**Online Appendix**

As we note in the main text, selecting the appropriate number of topics is an open question in natural language processing (NLP) research. Here, we rely on a number of criteria to evaluate model fit across different specifications of the number of topics. In Figure A1, we plot a series of diagnostic values for models estimated with 4, 6, 8, and 10 topics, respectively. By way of interpretation, higher values of held-out likelihood, semantic coherence, and the lower bound are considered better, while lower values for residuals are considered better. There is a tradeoff, however, in that most models can explain more variation by (i.e., perform better on tasks like reducing residuals) by increasing the number of topics, while they are also likely to perform worse on other criteria (i.e., the semantic coherence of the estimated topics) by virtue of the increase in the number of topics. Therefore, the choice is a tradeoff between fit and interpretability. Here, we select 6 topics, given the strong performance on held-out likelihood, and the associated improvement in other criteria.



***Figure A1:*** *Diagnostic Plot for 6 Topic STM Model of* The Advocate’s *Articles*