Model Appendix:
Ideological Segregation Online and Offline

Matthew Gentzkow Jesse M. Shapiro
Chicago Booth and NBER

January 2011

Model and Counterfactual Experiments

The analysis in the paper suggests that (i) if news sites were less vertically differentiated, segregation would increase considerably, and (ii) if users did not visit multiple sites, more users would have extremely high or low conservative exposure. To make these hypotheses more precise, we specify and estimate a simple model of Internet news demand, and use it to simulate segregation and conservative exposure under various counterfactual assumptions.

We assume that the number of days in the year on which consumer i will visit at least one news site is drawn from a discrete distribution η. On each day consumer i visits at least one news site, the number of visits she makes, $\kappa_i \in \{1,...K\}$, is drawn from a discrete distribution π. The number of days is drawn independently across consumers and the number of visits is drawn independently across consumers and days.

Consumer i's utility from choosing online outlet j on visit k is

$$u_{ijk} = \alpha_j + (2c_i - 1)\gamma_j + \varepsilon_{ijk}$$

where α_j is outlet j's quality, γ_j is outlet j's ideology, c_i is a dummy equal to 1 if consumer i is a conservative and equal to 0 if i is a liberal, and ε_{ijk} is a type-I extreme value error, drawn independently across consumers, sites, and visits.

Consumer i chooses website j on visit k if and only if $u_{ijk} \geq u_{irk} \forall r \neq j$. The probability that a specific visit by i is to site j is therefore

$$p_{ij} = \frac{exp\left[\alpha_j + (2c_i - 1)\gamma_j\right]}{\sum_{r=1}^{J_{web}} exp\left[\alpha_r + (2c_i - 1)\gamma_r\right]}.$$

where J_{web} is the number of online outlets. Observe that the probability of visiting site j on the k-th visit is independent of the sites chosen on visits before or after k.

We estimate the parameters of this model in two steps. First, we estimate the distributions of visit-days per year (η) and of daily visits (π) nonparametrically from the comScore micro data. (We set $K = 10$ and topcode the small share of cases in which a consumer makes more than 10 news site visits on a given day.)

Second, we estimate the remaining parameters $\{\alpha_j\}$ and $\{\gamma_j\}$ by GMM, fitting to $\hat{cons}_j$ and $\hat{lib}_j$ as defined in section the paper and treating the overall size of the conservative and liberal populations as parameters. We then simulate choices from the model for 1.2 million consumers over one year.

Simulated aggregate site size and share conservative match their empirical analogues almost perfectly, with differences that are plausibly attributable to simulation error. We report estimated parameters and standard errors in table 3.

To illustrate the importance of vertical differentiation, figure 1 plots for each Internet news site j the average utility of conservatives ($\alpha_j + \gamma_j$) against the average utility of liberals ($\alpha_j - \gamma_j$). We can think of vertical differentiation as driving positive correlation in this figure, leading points to lie close to the 45-degree line. We can think of horizontal (political) differentiation as driving negative correlation, leading to dispersion around the 45-degree line. The figure shows that the vertical force is dominant.

To see the role of multiple visiting in the model, we can compute the relationship between the conservative exposure of a site's average daily visitor and the share conservative on the site. The relationship in the simulated data, like the one in the observed data shown in the paper, is much flatter than the 45-degree line. This confirms that multiple visits per consumer naturally generate a pattern of many extreme sites but few users with extreme news diets.

Finally, table 1 presents counterfactuals that illustrate the quantitative importance of these economic forces. For each counterfactual, we report the isolation index, the share of conservatives whose news diet is at least as extreme as someone who only visits foxnews.com, and the share of liberals whose news diet is at least as extreme as someone who only visits nytimes.com.

In the first row, we repeat the estimates reported earlier for the observed data. In the second row, we report the simulated analogues of these estimates.

The next two rows present counterfactual experiments designed to illustrate the role of vertical differentiation and multiple visiting, respectively.

First, we simulate a world in which all sites are equal in size, by adjusting the $\{\alpha_j\}$ quality parameters so that the number of unique visitors for all sites is equal to the unique visitors of the median observed site. The isolation index increases to 17 percentage points. Eliminating vertical differentiation dramatically increases the extent of segregation. A simple economic intuition for this result is that the low fixed costs of operating news sites online permits niche outlets to survive. This explanation predicts that eliminating vertical differentiation would matter much less for media such as television with higher fixed costs of operation. In table 2, we show that a version of our model estimated on offline media predicts little increase in segregation due to eliminating vertical differentiation, except in the case of news magazines, which arguably have lower fixed costs and a "longer tail" than television and daily newspapers.

Second, we simulate a world in which each user is allowed to visit one and only one site during the year. As we would expect, this does not cause a large change in the isolation index, because it does not change the aggregate size and ideological compositions of websites. It does, however, significantly increase the share of liberals and conservatives with relatively extreme news diets. The share of conservatives whose average site is as extreme as foxnews.com increases from 0.01 to 0.06, while the share of liberals whose average site is as extreme as nytimes.com increases from 0.02 to 0.10. Multiple visits thus limit the extent of segregation

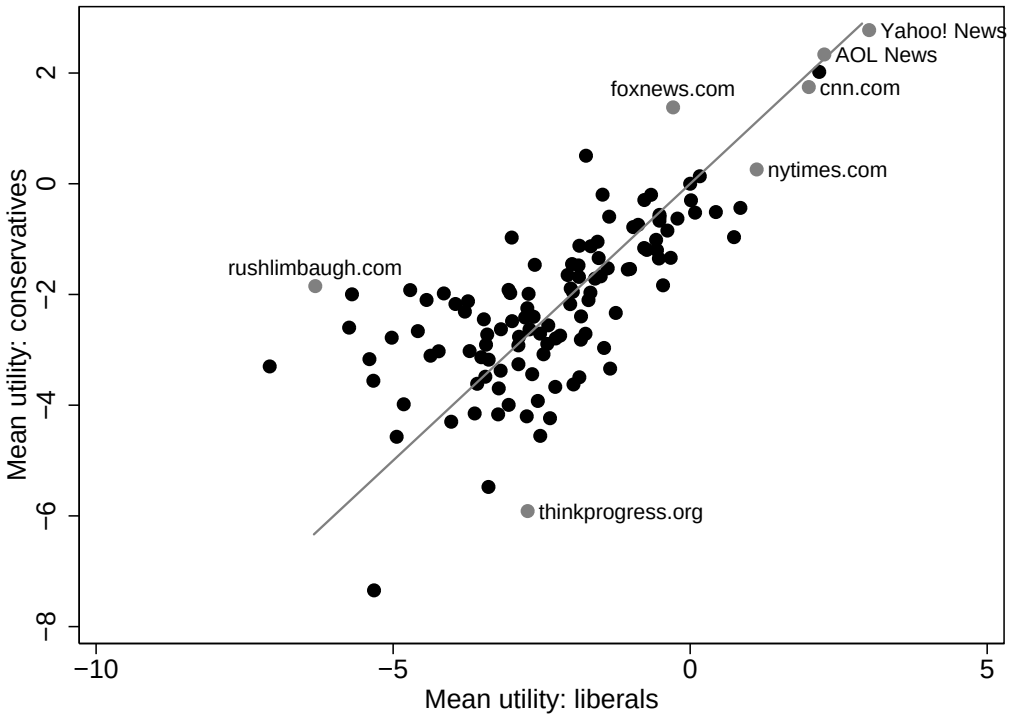
in the tails of the distribution of conservative exposure.

Finally, we can conduct an experiment to test Mullainathan and Shleifer's (2005) hypothesis that increasing the size of the choice set increases polarization. We compare the current set of Internet news sites to a hypothetical world in which only the top 10 sites are available. As predicted, the isolation index falls in this case, due to the fact that smaller sites tend to be more polarized. However, the quantitative impact of this change is small: isolation falls from 7.5 to 6.3 percentage points.

References

Mullainathan, Sendhil and Andrei Shleifer. 2005. The market for news. *American Economics Review* 95, no. 4, 1031-1053.

Figure 1: Estimated Site Utility: Conservatives vs. Liberals



Notes: Figure plots estimated mean utility for conservatives $(\alpha_j + \gamma_j)$ on the y-axis against estimated mean utility for liberals $(\alpha_j - \gamma_j)$ on the x-axis. The unit of observation is an individual Internet news site. See text for details.

Table 1: Model Predictions and Counterfactual Experiments

	Isolation Index	Share with Conservative Exposure	
		Left of nytimes.com	Right of foxnews.com
Observed	.075	.039	.013
Baseline Model	.075	.022	.010
	(.001)	(.002)	(.001)
Counterfactual Experiment			
All Sites the Same Size	.166	.045	.019
	(.002)	(.002)	(.001)
All Users Visit One Site Only	.070	.104	.061
	(.010)	(.004)	(.003)
Eliminate all but Top 10 Sites	.063	.019	.012
	(.001)	(.002)	(.001)

Notes: Data for observed moments are from comScore. Data for model and counterfactual experiments are produced by simulation, with statistics computed to match empirical analogues. Share with conservative exposure to the left of nytimes.com is the share of users whose conservative exposure is less than a user who visits nytimes.com exclusively. Share with conservative exposure to the right of foxnews.com is the share of users whose conservative exposure is greater than a user who visits foxnews.com exclusively. Standard errors in parentheses are from a parametric bootstrap with 100 replications. At each replication of the parametric bootstrap, we simulate a sample of $12,000$ users for one year (the approximate size of the Plan Metrix sample) and recompute the statistics reported. In the "all sites the same size" counterfactual experiment we adjust each site's α_j parameter so that all sites have the same number of daily unique visitors as the median site. In the "all users visit one site only" counterfactual experiment we restrict all users to visit at most one site during the year. In the "eliminate all but top 10 sites" counterfactual we remove all but the top 10 sites from the choice set.

Table 2: Model Predictions and Counterfactual Experiments: Offline Media

Isolation Index	Baseline Model	Counterfactual Experiment: All Outlets the Same Size
Offline Media		
Broadcast_News	0.022	0.040
Cable	0.035	0.025
Magazines	0.040	0.092
National_Newspapers	0.114	0.117

Notes: The table presents simulated moments from a model in which a population of consumers make one daily visit to an outlet in each listed medium with choice probabilities given by the model specified in the model section of the paper. The model is estimated via GMM and behavior is simulated for 12,000 consumers on a single day, both for the estimated parameters and for parameters adjusted so that all outlets have the same number of daily visitors as the median outlet in the medium.

Table 3: Parameters of Structural Model

	α_j	γ_j
ABC News	6.3334	.0318
	(0.0000)	(0.0000)
AOL News	8.6302	.0703
	(.0152)	(.0162)
BBC News	6.2205	-.8198
	(.0288)	(.0276)
Canoe	4.5106	-.1121
	(.0449)	(.0401)
Capitol Advantage	3.8142	.1482
	(.0613)	(.0675)
Sina News	3.2839	-.3587
	(.0727)	(.0734)
Sky News	3.0085	.2227
	(.0867)	(.0866)
The Mail Online	5.7190	-.1996
	(.0278)	(.0261)
Yahoo! News	9.2278	-.0870
	(.0153)	(.0149)
about.com news & issues	4.4782	.2387
	(.0449)	(.0423)
aclj.org	2.4312	1.1531
	(.1786)	(.1809)
aclu.org	2.1727	-.1070
	(.1171)	(.1338)
alternet.org	2.8573	-.6935
	(.1068)	(.1026)
ap.org	5.3655	-.1617
	(.0321)	(.0310)
australian broadcasting corp.	3.5587	-.2743
	(.0688)	(.0565)
azcentral.com	5.0577	-.2322
	(.0328)	(.0301)
barackobama.com	4.0052	-.4566
	(.0596)	(.0537)
billoreilly.com	1.1446	1.9206
	(.4494)	(.4432)

	α_j	γ_j
blackamericaweb.com	3.0898	-.6481
	(.0858)	(.0917)
blackvoices.com	5.5247	.0985
	(.0282)	(.0294)
blogcritics.org	2.7945	-.9839
	(.1130)	(.1217)
blogtalkradio.com	3.5384	-.7991
	(.0683)	(.0647)
boston.com	5.7941	.0083
	(.0261)	(.0264)
bostonherald.com	5.0341	-.2177
	(.0318)	(.0347)
breitbart.com	4.8412	.4039
	(.0362)	(.0403)
businessweek.com	5.0303	.2872
	(.0346)	(.0427)
bvblackspin.com	3.9903	-.9644
	(.0666)	(.0567)
cagle.com	1.5776	.2164
	(.1788)	(.1878)
canada.com	4.6563	.2337
	(.0422)	(.0459)
capitolconnect.com	3.0486	.1384
	(.0897)	(.0826)
cartoonstock.com	3.8387	.2782
	(.0585)	(.0623)
cbc.ca	4.2168	-.2482
	(.0534)	(.0466)
cbn.org	3.4268	.3101
	(.0686)	(.0776)
cbsnews.com	5.4628	.1189
	(.0260)	(.0303)
chicagotribune.com	5.4541	-.2897
	(.0273)	(.0290)
chron.com	4.9329	.3030
	(.0377)	(.0419)
cnbc.com	5.7771	-.0096
	(.0220)	(.0244)

	α_j	γ_j
cnn.com	8.2052	-.0935
	(.0166)	(.0163)
cnsnews.com	2.0498	1.1470
	(.1983)	(.2001)
csmonitor.com	3.6838	-.2127
	(.0598)	(.0677)
ctv.ca	2.7035	.6352
	(.1198)	(.1095)
dailykos.com	3.3633	-.6679
	(.0735)	(.0761)
democraticunderground.com	3.8677	-.2457
	(.0538)	(.0601)
drudgereport.com	5.7098	1.1617
	(.0345)	(.0349)
economist.com	4.2366	-.0461
	(.0478)	(.0491)
foxnews.com	6.8796	.8651
	(.0210)	(.0206)
freerepublic.com	3.8462	.6015
	(.0621)	(.0648)
ft.com	3.8321	.5587
	(.0689)	(.0661)
gallup.com	2.9659	.3764
	(.0861)	(.0922)
glennbeck.com	2.4883	1.8794
	(.2134)	(.2206)
gop.com	1.9306	.4505
	(.1683)	(.1555)
gopusa.com	1.8878	.9185
	(.1751)	(.1648)
heraldtribune.com	2.8664	.0125
	(.0989)	(.0944)
heritage.org	2.5950	.6639
	(.1107)	(.1228)
hotair.com	3.5111	.0901
	(.0680)	(.0685)
huffingtonpost.com	6.5378	-.6094
	(.0229)	(.0202)

	α_j	γ_j
humanevents.com	2.1638	1.6025
	(.2239)	(.2294)
latimes.com	6.1138	-.2707
	(.0223)	(.0228)
metafilter.com	4.7446	-.0517
	(.0393)	(.0368)
michellemalkin.com	2.7105	.9924
	(.1265)	(.1325)
moveon.org	3.0363	-.9064
	(.0924)	(.0985)
msnbc.com	8.4297	-.0446
	(.0158)	(.0160)
myfoxla.com	3.0498	-.0616
	(.0924)	(.0717)
nationalreview.com	3.2715	.9212
	(.0969)	(.0890)
newsbusters.org	3.3727	.5437
	(.0750)	(.0775)
newsmax.com	4.3453	1.0474
	(.0586)	(.0552)
newsobserver.com	3.5933	.2893
	(.0652)	(.0688)
newsrunner.com	0.0000	-.9799
	(.4372)	(.4404)
newsvine.com	4.1257	-.7267
	(.0493)	(.0496)
newsweek.com	5.5404	-.1885
	(.0231)	(.0286)
newyorker.com	2.7341	.0173
	(.0897)	(.0999)
npr.org	5.1877	-.6576
	(.0315)	(.0345)
nydailynews.com	5.8009	.2705
	(.0242)	(.0270)
nypost.com	5.4977	.6702
	(.0302)	(.0328)
nytimes.com	7.0225	-.3994
	(.0168)	(.0182)

	α_j	γ_j
pbs.org	5.3545	.4158
	(.0343)	(.0308)
philly.com	4.6139	.3004
	(.0398)	(.0461)
politico.com	4.8922	.1305
	(.0336)	(.0332)
politicsdaily.com	5.5006	-.4737
	(.0315)	(.0273)
politifact.com	3.2581	-.1531
	(.0697)	(.0725)
postchronicle.com	4.8765	-.0378
	(.0338)	(.0327)
propeller.com	3.7361	.2125
	(.0581)	(.0621)
realclearpolitics.com	3.2694	1.1144
	(.0947)	(.0932)
reddit.com	3.6528	-.7839
	(.0642)	(.0643)
rense.com	3.1616	.2958
	(.0875)	(.0789)
reuters.com	5.7413	-.0432
	(.0237)	(.0251)
rollingstone.com	4.3820	.0911
	(.0472)	(.0423)
rushlimbaugh.com	2.2538	2.2625
	(.3023)	(.3053)
salon.com	4.5407	-.5091
	(.0383)	(.0454)
sfgate.com	5.3949	-.3802
	(.0290)	(.0306)
slate.com	5.3712	-.2011
	(.0321)	(.0289)
stuff.co.nz	2.6342	-.4353
	(.1110)	(.1096)
tampabay.com	4.3698	.0434
	(.0434)	(.0452)
tbo.com	4.4267	-.1634
	(.0502)	(.0456)

	α_j	γ_j
technorati.com	4.0975	-.4426
	(.0542)	(.0514)
theatlantic.com	3.6661	.0689
	(.0622)	(.0594)
theglobeandmail.com	3.2823	.7728
	(.0929)	(.0819)
thehill.com	3.2635	.3781
	(.0871)	(.0791)
thenation.com	2.4449	-.2294
	(.1227)	(.1106)
theolympian.com	1.8973	-1.0093
	(.1644)	(.1755)
thestate.com	3.0649	1.2005
	(.1213)	(.1142)
thinkprogress.org	2.0092	-1.5568
	(.1984)	(.1947)
time.com	5.9117	-.1761
	(.0246)	(.0239)
today.com	3.4280	.0176
	(.0686)	(.0643)
topix.com	6.1910	-.1250
	(.0197)	(.0234)
topnews.in	3.4019	.8399
	(.0899)	(.0933)
townhall.com	3.0164	1.4272
	(.1448)	(.1456)
treehugger.com	3.8042	-.2337
	(.0580)	(.0558)
upi.com	3.7166	-.0588
	(.0646)	(.0657)
usatoday.com	6.4822	.0183
	(.0235)	(.0216)
usnews.com	4.5561	.1261
	(.0441)	(.0420)
villagevoice.com	2.8742	-.2060
	(.0907)	(.0955)
voanews.com	3.8603	-.0529
	(.0582)	(.0558)

	α_j	γ_j
washingtonpost.com	6.2941	-.4405
	(.0195)	(.0210)
washingtontimes.com	3.9793	.3973
	(.0605)	(.0570)
whitehouse.gov	4.6761	-.0246
	(.0361)	(.0375)
wn.com	2.8085	-.4382
	(.0888)	(.0940)
wnd.com	4.2932	.6072
	(.0500)	(.0517)
wsj.com	5.9043	.2610
	(.0250)	(.0226)

Note: The table presents GMM estimates of model parameters with standard errors in parentheses obtained via a parametric bootstrap with 100 replications. At each replication of the parametric bootstrap, we simulate a sample of $12,000$ users for one year (the approximate size of the Plan Metrix sample) and re-estimate the model. Parameters are normalized so that the smallest α_j and the size-weighted mean of the γ_js are 0.