ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Buchot, Tom; Couttenier, Mathieu; Laugerette, Lucile; Mougin, Elisa; Verlet, Alexandre

Working Paper

A Comment on Bai, Jia &, Yang (2023) Web of Power: How Elite Networks Shaped War and Politics in China

I4R Discussion Paper Series, No. 115

Provided in Cooperation with:
The Institute for Replication (I4R)

Suggested Citation: Buchot, Tom; Couttenier, Mathieu; Laugerette, Lucile; Mougin, Elisa; Verlet, Alexandre (2024) : A Comment on Bai, Jia &, Yang (2023) Web of Power: How Elite Networks Shaped War and Politics in China, I4R Discussion Paper Series, No. 115, Institute for Replication (I4R), s.l.

This Version is available at:
https://hdl.handle.net/10419/290341

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

INSTITUTE for REPLICATION

No. 115
I4R DISCUSSION PAPER SERIES

A Comment on Bai, Jia &, Yang (2023) Web of Power: How Elite Networks Shaped War and Politics in China

Tom Buchot
Mathieu Couttenier
Lucile Laugerette
Elisa Mougin
Alexandre Verlet

April 2024

INSTITUTE for REPLICATION

I4R DISCUSSION PAPER SERIES

I4R DP No. 115

A Comment on Bai, Jia &, Yang (2023) Web of Power: How Elite Networks Shaped War and Politics in China

Tom Buchot[1], Mathieu Couttenier[1], Lucile Laugerette[1], Elisa Mougin[1], Alexandre Verlet[1]

[1]École Normale Supérieure (CNS), and Center for Economic Research on Governance, Inequality and Confict (CERGIC), Lyon/France

APRIL 2024

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and meta-scientific work in the social sciences. Provided in cooperation with EconStor, a service of the ZBW – Leibniz Information Centre for Economics, and RWI – Leibniz Institute for Economic Research, I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Editors

Abel Brodeur
University of Ottawa

Anna Dreber
Stockholm School of Economics

Jörg Ankel-Peters
RWI – Leibniz Institute for Economic Research

E-Mail: joerg.peters@rwi-essen.de
RWI – Leibniz Institute for Economic Research
Hohenzollernstraße 1-3
45128 Essen/Germany
www.i4replication.org

ISSN: 2752-1931

A comment on Bai, Jia &, Yang (2023) Web of Power: How Elite Networks Shaped War and Politics in China

Tom Buchot, Mathieu Couttenier, Lucile Laugerette, Elisa Mougin & Alexandre Verlet*

January 18, 2024

Abstract

Bai et al. (2023) examine the impact of individual networks on state building, focusing on the role of the leader Zeng Guofan during the Taiping Revolution in China between 1850 and 1864. In their main results, the authors demonstrate that being connected to Zeng increases the number of fatalities during the war after his assumption of power, with point estimates being significant at the 1% or 5% level. They also find a positive and significant effect of connections to Zeng among Hunan people on the number of national-level office positions, with point estimates significant at the 1% level. First, we reproduce the paper's main findings and identify minor inaccuracies in the codes that need fixing for the proper reproduction of some tables. However, these issues do not significantly impact the overall results. Second, we conduct additional checks and argue that the results are robust to variations in the number of fixed effects but highly dependent on the choice of econometric specification. We employ alternative models more suitable for data with a substantial number of zeros, revealing a decrease in the magnitude and significance of the estimates. Last, we perform spatial robustness checks, confirming the absence of spatial correlation between Hunan county and its neighboring regions, as suggested by the authors.

*ENS de Lyon, Center for Economic Research on Governance, Inequality and Conflict (CERGIC).

1 Introduction

In the paper "Web of Power: How Elite Networks Shaped War and Politics in China", Bai, Jai, and Yang (2023), henceforth BJY, contribute significantly to the political economy literature, focusing on leaders' identity and their role in policymaking through the case study of the Taiping Rebellion, which took place in China from 1850 to 1864. They explore the role of one specific leader, Zeng Guofan, who commanded the Hunan Army as of 1853 and "mobilized" his network during the war to recruit soldiers. They question the possible effects on outcomes after the war of the leader's supremacy and network construction on post-war outcomes, such as the number of political elites coming from Hunan county. In the first part of the paper, BJY implement a Difference-in-Differences strategy to stress the link between Zeng's network and soldier deaths using variations between Hunan counties, considering those more or less connected to Zeng (mostly measuring the network from exam relationships or blood connections). They argue that more connected counties to Zeng experienced a higher number of fatalities: "*counties with one additional elite directly connected to Zeng experienced 21% more soldier deaths.*" The second set of analyses focuses on the post-war distribution of political power. Since non-Hunan counties didn't provide any soldiers for the battle, they are used by BJY as the control group to examine the effect of connections to Zeng on the number of national-level offices. They make use of a difference-in-differences strategy from which they conclude that more elite connections are associated with a higher number of national high-power positions. In other words, after the war, the power distribution closely followed the structure of power of the Hunan army. They conclude that the role of individual-level networks can shape the structure of war processes and can have long-term effects on the functioning of the state and the distribution of power.

In the present paper, we test the reproducibility of the results and conduct a series of robustness exercises, including: (1) suggesting more coherence into the inclusion of fixed effects, (2) discussing the validity of the econometric specification due to the specific distribution of the dependent variables (large number of zero), and (3) conducting additional spatial analysis to explore further spatial spillovers across regions.

Regarding reproducibility, our starting point is the replication package made available by the authors[1]. We notice both some inconsistencies and (small)

[1]Accessed here: `https://doi.org/10.7910/DVN/0H4HG2`

mistakes in some parts of the codes that hindered proper reproduction of the Tables and Figures. We detail the different issues in Section 2. Turning to the robustness checks, we first find that homogenizing the regression framework from Table II to Table III, i.e. by including the same set of fixed effects in both Tables, sustain the coherence of the results and ease the comparability across estimates (Section 3.1). Second, we express some concerns regarding the large number of zero values in the dependent variables and the authors' decision to resort to a $ln(1+y)$ transformation. Following Mullahy and Norton (2022) and Winkelmann (2015), we use a standard OLS model, a Poisson model, and a negative binomial functional form (Section 3.2). Lastly, we explore the possible role of distance to Hunan counties (in the spirit of the results displayed in Table V), i.e., the positive effect of being from Hunan on the probability of being nominated as a top official (Section 3.3).

2 Reproducibility

We describe in this section some inaccuracies or small errors noticed by running the replication package provided by the authors.

- There is a mismatch in the replication package between the names of the codes and the figures it replicates: starting as of Figure 3, the code entitled Figure F replicates Figure F+1.

- There are some small inconsistencies between Table 1 and the reproducing code. First, the variable $Zeng_all0_invdist_pc$, labeled as "Elite connections per capita", is missing in the code for Panel B.
  Also in Table 1, we noticed that the sample used to produce the statistics was not corresponding to the cited sample in Panel B "All countries 1820-1910". The variables being time-invariant, the summary statistics are not affected, yet the number of observations is different ($N = 182,706$): if we consider the sample 1820-1910, one needs to add the condition "$if\ year > 1819$" in the code to have $N = 149,786$ observations.

- In Table III, the code doesn't provide the same number of observations ($N = 1,110$) as in the paper ($N = 1,125$). This difference is due to 15 singleton observations that are dropped from the command `reghdfe` in Stata (that could have been updated since the publication of the replication package). The same mismatch holds for Appendix Table B2 as well.

- The replication package didn't include the data, the shapefile and the source of the shape file to reproduce Figure III. We asked the authors who kindly sent us the county coordinates and the corresponding data that allow to replicate Figure III.[2]

3 Robustness Reproduction

In this Section, we estimate the sensitivity of the results to three different exercises.

3.1 Fixed effects

Tables II and III in the paper aim to document the effects of connections on the number of soldier deaths: the former utilizes a general definition of connections, while the latter deepens the analysis by incorporating various types of connections (expanded network, blood and friends, ...). Nevertheless, the level of fixed effects varies across the Tables, hindering direct comparison. Table II incorporates year and county fixed effects, whereas Table III builds upon these by adding a fixed effect for prefecture×year. To maintain consistency and ease the comparison across estimate, it may be beneficial to include the same fixed effects in the main specifications, while introducing the additional level in the Appendix. Our replication exercise underscores that the results presented in Table II remain robust even with the inclusion of prefecture×year fixed effects (Table A4). Furthermore, the results remain consistent when prefecture×year fixed effects are excluded from Table III (Table A5).

3.2 Econometric specification

In Tables II, III and Figure V, the authors make use as dependent variable the following log transformation: ln(1 + number of soldier deaths). The authors motivate this choice by doing it to "facilitate interpretation of the coefficients" (page 20). Note that a transformation with the inverse hyperbolic of soldier deaths is also implemented (Appendix Table B.1).
When the outcome is strictly positive, the $ln(1 + y)$ and inverse hyperbolic yield similar marginal effects as those obtained from untransformed outcomes (Aihounton and Henningsen, 2021; Norton, 2022). However, Mullahy and Norton

[2]In order to estimate the sensitivity of the results to alternative network measurements, we also asked the authors the original databases (only for replication purposes and by pre-registering those robustness checks) but no answer was given to this request.

(2022) show that the latter is not true when the outcome has a mass of zeros (which is the case here with roughly 50% of zeros, see Figure 1). Instead, it is recommended to use two-part models, an untransformed dependent variable or a Poisson, which yield similar marginal effects in the presence of a mass of zeros and skewed positive values (Santos Silva and Tenreyro, 2006; Bellemare and Wichman, 2020; Chen and Roth, 2023).

Prompted by the recent literature, we replicate Table II, Table III and Figure V of the paper, using alternative methodologies. We first replicate Table II by considering an untransformed dependent variable (Table 1) and a Poisson model (Table 2). In Table 1, all coefficients are statistically insignificant at the conventional levels. In Table 2, the results are mixed, with around 50% of coefficients that are statistically significant at the conventional levels. We also provide estimates considering a Poisson model with heteroskedasticity-robust standard errors and a negative binomial regression (Santos Silva and Tenreyro, 2006, 2010; Wooldridge, 2010). More than 50% of coefficients are statistically insignificant (Table A1 and Table A2 in Appendix). Second, we replicate Table III by considering again an untransformed dependent variable (Table 3) and a Poisson model (Table 4). Most of the coefficients are statistically significant. Estimating a Poisson model with heteroskedasticity-robust standard errors yields similar results (Table A3). Finally, we replicate the Figure V of the papers with the above-mentioned methodologies (Figure 2 and Figure 3). The results corroborate previous exercises in which a very large majority of the estimated coefficients are statistically insignificant at the conventional levels.

As Figure 1 illustrates, approximately 50% of the values for the variable representing soldier deaths are zero. It is likely that the zero values originate from multiple sources. Some may stem from the absence of battles, while others may arise from battle events with no reported fatalities. In setting, with two types of zeros, one stemming from the "never" and the other from the "standard" observation units, Winkelmann (2015) suggests making use of zero-inflated count data models, while Bellemare and Wichman (2020) recommend 1/3 of zeros as a general rule of thumb to use two-part models.[3] Using a two-part model requires information on the data generating process of the count variable, which is beyond the scope of this paper.

[3]Table C4 in the Appendix already includes a zero-inflated negative binomial for the national-offices outcome (92% of null value)

3.3 Spatial diffusion

We further investigate the potential impact of geographical proximity to Hunan on individuals holding national-level offices. In Table V, the authors make use of a difference-in-difference (and triple-difference) model to estimate the effect of elite connections in Hunan on the national-level office distribution before and after Zeng took power. They do find that having elite connection in a Hunan county leads to more national-level offices but fail to identify an effect for counties outside of Hunan. In the following exercises, we estimate whether some spatial spillovers occur for counties close to Hunan by considering the column (4) of Table V. We estimate the potential spatial diffusion by using three alternative measures of distance: i) the (log-) mean distance of the geodesic distances between a non-Hunan county and all Hunan counties, ii) the (log) of the minimum distance to reach the closest Hunan county, and iii) the (log-) mean distance of the topographic distances between a non-Hunan county and all Hunan counties.[4] Table 5 displays the results.

In the odd columns, the log distance is not significantly correlated. In the even-numbered columns, we adopt an alternative approach. We analyze the distance distribution, assigning a dummy coded 1 to counties within the top 10% of the nearest counties and 0 otherwise. By this means, we estimate the effect of Baseline connections × 1854-1910 for counties in the nearest 10% of the distribution and for counties in the furthest 90% of the distribution[5]. As column (6) of Table 5 shows, having a direct elite connection in counties close to Hunan counties after 1853 is associated with more national level offices.

4 Conclusion

This short paper aims to replicate the findings of "Web of Power" by Bai et al. (2023). We identify several minor inaccuracies in the replication package that can be easily rectified to ensure proper replication of most tables and figures from the original paper. Additionally, we conduct a series of robustness analyses, specifically examining the consistency between the controls and fixed effects

[4]Topographic distance is computed as: $(1.6 \times alt) + Distance$. To compute alt for a non-Hunan county c, we do as follows. First, we compute S, the sum of the altitudes on the line between c and a Hunan county. Then we compute the mean of S between c and all Hunan counties.

[5]Considering the 10% closest to Hunan counties in terms of topographical distance, they represent 158 counties.

included in the main models. We also discuss the choice of the econometric specification, considering that most of the used outcomes are variables with a substantial number of zero values. This part of the discussion aims to provide insights into whether the results are dependent on the choices made in the model specification. Furthermore, we undertake additional spatial robustness checks, the conclusions of which align with the narratives presented by the original authors. The collective findings contribute to a comprehensive understanding of the original study's robustness and offer valuable insights into the potential impact of different modeling choices on the results.

References

Aihounton, G. B. D. and Henningsen, A. (2021). Units of measurement and the inverse hyperbolic sine transformation. *The Econometrics Journal*, 24(2):334–351.

Bai, Y., Jia, R., and Yang, J. (2023). Web of Power: How Elite Networks Shaped War and Politics in China*. *The Quarterly Journal of Economics*, 138(2):1067–1108.

Bellemare, M. F. and Wichman, C. J. (2020). Elasticities and the Inverse Hyperbolic Sine Transformation. *Oxford Bulletin of Economics and Statistics*, 82(1):50–61.

Chen, J. and Roth, J. (2023). Logs with zeros? Some problems and solutions.

Mullahy, J. and Norton, E. C. (2022). Why Transform Y? A Critical Assessment of Dependent-Variable Transformations in Regression Models for Skewed and Sometimes-Zero Outcomes.

Norton, E. C. (2022). The Inverse Hyperbolic Sine Transformation and Retransformed Marginal Effects.

Santos Silva, J. and Tenreyro, S. (2010). On the existence of the maximum likelihood estimates in Poisson regression. *Economics Letters*, 107(2):310–312. Publisher: Elsevier.

Santos Silva, J. M. C. and Tenreyro, S. (2006). The Log of Gravity. *The Review of Economics and Statistics*, 88(4):641–658.

Winkelmann, R. (2015). Counting on count data models. *IZA World of Labor*.

Wooldridge, J. M. (2010). Econometric Analysis of Cross Section and Panel Data. *MIT Press Books*, 1. Publisher: The MIT Press.

5 Figures

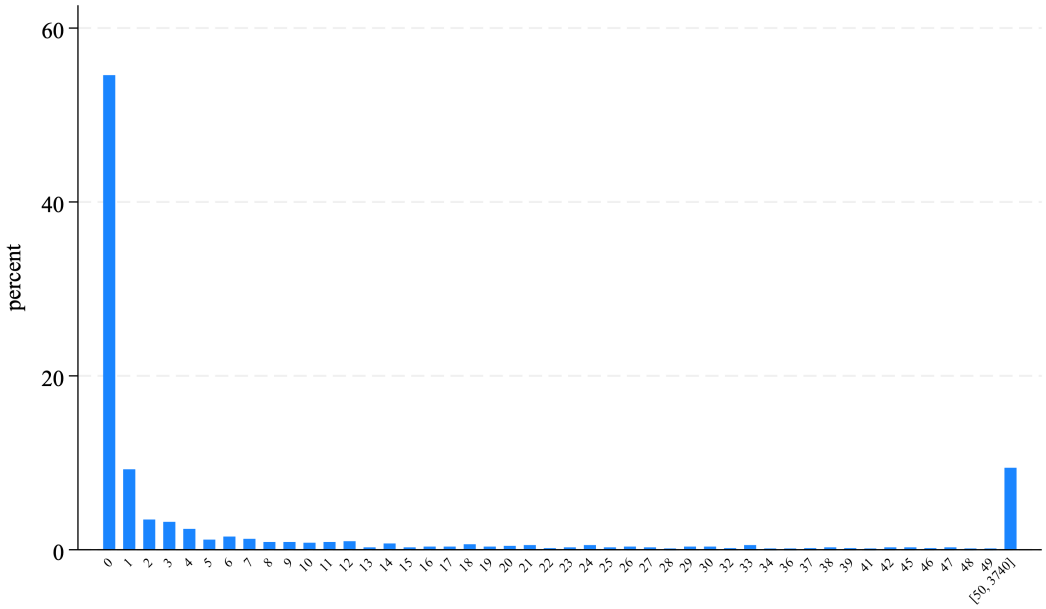
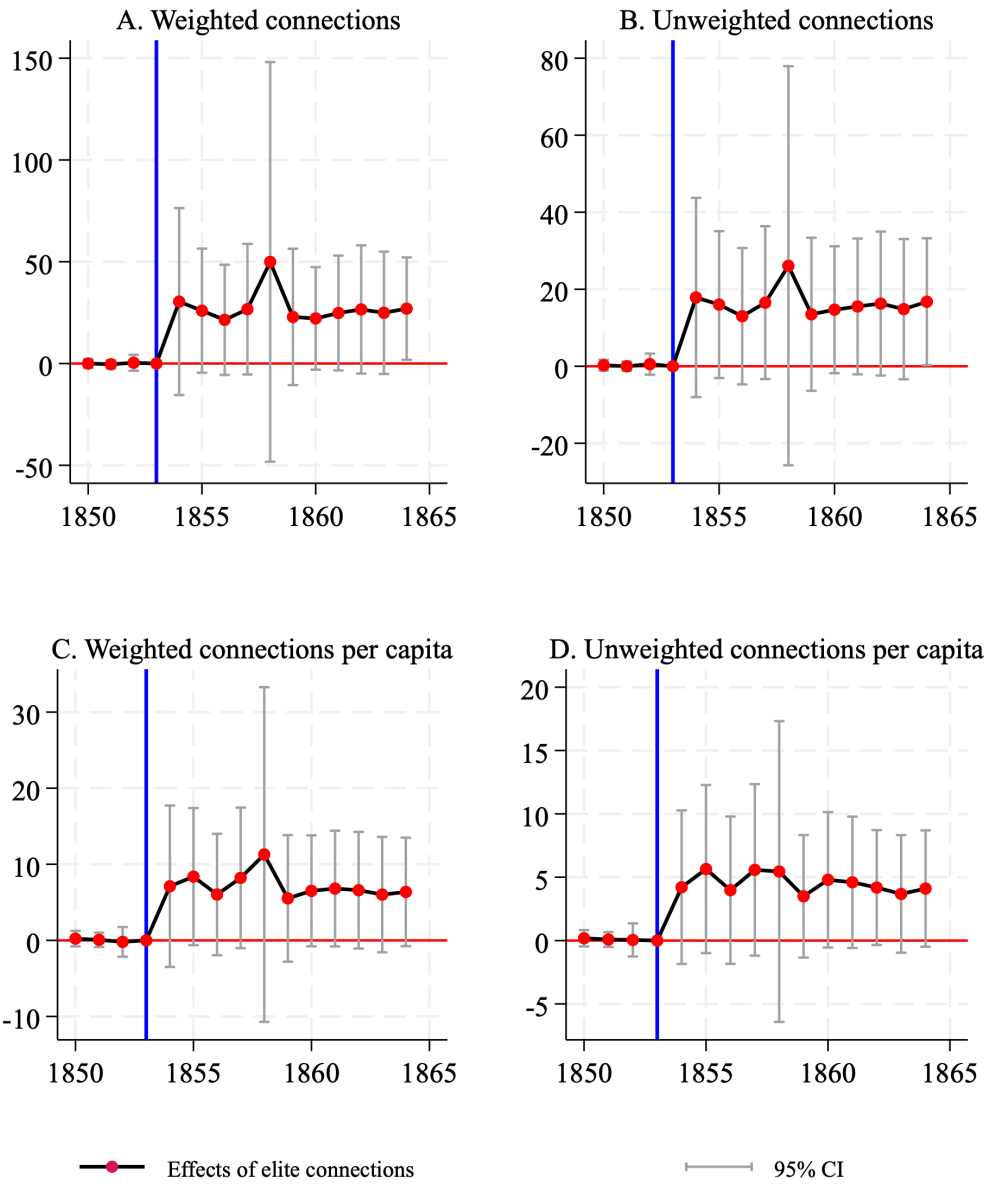
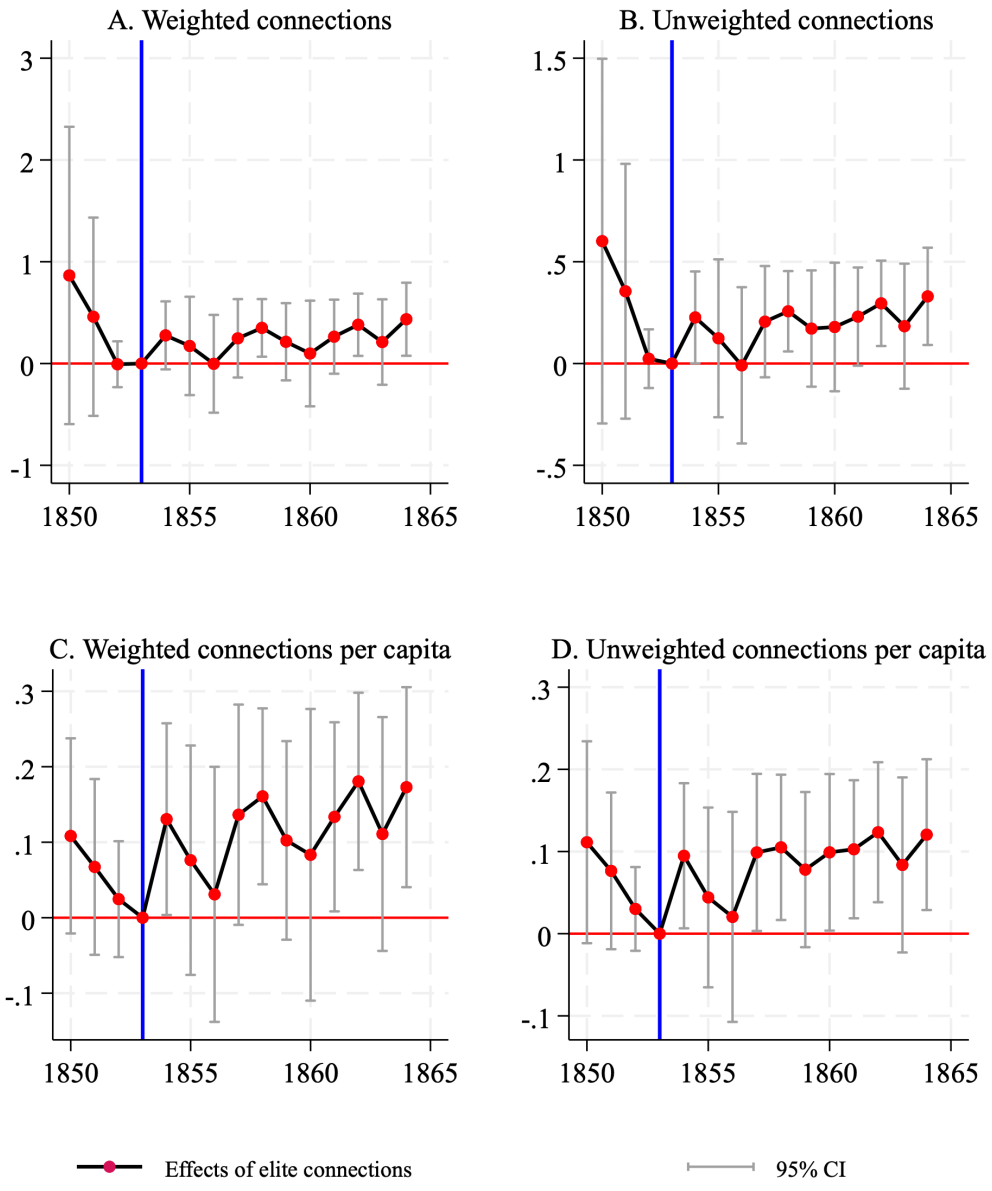


Figure 1: Distribution of soldier deaths (soldier deaths ≤ 50, Hunan counties, $N = 1,125$).



Notes: This Figure replicates Figure V of the paper, using an OLS model. Dependent variable is the untransformed outcome (soldier deaths).

Figure 2: The Effect of Elite Connections on Soldier Deaths, Year-by-Year Estimates (Figure V) – OLS.



Notes: This Figure replicates Figure V of the paper, using a Poisson distribution model. Dependent variable is the untransformed outcome (soldier deaths).

Figure 3: The Effect of Elite Connections on Soldier Deaths, Year-by-Year Estimates (Figure V) – Poisson model.

6 Tables

Table 1: The Effect of Elite Connections on Soldier Death (Table II) - OLS

	Dep. Var.: Soldier deaths									
	Connections measured by: $\sum_{n=1}^{N_c} \frac{1}{d_{c,n}}$						Connections measured by: N_c			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Baseline connections × post	12.147	14.342	19.027	19.417						
	(0.283)	(0.255)	(0.180)	(0.167)						
Baseline connections per capita × post					3.616	4.932				
					(0.241)	(0.188)				
Baseline connections × post							7.482	11.661		
							(0.278)	(0.164)		
Baseline connections per capita × post									2.503	3.348
									(0.245)	(0.172)
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
County FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Geographic-economic var. × post		Y	Y	Y		Y		Y		Y
Political var. × post			Y	Y		Y		Y		Y
Taiping var. × post				Y		Y		Y		Y
Observations	1,125	1,125	1,125	1,125	1,125	1,125	1,125	1,125	1,125	1,125
pseudo R-squared	0.334	0.346	0.353	0.356	0.332	0.352	0.333	0.353	0.332	0.351

Notes: This table replicates Table II of the paper using an OLS model. Dependent variable is the untransformed outcome variable (soldier deaths). Standard errors are clustered at the county level. p-values are displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: The Effect of Elite Connections on Soldier Deaths (Table II) - Poisson model

	Dep. Var.: Soldier deaths									
	Connections measured by: $\sum_{n=1}^{N_c} \frac{1}{d_{c,n}}$						Connections measured by: N_c			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Baseline connections × post	0.285*	0.215**	0.136	0.138*						
	(0.069)	(0.014)	(0.131)	(0.089)						
Baseline connections per capita × post					0.104*	0.047				
					(0.069)	(0.188)				
Baseline connections × post							0.205	0.086		
							(0.114)	(0.185)		
Baseline connections per capita × post									0.075*	0.029
									(0.097)	(0.286)
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
County FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Geographic-economic var. × post		Y	Y	Y		Y		Y		Y
Political var. × post			Y	Y		Y		Y		Y
Taiping var. × post				Y		Y		Y		Y
Observations	1,110	1,110	1,110	1,110	1,110	1,110	1,110	1,110	1,110	1,110
pseudo R-squared	0.689	0.702	0.704	0.705	0.689	0.705	0.689	0.704	0.689	0.704

Notes: This table replicates Table II of the paper using a Poisson distribution model. Dependent variable is the untransformed outcome variable (soldier deaths). Standard errors are clustered at the county level. p-values are displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: The Effect of Elite Connections on Soldier Death, Type of Links (Table III) - OLS

	Dep. Var.: Soldier deaths						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Expanded Network × post	33.954***						
	(0.004)						
Blood,mariage and friends × post		51.020***					
		(0.000)					
Provincial-level exam connections × post			-2.381				
			(0.842)				
National-level exam connections × post				11.096			
				(0.663)			
Same-surname baseline connections × post					1.399**	1.551**	1.346**
					(0.019)	(0.020)	(0.023)
Diff-surname baseline connections × post						0.338***	
						(0.008)	
County FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
Controls × post	Y	Y	Y	Y	Y	Y	
Pref FE × year FE	Y	Y	Y	Y	Y	Y	
Year FE × surname FE					Y	Y	Y
County FE × surname FE					Y	Y	Y
Year FE × county FE							Y
Observations	1,110	1,110	1,110	1,110	48,495	48,495	48,480
R-squared	0.465	0.487	0.421	0.421	0.318	0.320	0.398

Notes: This table replicates Table III of the paper using an OLS model. Dependent variable is the untransformed outcome variable (soldier deaths). Standard errors are clustered at the county level. p-values are displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: The Effect of Elite Connections on Soldier Death, Type of Links (Table III) - Poisson distribution model

	Dep. Var.: Soldier deaths						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Expanded Network × post	0.122**						
	(0.016)						
Blood, marriage and friends × post		0.096*					
		(0.050)					
Provincial-level exam connections × post			0.308				
			(0.140)				
National-level exam connections × post				1.995***			
				(0.006)			
Same-surname baseline connections × post					0.438***	0.629***	0.203*
					(0.003)	(0.002)	(0.082)
Diff-surname baseline connections × post						0.310***	
						(0.010)	
County FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
Controls × post	Y	Y	Y	Y	Y	Y	
Pref FE × year FE	Y	Y	Y	Y	Y	Y	
Year FE × surname FE					Y	Y	Y
County FE × surname FE					Y	Y	Y
Year FE × county FE							Y
Observations	910	910	910	910	35,341	35,341	24,363
pseudo R-squared	0.853	0.853	0.853	0.854	0.702	0.702	0.769

Notes: This table replicates Table III of the paper using Poisson distribution model. Dependent variable is the untransformed outcome variable (soldier deaths). Standard errors are clustered at the county level. p-values are displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: The Effect of Elite Connections on Elite Power (Table V) - Distance with Hunan by deciles of distance (Sample: Non-Hunan Counties)

	Dep. Var.: National-level offices					
	Min. distance		Mean distance		Topographic distance	
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline x Post x Distance	-0.007		-0.001		-0.027	
	(0.767)		(0.982)		(0.272)	
Baseline x Post	0.058		0.017		0.345	
	(0.719)		(0.948)		(0.247)	
Distance x Post	-0.001		-0.007		-0.011	
	(0.971)		(0.845)		(0.683)	
Baseline x Post x 10% Close		0.047		0.045		0.029***
		(0.100)		(0.114)		(0.000)
Baseline x Post x 90% Far		0.010		0.010		0.001
		(0.385)		(0.382)		(0.970)
10% Close x Post		-0.024		-0.018		-0.035
		(0.202)		(0.298)		(0.303)
Year FE	Y	Y	Y	Y	Y	Y
County FE	Y	Y	Y	Y	Y	Y
Controls × Period	Y	Y	Y	Y	Y	Y
Observations	142,961	142,961	142,961	142,961	142,961	142,961
R-squared	0.388	0.389	0.388	0.389	0.389	0.389

Notes: This table replicates column (4) of Table V of the paper by adding interactions with distance. Dependent variable is the untransformed outcome variable (soldier deaths). Post is a dummy variable equals to 1 after 1853. Distance is computed as the mean of the geodesic distances between a non-Hunan county and all Hunan counties using county coordinates from the original data. Topographic distance is computed as: $(1.6 \times alt) + Distance$. To compute alt for a non-Hunan county c, we do as follows. First, we compute S, the sum of the altitudes on the line between c and a Hunan county. Then we compute the mean of S between c and all Hunan counties.
Column (1) displays the results using the log of the distance to the closest Hunan county, column (3) displays using the log of average distance to Hunan counties and column (5) using the log of the average topographic distance to Hunan counties. Column (2) displays the results using a dummy equals to 1 if the county is in the first decile (10% Close) vs. the 9 last deciles (90 % Far) of minimum distance to Hunan. Columns (4) and (6) do the same respectively with deciles of mean distance to Hunan and topographic distance to Hunan.
Standard errors are clustered at the prefecture level. p-values are displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A Appendix

Table A1: The Effect of Elite Connections on Soldier Deaths (Table II) - Poisson model (robust)

	Dep. Var.: Soldier deaths									
	Connections measured by: $\sum_{n=1}^{N_c} \frac{1}{d_{c,n}}$						Connections measured by: N_c			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Baseline connections × post	0.285**	0.215*	0.136	0.138						
	(0.048)	(0.051)	(0.196)	(0.181)						
Baseline connections per capita × post					0.104**	0.047				
					(0.042)	(0.261)				
Baseline connections × post							0.205	0.086		
							(0.114)	(0.280)		
Baseline connections per capita × post									0.075*	0.029
									(0.097)	(0.368)
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
County FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Geographic-economic var. × post		Y	Y	Y		Y		Y		Y
Political var. × post			Y	Y		Y		Y		Y
Taiping var. × post				Y		Y		Y		Y
Observations	1,110	1,110	1,110	1,110	1,110	1,110	1,110	1,110	1,110	1,110
pseudo R-squared	0.689	0.702	0.704	0.705	0.689	0.705	0.689	0.704	0.689	0.704

Notes: This table replicates Table II of the paper using a Poisson distribution model with robust standard errors. Dependent variable is the untransformed outcome variable (soldier deaths). p-values are displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: The Effect of Elite Connections on Soldier Deaths (Table II) - Negative Binomial

	Dep. Var.: Soldier deaths									
	Connections measured by: $\sum_{n=1}^{N_c} \frac{1}{d_{c,n}}$						Connections measured by: N_c			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Soldier deaths										
Baseline connections × post	0.138*	0.270	0.306	0.314						
	(0.055)	(0.107)	(0.131)	(0.139)						
Baseline connections per capita × post					0.028	0.077				
					(0.220)	(0.244)				
Baseline connections × post							0.090*	0.226		
							(0.087)	(0.148)		
Baseline connections per capita × post									0.022	0.062
									(0.169)	(0.228)
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
County FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Geographic-economic var. × post		Y	Y	Y		Y		Y		Y
Political var. × post			Y	Y		Y		Y		Y
Taiping var. × post				Y		Y		Y		Y
Observations	1,125	1,125	1,125	1,125	1,125	1,125	1,125	1,125	1,125	1,125
pseudo R-squared	0.108	0.111	0.114	0.114	0.108	0.113	0.108	0.114	0.108	0.113

Notes: This table replicates Table II of the paper using a negative binomial model. Dependent variable is the untransformed outcome variable (soldier deaths). Standard errors are clustered at the county level. p-values are displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: The Effect of Elite Connections on Soldier Deaths: Type of Links (Table III) - Poisson model (robust)

	Dep. Var.: Soldier deaths						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Expanded Network × post	0.122*						
	(0.066)						
Blood, marriage and friends × post		0.096					
		(0.186)					
Provincial-level exam connections × post			0.308				
			(0.114)				
National-level exam connections × post				1.995***			
				(0.005)			
Same-surname baseline connections × post					0.438**	0.629***	0.203
					(0.025)	(0.001)	(0.120)
Diff-surname baseline connections × post						0.310***	
						(0.000)	
County FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
Controls × post	Y	Y	Y	Y	Y	Y	
Pref FE × year FE	Y	Y	Y	Y	Y	Y	
Year FE × surname FE					Y	Y	Y
County FE × surname FE					Y	Y	Y
Year FE × county FE							Y
Observations	910	910	910	910	35,341	35,341	24,363
pseudo R-squared	0.853	0.853	0.853	0.854	0.702	0.702	0.769

Notes: This table replicates Table III of the paper using Poisson distribution model with robust standard errors. Dependent variable is the untransformed outcome variable (soldier deaths). p-values are displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: The Effect of Elite Connections on Soldier Deaths (Table II) - ln(Soldier deaths + 1) with Pref FE × Year FE.

	Dep. Var.: ln(Soldier deaths +1)									
	Connections measured by: $\sum_{n=1}^{N_c} \frac{1}{d_{c,n}}$						Connections measured by: N_c			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Baseline connections × post	0.167**	0.174**	0.283***	0.261***						
	(0.017)	(0.032)	(0.001)	(0.001)						
Baseline connections per capita × post					0.065**	0.088***				
					(0.032)	(0.001)				
Baseline connections × post							0.114**	0.182***		
							(0.028)	(0.002)		
Baseline connections per capita × post									0.045*	0.063***
									(0.081)	(0.006)
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
County FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Pref FE × year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Geographic-economic var. × post		Y	Y	Y		Y		Y		Y
Political var. × post			Y	Y		Y		Y		Y
Taiping var. × post				Y		Y		Y		Y
Observations	1110.000	1110.000	1110.000	1110.000	1110.000	1110.000	1110.000	1110.000	1110.000	1110.000
R-squared	0.635	0.651	0.667	0.669	0.637	0.670	0.635	0.668	0.636	0.669

Notes: This table replicates Table II of the paper using an OLS model, adding prefecture-by-year fixed effects. Dependent variable is ln(Soldier deaths + 1). Standard errors are clustered at the county level. p-values are displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: The Effect of Elite Connections on Soldier Death, Type of Links (Table III) – ln(Soldier deaths + 1) without Pref FE × Year FE.

	Dep. Var.: Soldier deaths						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Expanded Network × post	0.152***						
	(0.000)						
Blood, marriage and friends × post		0.160***					
		(0.000)					
Provincial-level exam connections × post			0.239***				
			(0.010)				
National-level exam connections × post				0.352**			
				(0.028)			
Same-surname baseline connections × post					0.228***	0.222***	0.214***
					(0.002)	(0.004)	(0.001)
Diff-surname baseline connections × post						0.039**	
						(0.011)	
County FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
Controls × post	Y	Y	Y	Y	Y	Y	
Year FE × surname FE					Y	Y	Y
County FE × surname FE					Y	Y	Y
Year FE × county FE							Y
Observations	1,125	1,125	1,125	1,125	48,495	48,495	48,480
R-squared	0.479	0.475	0.475	0.471	0.392	0.396	0.616

Notes: This table replicates Table III of the paper using an OLS model, removing prefecture-by-year fixed effects. Dependent variable is ln(Soldier deaths + 1). Standard errors in parentheses are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.