LGAs with at least 10,000 inhabitants. Further, we observe that about 36% (20/55) of the "monopolistic" news outlet markets are dominated by a "conglomerate" owner (dark red). Moreover, Figure 2.2 reveals that "monopolistic" market structures are not insulated to regional or remote areas but are also present in metropolitan regions such as in and round Adelaide (zoomed in display in top right corner).

Figure 2.2: Geographic distribution of "monopolistic" and "competitive" news outlet markets



Notes: The Australian mainland LGAs are displayed. LGAs with a population below 10,000 are not considered in the analysis and depicted in (light) grey. A zoomed in display of LGAs in and around Adelaide is presented in the top right corner.

Next, we will again examine how socio-demographic and local economic factors differ between LGAs with a monopolistic and compared to a competitive local news publisher market using balance tables. We determine the average and standard deviation for each variable within the two categories of LGAs (monopolistic = 1 owner; competitive = 2+ owner). Following this, we calculate the difference in these averages. To assess whether these differences in average values are statistically meaningful, we conduct a conventional t-test.

Table 2.4 corroborates this visual observation. While remote areas and area with a larger share of Indigenous population are on average more likely to be dominated by a single news publisher corporation, news outlet market structures on average do not significantly differ for regional and metropolitan areas. On the other side, "competitive" news outlet markets are on average more populous, "cosmopolitan" ("birthplace outside AUS", "foreign language used at home", and "parent born overseas"), and characterised by a larger share of the population with tertiary education.

Table 2.4: Difference in means for LGAs with "monopolistic" vs. "competitive" news outlet markets (Part I)

	1 owner (N = 55)		2+ owner (N = 243)			
	Mean	SD	Mean	SD	Diff. in Mean	p-value
Average household size	2.4	0.2	2.5	0.2	0.1	0.035
Birthplace outside AUS (% Population)	16.8	9.9	22.0	13.4	5.2	0.001
Dependency ratio	62.2	13.7	60.4	12.8	-1.8	0.381
Foreign language used at home	11.3	10.3	15.2	15.4	3.9	0.023
Indigenous (% Population)	5.9	6.2	3.7	3.4	-2.2	0.012
Major City (%)	29.1	45.8	34.2	47.5	5.1	0.464
Median total household income ($/weekly)	1574.9	457.3	1680.3	450.9	105.4	0.126
Parent born overseas (% Population)	6.1	1.6	6.8	1.9	0.7	0.004
Regional (%)	52.7	50.4	60.1	49.1	7.4	0.329
Remote (%)	18.2	38.9	5.8	2.3	-12.4	0.026
Tertiary education (% Population)	29.7	5.6	31.5	6.6	1.8	0.039
Total population	35907.9	33069	92895.5	119260.8	56987.6	0.000

Notes: Grey colored rows indicate variables that are significantly different between LGAs with and without coverage at the 5% level. The "dependency ratio" is defined as the share of the population that is younger than 15 years and older than 64 relative to the working age population of 15 to 64 year old. "Parent born overseas" comprises all individuals who have *at least* one parent born overseas and "tertiary education" comprises all individuals who hold *at least* a certificate or a higher level degree. LGAs are classified as "remote", "regional", or "major city" if the majority of their area lies within remoteness area class {"Remote Australia";"Very Remote Australia"}, {"Inner Regional Australia";"Outer Regional Australia"}, or {"Major Cities of Australia"}, respectively. Variables are ordered alphabetically.

Table 2.5 provides more information on the difference in labor market characteristics. On average, labor markets across "competitive" and "monopolistic" news outlet market LGAs are fairly homogeneous. We observe a significant differences in the mean of the labor force employed in financial, professional, and real estate services as well as in wholesale and information, media and telecommunication, which are all more prominent in "competitive" news outlet markets. Conversely, "monopolistic" news outlet markets have on average less employment in electricity, gas, water, and waste service.

Table 2.5: Difference in Means for LGAs with "monopolistic" vs. "competitive" news publisher markets (Part II)

	1 owner (N = 55)		2+ owner (N = 243)			
	Mean	SD	Mean	SD	Diff. in Mean	p-value
Avg. unemployment rate (Mar 2022 to Mar 2023)	3.9	2.1	3.8	1.5	-0.1	0.836
Working in health care & social assistance (% Labor force)	14.0	3.4	13.6	2.8	-0.4	0.491
Working in retail trade (% Labor force)	8.6	1.4	8.7	1.3	0.1	0.826
Working in professional, scientific & technical services (% Labor force)	4.4	3.0	6.3	4.2	1.9	0.000
Working in construction (% Labor force)	8.0	2	8.4	2.4	0.4	0.197
Working in accommodation & food services (% Labor force)	6.7	2.0	6.4	1.9	-0.3	0.303
Working in administrative & support services (% Labor force)	3.0	0.6	3.0	0.6	-0.0	0.526
Working in manufacturing (% Labor force)	5.7	2.8	6.1	2.9	0.4	0.373
Working in transport, postal & warehousing (% Labor force)	4.1	1.4	4.0	1.5	-0.1	0.685
Working in wholesale trade (% Labor force)	2.0	0.5	2.3	0.7	0.3	0.001
Working in financial & insurance services (% Labor force)	1.7	1.1	2.8	2.5	1.1	0.000
Working in education & training (% Labor force)	8.2	1.5	8.3	1.6	0.1	0.629
Working in rental, hiring & real estate services (% Labor force)	1.2	0.4	1.4	0.6	0.2	0.003
Working in agriculture, forestry & fishing (% Labor force)	6.4	7.0	4.7	6.0	-1.7	0.111
Working in mining (% Labor force)	4.3	7.1	2.4	4.0	-1.9	0.063
Working in information, media & telecommunications (% Labor force)	0.7	0.4	1.1	0.8	0.4	0.000
Working in electricity, gas, water & waste services (% Labor force)	1.4	0.6	1.2	0.6	-0.2	0.012
Working in public administration & safety (% Labor force)	6.7	3.1	5.9	2.8	-0.8	0.061

Notes: Grey colored rows indicate variables that are significantly different between LGAs with and without coverage at the 5% level. Industry divisions are in descending order of their share of total employment in Australia in June 2021.

Again, simply looking at the mean differences disregards correlations between variables and other socioeconomic factors. Table 2.6 presents the results of a number of multivariate regression models to isolate the association between local news market structure and a specific variable, while maintaining all other parameters constant. To determine the correlation coefficient of a variable, we again limit socioeconomic characteristics to those that are not substantially associated. Figures 5.4 and 5.5 show the pairwise correlation for a pre-selected collection of variables. The final variable selection is based on these factors in Table 2.2. Additionally, we provide pairwise correlation coefficients for Lasso's top-5 coverage predictors (bold). Predictors are either already in the selected collection or substantially linked with specified variables. We may fairly assume that the final regression models address the key socioeconomic predictors in our data set.

In Table 2.6, we introduce different groups of variables in a step-wise manner. Column 1 limits variables to demographics. Just three variables account for 9.6% of variation in local news market structure (R^2). The bracketed Shapley value underneath the regression coefficient indicates the extent of variance explained by each variable. We add variables representing an LGA's labour market structure, including employment share in primary, secondary, and tertiary sectors, in columns 2 and 3. In columns 4 and 5, we emphasise labour supply over demand. We found that focusing on company structure in LGAs rather than labour force composition slightly enhances the model's predictive power.

A LGA's total population is most consistently correlated with the likelihood that an LGA has a more diverse local news publisher market. The results in column 5, further suggest that LGAs with a higher fraction of manufacturing companies are more likely to have a more diverse local news publisher landscape, while LGAs with a higher fraction of administrative and support services seem to be more likely to have a monopolistic market structure. A higher fraction of businesses in the professional, scientific and technical services sector is also positively correlated with a more diverse local news publisher market.

Table 2.6: Conditional correlation between socioeconomic characteristics and "competitive" news publisher markets

		Labor force		Businesses	
	Demographic Variables (1)	Primary & Secondary Sector (2)	Tertiary Sector (3)	Primary & Secondary Sector (4)	Tertiary Sector (5)
Average household size	0.049	0.003	0.231	0.026	0.204
	[0.005]	[0.005]	[0.006]	[0.004]	[0.006]
Avg. unemployment rate	-0.007	-0.008	-0.005	-0.006	0.011
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Indigenous	-0.012*	-0.008	0.007	-0.011	-0.006
	[0.028]	[0.022]	[0.012]	[0.026]	[0.016]
Median total household income	-0.072	0.014	-0.462	-0.038	
	[0.004]	[0.005]	[0.006]	[0.004]	
Total population	0.093***	0.092***	0.124***	0.088***	0.112***
	[0.055]	[0.051]	[0.052]	[0.052]	[0.046]
Mining		-0.004	0.002	-0.033	-0.054
		[0.011]	[0.009]	[0.008]	[0.010]
Manufacturing		0.009	0.015	0.000	0.044*
		[0.003]	[0.005]	[0.001]	[0.006]
Accommodation & food services			0.010		-0.007
			[0.001]		[0.006]
Administrative & support services			-0.086		-0.087**
			[0.005]		[0.015]
Construction			0.019		-0.009
			[0.006]		[0.003]
Education & training			0.002		-0.018
			[0.001]		[0.010]
Electricity, gas, water & waste services			-0.048		0.008
			[0.010]		[0.000]
Health care & social assistance			-0.021*		-0.021
			[0.008]		[0.006]
Professional, scientific & technical services			0.030		0.030***
			[0.015]		[0.030]
Public administration & safety			-0.013		0.119
			[0.011]		[0.002]
Retail trade			-0.013		0.011
			[0.002]		[0.001]
Transport, postal & warehousing			-0.030*		0.003
			[0.005]		[0.003]
Wholesale trade			0.006		-0.056*
			[0.006]		[0.006]
R^2	0.093	0.098	0.160	0.096	0.168

Notes: No. of obs. 298. The outcome variable is a binary indicator that takes the value of one if the LGA has a competitive local news publisher market (multiple news publishers from different owners) and zero if the local news publisher market is monopolistic (only one news publisher or multiple news publishers owned by a conglomerate). "Median total household income" is excluded in column 5 due to its high correlation with "Businesses in Professional, scientific & technical services (% Total Businesses)". Shapley values for R^2 decomposition are presented in brackets. Robust standard errors: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

2.3 Discussion & Conclusions

This report offers a preliminary, empirical analysis aimed at understanding the socio-economic factors driving local public interest journalism in Australia, particularly focusing on how these elements affect the availability and nature of news publisher coverage in local government areas (LGAs). We constructed a cross-sectional dataset that amalgamates data on the availability and market structure of local news publishers with demographic, economic, and labor market characteristics across 540 Australian LGAs, spanning all states and territories.

Our empirical methodology includes a comparative analysis of means to identify socio-economic disparities among LGAs with varying levels of news media coverage and market competition. This is further enriched by linear regression models, which help isolate the relationships between variables, thus offering a more robust examination of the influences on public interest journalism. This initial exploratory analysis has also highlighted certain limitations and illuminated potential avenues for future empirical research in this field. Two primary data constraints need addressing:

First, the cross-sectional nature of the dataset, coupled with the absence of exogenous variation in key variables (i.e., local economic structure, demographic factors), limits our analysis to largely descriptive methods. This constraint hinders the development of an empirical identification strategy for causal inference. A significant enhancement would involve compiling panel data on local news publishers, including information on the timing of their entry and exit. Second, acquiring more detailed information about the structure of the local news market, the business models of local news publishers, and the content of local news would enable a more nuanced empirical analysis. Specifically, this would facilitate an examination of the mechanisms underlying the entry and exit of local news publishers in the market and their impact on Australian society and politics. Third, an absence of data about the news production of radio and television broadcast news means these outlets were excluded from the study. Reliable access to transcript data from across the entire sector would enable these outlets to be included in future research as well.

Once these data requirements are met, future research could explore several areas: 1. Investigating the relationship between local news outlet availability and voter information and behavior. 2. Examining the role of public interest (PI) journalism in democratic participation and voting outcomes. 3. Assessing the impact of local news outlet circulation on election results. 4. Analyzing the willingness to pay for tax deductions supporting charitable contributions to local news outlet apprenticeships. 5. Exploring the potential contagion effects of news media content and media slant.

Part 3

Technical Appendix

3.1 Lasso

We use an *logit–lasso* model to select the strongest predictors. The logistic lasso estimator $\hat{\beta} = \hat{\beta}_1, ..., \hat{\beta}_p$ is defined as the minimizer of the penalized negative log likelihood (also called "logistic loss" function):

$$\hat{\beta} = \arg\min_{\beta} \Bigg\{ \underbrace{-\left[\frac{1}{N} \sum_{i=1}^{N} y_i \cdot (X_i^T \beta) - \log\left(1 + e^{X_i^T \beta}\right) \right]}_{\text{"logistic loss function"}} + \underbrace{\lambda \sum_{j=1}^{p} |\beta_j|}_{\text{L1 penalty}} \Bigg\},$$

where y_i is the binary class indicator for coverage, respectively, competition and λ is the "L1 penalty" (or "L1 nrom"). Intuitively, the higher the penalty λ the more coefficients of irrelevant features are pushed towards zero, with the weakest predictors being pushed all the way to zero. The model thus conducts *automated feature selection*. We use the `glmnet` (Friedman et al., 2010) **R** package to implement the *logit–lasso* model.

The optimal value of λ (λ^*) is identified to maximize the out-of-sample *Brier score* using 10-fold cross-validation repeated 10 times.[1] The probability of class labels is computed as the predicted value from the fitted coefficients $\hat{\beta}$ at λ^*.

3.2 The Shapley Value

The *Shapley value* is defined via a value function val of players in S. The *Shapley value* of a feature value is its contribution to the payout, weighted and summed over all possible feature value combinations:

$$\phi_j(val) = \sum_{S \subseteq \{1,\ldots,p\} \setminus \{j\}} \frac{|S|!\,(p - |S| - 1)!}{p!} \left(val\left(S \cup \{j\}\right) - val(S) \right),$$

where S is a subset of predictors used in the model, x is the vector of feature values of the instance to be explained and p the number of v. $val(S)$ is the prediction for independent variables in set S that are marginalized over variables that are not included in set S:

$$val_x(S) = \int \hat{f}(x_1, \ldots, x_p) d\mathbb{P}_{x \notin S} - E_X(\hat{f}(X))$$

[1] The repetitions ensure that the stochasticity in the 10-fold cross-validation split does not drive the results. We use the `vfold_cv` function of the `tidymodels` (Kuhn and Wickham, 2020) **R** package to implement the k-fold cross-validation.

You actually perform multiple integrations for each variable that is not contained in S. For instance, the linear regression model includes 4 predictors x_1, x_2, x_3 and x_4 and we evaluate the prediction for the coalition S consisting of feature values x_1 and x_3:

$$val_x(S) = val_x(\{1,3\}) = \int_{\mathbb{R}} \int_{\mathbb{R}} \hat{f}(x_1, X_2, x_3, X_4) d\mathbb{P}_{X_2 X_4} - E_X(\hat{f}(X))$$

For intuition, consider the following example: The predictors enter a room in *random* order. All predictors in the room participate in the game = contribute to the prediction. The Shapley value of a predictor is the average change in the prediction that the coalition already in the room receives when the predictor joins them.

For more details, the reader is referred to the excellent summary in Molnar (2022)

Part 4

Additional Tables

Table 4.1: Variable definitions

Formula, resp., Definition	Variable
Panel A: 2021 Census of Population and Housing	
p_tot_adf_ev_ser_tot / tot_p_p	Ever served in Australian Defence Force (% Population)
p_tot_volunteer / tot_p_p	Volunteers (% Population)
median_age_persons	Median age of population
median_mortgage_repay_monthly	Median mortgage repayment ($/monthly)
median_rent_weekly	Median rent ($/weekly)
median_tot_fam_inc_weekly	Median total family income ($/weekly)
median_tot_hhd_inc_weekly	Median total household income ($/weekly)
median_tot_prsnl_inc_weekly	Median total personal income ($/weekly)
(age_15_19_yr_f + age_20_24_yr_f + age_25_34_yr_f) / tot_p_p	Women aged 15-34 years (% Population)
(age_35_44_yr_f + age_45_54_yr_f + age_55_64_yr_f) / tot_p_p	Women aged 35-64 years (% Population)
(age_65_74_yr_f + age_75_84_yr_f + age_85ov_f) / tot_p_p	Women aged 65 years and older (% Population)
(age_15_19_yr_m + age_20_24_yr_m + age_25_34_yr_m) / tot_p_p	Men aged 15-34 years (% Population)
(age_35_44_yr_m + age_45_54_yr_m + age_55_64_yr_m) / tot_p_p	Men aged 15-64 years (% Population)
(age_65_74_yr_m + age_75_84_yr_m + age_85ov_m) / tot_p_p	Men aged 65 years and older (% Population)
tot_p_p	Total population
average_household_size	Average household size
average_num_psns_per_bedroom	Average number of persons per bedroom
christianity_tot_p / tot_p_p	Profession of Christianity (% Population)
high_yr_schl_comp_yr_12_eq_p / tot_p_p	Year 12 completed (% Population)
(high_yr_schl_comp_yr_12_eq_p + high_yr_schl_comp_yr_11_eq_p + high_yr_schl_comp_yr_10_eq_p) / tot_p_p	Year 10 completed (% Population)

Table 4.1: Variable Definitions *(continued)*

Formula, resp., Definition	Variable
(high_yr_schl_comp_yr_8_belw_p + high_yr_schl_comp_d_n_g_sch_p) / tot_p_p	Year 8 or below completed
(p_tot_pd + p_tot_gd_gc + p_tot_bd + p_tot_add_d + p_tot_cert) / tot_p_p	Tertiary education (% Population)
(p_tot_pd + p_tot_gd_gc + p_tot_bd) / tot_p_p	AQF level 8-10 (% Population)
(aust_bp_b_os + aust_fo_b_os + aust_mo_b_os) / (tot_p_p-aust_birthplace_not_stated)	Parent born overseas (% Population)
birthplace_elsewhere_p / tot_p_p	Birthplace outside AUS (% Population)
(p_china_tot + p_hong_kong_sar_ch_tot + p_taiwan_tot) / tot_p_p	Chinese (% Population)
tot_p_m/tot_p_f	Sex ratio
(age_0_4_yr_p + age_5_14_yr_p + age_65_74_yr_p + age_75_84_yr_p + age_85ov_p) / (age_15_19_yr_p+ age_20_24_yr_p + age_25_34_yr_p + age_35_44_yr_p + age_45_54_yr_p + age_55_64_yr_p)	Dependency ratio
p_tot_married / tot_p_p	Married (% Population)
indigenous_p_tot_p / tot_p_p	Indigenous (% Population)
lang_used_home_oth_lang_p / tot_p_p	Foreign language used at home
p_accom_food_tot / p_tot_lf_tot	Working in accommodation & food services (% Labor force)
p_admin_supp_tot / p_tot_lf_tot	Working in administrative & support services (% Labor force)
p_ag_for_fshg_tot / p_tot_lf_tot	Working in agriculture, forestry & fishing (% Labor force)
p_constru_tot / p_tot_lf_tot	Working in construction (% Labor force)
p_educ_trng_tot / p_tot_lf_tot	Working in education & training (% Labor force)
p_el_gas_wt_waste_tot / p_tot_lf_tot	Working in electricity, gas, water & waste services (% Labor force)
p_fin_insur_tot / p_tot_lf_tot	Working in financial & insurance services (% Labor force)
p_hlthcare_socas_tot / p_tot_lf_tot	Working in health care & social assistance (% Labor force)
p_info_media_teleco_tot / p_tot_lf_tot	Working in information, media & telecommunications (% Labor force)
p_manufact_tot / p_tot_lf_tot	Working in manufacturing (% Labor force)
p_mining_tot / p_tot_lf_tot	Working in mining (% Labor force)
p_pro_scien_tec_tot / p_tot_lf_tot	Working in professional, scientific & technical services (% Labor force)
p_public_admin_sfty_tot / p_tot_lf_tot	Working in public administration & safety (% Labor force)
p_rettde_tot / p_tot_lf_tot	Working in retail trade (% Labor force)
p_rtnhir_rest_tot / p_tot_lf_tot	Working in rental, hiring & real estate services (% Labor force)

Table 4.1: Variable Definitions *(continued)*

Formula, resp., Definition	Variable
p_trans_post_wrehsg_tot / p_tot_lf_tot	Working in transport, postal & warehousing (% Labor force)
p_whlesaletde_tot / p_tot_lf_tot	Working in wholesale trade (% Labor force)
p_tot_lf_tot	Labor force
Panel B: Counts of Australian Businesses, including Entries and Exits	
"Non employing"/Total	Businesses: Self-employed (% Total businesses)
"1-19 Employees"/Total	Businesses: 0-19 employees (% Total businesses)
"20-199 Employees"/Total	Businesses: 20-199 employees (% Total businesses)
"200+ Employees"/Total	Businesses: 200+ employees (% Total businesses)
A/Total	Businesses in agriculture, forestry & fishing (% total businesses)
B/Total	Businesses in mining (% total businesses)
C/Total	Businesses in manufacturing (% total businesses)
D/Total	Businesses in electricity, gas, water & waste services (% total businesses)
E/Total	Businesses in construction (% total businesses)
F/Total	Businesses in wholesale trade (% total businesses)
G/Total	Businesses in retail trade (% total businesses)
H/Total	Businesses in accommodation & food services (% total businesses)
I/Total	Businesses in transport, postal & warehousing (% total businesses)
J/Total	Businesses in information media & telecommunications (% total businesses)
K/Total	Businesses in financial & insurance services (% total businesses)
L/Total	Businesses in rental, hiring & real estate services (% total businesses)
M/Total	Businesses in professional, scientific & technical services (% total businesses)
N/Total	Businesses in administrative & support services (% total businesses)
O/Total	Businesses in public administration & safety (% total businesses)
P/Total	Businesses in education & training (% total businesses)
Q/Total	Businesses in health care & social assistance (% total businesses)

Table 4.1: Variable Definitions *(continued)*

Formula, resp., Definition	Variable
R/Total	Businesses in arts & recreation services (% total businesses)
S/Total	Businesses in other services (% total businesses)
X/Total	Businesses with unknown sector
Panel C: Jobs and Skills Australia	
(Mar-22 + Jun-22 + Sep-22 + Dec-22 + Mar-23) / 5	Avg. unemployment rate (Mar 2022 to Mar 2023)
Panel D: Remoteness Structure	
{"Major Cities of Australia"}	Major City
{"Inner Regional Australia";"Outer Regional Australia"}	Regional
{"Remote Australia";"Very Remote Australia"}	Remote

Table 4.2: Conditional correlation between socioeconomic characteristics and news outlet coverage

		Labor force		Businesses	
	Demographic Variables (1)	Primary & Secondary Sector (2)	Tertiary Sector (3)	Primary & Secondary Sector (4)	Tertiary Sector (5)
Indigenous	-0.004***	-0.003***	-0.004	-0.004**	-0.001
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Median total household income	0.057	0.037	0.068	0.063	0.045
	(0.087)	(0.083)	(0.111)	(0.095)	(0.087)
Total population	0.036**	0.055**	0.066*	0.041*	0.041*
	(0.018)	(0.025)	(0.034)	(0.022)	(0.023)
Agriculture, forestry & fishing		0.003	0.002	0.001	-0.002
		(0.002)	(0.006)	(0.002)	(0.002)
Mining		0.003	0.002	0.013**	0.010*
		(0.002)	(0.006)	(0.006)	(0.005)
Manufacturing		0.000	-0.002	0.007	0.003
		(0.004)	(0.008)	(0.008)	(0.009)
Accommodation & food services			0.007		0.002
			(0.009)		(0.004)
Administrative & support services			0.005		-0.029***
			(0.021)		(0.010)
Construction			-0.005		-0.006
			(0.012)		(0.004)
Education & training			-0.011		-0.024***
			(0.007)		(0.006)
Electricity, gas, water & waste services			-0.071**		-0.004
			(0.033)		(0.016)
Health care & social assistance			0.009		0.028***
			(0.010)		(0.009)
Professional, scientific & technical services			-0.011		-0.016***
			(0.018)		(0.005)
Public administration & safety			0.002		-0.004
			(0.008)		(0.011)
Retail trade			-0.005		0.000
			(0.015)		(0.004)
Transport, postal & warehousing			-0.003		0.000
			(0.013)		(0.008)
Wholesale trade			0.019		-0.012***
			(0.019)		(0.003)
Financial & insurance services					-0.006***
					(0.002)
Information, media & telecommunications					-0.013***
					(0.003)
Rental, hiring & real estate services					0.003
					(0.004)
R^2	0.174	0.186	0.245	0.151	0.340

Notes: No. of obs. 242. The outcome variable is a binary indicator that takes the value of one if the LGA has at least one local news outlet and zero if the LGA does not have a local news outlet. Robust standard errors are presented in parantheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.3: Conditional correlation between socioeconomic characteristics and "competitive" news outlet markets

		Labor force		Businesses	
	Demographic Variables (1)	Primary & Secondary Sector (2)	Tertiary Sector (3)	Primary & Secondary Sector (4)	Tertiary Sector (5)
Average household size	0.049	0.003	0.231	0.026	0.204
	(0.079)	(0.088)	(0.167)	(0.090)	(0.146)
Avg. unemployment rate	-0.007	-0.008	-0.005	-0.006	0.011
	(0.019)	(0.019)	(0.022)	(0.019)	(0.020)
Indigenous	-0.012*	-0.008	0.007	-0.011	-0.006
	(0.007)	(0.008)	(0.009)	(0.007)	(0.008)
Median total household income	-0.072	0.014	-0.462	-0.038	
	(0.111)	(0.128)	(0.299)	(0.125)	
Total population	0.093***	0.092***	0.124***	0.088***	0.112***
	(0.024)	(0.026)	(0.028)	(0.024)	(0.032)
Mining		-0.004	0.002	-0.033	-0.054
		(0.006)	(0.009)	(0.040)	(0.037)
Manufacturing		0.009	0.015	0.000	0.044*
		(0.009)	(0.012)	(0.019)	(0.023)
Accommodation & food services			0.010		-0.007
			(0.020)		(0.029)
Administrative & support services			-0.086		-0.087**
			(0.057)		(0.040)
Construction			0.019		-0.009
			(0.012)		(0.006)
Education & training			0.002		-0.018
			(0.017)		(0.092)
Electricity, gas, water & waste services			-0.048		0.008
			(0.050)		(0.079)
Health care & social assistance			-0.021*		-0.021
			(0.011)		(0.013)
Professional, scientific & technical services			0.030		0.030***
			(0.021)		(0.010)
Public administration & safety			-0.013		0.119
			(0.010)		(0.186)
Retail trade			-0.013		0.011
			(0.032)		(0.029)
Transport, postal & warehousing			-0.030*		0.003
			(0.018)		(0.009)
Wholesale trade			0.006		-0.056*
			(0.047)		(0.031)
R^2	0.093	0.098	0.160	0.096	0.168

Notes: No. of obs. 298. The outcome variable is a binary indicator that takes the value of one if the LGA has a competitive local news outlet market (multiple news outlets from different owners) and zero if the local news outlet market is monopolistic (only one news outlet or multiple outlets owned by a conglomerate). "Median total household income" is excluded in column 5 due to its high correlation with "Businesses in Professional, scientific & technical services (% Total Businesses)". Robust standard errors are presented in parantheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Part 5

Additional Figures

Figure 5.1: Population in LGAs with 0 outlets

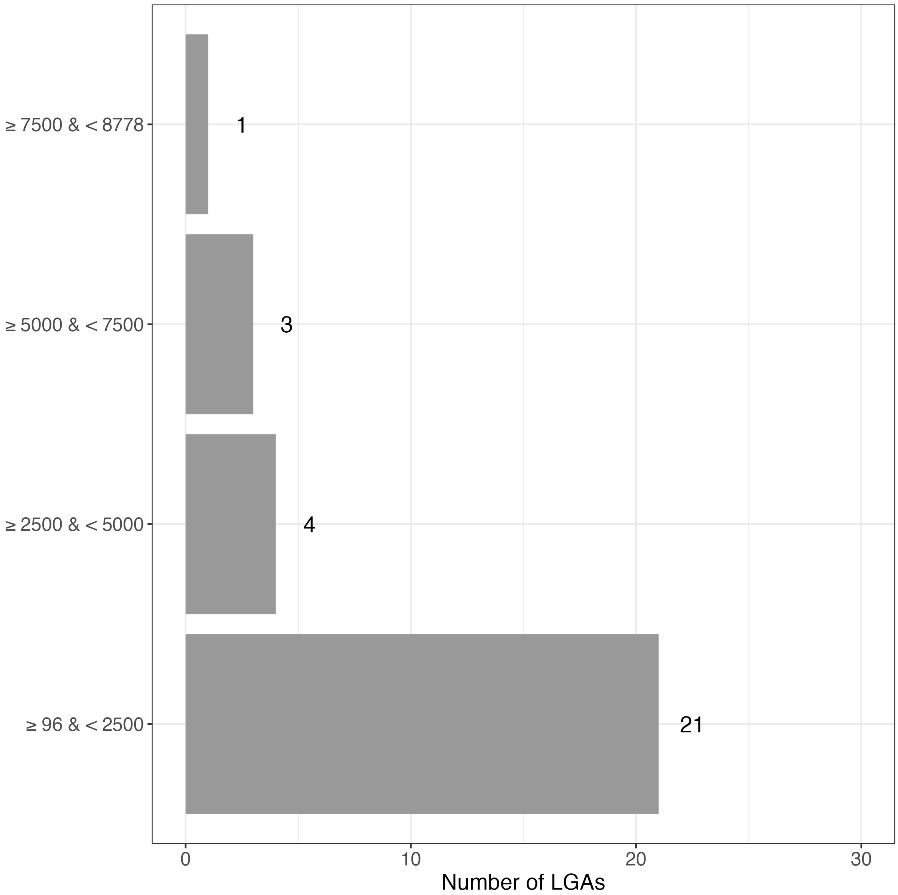
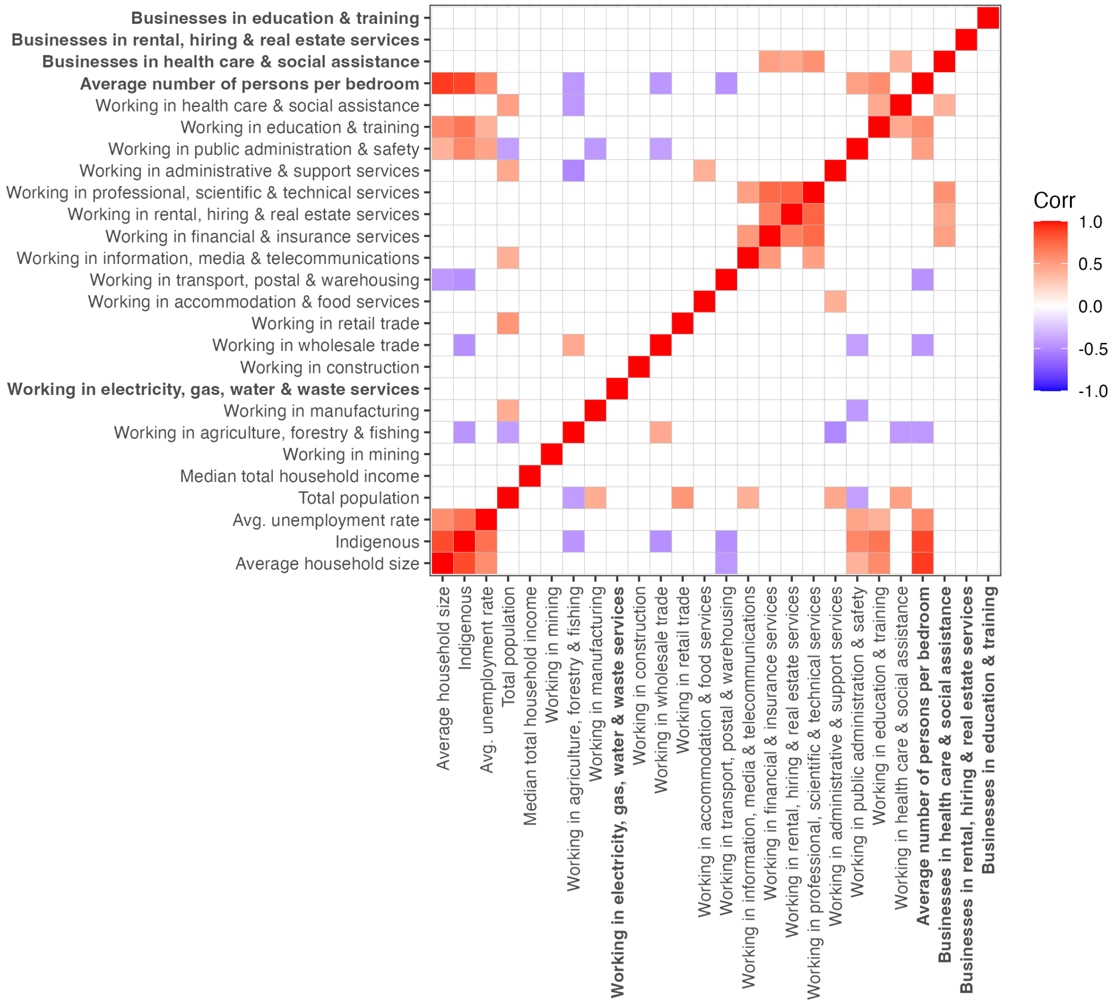
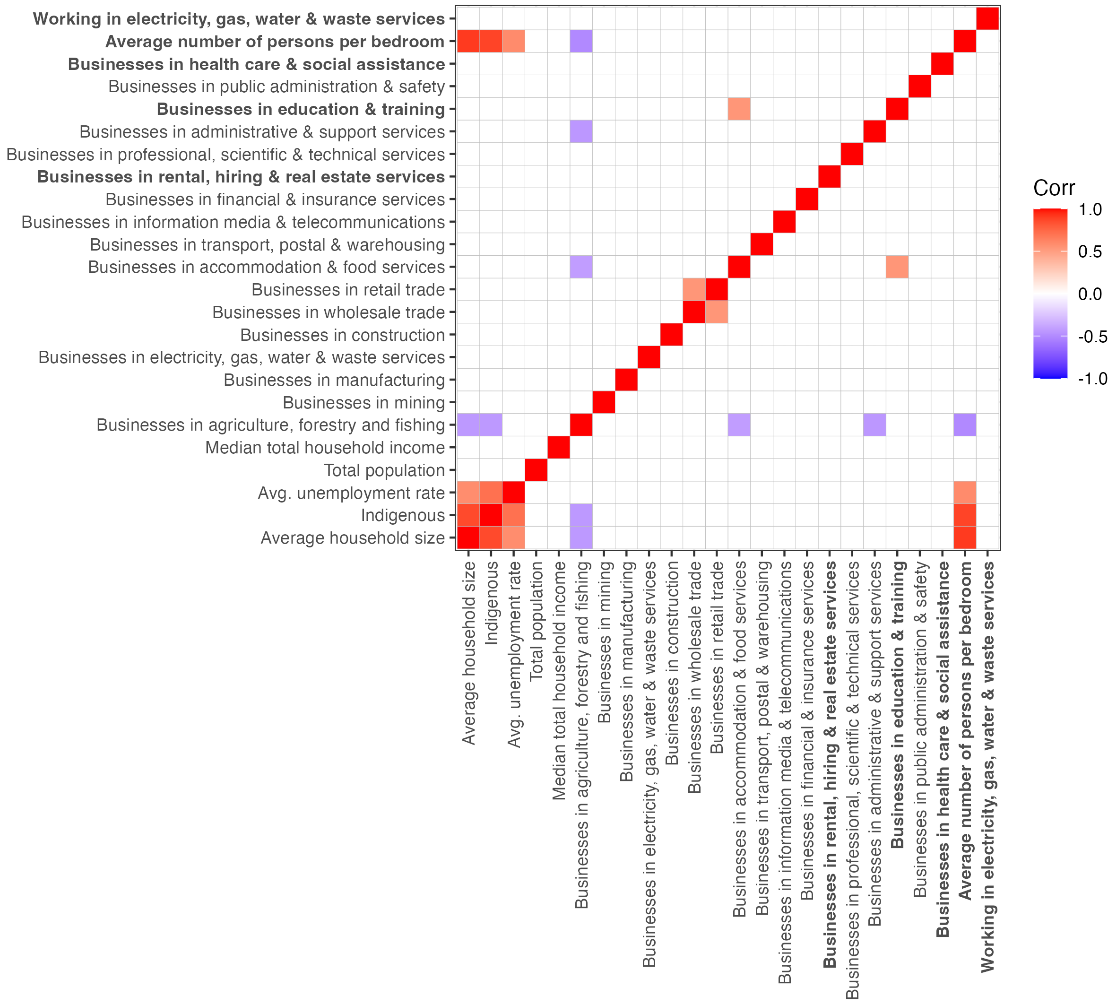


Figure 5.2: Correlation of Top-5 variables selected by Lasso with variables in column 3 of Table 2.3



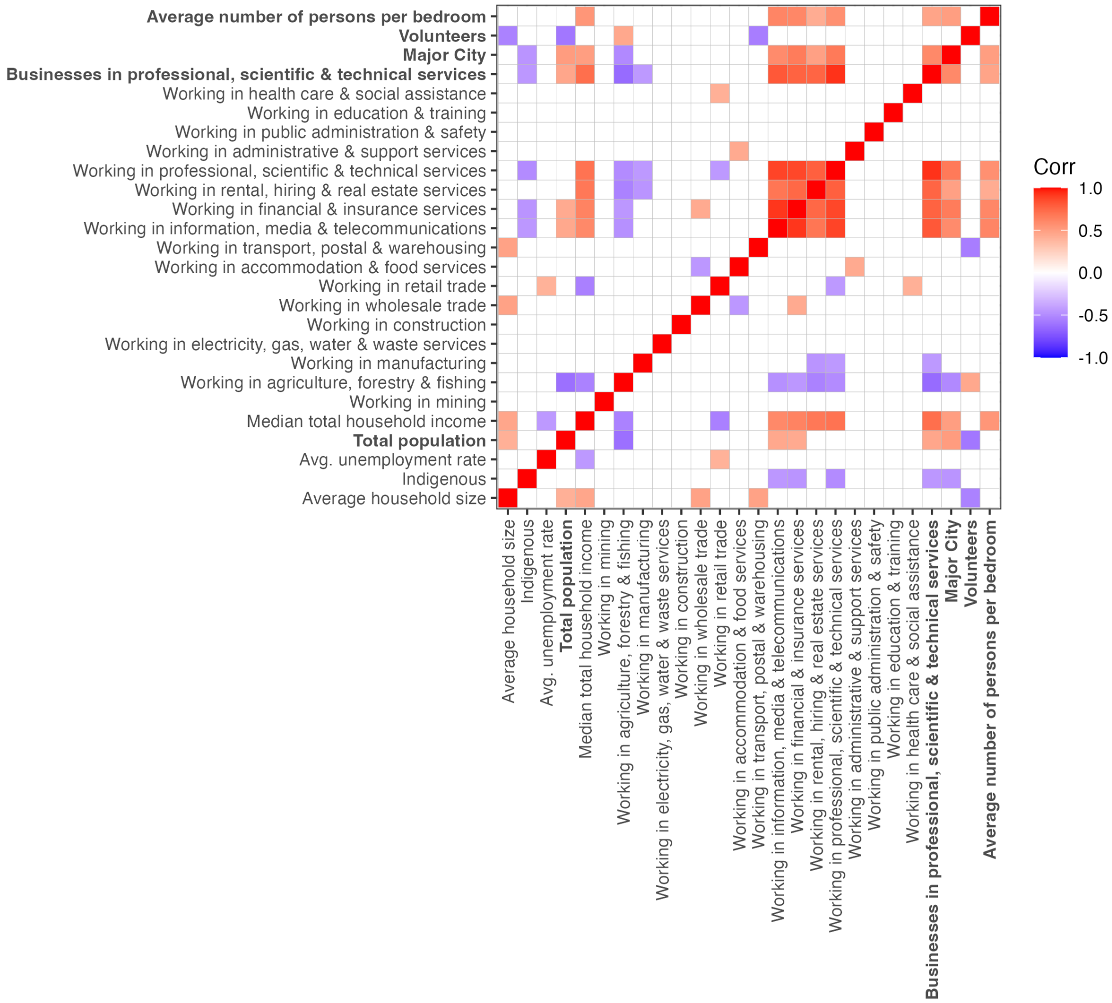
Notes: Pairwise correlation coefficients in the range of -0.4 to 0.4 are blanked out. The Top-5 variables selected by Lasso are highlighted in bold.

Figure 5.3: Correlation of Top-5 variables selected by Lasso with variables in column 5 of Table 2.3



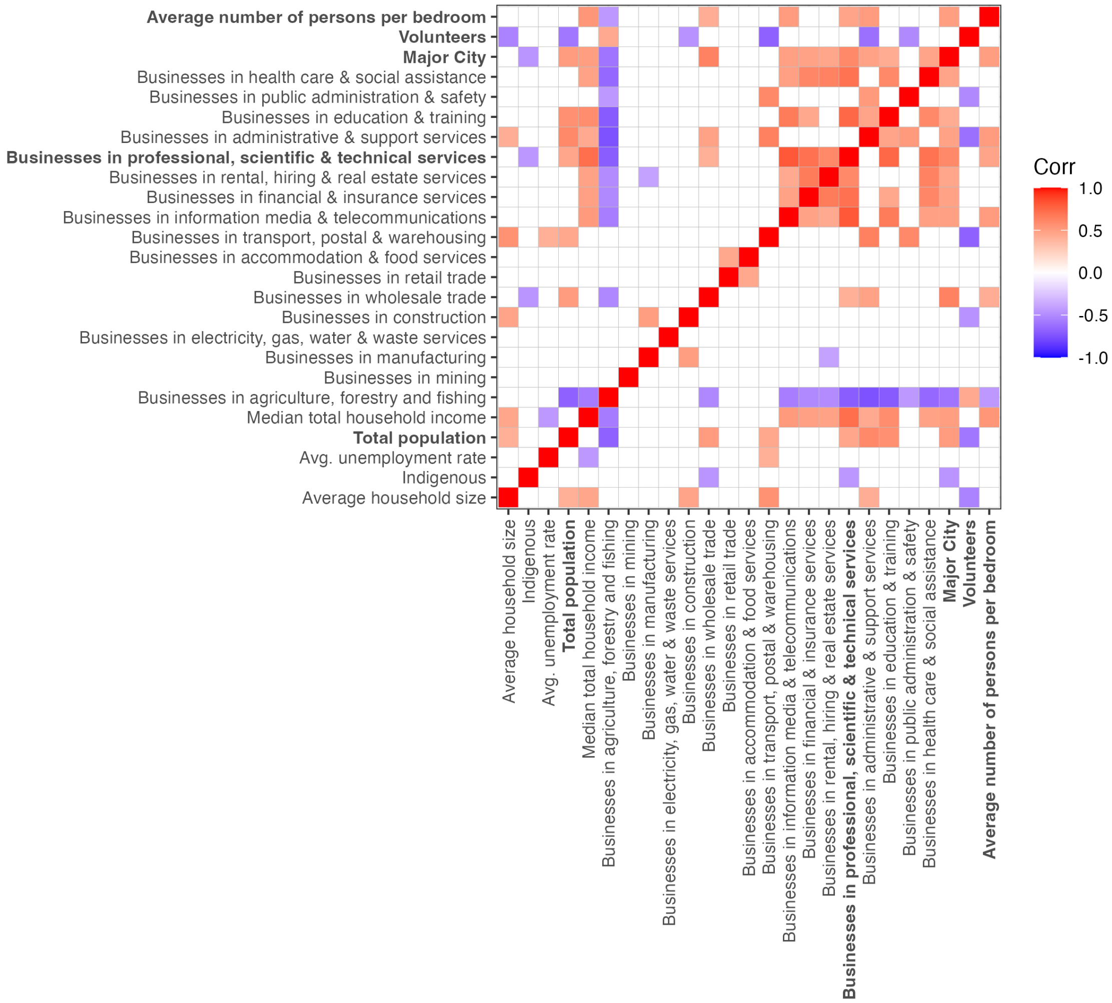
Notes: Pairwise correlation coefficients in the range of -0.4 to 0.4 are blanked out. The Top-5 variables selected by Lasso are highlighted in bold.

Figure 5.4: Correlation of Top-5 variables selected by Lasso with variables in column 3 of Table 2.6



Notes: Pairwise correlation coefficients in the range of -0.4 to 0.4 are blanked out. The Top-5 variables selected by Lasso are highlighted in bold.

Figure 5.5: Correlation of Top-5 variables selected by Lasso with variables in column 5 of Table 2.6



Notes: Pairwise correlation coefficients in the range of -0.4 to 0.4 are blanked out. The Top-5 variables selected by Lasso are highlighted in bold.

Bibliography

ACCC (2018). Digital platforms inquiry. Technical report, Australian Competition and Consumer Commission.

Besley, T. and Burgess, R. (2002). The political economy of government responsiveness: Theory and evidence from india. *Quarterly Journal of Economics*, 117(4):1415–1451.

Bhuller, M., Havnes, T., McCauley, J., and Mogstad, M. (2023). How the internet changed the market for print media. *American Economic Journal: Applied Economics*, forthcoming.

Dickson, G. (2023). Australian news data report: June 2023. Technical report, Public Interest Journalism Initiative.

Djourelova, M, R. D. and Martin, G. (2023). The impact of online competition on local newspapers: Evidence from the introduction of craigslist. *Review of Economic Studies*, forthcoming.

Friedman, J., Tibshirani, R., and Hastie, T. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22.

Kuhn, M. and Wickham, H. (2020). *Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles.*

Magasic, M., H. K. . F. J. (2023). Examining the social, civic, and political impact of local newspaper closure in outback australia. *Media and Communication*, 11(13):404–413.

Molnar, C. (2022). *Interpretable Machine Learning*. 2 edition.

Simons, M. and Dickson, G. (2019). Availability of local news and information. Technical report, Public Interest Journalism Initiative.

Snyder, J. M. and Strömberg, D. (2010). Press coverage and political accountability. *Journal of Political Economy*, 118(2):355–408.