Google Politics: Online Appendix

1 Robustness Checks

Table 1 contains additional robustness checks. Models 6 and 7 present specifications of model 4 using robust standard errors, clustered on countries, as an alternative to the random intercepts we include in model 4. Results are substantively identical across the two methods. Models 8 and 9 present random effects models that use a dummy indicating single member district systems as an alternate measure of electoral system. These models generate identical substantive conclusions to models that include logged district magnitude albeit, with a slightly lower level of statistical significance for the coefficient for electoral system (p-value = 0.08 and 0.05, respectively). Given that an SMD dummy is a significantly less precise operationalization of our concept, this is not surprising. Model 10 uses random effects and an alternate events variable, taken from the Cross-National Time-Series data archive, to measure instability. We use their domestic9 variable, which includes assassinations, general strikes, guerrilla warfare, major government crises, purges, riots, revolutions and anti-government demonstrations and weights those events by their severity to create an index variable (Banks & Wilson 2013). We log this variable due to its skewness. Model results are identical to those using other measures. Model 11, finally, is a random effects specification that includes dummy variables indicating the basis of national legal systems: civil, common, religious, customary, and hybrid law systems, coded from the CIA World Factbook (*CIA World Factbook: Legal Systems* 2015). We coded hybrid systems as combinations of pure types. Civil law systems, the most common, are the baseline category in our analysis. It is conceivable that variation in the use of private points of control reflects variations in legal culture that may covary with our core independent variables. Yet none of the particular types of legal system have a statistically significant effect on number of takedown attempts and other results are fully robust to their inclusion.

While section 3 of this appendix contains our substantive justification for sampling only democ-

Table 1: Predicting Google Content Removal Requests: Robustness Checks

	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Log Patents Per Capita	0.45*	0.67*	0.42	0.42	0.49*	0.46*
	(0.22)	(0.21)	(0.23)	(0.23)	(0.22)	(0.22)
Log Terror Incidents	0.71*		0.31*			0.29*
	(0.20)		(0.11)			(0.11)
WGI Stability/Violence		-1.56*		-0.81*		
		(0.53)		(0.39)		
Log CNTS Instability Index					0.07*	
					(0.04)	
Time to Start a Business	0.05*	0.04*	0.03*	0.02*	0.03*	0.03*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Log District Magnitude	-0.35*	-0.53*			-0.40*	-0.40*
	(0.18)	(0.18)			(0.18)	(0.18)
Single Member District Dummy			1.17	1.32		
			(0.67)	(0.68)		
Log GDP Per Capita	1.12*	1.45*	0.12	0.34	0.18	0.10
	(0.56)	(0.52)	(0.49)	(0.52)	(0.49)	(0.50)
Internet Users Per Capita	-1.94	-3.52	2.00	2.27	0.79	1.62
	(2.62)	(2.61)	(2.08)	(2.20)	(2.05)	(2.06)
Google Search Share	-0.01	0.02	-0.01	-0.01	0.00	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Common Law Legal System						-0.05
						(0.63)
Religious Legal System						0.27
						(1.05)
Customary Legal System						-1.06
						(0.69)
Intercept	-3.59	-4.93	2.14	0.11	3.03	4.05
	(4.46)	(4.02)	(5.11)	(5.46)	(5.05)	(5.23)

The dependent variable is content requests issued. The observation level is country-half-year. We estimated all models using negative binomial regression with dummies for half year time period (not shown). We used chained multiple imputation for missing data; N=322.

* p-value less than .05

racies that sent at least single request in our main models, we recognize the possibility that our statistical specifications could be sensitive to that choice. In order to address concerns about our sampling criteria, we ran a zero inflated model with random effects on a larger dataset containing all democracies, even those that never sent a takedown request, and present it in Table 2, model 12. Zero inflated negative binomial regressions are mixture models that allow different processes to generate zero count observations (logit process) and positive count observations (negative binomial

Table 2: Predicting Google Content Removal Requests: Zero Inflated Model

	Model 12	
	Logit (Zero) Stage	Negative Binomial
Log Patents Per Capita	-0.83*	0.32
	(0.44)	(0.20)
Log Terror Incidents	-0.95*	0.27*
	(0.28)	(0.10)
Time to Start a Business	0.01	0.02*
	(0.01)	(0.01)
Log District Magnitude	-0.09	-0.27
	(0.31)	(0.15)
Log GDP Per Capita	-0.24	0.19
	(0.68)	(0.44)
Internet Users Per Capita	-6.94*	-0.12
	(3.43)	(2.07)
Google Search Share	-0.10	-0.02
	(0.05)	(0.01)
Intercept	8.38	4.42
	(7.57)	(4.41)

The dependent variable is content requests issued. The observation level is country-half-year. Model estimated using zero inflated negative binomial regression with dummies for half year time period (not shown). We used chained multiple imputation for missing data; N=322.

* 95% HPD interval does not contain 0

process). Positive coefficients in the logit stage indicate a negative relationship between the given independent variable and the use of Google takedown requests. Because software to fit zero inflated negative binomial models with random intercepts in the regression equations are not widely available, we use a Bayesian framework and Gibbs sampling to simulate draws from the posterior distributions of the model's parameters.[1]

The findings that we present in the main text are robust to this model of the selection process. Negative binomial coefficients remain in the hypothesized direction with a high degree of confidence. While coefficients for log of district magnitude and patent applications per capita now fall slightly

[1]We use diffuse normal priors, with mean 0 and precision 0.00001 for the regression coefficients in each equation, assume that random effect precision parameters are distributed according to gamma distributions with shape and rate parameters of 1.8 and 1, and assume a uniform distribution between 0.2 and 501 for the dispersion parameter in the negative binomial portion of the model. Results are robust to varying these priors. We ran the four chains of the Gibbs sampler for 20,000 burn-in iterations each and then sampled every 150th draw from each chain for 150,000 subsequent iterations. We repeated this process for each of the 20 multiply imputed datasets (described below) and combined posterior information across those 20 runs to draw inferences.

outside the 0.95 highest posterior density (HPD) interval, their intervals are, nevertheless, extremely close to the 0.95 and 0.90 levels respectively. Furthermore, considering the issue from a Bayesian perspective, the posterior probability that the coefficient for patents per capita is greater than zero is 0.95 while the probability that the coefficient for log of district magnitude is less than zero is 0.96. Finally, examining the first stage coefficients (predicting 0s) more closely, it appears that countries with low internet connectivity, and those that do not use Google services at high rates are likely to send zero requests. Moreover, both of the demand factors identified by our theory—internal unrest and patent production—are strong predictors of the use of Google takedown requests. Thus, they both predict the use of private points of control (logit stage) and the extent of that use (negative binomial stage).

2 Descriptive Statistics

Table 3: Descriptive Statistics

	Mean	Std Deviation	N
Content Removal Requests	23.02	59.50	286*
Internet Users Per Capita	0.59	0.24	322
Log GDP Per Capita	9.39	1.17	310
Google Search Share	92.70	9.26	322
Log Terror Incidents	0.96	1.45	280
Log Patents Per Capita	-8.92	1.42	288
Time to Start a Business	20.03	19.67	319
Log District Magnitude	2.00	1.45	316

* 36 additional observations are described as 1–10 requests.

3 Sample Selection

Our core analysis focuses on democracies which lodged at least one takedown request between July 2009 and June 2012. Therefore the sample contains repeated observations of 53 democracies spanning 6 half-year time periods. Figure 1 displays aggregate variation in Google requests, within our sample, over time.

Our decision to limit our core analysis to countries that actually lodged a Google takedown

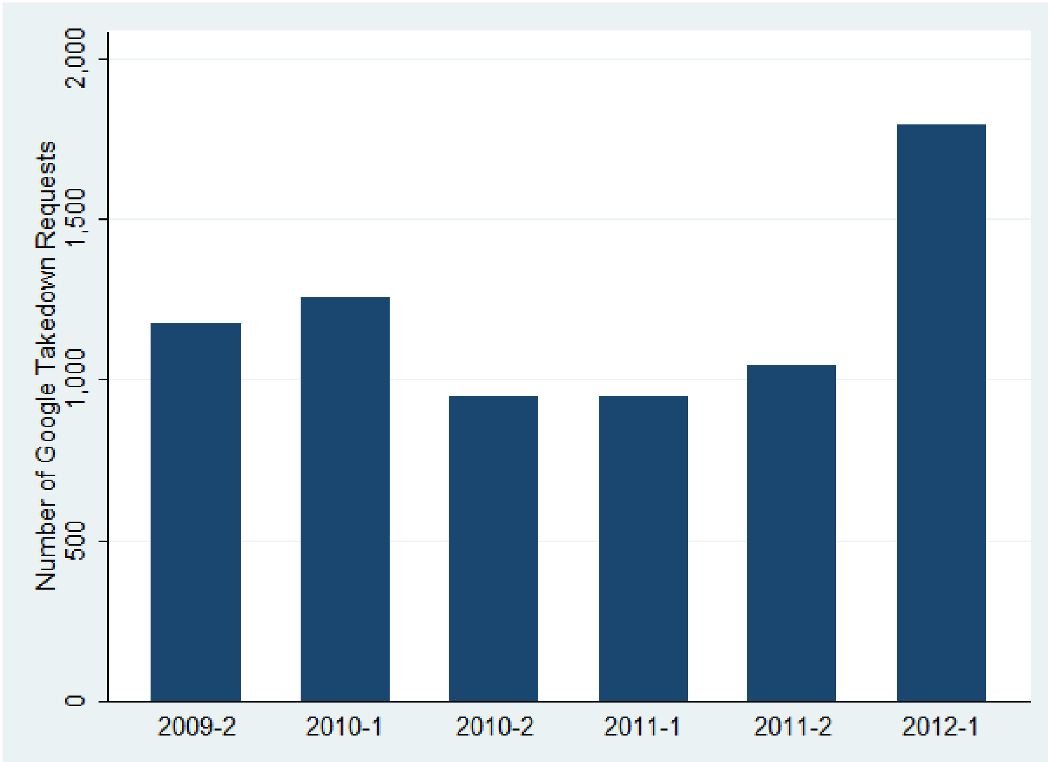
Table 4: States in Sample

Argentina	Australia	Austria	Belgium
Bolivia	Bosnia & Herzegovina	Brazil	Canada
Colombia	Croatia	Cyprus	Czech Republic
Denmark	Estonia	Finland	France
Germany	Greece	Hungary	India
Indonesia	Ireland	Israel	Italy
Japan	Lithuania	Macedonia	Malaysia
Mauritius	Mexico	Netherlands	New Zealand
Norway	Panama	Peru	Philippines
Poland	Portugal	Slovak Republic	Slovenia
Solomon Islands	South Africa	South Korea	Spain
Sri Lanka	Sweden	Switzerland	Thailand
Trinidad & Tobago	Turkey	Ukraine	United Kingdom
United States			

request during the observation period has important implications for how one should interpret our findings. In particular, our the results in the main text tell us little about whether or not a particular country uses private points of control—and leans on Google in particular—in order to regulate the digital sphere. Rather, given that a country does employ this particular mechanism, we ask how intensively they use this resource. The decision to use private points of control, and to target Google in particular, is a potentially interesting one, and our robustness checks show that it is driven both by internet use and demand factors. Yet, because the results of the zero-inflated negative binomial model mirror those for the reduced sample, we limit our discussion in the main text to the simpler model and focus our attention on countries that use private points of control in the first place.

Our second key sampling decision is to focus exclusively on democracies. We measure democracy using Unified Democracy Scores (UDS) (Pemstein, Meserve & Melton 2010). These scores use a Bayesian measurement model to aggregate democracy ratings by multiple scholars into a single composite scale and provide a mapping between the composite scale and included raters' original metrics. In particular, Przeworski, Alvarez, Cheibub & Limongi (2000) are a constituent measure of the UDS, and classify cases along an autocracy/democracy dichotomy. The Bayesian framework employed by the UDS allows us to calculate the probability that a particular case meets Przeworski et al.'s (2000) democracy threshold, while taking information provided by other democracy raters

Figure 1: Google Takedown Requests Over Time



into account. We classify cases as democracies if that probability exceeds 0.5.[2]

We focus on democracies for both theoretical and practical reasons. First, we are specifically interested in how liberal democracies regulate digital content and how politics interact with that regulatory process. Thus, democratic states represent our target population. Second, while authoritarian regimes may also make use of private points of control, they have a wider variety of tools at their disposal, often more draconian in nature, than takedown requests. Authoritarian regimes can trace individual dissidents, shut down internet infrastructure, and arrest producers of digital information with relative impunity (Deibert, Palfrey, Rohozinski & Zittrain 2008, Wright, de Souza & Brown 2011, King, Pan & Roberts 2013). By contrast, democracies are generally more constrained by domestic law and citizens' expectations. Thus, to the extent that non-democratic regimes lodge takedown requests with Google, they are likely to be generated by a different process than they are in democracies. This process is certainly interesting, but is beyond the scope of this work.

[2]Our conclusions are broadly robust to perturbing this cutoff. In particular, adopting the conventional approach of using the raw Przeworski et al. (2000) classifications yields substantively similar results.

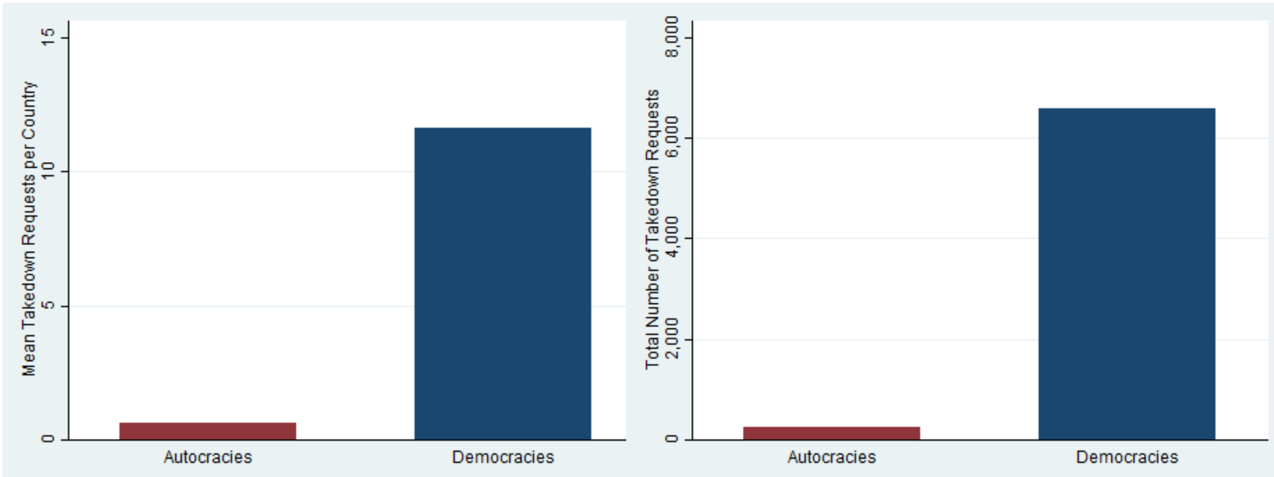


Figure 2: Google Takedown Requests By Regime Type 2009-2 to 2012-1

As a practical matter, autocracies simply do not appear to make use of private points of control with any regularity. Figure 2 shows the breakdown of Google takedown requests by type of regime, generated from a cross-national dataset of 178 countries from the 2nd half of 2009 to the end of the first half of 2012.[3] It shows the huge disparity between democracies and autocracies in both the average number of requests sent per country—nearly 12 for democracies and less than one for autocracies—and total number, 258 autocratic requests and 6583 democratic requests. The pattern fits neatly with the broader literature's emphasis on the more invasive tools that autocracies possess for censorship purposes. As a consequence, our project focuses solely on democratic use of points of control, examining the determinants of censorship only in democratic countries. While the results will thus provide a less general examination of private points of control use, this design decision allows us to focus on observable characteristics unique to democracies (i.e. electoral system design) that we argue should modulate censorship.

4 Missing Data

For two periods (2nd half 2009, 1st half 2010), Google reports observations with fewer than 10 takedown requests simply as < 10, omitting specific numbers of content requests, for case identifiability reasons (36/322 observations). Furthermore, we face limited missing data in our independent

[3]These data include countries that never made a content removal request during the observation period.

Table 5: Average Pre-Treatment Predictor Values

	Turkey	Synthetic Control	Sample Mean
Internet Users Per Capita	0.398	0.420	0.597
Log GDP Per Capita	8.965	8.960	9.512
Log Patents Per Capita	-9.976	-9.894	-8.765
Log Terror Incidents	2.233	2.232	0.981
Time to Start a Business	6.000	11.936	19.933
Log District Magnitude	1.946	1.951	2.008

Table 6: Synthetic Control Contributions (> 1%)

Country	Weight
Colombia	0.39
India	0.08
Italy	0.04
Mexico	0.03
Portugal	0.44

variables. We therefore used chained multiple imputation (m=20) to address missing data in our dataset, generating numbers between 1 and 10 for the requests variable and imputing missing values for any other variables where necessary. We included every variable incorporated in the analysis in our multiple imputation equation. We also include UDS score, population, alternative measures of district magnitude and measures of government type (parliamentary or presidential) and legislative election type (majoritarian, proportional, or mixed) to improve the quality of our imputations. We imputed most variables using linear regression, while we imputed requests with negative binomial regression and used truncated regression to impute business start times.

5 Details of the Synthetic Case-Control Study

Table 5 shows that our synthetic control closely approximates Turkey's pre-treatment covariate profile,[4] while Table 6 lists the cases that contributed substantially to the synthetic control, along with their contribution weights.

[4]We dropped Google market share from the analysis because including it caused estimation errors, likely because of uniformly high market penetration across our sample.

References

Banks, Arthur S & Kennth A. Wilson. 2013. "Cross-National Times-Series Data Archive." Databanks International. Jerusalem, Israel. http://www.databanksinternational.com.

CIA World Factbook: Legal Systems. 2015. Central Intelligence Agency. https://www.cia.gov/library/publications/the-world-factbook/fields/2100.html#download (Last Accessed 6/9/2015).

Deibert, R.J., J.G. Palfrey, R. Rohozinski & J. Zittrain. 2008. *Access Denied: The Practice and Policy of Global Internet Filtering*. Cambridge, MA: MIT Press.

King, Gary, Jennifer Pan & Margaret E. Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review* 107(02):326–343.

Pemstein, Daniel, Stephen A. Meserve & James Melton. 2010. "Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type." *Political Analysis* 18(4):426–449.

Przeworski, Adam, Michael Alvarez, José Cheibub & Fernando Limongi. 2000. *Democracy and Development: Political Regimes and Economic Well-being in the World, 1950–1990*. Cambridge: Cambridge University Press.

Wright, Joss, T de Souza & Ian Brown. 2011. Fine-grained censorship mapping information sources, legality and ethics. In *Freedom of Communications on the Internet Workshop*.