**Supplementary Information**

**Data Appendix**

This Data Appendix provides a more technical discussion of aspects of the research design with particular reference to the topic modelling.

**Data Collection**

The reporting on a policy issue by media organisations is responsive to the logics and norms of the media system in which it operates. In an otherwise liberal pluralist media system, we would expect a norm of balanced coverage on a controversial policy topic, such as Coal Seam Gas (CSG). In Australia, the media system combines a liberal system with some degree of polarisation (Hallin and Mancini 2004; Jones and Pusey 2010). These background considerations about the nature of the media system in Australia informed our decision to select four newspapers for this study. These four newspapers include the two national newspapers with the widest circulation in Australia and the two regional newspapers with the widest circulation in the two states (NSW and Queensland) where CSG exploration and/or production is currently underway (CSG is largely regulated by the states in Australia where different states have adopted more or less favourable policy settings towards CSG, including a moratorium in some cases) (Roy Morgan Research 2016). Further, these papers include both a Fairfax and News Corporation newspaper (the two largest commercial news organisations that have played a significant and competing role in Australia) (Griffen-Foley 2002) as well as three broadsheets and a tabloid, which helps to capture meaningful differences in reporting (depth and tone) (Connell 1998; Boykoff 2008).

The search term “coal seam gas” was selected because it is a term that has been used by all sides in the Australian debate. Other terms, such as ‘fracking’, are more problematic because they are far more commonly used when reporting on contexts outside Australia and by one particular side in the debate; as such, selecting terms such as these would undermine the contextual coherence of our data. We also searched a variety of keywords (e.g. CSG, unconventional gas, coal bed methane, fracking, hydraulic fracturing) but found that most alternatives returned a nil result. We included every news article that matched the search term except Letters to the Editor, which we are able to exclude at source by using a function within the Lexis/Nexis database. We decided to exclude Letters to the Editor from the search because they neither represent the position of the paper nor do they conform to the journalistic practices that appear elsewhere in the paper.

Lexis/Nexis was selected as a database because it excludes newswire reporting (Weaver and Bimber 2008) whose inclusion would pose a greater risk of including reporting outside our geographic area of interest. Additionally, alternative databases, such as Factiva, provide less reproducible results as the university subscriptions are highly varied (Driedger and Weimer 2015). We selected the 1 January 2008 as the start date because there was a general rise in the number of media articles that matched the search criteria after this date and a greater interest in the CSG debate more generally. We extended the date range up until the latest date possible (allowing for updates to the Lexis/Nexis database) because the debate was ongoing at the time of data collection.

**Text Pre-Processing**

The topic models were conducted on 1,789 articles. By using the utility function prepDocuments in R, English language "stop words" or very common words such as articles which do not convey significant semantic meaning were removed (using the standard TM library stop words dictionary (Feinerer and Hornik 2017)). A lower threshold for word inclusion was also set at 15 documents (the minimum number of documents a word needs to appear in order for the word to be kept within the model vocabulary). The model vocabulary contained 15806 words of which 9438 words were removed due to their low incidence (these words appeared in less than 0.6% of articles). 625 articles were completely removed during the data preparation because they contained only rare or infrequently used terms. Removing documents that only contain infrequently used terms and rare words avoids the creation of topics that only germane to one particular article. In this particular case, removed articles were typically very short and often conveyed industry press releases rather than policy concerns. The LDA topic model converged after 316 iterations.

**Topic Modelling**

Table 1 presents the semantic coherence and exclusivity results. K denotes the number of topics, semantic coherence indicates the mean level of co-occurrence in the topic words, and exclusivity refers to the level of differentiation between the topics.

**Table 1: Semantic Coherence and Exclusivity (10-20 topics)**

|  |  |  |
| --- | --- | --- |
| **K** | **Semantic Coherence** | **Exclusivity** |
| 10 | -83.94270 | 9.442638 |
| 11 | -88.3428 | 9.482298 |
| 12 | -86.13686 | 9.555781 |
| 13 | -86.21756 | 9.534682 |
| 14 | -87.25693 | 9.596924 |
| 15 | -85.10190 | 9.623557 |
| 16 | -91.18893 | 9.644231 |
| 17 | -90.42031 | 9.678385 |
| 18 | -94.41738 | 9.694039 |
| 19 | -95.05189 | 9.703864 |
| 20 | -93.77859 | 9.712201 |

Table 2 and Figure 1 present the results from the 15 topic model and the expected topic proportions for this model respectively. The six highest probability terms are listed after the topic number in Figure 1. As discussed in the main article, we used the word lists in Table 1 to label each topic and identify those topics that were most likely to cover the ‘policy-related’ issues. Topics 8, 9, 11, and 14 were identified as being “policy-related”. Topic 9 was selected because it was a policy-related topic that also exhibited the highest expected topic proportions (see Figure 1). Topics 8, 11, and 14 all exhibited lower expected topic proportions than Topic 9. Moreover, the mining tax (Topic 14), was repealed in 2014. This would have had the effect of limiting our observation period compared with the selected topic. These factors, plus those identified in the main article, informed our selection of the risk and regulation topic.

**Table 2: Topic Model Results (15 topic)**

|  |  |
| --- | --- |
| **Topic** | **Topic Name** |
| Topic 1 Top Words:Highest Prob: lng, project, gas, said, arrow, santo, shell FREX: curti, lng, shell, petrochina, knox, aplng, gladston  | CSG Projects |
| Topic 2 Top Words:Highest Prob: million, billion, bank, asset, per, cent, interest FREX: chinalco, asciano, rio, feb, kirin, debt, martab  | Investment in CSG |
| Topic 3 Top Words:Highest Prob: compani, per, year, cent, share, fund, manag FREX: mig, acorn, fye, damato, dyno, hosk, consmin  | Industry Financial Situation |
| Topic 4 Top Words:Highest Prob: compani, oil, year, said, sector, develop, manag FREX: oil, engin, barrel, opportun, servic, space, contractor  | Oil |
| Topic 5 Top Words:Highest Prob: queensland, water, coal, basin, said, will, mine FREX: chinchilla, down, underground, surat, toowoomba, dalbi, basin  | Coal Mining Effects on Water |
| Topic 6 Top Words:Highest Prob: gas, coal, seam, price, queensland, coalseam, australia FREX: domest, gas, methan, coalseam, shale, east, suppli  | Domestic Supply |
| Topic 7 Top Words:Highest Prob: per, cent, share, million, market, price, stock FREX: rose, fullyear, lihir, index, fell, wds, inflat  | Company Stocks |
| Topic 8 Top Words:Highest Prob: energi, power, emiss, carbon, electr, coal, cent FREX: greenhous, emiss, solar, electr, carbon, coalfir, fuel  | Greenhouse Effect |
| Topic 9 Top Words:Highest Prob: said, gas, coal, seam, govern, nsw, water FREX: frack, pilliga, buckingham, hydraul, okan, epa, protest  | Risk/Regulation |
| Topic 10 Top Words:Highest Prob: peopl, one, labor, parti, year, say, govern FREX: pratzki, democraci, letter, daytim, children, email, compendium  | Labor Proposed Ban |
| Topic 11 Top Words:Highest Prob: will, australia, govern, invest, new, industri, state FREX: economi, econom, job, crisi, boom, import, benefit  | Jobs/Economic Benefits |
| Topic 12 Top Words:Highest Prob: yesterday, year, million, group, compani, one, will FREX: diarist, bizzel, criterion, suncorp, reader, gud, bizzoid  | Company Investments in CSG |
| Topic 13 Top Words:Highest Prob: resourc, mine, gold, compani, per, cent, project FREX: uranium, macarthur, african, brombyrtheaustraliancomau, phosphat, gold, copper  | Other Mining |
| Topic 14 Top Words:Highest Prob: tax, govern, mine, resourc, miner, industri, compani FREX: rspt, tax, super, prrt, mrrt, swan, rudd  | Mining Tax |
| Topic 15 Top Words:Highest Prob: origin, share, per, bid, offer, cent, arrow FREX: bid, origin, bgs, chapman, sunshin, qgc, offer  | Company Merger |

**Figure 1: Expected Topic Proportions: Highest Probability Topic Terms**



**Problem Windows**

Figures 2a-2d indicate the daily topic prevalence within the four selected problem windows. The y-axis denotes the summated topic prevalence across all articles on those dates.

**Figure 2a: Expected Topic Prevalence Problem Window 1**



**Figure 2b: Expected Topic Prevalence Problem Window 2**



**Figure 2c: Expected Topic Prevalence Problem Window 3**



**Figure 2d: Expected Topic Prevalence Problem Window 4**



**References**

Boykoff M. T. 2008. The Cultural Politics of Climate Change Discourse in UK Tabloids. *Political Geography* 27(5): 549–569.

Connell I. 1998. Mistaken Identities: Tabloid and Broadsheet News Discourse. *Javnost - The Public* 5, 11–31.

Driedger S. M. and Weimer J. 2015. Factiva and Canadian Newsstand Major Dailies: Comparing Retrieval Reliability between Academic Institutions. *Online Information Review* 39(3): 346–359.

Feinerer I. and Hornik K. 2017. *Tm: Text Mining Package (Version 0.7-3).* https://CRAN.R-project.org/package=tm (29 December 2017).

Griffen-Foley B. 2002. The Fairfax, Murdoch and Packer Dynasties in Twentieth-Century Australia. *Media History* 8(1): 89–102.

Hallin D. C. and Mancini P. 2004. *Comparing Media Systems: Three Models of Media and Politics*. New York: Cambridge University Press.

Jones P. K. and Pusey M. 2010. Political Communication and ‘media System’: The Australian Canary. *Media, Culture & Society* 32(3): 451–471.

Roy Morgan Research. 2016. Australian Newspaper Readership, 12 Months to September 2016. *Roy Morgan*. http://www.roymorgan.com/industries/media/readership/newspaper-readership (accessed December 19, 2016).

Weaver D. A. and Bimber B. 2008. Finding News Stories: A Comparison of Searches Using LexisNexis and Google News. *Journalism & Mass Communication Quarterly* 85(3): 515–530.