Replication and extension of How Chinese Officials Use the Internet to Construct their Public Image by Jennifer Pan

Alexander Klueber

4/7/2020

Contents

1 Abstract

Pan (2017) shows that the emphasis on Chinese local government websites on either the competence or benevolence of county executives depends on where they are in the political tenure cycle. Early tenure county executives project images of benevolence by emphasizing their attentiveness and concern toward citizens. Late tenure executives project images of competence by highlighting their achievements. These findings shift the nature of debates concerning the role of the Internet in authoritarian regimes from a focus on regime-society interactions to an examination of dynamics among regime insiders.

I was largely able to replicate the statistical models that the paper used to suggest the above relationships. My extension confirms that this is the most likely explanation for the observed effect by introducing a series of models that could support alternative explanations for the observed effect (e.g. cultural differences among regions, gender differences, etc.) and comparing the explanatory power of these models through the leave one out method. In addition I could validate the randomness of the sample selected to draw the underlying conclusions from a geographical perspective through a repeated sampling simulation and the construction of confidence intervals and their comparison with the observed provinces in the sample of 100 and the sample of 48 that was ultimately used to construct the models. This corraborates the findings of the paper by confirming the geographic randomness of the sample and the relative explanatory power of the model using political tenure cycle to explain website content.

2 Introduction

!!!! The first paragraph is a review of the paper you are replicating. Flesh out the details. Tell us about the data and the model. Place it within the relevant literature, via a key citation or two. Highlight implications and caveats. Again, it is hard to summarize a 25 page paper in a paragraph. Do your best. Note that the paper's own abstract is often a useful guide.

To replicate this paper, I used R (R Core Team 2019). Original data and code of the replication paper are available in the Harvard dataverse.[1] The code for the extension is available at my repo.[2]

In addition to replicating the resulty, I am interested to explore whether other explanations than the signalling function within authoritarian regimes may plausibly explain the alterations in competence / benevolence patterns described in the paper. I will explore these alternatives by comparing the explanatory power of the variable categories employed in the paper and extending them with a new category (culture which will include the macro-region and the county type). Alternative hypothesis therefore are: 1. The benevolence/competence patterns may be explained by regional cultural variations 2. The benevolence/competence patterns may be explained by the resources at disposal to the official 3. The benevolence/competence patterns may be explained by internal peer preferences
4. The benevolence/competence patterns may be explained by characteristics of the prefecture 5. The benevolence/competence patterns may be explained by the individual abilities of the county officials 6. The benevolence/competence patterns may be explained by the immediate career success of county officials

Proving that it is in fact not tenure, but one of the other factors that have nothing to do with the career path of the individuals within the party, would weaken that connection. This comparison between models seems relevant as in the original paper the variables that are statistically significant vary between the models. In the model with the most controls, for competence "mayor education levels" and "whether a county party secretary is in first year of office" are statistically significant, for benevolence no variables are statistically significant. Between the regressions the number of observations included also varies, which initially makes a comparison between the models and the variables in the models difficult.

This endevour is constrained by the data available in the dataverse of the replication paper. Only the pre-selected sample of 100 contains all variables relevant for the analysis on a county level. Therefore a

[1] https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/2HTWSU
[2] https://github.com/Alex1005-stack/Gov_1006_final_project

sub-division of that data by provinces will leave us with very small sample sizes. Other examples are that the sample only contains 4 female mayors, 1 person with education level 6, 4 people with education level 5, 3 people with education levle 2 and 2 people with education level 0.

Building on this, the other part of my extension is around validating the sample selection process. Initially, it seems surprising that the author has sub-selected 100 of the 2,787 counties with website, and we immediately dismiss another 29 counties in our modelling because of data inavailability and subsequently another ~23 as our regressions get more ambitious (include more variables). The absence of any website from Tibet is furthermore conspicuous.

3 Literature review

The replication paper aims to contribute to the larger discusison around role of the Internet in authoritarian regimes, especially in China. It is part of a larger shift in embracing the complexity of the role of the internet within these regimes. It moved from a prevailing assumption of the Internet's inherent democratic nature, and its ascribed power to undermine authoritarian regimes to a much more nuanced view that emphasises the utilitization of the internet in relation to how it is playing out.(Kalathil and Boas 2003)

While transparency laws are implemented by central authorities with the intention to obtain more information about the performance of local officials, they turn out to be political instruments for self-promotion among regime insiders. Research shows that the desired transparency does not come to frutition because grievances submitted through these online forums are systematically concealed from upper-level authorities when they implicate lower-tier officials or associates connected to lower-tier officialsthrough patronage ties. Information manipulation occurs primarily through omission of wrongdoingrather than censorship or falsification, suggesting that even in the digital age, in a highly determined andcapable regime where reports of corruption are actively and publicly voiced, monitoring the behavior ofregime agents remains a challenge.(Pan and Chen 2018)

There is however some disagreement whether creating transparency on local government performance is the primary function of these laws to begin with, or whether they are primarily intended as subtle instruments of online social control through information delivery, agenda setting, and containment of public dissent.(**???**)

The emphasis of local officials on the communication among insiders are also reflected by field experiments testing the responsiveness of local officials. Tattling to upper levels of government made county governments considerably more responsive to citizen's demands.(Chen, Pan, and Xu 2016) This challenges to some degree earlier findings that the capacity of the central state to monitor and control lower level agents has increased in China as it suggests that some of the efforts to do so have altered the nature of the interaction rather than shifting the power relationships within it.(Edin 2003)

This isn't necessarily something negative in of itself, as empirical findings suggest that China uses personnel control to induce desirable outcomes, especially economic gains.(Li and Zhou 2005) This alludes to an associated discussion around the relevance of image building in authoritarian regimes and the missapropriation of resources in the service of that. China is mentioned as an example of that where the political arrangements have greacted the incentive and the opportunities for irresponsible behaviour among state agents to do so.(Cai 2004)

4 Paper Review

The paper analyses how local government officials use the websites they are required to put up to meet the central government's transparency requirements, to engineer their public image. It concludes that the internet becomes a tool for self-promotion in authoritarian regimes. The analysis demonstrates that the websites tend to highlight the competence (achievements) or the benevolence (concern for citizens) of these county executives which changes based on the tenure of their cycle. The former is more important to officials that are later in their cycle, while the latter is more important to those earlier in the cycle.

The analysis does so by focusing on a random sample of 100 Chinese counties and their websites. They represent 29 Chinese provinces. Based on this sample they estimate the proportion of posts for competence and benevolence by year in office, where group refers to the year, with 95% confidence intervals in lwr and upr. Generated using Hopkins and King (2010) ReadMe software. The authors then conduct a linear regression that regresses a mayor first year or last year binary indicator on the number of competence metnions on the websites by county which they extend gradually for additional variables in alternative models. The same is then done for benevolence. These variable include: county resources and environmental factors, incentives of county party secretaries, incentives of prefecture party secretary and other prefecture-level characteristics, county executive's ability, age, and gender, post-treatment variable of whether the county executive was promoted. The authors conclude from the regression that When a county executive is in the last year of office, an additional 15% or so of website content is on average dedicated to claims of competence. Similarly they conclude that When a county executive is in the first year of office, an additional 10% to 15% of website content is on average dedicated to claims of benevolence. Both results are statistically significant at the 0.05 level.

5 Replication

5.1 Section 3.1: Website Content

5.1.1 Figure 1

I have not replicated the figure 1, a map of the various provinces in China and the availability of county government websites in them. This is because this was a manual step in the construction of the paper. I have replicated the data that was used as a basis for that map.

5.2 Section 4.1: Topics

5.2.1 Table 1

I have generated the underlying data for the table. I have not transferred it into the same table form as outlined in the paper.

5.3 Section 5.3: Predictive Inference

5.3.1 Table 3: Regression Results: Competence

Replicated the table, except positioning the constant at the bottom of the table

5.3.2 Table 4: Regression Results: Benevolence

Same as above

6 Extension

6.1 Comparison of alternative regression models in explaining competence and benevolence

One challenge with the regressions is that they are actually based on different underlying data. This is because of the data availability in the sample of 100 counties and then the prediction file. As we employ the various regressions, the number of observations used to fit these regressions decreases from 71 to 48. They are therefore fit based on different underlying data.

In a first step I therefore harmonize the regressions by basing them on the same number of observations throughout - the 48 observations that have all data available. I then compare whether the new regressions yield similar results as the regressions in the paper.

6.1.1 Competence

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Fri, Apr 17, 2020 - 11:46:16

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Beginning Tenure	4,000	0.169	0.028	0.069	0.151	0.188	0.280
End Tenure	4,000	0.008	0.049	−0.173	−0.025	0.042	0.209
mayor_last	4,000	0.136	0.061	−0.140	0.096	0.176	0.347
sigma	4,000	0.145	0.016	0.102	0.134	0.154	0.214

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Fri, Apr 17, 2020 - 11:46:16

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Beginning Tenure	4,000	0.187	0.054	0.010	0.151	0.224	0.393
End Tenure	4,000	−0.006	0.056	−0.233	−0.043	0.031	0.160
mayor_last	4,000	0.130	0.063	−0.154	0.089	0.171	0.381
X2009_gdppc_cny	4,000	0.00000	0.00000	−0.00000	0.00000	0.00000	0.00000
X2010_illiterateprop	4,000	−0.003	0.004	−0.017	−0.006	−0.001	0.012
itemploy	4,000	−0.0001	0.0001	−0.0003	−0.0001	−0.0001	0.0001
linksall	4,000	0.00000	0.00000	−0.00000	0.00000	0.00000	0.00000
county_mediaexp	4,000	−0.000	0.00000	−0.00000	−0.00000	0.00000	0.00000
sigma	4,000	0.145	0.017	0.097	0.133	0.155	0.241

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Fri, Apr 17, 2020 - 11:46:16

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Beginning Tenure	4,000	0.158	0.056	−0.090	0.122	0.195	0.343
End Tenure	4,000	−0.015	0.056	−0.224	−0.052	0.023	0.226
mayor_last	4,000	0.128	0.069	−0.178	0.081	0.173	0.403
X2009_gdppc_cny	4,000	0.00000	0.00000	−0.00000	0.00000	0.00000	0.00000
X2010_illiterateprop	4,000	−0.003	0.004	−0.019	−0.006	−0.001	0.015
itemploy	4,000	−0.0001	0.0001	−0.0003	−0.0001	−0.0001	0.0001
linksall	4,000	0.00000	0.00000	−0.00000	0.00000	0.00000	0.00000
county_mediaexp	4,000	−0.00000	0.00000	−0.00000	−0.00000	0.00000	0.00000
sec_first	4,000	0.126	0.059	−0.103	0.086	0.165	0.359
sec_last	4,000	0.039	0.057	−0.162	0.001	0.076	0.236
sigma	4,000	0.140	0.016	0.098	0.129	0.149	0.226

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Fri, Apr 17, 2020 - 11:46:16

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Beginning Tenure	4,000	0.262	0.149	−0.324	0.164	0.362	0.788
End Tenure	4,000	0.010	0.064	−0.241	−0.033	0.053	0.241
mayor_last	4,000	0.166	0.073	−0.101	0.116	0.214	0.444
X2009_gdppc_cny	4,000	0.00000	0.00000	−0.00000	0.00000	0.00000	0.00001
X2010_illiterateprop	4,000	−0.002	0.005	−0.020	−0.006	0.001	0.020
itemploy	4,000	−0.0001	0.0001	−0.0003	−0.0001	−0.00001	0.0002
linksall	4,000	0.00000	0.00000	−0.00000	0.00000	0.00000	0.00000
county_mediaexp	4,000	0.00000	0.00000	−0.00000	−0.00000	0.00000	0.00000
sec_first	4,000	0.112	0.061	−0.140	0.071	0.154	0.306
sec_last	4,000	0.049	0.058	−0.163	0.011	0.086	0.283
pref_ps_first	4,000	0.076	0.076	−0.181	0.026	0.126	0.356
pref_ps_last	4,000	0.078	0.070	−0.159	0.032	0.125	0.300
pref_ps_edulevel	4,000	−0.033	0.029	−0.151	−0.052	−0.014	0.068
pref_2010_gdppc	4,000	−0.00000	0.00000	−0.00001	−0.00000	−0.00000	0.00000
sigma	4,000	0.140	0.017	0.094	0.128	0.150	0.232

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Fri, Apr 17, 2020 - 11:46:17

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Beginning Tenure	4,000	0.429	0.299	−0.693	0.235	0.628	1.413
End Tenure	4,000	−0.046	0.070	−0.337	−0.093	0.003	0.180
mayor_last	4,000	0.153	0.073	−0.136	0.104	0.203	0.414
X2009_gdppc_cny	4,000	0.00000	0.00000	−0.00000	0.000	0.00000	0.00000
X2010_illiterateprop	4,000	−0.001	0.005	−0.022	−0.004	0.002	0.018
itemploy	4,000	0.00001	0.0001	−0.0003	−0.00004	0.0001	0.0004
linksall	4,000	0.00000	0.00000	−0.00000	−0.00000	0.00000	0.00000
county_mediaexp	4,000	−0.00000	0.00000	−0.00000	−0.00000	0.00000	0.00000
sec_first	4,000	0.150	0.062	−0.074	0.108	0.191	0.398
sec_last	4,000	0.029	0.058	−0.174	−0.009	0.067	0.259
pref_ps_first	4,000	0.070	0.080	−0.243	0.016	0.123	0.419
pref_ps_last	4,000	0.104	0.072	−0.227	0.058	0.151	0.371
pref_ps_edulevel	4,000	−0.050	0.030	−0.159	−0.071	−0.031	0.062
pref_2010_gdppc	4,000	−0.00000	0.00000	−0.00001	−0.00000	0.00000	0.00001
mayor_age	4,000	−0.005	0.006	−0.028	−0.010	−0.001	0.016
mayor_genderM	4,000	−0.078	0.100	−0.431	−0.144	−0.012	0.338
mayor_edulevel	4,000	0.063	0.031	−0.048	0.042	0.084	0.180
sigma	4,000	0.135	0.017	0.086	0.123	0.146	0.220

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Fri, Apr 17, 2020 - 11:46:17

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Beginning Tenure	4,000	0.455	0.311	−0.686	0.250	0.663	1.657
End Tenure	4,000	−0.050	0.071	−0.369	−0.096	−0.003	0.212
mayor_last	4,000	0.170	0.091	−0.150	0.110	0.232	0.483
X2009_gdppc_cny	4,000	0.00000	0.00000	−0.00000	−0.00000	0.00000	0.00000
X2010_illiterateprop	4,000	−0.001	0.005	−0.021	−0.004	0.002	0.016
itemploy	4,000	0.00001	0.0001	−0.0004	−0.00004	0.0001	0.0003
linksall	4,000	0.00000	0.00000	−0.00000	−0.00000	0.00000	0.00000
county_mediaexp	4,000	−0.00000	0.00000	−0.00000	−0.00000	0.000	0.00000
sec_first	4,000	0.149	0.066	−0.100	0.106	0.191	0.438
sec_last	4,000	0.030	0.059	−0.237	−0.009	0.070	0.265
pref_ps_first	4,000	0.072	0.078	−0.244	0.022	0.127	0.353
pref_ps_last	4,000	0.105	0.071	−0.188	0.057	0.151	0.332
pref_ps_edulevel	4,000	−0.050	0.030	−0.177	−0.070	−0.030	0.077
pref_2010_gdppc	4,000	−0.00000	0.00000	−0.00001	−0.00000	0.00000	0.00000
mayor_age	4,000	−0.006	0.007	−0.030	−0.010	−0.002	0.020
mayor_genderM	4,000	−0.076	0.098	−0.412	−0.141	−0.009	0.360
mayor_edulevel	4,000	0.064	0.031	−0.048	0.045	0.085	0.189
mayor_promote	4,000	−0.023	0.072	−0.311	−0.070	0.024	0.334
sigma	4,000	0.137	0.018	0.092	0.125	0.148	0.237

6.1.2 Benevolence

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Fri, Apr 17, 2020 - 11:46:19

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Beginning Tenure	4,000	0.184	0.032	0.066	0.164	0.205	0.307
End Tenure	4,000	0.089	0.057	−0.141	0.051	0.128	0.315
mayor_last	4,000	−0.017	0.071	−0.277	−0.066	0.032	0.253
sigma	4,000	0.169	0.018	0.122	0.156	0.180	0.268

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Fri, Apr 17, 2020 - 11:46:19

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Beginning Tenure	4,000	0.154	0.062	−0.092	0.112	0.194	0.377
End Tenure	4,000	0.146	0.066	−0.151	0.101	0.190	0.425
mayor_last	4,000	0.007	0.073	−0.240	−0.044	0.056	0.255
X2009_gdppc_cny	4,000	−0.00000	0.00000	−0.00001	−0.00000	−0.00000	0.00000
X2010_illiterateprop	4,000	0.003	0.005	−0.017	0.0002	0.007	0.021
itemploy	4,000	0.00001	0.0001	−0.0003	−0.00004	0.0001	0.0003
linksall	4,000	0.00000	0.00000	−0	0.000	0.000	0
county_mediaexp	4,000	0.00000	0.00000	−0.00000	−0.00000	0.00000	0.00000
sigma	4,000	0.168	0.019	0.122	0.154	0.179	0.257

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Fri, Apr 17, 2020 - 11:46:20

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Beginning Tenure	4,000	0.145	0.071	−0.143	0.097	0.193	0.424
End Tenure	4,000	0.144	0.069	−0.124	0.099	0.189	0.391
mayor_last	4,000	0.010	0.085	−0.326	−0.048	0.068	0.291
X2009_gdppc_cny	4,000	−0.00000	0.00000	−0.00001	−0.00000	−0.00000	0.00000
X2010_illiterateprop	4,000	0.004	0.005	−0.013	0.0003	0.007	0.022
itemploy	4,000	0.00001	0.0001	−0.0003	−0.00005	0.0001	0.0004
linksall	4,000	0.00000	0.00000	−0.00000	0.00000	0.00000	0.00000
county_mediaexp	4,000	0.00000	0.00000	−0.00000	−0.00000	0.00000	0.00000
sec_first	4,000	0.046	0.074	−0.209	−0.003	0.097	0.290
sec_last	4,000	0.006	0.069	−0.229	−0.038	0.053	0.247
sigma	4,000	0.172	0.020	0.120	0.158	0.184	0.287

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Fri, Apr 17, 2020 - 11:46:21

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Beginning Tenure	4,000	0.214	0.190	−0.558	0.088	0.342	0.823
End Tenure	4,000	0.142	0.080	−0.238	0.089	0.196	0.452
mayor_last	4,000	0.016	0.096	−0.334	−0.047	0.079	0.372
X2009_gdppc_cny	4,000	−0.00000	0.00000	−0.00001	−0.00000	−0.00000	0.00000
X2010_illiterateprop	4,000	0.003	0.006	−0.021	−0.001	0.007	0.025
itemploy	4,000	0.00002	0.0001	−0.0003	−0.00004	0.0001	0.0003
linksall	4,000	0.00000	0.00000	−0.00000	0.00000	0.00000	0.00000
county_mediaexp	4,000	0.00000	0.00000	−0.00000	−0.00000	0.00000	0.00000
sec_first	4,000	0.044	0.081	−0.294	−0.009	0.096	0.366
sec_last	4,000	0.010	0.075	−0.255	−0.039	0.061	0.292
pref_ps_first	4,000	0.026	0.099	−0.373	−0.040	0.091	0.390
pref_ps_last	4,000	−0.013	0.091	−0.460	−0.074	0.050	0.288
pref_ps_edulevel	4,000	−0.018	0.038	−0.167	−0.043	0.007	0.137
pref_2010_gdppc	4,000	−0.00000	0.00000	−0.00001	−0.00000	0.00000	0.00000
sigma	4,000	0.180	0.023	0.122	0.164	0.192	0.286

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Fri, Apr 17, 2020 - 11:46:21

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Beginning Tenure	4,000	0.191	0.414	−1.367	−0.085	0.458	2.296
End Tenure	4,000	0.144	0.093	−0.208	0.082	0.207	0.480
mayor_last	4,000	0.026	0.099	−0.293	−0.039	0.091	0.393
X2009_gdppc_cny	4,000	−0.00000	0.00000	−0.00001	−0.00000	−0.00000	0.00000
X2010_illiterateprop	4,000	0.004	0.007	−0.020	−0.001	0.008	0.034
itemploy	4,000	0.00002	0.0001	−0.0004	−0.00005	0.0001	0.0005
linksall	4,000	0.00000	0.00000	−0.00000	0.00000	0.00000	0.00000
county_mediaexp	4,000	0.00000	0.00000	−0.00000	−0.00000	0.00000	0.00000
sec_first	4,000	0.040	0.086	−0.274	−0.017	0.100	0.422
sec_last	4,000	0.007	0.079	−0.291	−0.044	0.059	0.295
pref_ps_first	4,000	0.014	0.104	−0.333	−0.053	0.083	0.420
pref_ps_last	4,000	−0.018	0.096	−0.369	−0.082	0.045	0.388
pref_ps_edulevel	4,000	−0.016	0.041	−0.171	−0.043	0.011	0.143
pref_2010_gdppc	4,000	0.00000	0.00000	−0.00001	−0.00000	0.00000	0.00001
mayor_age	4,000	−0.002	0.009	−0.037	−0.008	0.004	0.029
mayor_genderM	4,000	0.108	0.135	−0.392	0.017	0.197	0.690
mayor_edulevel	4,000	−0.004	0.041	−0.168	−0.030	0.024	0.143
sigma	4,000	0.186	0.024	0.126	0.169	0.200	0.331

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Fri, Apr 17, 2020 - 11:46:22

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Beginning Tenure	4,000	0.069	0.406	−1.264	−0.201	0.335	1.705
End Tenure	4,000	0.169	0.093	−0.178	0.109	0.231	0.495
mayor_last	4,000	−0.066	0.122	−0.649	−0.147	0.014	0.395
X2009_gdppc_cny	4,000	−0.00000	0.00000	−0.00001	−0.00000	−0.00000	0.00000
X2010_illiterateprop	4,000	0.004	0.006	−0.023	−0.0004	0.008	0.026
itemploy	4,000	0.00001	0.0001	−0.0004	−0.0001	0.0001	0.001
linksall	4,000	0.00000	0.00000	−0.00000	0.00000	0.00000	0.00000
county_mediaexp	4,000	0.00000	0.00000	−0.00000	−0.00000	0.00000	0.00000
sec_first	4,000	0.055	0.084	−0.357	−0.001	0.111	0.419
sec_last	4,000	0.002	0.078	−0.299	−0.048	0.053	0.304
pref_ps_first	4,000	−0.009	0.107	−0.446	−0.079	0.062	0.476
pref_ps_last	4,000	−0.021	0.094	−0.360	−0.083	0.038	0.358
pref_ps_edulevel	4,000	−0.013	0.040	−0.172	−0.038	0.012	0.161
pref_2010_gdppc	4,000	0.00000	0.00000	−0.00001	−0.00000	0.00000	0.00001
mayor_age	4,000	0.0004	0.009	−0.038	−0.005	0.006	0.033
mayor_genderM	4,000	0.119	0.131	−0.417	0.033	0.207	0.631
mayor_edulevel	4,000	−0.013	0.040	−0.177	−0.039	0.013	0.166
mayor_promote	4,000	0.124	0.095	−0.224	0.063	0.186	0.474
sigma	4,000	0.183	0.024	0.119	0.166	0.197	0.308

In a second step I go through the various regressions seeking to understand whether any of the tested variable classes (resources, peers, prefecture, ability, career path) explain the observed phenomena better than the ones around tenure. I do so by creating a series of new regressions that include only the variables in the respective variable classes. In addition I introduce a new series of classes: culture that includes the variables macro-region and county type.

I then compare all the available models with the leave-one-out method to see which one of these is best suited to explain the observed phenomena.

6.1.3 Competence

```
##           elpd_diff se_diff
## linear_1   0.0       0.0
## linear_11 -0.1       1.5
## linear_10 -0.4       3.2
## linear_8  -0.4       4.1
## linear_12 -2.2       2.8
## linear_2  -2.6       2.1
## linear_7  -2.7       3.7
## linear_3  -3.0       4.3
## linear_9  -3.9       3.2
## linear_5  -7.0       4.8
## linear_4  -7.2       4.3
## linear_6  -7.8       4.9
```

6.1.4 Benevolence

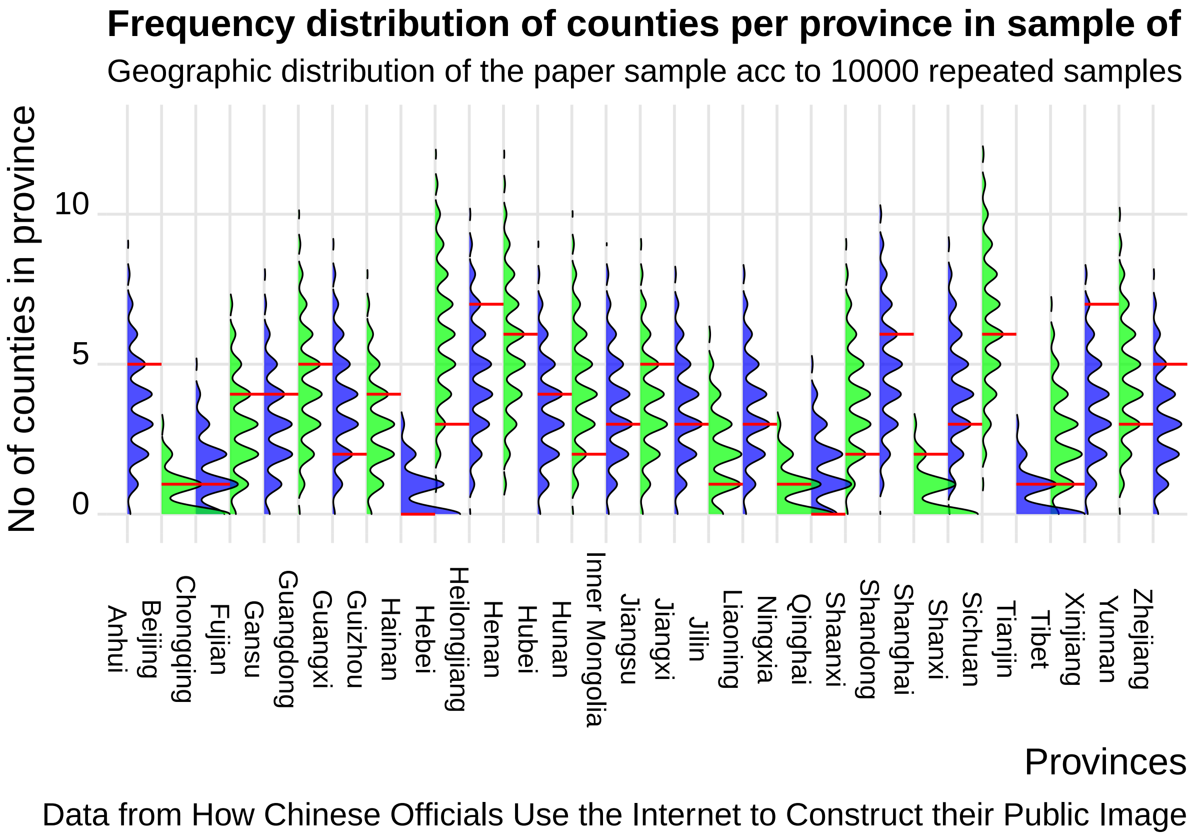
```
##           elpd_diff se_diff
## linear.1    0.0       0.0
## linear.11  -0.5       2.1
## linear.12  -1.3       2.0
## linear.8   -1.6       1.8
## linear.9   -2.1       2.4
## linear.10  -2.7       2.1
## linear.2   -2.7       2.0
## linear.7   -4.0       2.4
## linear.3   -5.0       2.4
## linear.4   -9.9       2.7
## linear.5  -13.2       2.9
## linear.6  -13.6       3.0
```

The comparison suggests that the best models for how competent an official is described as, are the tenure based model employed in the paper and whether officials were promoted in the two ensuing years. The later is a post-treatment variable. Rather than delviering additional insights on what the factors around an official are that determine the website content, it suggests the relative importance of the website content in determining whether an individual is promoted. In combination this supports the hypothesis of the author that the websites have an important signalling function within the Chinese state apparatus to determine who gets promoted.

6.2 Validating geographic split of sample

Subsequently I am simulating repeated sampling. I do so by creating a function that allows me to draw 100 and 48 random samples from the underlying countywebsites (countyweb). I repeat this step 10000 times, counting the number of counties from each province. I then compare that count with the count of counties in the paper samples (the sample of 100 and the 48 counties we actually end up constructing a model with).

6.2.1 Random sample of 100



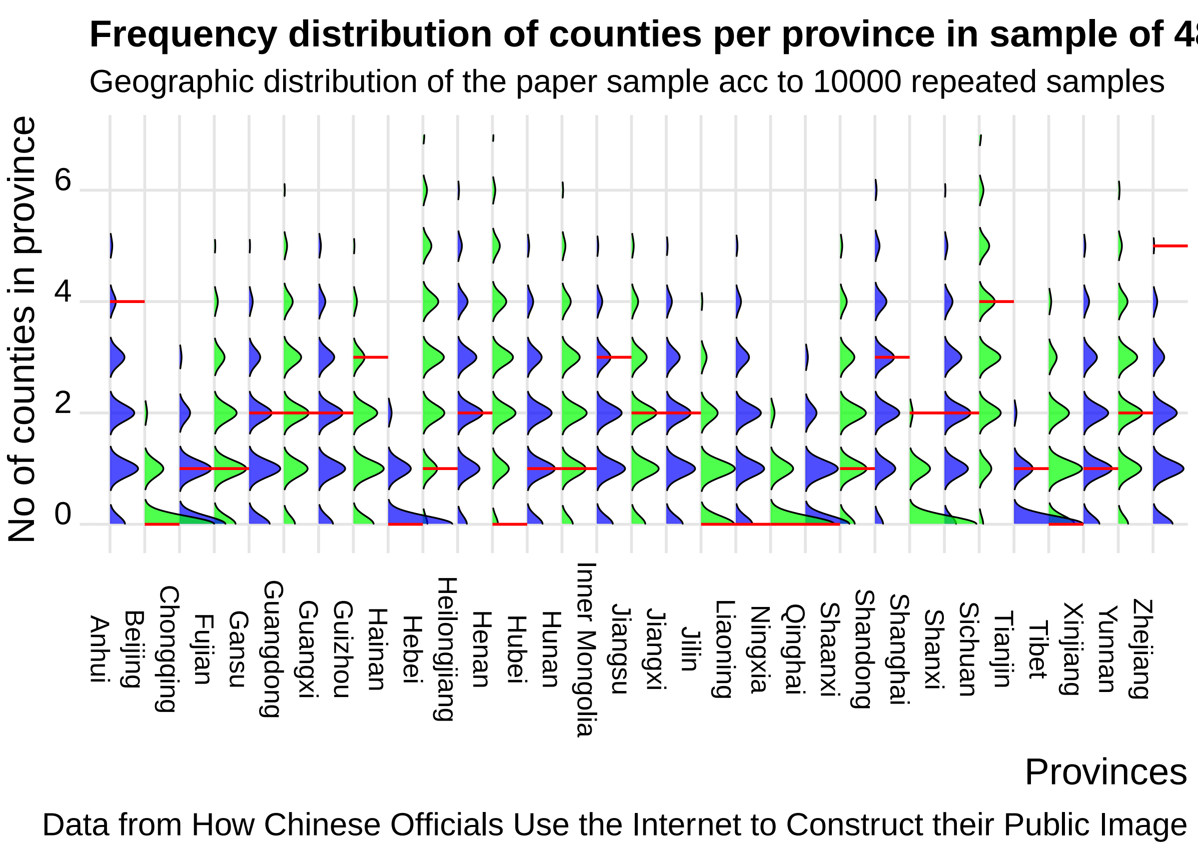
Extension Table 1: Frequency of provinces in sample of 100
Based on simulation of 10000 samples of 100

Provinces	2.5 Percentile	97.5 Percentile	Frequency in paper sample	Outside CI
Anhui	0	8	5	No
Beijing	0	2	1	No
Chongqing	0	4	1	No
Fujian	0	7	4	No
Gansu	0	7	4	No
Guangdong	1	9	5	No
Guangxi	1	8	2	No
Guizhou	0	7	4	No
Hainan	0	3	0	No
Hebei	2	11	3	No
Heilongjiang	1	9	7	No
Henan	2	10	6	No
Hubei	1	7	4	No
Hunan	1	9	2	No
Inner Mongolia	1	8	3	No
Jiangsu	1	8	5	No
Jiangxi	0	7	3	No
Jilin	0	5	1	No
Liaoning	0	7	3	No
Ningxia	0	3	1	No

Qinghai	0	4	0	No
Shaanxi	1	8	2	No
Shandong	1	9	6	No
Shanghai	0	2	2	No
Shanxi	1	8	3	No
Sichuan	2	11	6	No
Tianjin	0	2	1	No
Tibet	0	6	1	No
Xinjiang	1	7	7	No
Yunnan	1	9	3	No
Zhejiang	0	7	5	No

Repeated simulated sampling allows us to conclude that the sample of 100 in the paper is random and thereby representative in terms of geographic sampling. This is because the number of counties from a province in no case is outside the 95% CI interval that we constructed. The graph shows that in some counties, s.a. Heilongjiang or Hennan the county occurences are rather on the margins of what we would expect to see.

6.2.2 Sample of 48 for modelling



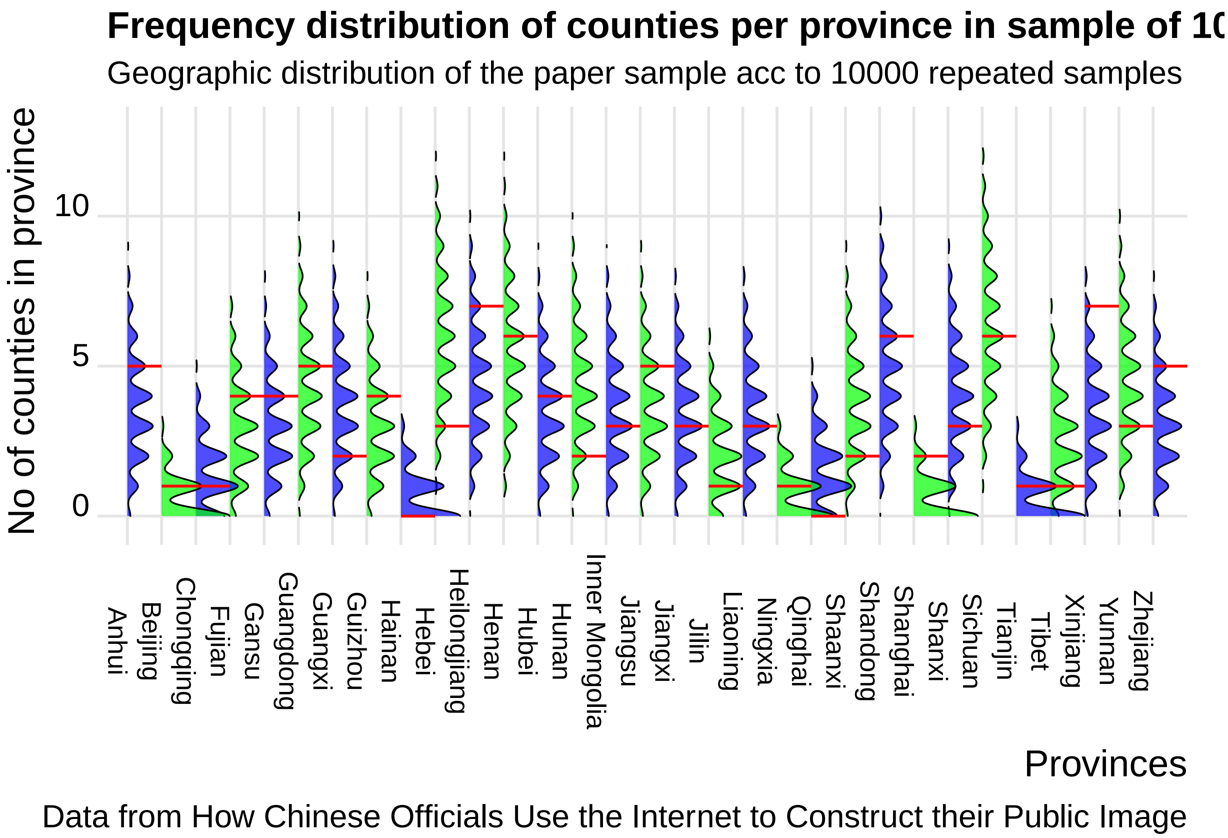
Extension Table 2: Frequency of provinces in sample of 48
Based on simulation of 10000 samples of 48

Provinces	2.5 Percentile	97.5 Percentile	Frequency in paper sample	Outside CI	On lower CI boundary
Anhui	0	5	4	No	No

Beijing	0	2	0	No	Yes
Chongqing	0	3	1	No	No
Fujian	0	4	1	No	No
Gansu	0	4	2	No	No
Guangdong	0	5	2	No	No
Guangxi	0	5	2	No	No
Guizhou	0	4	3	No	No
Hainan	0	2	0	No	Yes
Hebei	0	6	1	No	No
Heilongjiang	0	5	2	No	No
Henan	0	6	0	No	Yes
Hubei	0	5	1	No	No
Hunan	0	5	1	No	No
Inner Mongolia	0	4	3	No	No
Jiangsu	0	5	2	No	No
Jiangxi	0	4	2	No	No
Jilin	0	3	0	No	Yes
Liaoning	0	4	0	No	Yes
Ningxia	0	2	0	No	Yes
Qinghai	0	3	0	No	Yes
Shaanxi	0	5	1	No	No
Shandong	0	6	3	No	No
Shanghai	0	2	2	No	No
Shanxi	0	5	2	No	No
Sichuan	0	7	4	No	No
Tianjin	0	2	1	No	No
Tibet	0	4	0	No	Yes
Xinjiang	0	5	1	No	No
Yunnan	0	5	2	No	No
Zhejiang	0	4	5	Yes	No

Repeated simulated sampling allows us to conclude that the sample of 48 in the paper is likely random and thereby representative in terms of geographic sampling. This is because the number of counties from a province in only one case (Zhejiang) is outside the 95% CI interval that we constructed. The graph shows that there is a surprising amount of states at the lower boundary of 0. This seems plausible due to the small size of the sample (Beijing, Hainan, Henan, Jilin, Liaoning, Ningxia, Qinghai and Tibet). There seems to be no regional pattern among these states (3 East, 2 Central, 3 West).

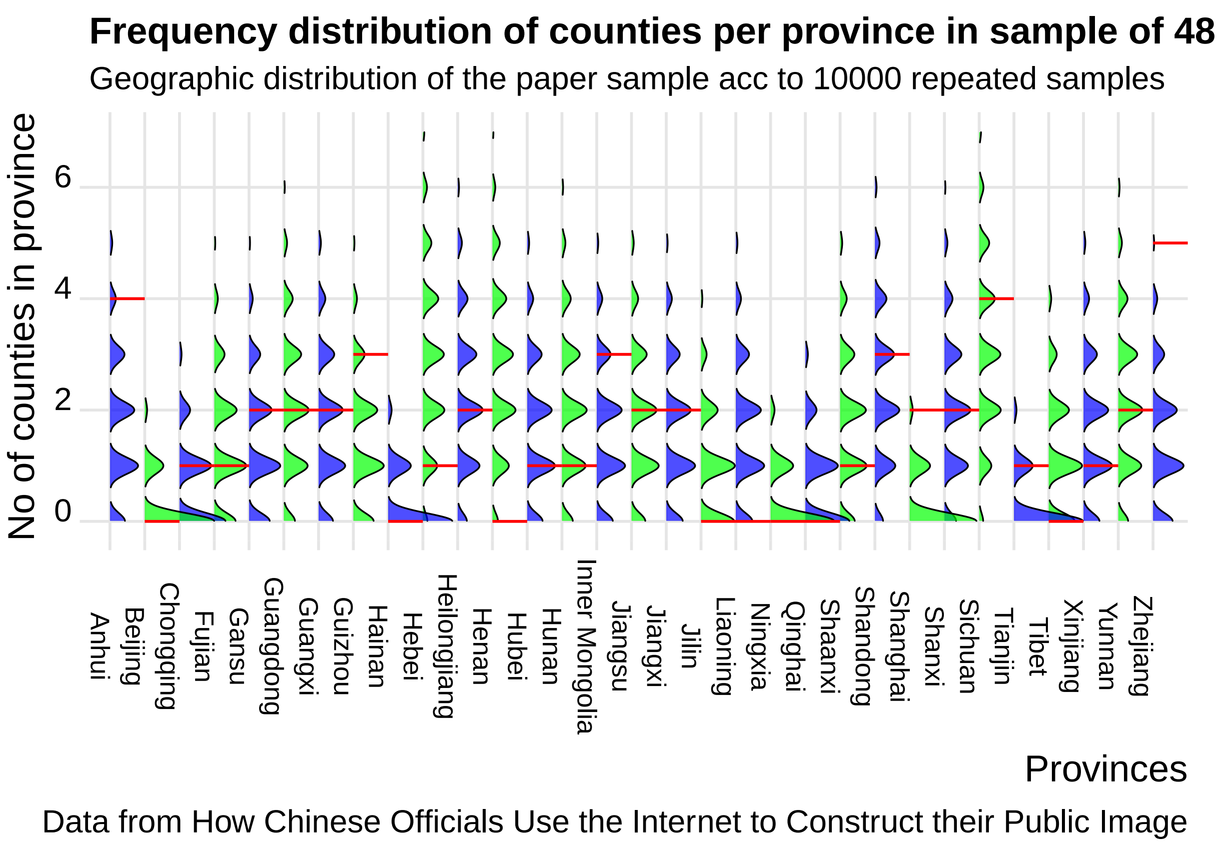
7 Tables and Figures



Extension Table 1: Frequency of provinces in sample of 100
Based on simulation of 10000 samples of 100

Provinces	2.5 Percentile	97.5 Percentile	Frequency in paper sample	Outside CI
Anhui	0	8	5	No
Beijing	0	2	1	No
Chongqing	0	4	1	No
Fujian	0	7	4	No
Gansu	0	7	4	No
Guangdong	1	9	5	No
Guangxi	1	8	2	No
Guizhou	0	7	4	No
Hainan	0	3	0	No
Hebei	2	11	3	No
Heilongjiang	1	9	7	No
Henan	2	10	6	No
Hubei	1	7	4	No
Hunan	1	9	2	No
Inner Mongolia	1	8	3	No
Jiangsu	1	8	5	No
Jiangxi	0	7	3	No
Jilin	0	5	1	No
Liaoning	0	7	3	No

Ningxia	0	3	1	No
Qinghai	0	4	0	No
Shaanxi	1	8	2	No
Shandong	1	9	6	No
Shanghai	0	2	2	No
Shanxi	1	8	3	No
Sichuan	2	11	6	No
Tianjin	0	2	1	No
Tibet	0	6	1	No
Xinjiang	1	7	7	No
Yunnan	1	9	3	No
Zhejiang	0	7	5	No



Extension Table 2: Frequency of provinces in sample of 48
Based on simulation of 10000 samples of 48

Provinces	2.5 Percentile	97.5 Percentile	Frequency in paper sample	Outside CI	On lower CI boundary
Anhui	0	5	4	No	No
Beijing	0	2	0	No	Yes
Chongqing	0	3	1	No	No
Fujian	0	4	1	No	No
Gansu	0	4	2	No	No
Guangdong	0	5	2	No	No
Guangxi	0	5	2	No	No
Guizhou	0	4	3	No	No

Hainan	0	2	0	No	Yes
Hebei	0	6	1	No	No
Heilongjiang	0	5	2	No	No
Henan	0	6	0	No	Yes
Hubei	0	5	1	No	No
Hunan	0	5	1	No	No
Inner Mongolia	0	4	3	No	No
Jiangsu	0	5	2	No	No
Jiangxi	0	4	2	No	No
Jilin	0	3	0	No	Yes
Liaoning	0	4	0	No	Yes
Ningxia	0	2	0	No	Yes
Qinghai	0	3	0	No	Yes
Shaanxi	0	5	1	No	No
Shandong	0	6	3	No	No
Shanghai	0	2	2	No	No
Shanxi	0	5	2	No	No
Sichuan	0	7	4	No	No
Tianjin	0	2	1	No	No
Tibet	0	4	0	No	Yes
Xinjiang	0	5	1	No	No
Yunnan	0	5	2	No	No
Zhejiang	0	4	5	Yes	No

!!! Many people, after looking at the abstract, will ignore the prose and just go straight for the tables and figures. This means that you need captions which are, on average, a paragraph long. (Consult almost any published paper for an example.) The caption explains most of what you need to know to understand the table/graphic. It is very detailed. It is also, when considered in conjunction with the paper itself, somewhat redundant. And that is OK! It must be possible for the reader, looking at just your abstract and your tables/graphics, to have a fairly complete understanding of your paper.

The converse holds as well. The reader should be able to read your paper, without consulting the tables/graphics, and understand everything.

One reason we use pdf_document2 as our document type is that, in conjunction with the bookdown package, it makes the creation of long captions for numbered figures/tables much easier.

8 Appendixes

8.1 Section 3.1: Website Content

8.1.1 Figure 1: County Government Website Availability by Province

8.2 Section 4.1: Topics

8.2.1 Table 1: LDA topics and OGI Requirements

8.3 Section 5.1: Measuring Tenure

8.3.1 Table 2: Distribution of Year in office

8.4 Section 5.2: Descriptive Results

8.4.1 Figure 2: Proportion of web pages with content focused on competence by year in office. . .

8.4.2 Figure 3: Proportion of web pages with content focused on belevolence by year in office. . .

8.5 Section 5.3: Predictive Inference

8.5.1 Table 3: Regression Results: Competence

8.5.2 Table 4: Regression Results: Benevolence

References

Cai, Yongshun. 2004. "Irresponsible State: Local Cadres and Image-Building in China." *Journal of Communist Studies and Transition Politics* 20 (4): 20–41. https://doi.org/10.1080/1352327042000306039.

Chen, Jidong, Jennifer Pan, and Yiqing Xu. 2016. "Sources of Authoritarian Responsiveness: A Field Experiment in China." *American Journal of Political Science* 60 (2): 383–400. http://www.jstor.org/stable/24877628.

Edin, Maria. 2003. "State Capacity and Local Agent Control in China: CCP Cadre Management from a Township Perspective." *The China Quarterly*, no. 173: 35–52. http://www.jstor.org/stable/20058957.

Kalathil, Shanthi, and Taylor Boas. 2003. "Open Networks, Closed Regimes: The Impact of the Internet on Authoritarian Rule (Table of Contents)." *First Monday* 8 (1). https://doi.org/10.5210/fm.v8i1.1024.

Li, Hongbin, and Li-An Zhou. 2005. "Political Turnover and Economic Performance: The Incentive Role of Personnel Control in China." *Journal of Public Economics* 89 (9): 1743–62. https://doi.org/https://doi.org/10.1016/j.jpubeco.2004.06.009.

Pan, Jennifer, and Kaping Chen. 2018. *Concealing Corruption: How Chinese Officials Distort Upward Reporting of Online Grievances.* American Political Science Review 112.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.