

The Territorial Expansion of the Colonial State: Evidence  
from German East Africa 1890–1909

**Supplementary Information**

May 31, 2016

# Contents

<b>1</b>	<b>Historical Sources</b>	<b>3</b>
1.1	Violence . . . . .	3
1.2	Violence, Alternative . . . . .	3
1.3	Cotton and Rubber . . . . .	3
1.4	Ethnic group size and integration . . . . .	4
1.5	Security personnel . . . . .	4
1.6	Main roads and caravan routes . . . . .	4
1.7	Location of German stations . . . . .	5
1.8	Population density and presence of the Tsetse Fly . . . . .	5
1.9	Precipitation and Temperature . . . . .	5
1.10	Construction of the Territorial Value Variable . . . . .	7
<b>2</b>	<b>Summary Statistics</b>	<b>9</b>
<b>3</b>	<b>Distance to the Nearest Station</b>	<b>10</b>
<b>4</b>	<b>Alternative Station Type</b>	<b>12</b>
<b>5</b>	<b>Alternative Grid Cells</b>	<b>14</b>
<b>6</b>	<b>Ethnic Centralization</b>	<b>19</b>
<b>7</b>	<b>Christian Missions</b>	<b>24</b>
<b>8</b>	<b>Post-Treatment Bias</b>	<b>26</b>
<b>9</b>	<b>Battle Index Sub-Components</b>	<b>28</b>
<b>10</b>	<b>Fixed Effects</b>	<b>30</b>
<b>11</b>	<b>Lagged Dependent Variable</b>	<b>35</b>
<b>12</b>	<b>Instrumental Variables</b>	<b>37</b>

13 Plantations	39
14 Longterm Effects: Nightlights	41
15 Motives for state expansion according to the German administration	43
16 Cost considerations in the expansion of the German presence	45
17 State expansion after violent conflict – reactions to the Maji Maji rebellion	46

# 1 Historical Sources

This section provides information on the historical sources used for the measurement of the main outcome, explanatory and control variables. All maps, periodicals and monographs have been obtained from the “Berlin State Library – Prussian Cultural Heritage” (Staatsbibliothek zu Berlin – Preussischer Kulturbesitz), the “ZBW Leibniz Information Centre for Economics” (Leibniz- Informationszentrum Wirtschaft) in Hamburg and the German Federal Archives in Berlin. Geo-processing of maps has been made using QGIS (2.0.1).

## 1.1 Violence

In 1911, Ernst Nigmann, a former major in the German colonial force has published a book on the German military in German East-Africa. His account contains a detailed list of all violent encounters of the military between 1889 and 1910 (Nigmann 1911). We ge-located the events using the “German Colonial Atlas”. The atlas was published in 1920 by the Colonial Department of the Imperial Foreign Ministry (Kolonialabteilung des Auswärtigen Amtes). The atlas consists of nine over-sized map sheets (1:1.000.000). We geo-referenced the maps using a total of 129 markings of the coordinate grid. We were able to locate a total of 166 events using a complete village register accompanying the atlas.

## 1.2 Violence, Alternative

Nigmann has also identified and roughly located seven distinct periods of engagement of the German military, each comprising various local eruptions of violence. His book provides maps displaying the approximate location of all of these phases of violence as polygons sketched on hand-drawn maps. We geo-referenced these maps (using landmarks as the maps do not contain coordinate grids) and digitalized the polygons. Grids are coded “1” if their centroids are located within the boundaries of these conflict polygons.

## 1.3 Cotton and Rubber

Data on the location of cotton plantations and rubber forests stems from the “Economic Atlas of the German Colonies” (Wirtschaftsatlas der deutschen Kolonien) prepared by the “Colonial Economic Committee” (Kolonialwirtschaftliches Komitee) in 1906. The map number 7 of the atlas “Economic and Commerce Map of German East-Africa” (Wirtschafts- und Verkehrskarte von Deutsch-Ostafrika; 1:5.000.000) displays the location of the production sites of all main agricultural goods in the colony. We geo-referenced the map using all 43 visible markings of the coordinate grid and created two separate vector layers pin-

pointing the location of all rubber (89) and all cotton (102) producing sites displayed on the map.

#### **1.4 Ethnic group size and integration**

We use information from the Military Orientation Book for German East-Africa (“Militarisches Orientierungsheft für Deutsch-Ostafrika”), which was prepared by the German military and published in 1911. According to the introductory remarks by Kurt Freiherr von Schleinitz, commander of the German Schutztruppe from 1907/14, the book aims to provide “a picture of the colony as needed by soldiers in German East Africa.” It includes information on the size of ethnic groups. We used a map created by the British colonial government in 1956 to attribute ethnic groups to locations. It is highly congruent to a map provided by the German Karl Weule in 1906, probably the only German “ethnologue” who travelled to the colony (Rottland 2003). The former map has the advantage that it provides delineated settlement areas, whereas the latter only places names of ethnic groups on specific areas of the map. Information on the number of military personnel is taken from the statistical yearbooks of the German colonial administration and attributed to the respective groups main settlement areas (each grid cell is assigned the number of military personnel of the ethnic group whose settlement areas the grid cell’s centroid is located on).

#### **1.5 Security personnel**

The security apparatus consisted of a small number German officers and a much larger African fighting force. We consider the total number of African security personnel in early 1889 as a measure of ethnic group integration into the colonial administration. Information on the number of troops per district/German station stems from two sources: from the yearly report of the “Imperial Colonial Office” 1906/07 (data for 1905) and from a monograph written on the Maji Maji rebellion by the former governor of the colony Gustav Adolf Graf von Gtzen (Gtzen 1909). We have attributed these data to the exact locations of German stations and villages.

#### **1.6 Main roads and caravan routes**

Information on the location and length of main roads and traditional caravan routes stem from two colonial maps. The edited volume “The German Colonial Empire” (Das Deutsche Kolonialreich - Eine Länderkunde der deutschen Schutzgebiete) published by Hans Meyer in 1909/1910 is one of the most detailed historical descriptions of economic, natural and social conditions of the German colonies. We geo-referenced the “Transportation and Ad-

ministration Map of German East-Africa” (Verkehrs- und Verwaltungskarte von Deutsch-Ostafrika; 1:2.000.000) included in the book with 37 visible markings of the coordinate grid displayed on the map. Finally, we created a line layer representing all main roads and summed the road length per grid cell. We proceeded in a similar way for the caravan routes with a map prepared in 1892 (1:5.000.000), geo-referenced with 153 markings. The map has been published in the monograph “History of the Arab Uprising in East-Africa” (Geschichte des Araberaufstands in Ost-Afrika. Seine Entstehung, seine Niederwerfung und seine Folgen) by Rochus Schmidt (1892).

## 1.7 Location of German stations

German stations varied according to their status. Regular stations (“Bezirksämter” and “Militärstationen”) were the actual military and administrative hubs. They were assisted by smaller auxiliary stations and posts with more limited financial capacities and manpower (“Bezirksnebenstellen”) as well as small and temporary military posts (“Militärposten”). We geo-located regular and auxiliary stations in from 1890-1909 using the colonial atlas (see above).

## 1.8 Population density and presence of the Tsetse Fly

Rough estimates of the African population density are based on two maps taken from the “German Colonial Lexicon” (Deutsches Koloniallexikon) published in 1920 by the former governor of German East-Africa Heinrich Schnee; the maps themselves are not dated. We have georeferenced the map on the population density (1:15.000.000) using landmarks as the maps does not contain coordinate grids. The map displays rough polygons of four density levels (1-5; 5-10; 10-50; more than 50 people per square kilometer). While the map does not provide a reliable source for accurate numerical estimates of the actual population density it can be used to roughly indicate areas of relatively higher or lower density. We have attributed ascending numbers for the 4 density-levels to each grid according to spatial overlaps of grid centroids and aforementioned polygons. The map on Tsetse Fly presence (1:5.000.000) has been geo-referenced using 26 markings of the coordinate grid. A total of 1357 individual points on the map represent the Tsetse Fly density. We have digitized these points and calculated counts per grid cell.

## 1.9 Precipitation and Temperature

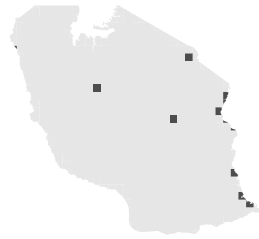
German statistical yearbooks contain information on precipitation and temperature for a number of weather stations. Unfortunately, reliable extrapolation of information across all

parts of the colony is not viable, as too few such stations existed. We compared climate information from a total of 41 stations across the colony with current climate information from the WorldClim project (1950–2000) for the exact same locations. Precipitation correlates at 0.85 and temperature at 0.9. While correlation is not perfect, we believe that using highly disaggregated current data provides a more accurate representation of spatial climate variations than extrapolations of historical data. We aggregated WorldClim data per grid cell.

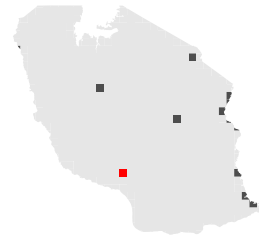
## 1.10 Construction of the Territorial Value Variable

Figure 1 visualizes the construction of the territorial control value variable.

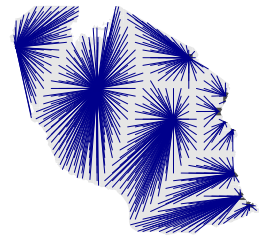




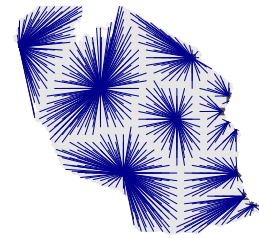
(a) Stations in 1890



(b) Adding a Station



(c) Actual Minimum Distances



(d) Hypothetical Minimum Distances

Figure 1: (a) Shows the actual station network in 1890. (c) Visualizes the minimum distances from each grid to the closest actual station. (b) Indicates the position of a hypothetical new station. (d) Visualizes how minimum distances adjust. Our measure is based on the average length of minimum distances from panel (c) minus the average length of distances from panel (d). As can be seen from the example figures, adding a station in the Southwestern part of the colony would dramatically reduce distances for many grid cells to the nearest station.

## 2 Summary Statistics

Table 1: Summary statistics

	count	mean	sd	min	max
Area	8550.00	0.17	0.06	0.00	0.20
Caravan Road Length	8550.00	0.35	0.44	0.00	2.20
Road Length	8550.00	0.22	0.29	0.00	1.16
Ethnic Centralization	5472.00	1.36	0.75	0.00	2.67
Population Density	8550.00	1.72	0.76	1.00	4.00
Mean Elevation	8550.00	1032.81	480.76	8.24	2412.21
Annual Precipitation	8550.00	999.08	248.72	489.15	2020.03
Annual Mean Temp	8550.00	223.82	23.09	138.30	272.50
Border Distance	8550.00	116.34	100.79	0.47	394.99
Coastal Distance	8550.00	511.32	307.89	0.47	1168.10
Station Presence	8550.00	0.07	0.25	0.00	1.00
Distance to Nearest Station	8550.00	120.23	84.70	2.24	573.77
Postal Station	8550.00	0.04	0.20	0.00	1.00
Major Battles	8550.00	0.02	0.16	0.00	3.00
Distance to Nearest Major Battle	8550.00	333.69	232.77	2.19	1423.89
Battle Index	8550.00	0.01	0.52	-2.34	2.21
Soil Quality	8550.00	5.35	1.02	0.00	7.00
Cotton Price	8550.00	54.89	11.44	37.00	73.00
Rubber Suitability	8550.00	0.30	0.34	0.00	1.00
Rubber Price	8550.00	98.37	17.78	79.50	148.40
Territorial Control Value	8550.00	0.19	1.32	-0.66	5.34

### 3 Distance to the Nearest Station

Table 2: Distance to Nearest Station, OLS

	(1)	(2)	(3)
	Distance to Nearest Station	Distance to Nearest Station	Distance to Nearest Station
Distance Nearest Station, Spatial Lag, 1yr lag	0.926*** (0.0131)	0.904*** (0.0168)	0.980*** (0.0152)
Area	-15.20 (15.61)	-4.142 (16.71)	-0.374 (17.02)
Caravan Road Length	-5.037** (1.698)	-5.081** (1.754)	-3.729* (1.788)
Road Length	-9.806*** (2.292)	-11.67*** (2.413)	-12.67*** (2.772)
Disease Environment	0.134 (0.138)	0.202 (0.141)	0.0674 (0.147)
Number of Troops, 1889	-0.00741 (0.00820)	-0.00340 (0.00863)	0.00908 (0.00859)
Ethnic Population	0.00000423*** (0.00000115)	0.00000282* (0.00000116)	0.00000318** (0.00000112)
Population Density	-6.539*** (1.522)	-4.631** (1.533)	-4.433** (1.490)
Mean Elevation	0.0108 <sup>+</sup> (0.00559)	0.0120 <sup>+</sup> (0.00619)	0.00450 (0.00656)
Annual Precipitation	-0.0163*** (0.00323)	-0.0173*** (0.00356)	-0.00673 <sup>+</sup> (0.00382)
Annual Mean Temp	0.224* (0.0965)	0.233* (0.107)	0.0808 (0.114)
Border Distance	-0.0288** (0.00967)	-0.0226* (0.00990)	-0.0289** (0.0109)
Coastal Distance	-0.00150 (0.00531)	0.000933 (0.00557)	-0.00602 (0.00585)
Soil Quality	-5.466 (6.149)	3.084 (3.902)	-1.410 (2.845)
Soil Quality × Cotton price, 1yr lag	0.108 (0.114)	-0.0734 (0.0656)	-0.00983 (0.0468)
Rubber Suitability	30.10*** (7.728)	31.81*** (7.817)	20.61** (6.858)
Rubber Suit × Rubber price, 1yr lag	-0.199** (0.0768)	-0.209** (0.0766)	-0.124 <sup>+</sup> (0.0646)
Battle Index, 1yr lag	-0.546 (1.001)		
Territorial Control Value, 1yr lag	-4.903*** (0.563)		
Battle Index, 2yr lag		-6.387*** (0.906)	
Territorial Control Value, 2yr lag		-4.834*** (0.525)	
Battle Index, 5yr lag			0.900 (1.021)
Territorial Control Value, 5yr lag			-3.122*** (0.420)
Constant	-54.14 (43.43)	-122.6** (37.45)	-44.55 (37.12)
Year FE	✓	✓	✓
Observations	7074	6681	5502
R <sup>2</sup>	0.834	0.825	0.815
Log-Likelihood	-34370.2	-32167.0	-25725.6
AIC	68814.5	64406.0	51517.1

Clustered standard errors in parentheses

<sup>+</sup>  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

## 4 Alternative Station Type

Table 3: Station Type, Ordered Logit, Alternate Coding

	(1)	(2)	(3)
	Station Type	Station Type	Station Type
Area	-2.547 (2.364)	-2.726 (2.371)	-3.307 (2.425)
Caravan Road Length	0.274 (0.269)	0.270 (0.266)	0.251 (0.262)
Road Length	0.715* (0.313)	0.741* (0.312)	0.812* (0.316)
Disease Environment	-0.00195 (0.0193)	-0.00135 (0.0194)	-0.000505 (0.0199)
Number of Troops, 1889	-0.00179+ (0.000939)	-0.00181+ (0.000941)	-0.00171+ (0.000959)
Ethnic Population	-0.00000297+ (0.00000175)	-0.00000294+ (0.00000174)	-0.00000288+ (0.00000173)
Population Density	0.694** (0.213)	0.697** (0.213)	0.712** (0.217)
Mean Elevation	-0.000803 (0.000864)	-0.000819 (0.000867)	-0.000979 (0.000902)
Annual Precipitation	0.000893+ (0.000531)	0.000882+ (0.000530)	0.000768 (0.000544)
Annual Mean Temp	-0.00458 (0.0159)	-0.00490 (0.0159)	-0.00817 (0.0163)
Border Distance	0.00356** (0.00131)	0.00359** (0.00131)	0.00358** (0.00134)
Coastal Distance	-0.000242 (0.000647)	-0.000256 (0.000653)	-0.000129 (0.000681)
Soil Quality	-0.283 (0.271)	-0.291 (0.314)	-0.292 (0.270)
Soil Quality × Cotton price, 1yr lag	0.00561+ (0.00304)	0.00603 (0.00393)	0.00673+ (0.00345)
Rubber Suitability	-0.566 (0.786)	-0.856 (0.795)	-1.668* (0.801)
Rubber Suit × Rubber price, 1yr lag	-0.00544 (0.00725)	-0.00260 (0.00735)	0.00559 (0.00724)
Battle Index, 1yr lag	0.231*** (0.0635)		
Territorial Control Value, 1yr lag	0.495*** (0.0492)		
Battle Index, 2yr lag		0.228*** (0.0563)	
Territorial Control Value, 2yr lag		0.479*** (0.0482)	
Battle Index, 5yr lag			0.252** (0.0859)
Territorial Control Value, 5yr lag			0.412*** (0.0454)
Year FE	✓	✓	✓
Observations	7074	6681	5502
Log-Likelihood	-1452.6	-1434.8	-1373.6
AIC	2983.2	2945.6	2817.2

Clustered standard errors in parentheses

+  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

## 5 Alternative Grid Cells

Table 4: Station Presence, Logit, Grid Alternative 1

	(1)	(2)	(3)
	Station Presence	Station Presence	Station Presence
Area	-8.920* (4.409)	-8.575* (4.356)	-7.902+ (4.224)
Caravan Road Length	0.452 (0.525)	0.389 (0.533)	0.298 (0.556)
Road Length	2.586*** (0.614)	2.649*** (0.614)	2.745*** (0.629)
Disease Environment	0.0334 (0.0483)	0.0363 (0.0478)	0.0402 (0.0474)
Number of Troops, 1889	-0.00500 (0.00325)	-0.00498 (0.00316)	-0.00399 (0.00272)
Ethnic Population	-0.00000266 (0.00000259)	-0.00000279 (0.00000256)	-0.00000288 (0.00000251)
Population Density	1.276** (0.410)	1.199** (0.402)	1.095** (0.390)
Mean Elevation	-0.00280* (0.00122)	-0.00272* (0.00124)	-0.00279* (0.00130)
Annual Precipitation	0.00968 (0.00808)	0.00799 (0.00796)	0.00517 (0.00868)
Annual Mean Temp	-0.0493* (0.0206)	-0.0478* (0.0207)	-0.0484* (0.0200)
Border Distance	0.00846** (0.00322)	0.00809* (0.00324)	0.00734* (0.00323)
Coastal Distance	0.000432 (0.00151)	0.000635 (0.00146)	0.00111 (0.00146)
Soil Quality	-0.743+ (0.442)	-0.603 (0.384)	-0.779+ (0.455)
Cotton price, 1yr lag	-0.0227 (0.0263)	0.00733 (0.0259)	-0.00399 (0.0477)
Soil Quality × Cotton price, 1yr lag	0.00908* (0.00427)	0.00671+ (0.00407)	0.0109 (0.00894)
Rubber Suitability	1.590 (2.439)	0.406 (2.450)	-0.700 (2.310)
Rubber price, 1yr lag	0.0121 (0.0148)	-0.0209 (0.0155)	-0.0238 (0.0302)
Rubber Suitability × Rubber price, 1yr lag	-0.0284 (0.0199)	-0.0165 (0.0205)	-0.00607 (0.0187)
Battle Index, 1yr lag	0.606*** (0.147)		
Control Value, 1yr lag	2.155* (0.857)		
Battle Index, 2yr lag		0.463*** (0.130)	
Control Value, 2yr lag		1.668* (0.666)	
Battle Index, 5yr lag			0.315 (0.254)
Control Value, 5yr lag			0.882*** (0.131)
Constant	9.562 (6.542)	10.99+ (6.266)	11.71+ (6.168)
Observations	6984	6596	5432
Log-Likelihood	-866.1	-883.0	-891.4
AIC	1804.1	1835.9	1846.8

Clustered standard errors in parentheses  
+  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$



Table 5: Station Presence, Logit, Grid Alternative 2

	(1)	(2)	(3)
	Station Presence	Station Presence	Station Presence
Area	-10.76* (4.622)	-10.89* (4.611)	-10.91* (4.671)
Caravan Road Length	1.000 (0.636)	0.947 (0.640)	0.761 (0.649)
Road Length	3.716*** (0.735)	3.826*** (0.746)	4.167*** (0.788)
Disease Environment	-0.0377 (0.0569)	-0.0365 (0.0576)	-0.0303 (0.0585)
Number of Troops, 1889	-0.00524+ (0.00279)	-0.00513+ (0.00276)	-0.00445 (0.00276)
Ethnic Population	-0.00000195 (0.00000273)	-0.00000225 (0.00000270)	-0.00000294 (0.00000269)
Population Density	0.925* (0.439)	0.915* (0.452)	0.897+ (0.487)
Mean Elevation	-0.00346* (0.00142)	-0.00339* (0.00144)	-0.00342* (0.00148)
Annual Precipitation	0.000945 (0.0111)	0.000112 (0.0113)	-0.00249 (0.0115)
Annual Mean Temp	-0.0638* (0.0257)	-0.0620* (0.0257)	-0.0619* (0.0246)
Border Distance	0.00750* (0.00293)	0.00742* (0.00301)	0.00700* (0.00320)
Coastal Distance	0.00175 (0.00187)	0.00189 (0.00186)	0.00237 (0.00189)
Soil Quality	-0.717 (0.609)	-0.866 (0.710)	-1.073 (0.663)
Cotton price, 1yr lag	-0.0309 (0.0448)	-0.0201 (0.0560)	-0.0211 (0.0482)
Soil Quality × Cotton price, 1yr lag	0.0110 (0.00765)	0.0138 (0.00958)	0.0172* (0.00873)
Rubber Suitability	-2.213 (1.806)	-3.053 (1.961)	-3.456 (2.174)
Rubber price, 1yr lag	-0.00169 (0.0159)	-0.0450* (0.0182)	-0.0507+ (0.0282)
Rubber Suitability × Rubber price, 1yr lag	0.00475 (0.0174)	0.0130 (0.0187)	0.0170 (0.0199)
Battle Index, 1yr lag	0.631*** (0.190)		
Control Value, 1yr lag	2.031*** (0.514)		
Battle Index, 2yr lag		0.589*** (0.148)	
Control Value, 2yr lag		1.692*** (0.424)	
Battle Index, 5yr lag			0.490* (0.238)
Control Value, 5yr lag			0.934*** (0.139)
Constant	15.19+ (8.108)	18.67* (7.958)	19.22** (7.404)
Observations	7074	6681	5502
Log-Likelihood	-754.5	-764.7	-769.3
AIC	1581.0	1599.3	1602.6

Clustered standard errors in parentheses  
+  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 6: Station Presence, Logit, 40x40km Grid

	(1)	(2)	(3)
	Station Presence	Station Presence	Station Presence
Area	-9.906 (6.360)	-10.13 (6.281)	-11.14 <sup>+</sup> (6.094)
Caravan Road Length	0.516 (0.639)	0.447 (0.637)	0.174 (0.648)
Road Length	3.527*** (0.743)	3.623*** (0.755)	3.787*** (0.776)
Disease Environment	0.0277 (0.0565)	0.0307 (0.0568)	0.0387 (0.0579)
Number of Troops, 1889	-0.00541 <sup>+</sup> (0.00296)	-0.00528 <sup>+</sup> (0.00287)	-0.00429 (0.00268)
Ethnic Population	-0.00000502 <sup>+</sup> (0.00000275)	-0.00000522 <sup>+</sup> (0.00000270)	-0.00000575* (0.00000253)
Population Density	1.628*** (0.396)	1.604*** (0.402)	1.536*** (0.399)
Mean Elevation	-0.00137 (0.00124)	-0.00136 (0.00124)	-0.00140 (0.00118)
Annual Precipitation	0.0142 <sup>+</sup> (0.00799)	0.0129 (0.00808)	0.00972 (0.00833)
Annual Mean Temp	-0.0164 (0.0197)	-0.0157 (0.0195)	-0.0134 (0.0170)
Border Distance	0.00844** (0.00277)	0.00830** (0.00281)	0.00761** (0.00290)
Coastal Distance	-0.000668 (0.00154)	-0.000492 (0.00153)	0.0000222 (0.00152)
Soil Quality	-1.450*** (0.366)	-1.655*** (0.463)	-1.771*** (0.483)
Cotton price, 1yr lag	-0.0623 <sup>+</sup> (0.0348)	-0.0675 <sup>+</sup> (0.0385)	-0.0549 (0.0349)
Soil Quality × Cotton price, 1yr lag	0.0172** (0.00574)	0.0212*** (0.00642)	0.0227*** (0.00595)
Rubber Suitability	-0.121 (2.566)	-0.789 (2.580)	-1.365 (2.320)
Rubber price, 1yr lag	0.000600 (0.0157)	-0.0288 <sup>+</sup> (0.0166)	-0.0341 (0.0285)
Rubber Suitability × Rubber price, 1yr lag	-0.0135 (0.0218)	-0.00688 (0.0222)	-0.00191 (0.0197)
Battle Index, 1yr lag	0.619*** (0.161)		
Control Value, 1yr lag	2.504*** (0.654)		
Battle Index, 2yr lag		0.526*** (0.145)	
Control Value, 2yr lag		2.172*** (0.563)	
Battle Index, 5yr lag			0.400 <sup>+</sup> (0.212)
Control Value, 5yr lag			1.197*** (0.184)
Constant	4.736 (5.934)	7.894 (5.908)	7.498 (5.511)
Observations	10728	10132	8344
Log-likelihood	-935.6	-947.9	-971.6
AIC	1943.1	1965.8	2007.2

Clustered standard errors in parentheses  
<sup>+</sup>  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 7: Station Presence, Logit, 60x60km Grid

	(1)	(2)	(3)
	Station Presence	Station Presence	Station Presence
Area	-10.63** (3.255)	-10.42** (3.228)	-9.885** (3.200)
Caravan Road Length	0.829+ (0.434)	0.796+ (0.432)	0.703 (0.437)
Road Length	2.304*** (0.615)	2.326*** (0.604)	2.408*** (0.611)
Disease Environment	0.0881+ (0.0501)	0.0850+ (0.0501)	0.0794 (0.0500)
Number of Troops, 1889	-0.00239 (0.00243)	-0.00223 (0.00236)	-0.00181 (0.00220)
Ethnic Population	2.91e-08 (0.000000297)	4.39e-09 (0.000000295)	-1.81e-08 (0.000000302)
Population Density	0.766+ (0.456)	0.775+ (0.450)	0.795+ (0.443)
Mean Elevation	-0.00204 (0.00137)	-0.00194 (0.00136)	-0.00179 (0.00133)
Annual Precipitation	0.0123 (0.0113)	0.0120 (0.0112)	0.0105 (0.0114)
Annual Mean Temp	-0.0567** (0.0205)	-0.0546** (0.0205)	-0.0510** (0.0191)
Border Distance	0.00815* (0.00333)	0.00807* (0.00329)	0.00755* (0.00328)
Coastal Distance	-0.000412 (0.00203)	-0.000397 (0.00198)	-0.000217 (0.00188)
Soil Quality	-1.179** (0.454)	-0.931* (0.436)	-1.036* (0.430)
Cotton price, 1yr lag	-0.0676+ (0.0355)	-0.0232 (0.0383)	-0.0186 (0.0465)
Soil Quality × Cotton price, 1yr lag	0.0170** (0.00596)	0.0128* (0.00619)	0.0159+ (0.00888)
Rubber Suitability	0.116 (2.152)	-0.962 (2.152)	-1.936 (2.191)
Rubber price, 1yr lag	0.0116 (0.0141)	-0.0267+ (0.0157)	-0.0375 (0.0300)
Rubber Suitability × Rubber price, 1yr lag	-0.0136 (0.0197)	-0.00273 (0.0207)	0.00757 (0.0200)
Battle Index, 1yr lag	0.650*** (0.167)		
Control Value, 1yr lag	1.704*** (0.503)		
Battle Index, 2yr lag		0.555*** (0.145)	
Control Value, 2yr lag		1.462*** (0.366)	
Battle Index, 5yr lag			0.476+ (0.250)
Control Value, 5yr lag			0.950*** (0.135)
Constant	14.30* (6.223)	15.18* (6.082)	14.62** (5.519)
Year FE	✓	✓	✓
Observations	5022	4743	3906
Log-Likelihood	-750.0	-758.7	-758.0
AIC	1572.0	1587.4	1580.1

Clustered standard errors in parentheses  
+  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

## 6 Ethnic Centralization

While the analysis of the Murdoch pre-colonial ethnic centralization variable indicates that motives for indirect rule also mattered, we also would like to caution against the use of this variable in all circumstances. Specifically, we believe that the literature on indirect rule, especially when drawing on the measures provided by Murdoch, provides only an *ex post* explanation of indirect rule. By that we mean that eventually colonial rulers might have settled on arrangements of indirect rule with local groups that were sufficiently stratified, but this was often preceded by a process of shifting political alliances and endogenous change of the groups themselves (Mamdani himself indicates as much). Also in the case of German East Africa, political bargains with ethnic groups are struck and broken with quite some frequency and the German's assessment of local groups' capabilities adjust accordingly. More importantly, the colonial administration did not have reliable information on local ethnic groups that could have sufficiently informed on these issues. The administration tried to collect information on pre-colonial social and political factors to inform its strategies. For example, the Military Orientation Book for German East-Africa was designed as a military guidebook and should therefore reflect the most accurate information that was available at the time. As such, it provides a unique glance into Germanys own appraisal of the colony. We exploit the fact that the book provides information on ethnic groups by district. Since the settlement areas of ethnic groups often extended across multiple districts, we can compare German appraisals of single ethnic groups across different districts and officials. We compared assessments of factors that have likely been particularly relevant to the German administration (i.e., ethnic groups' centralization and stance towards the colonial regime) but also particularly hard to assess from the outside. As expected, valuations are often heterogeneous. The Wazeguha are characterized as bellicose and difficult to handle in Morogoro, while officials in Tanga stress that its members easily submit to the authority of the German administration. Officials in Mahenge stress the Wabena's proven war-making skills and highlight internal organization under influential headmen, while in Iringa the same group is said to be peaceful and without effective internal organization. Finally, officials in Tanga are optimistic that the Wabondei will follow state agents willingly, while these agents are said to be without any influence among the Wabondei in Pangani.

These differences in evaluations suggest two things. First, shifting accounts underscore how difficult it was for the colonial administration to acquire reliable and consistent information on ethnic groups internal organization that could have shaped colonial strategies. Also note that the Military Orientation Book was written in 1911 after a more than 20-year-long German presence. Information was likely much patchier in the early phase of

Table 8: Station Presence, Logit and Ordered Probit, Ethnic Centralization

	(1)	(2)	(3)	(4)	(5)	(6)
	Station Presence	Station Presence	Station Presence	Station Type	Station Type	Station Type
Area	-13.79* (5.794)	-14.38* (5.865)	-15.54* (6.378)	-5.563 <sup>+</sup> (3.057)	-5.798 <sup>+</sup> (3.028)	-6.671* (3.077)
Caravan Road Length	1.332 <sup>+</sup> (0.760)	1.306 <sup>+</sup> (0.744)	1.335 <sup>+</sup> (0.721)	0.825* (0.329)	0.809* (0.327)	0.806* (0.325)
Road Length	1.328 (1.020)	1.400 (1.031)	1.534 (1.025)	0.584 (0.396)	0.609 (0.398)	0.654 (0.407)
Disease Environment	0.0307 (0.0641)	0.0280 (0.0637)	0.0175 (0.0633)	0.0133 (0.0328)	0.0135 (0.0329)	0.0138 (0.0340)
Number of Troops, 1889	-0.00436 (0.00515)	-0.00409 (0.00511)	-0.00236 (0.00447)	-0.00169 (0.00146)	-0.00164 (0.00148)	-0.00133 (0.00157)
Ethnic Population	-0.00000228* (0.00000101)	-0.00000237* (0.00000100)	-0.00000247* (0.000000986)	-0.000000797* (0.000000404)	-0.000000815* (0.000000410)	-0.000000870* (0.000000419)
Population Density	0.714 (0.552)	0.713 (0.545)	0.682 (0.560)	0.140 (0.267)	0.154 (0.268)	0.176 (0.273)
Mean Elevation	-0.00992** (0.00320)	-0.00979** (0.00313)	-0.00955** (0.00305)	-0.00320** (0.00119)	-0.00324** (0.00119)	-0.00356** (0.00122)
Annual Precipitation	0.000531 (0.00196)	0.000598 (0.00190)	0.000588 (0.00169)	0.000772 (0.000834)	0.000738 (0.000832)	0.000524 (0.000829)
Annual Mean Temp	-0.177** (0.0557)	-0.175** (0.0532)	-0.169*** (0.0508)	-0.0564* (0.0223)	-0.0572** (0.0221)	-0.0635** (0.0225)
Border Distance	0.0143** (0.00487)	0.0145** (0.00495)	0.0136** (0.00527)	0.00494** (0.00173)	0.00505** (0.00176)	0.00502** (0.00184)
Coastal Distance	0.00807* (0.00329)	0.00828* (0.00322)	0.00874** (0.00316)	0.00238* (0.000982)	0.00239* (0.000990)	0.00265** (0.00102)
Soil Quality	-0.719 (0.538)	-0.924 (0.704)	-0.960 (0.702)	-0.747* (0.350)	-0.760 <sup>+</sup> (0.431)	-0.699 <sup>+</sup> (0.405)
Soil Quality × Cotton price, 1yr lag	0.0138 (0.00911)	0.0182 <sup>+</sup> (0.0102)	0.0193* (0.00873)	0.0103* (0.00440)	0.0110 <sup>+</sup> (0.00602)	0.0105* (0.00534)
Rubber Suitability	1.797 (2.552)	1.148 (2.387)	-1.425 (2.377)	0.433 (1.041)	0.238 (1.004)	-0.954 (1.003)
Rubber Suit × Rubber price, 1yr lag	-0.0415 <sup>+</sup> (0.0233)	-0.0354 (0.0238)	-0.0106 (0.0248)	-0.0164 <sup>+</sup> (0.00917)	-0.0143 <sup>+</sup> (0.00862)	-0.00216 (0.00857)
Ethnic Centralization	-2.592* (1.191)	-2.662* (1.154)	-2.779* (1.081)	-0.795* (0.373)	-0.800* (0.377)	-0.843* (0.380)
Battle Index, 1yr lag	0.451** (0.149)			0.216** (0.0723)		
Territorial Control Value, 1yr lag	2.350** (0.830)			0.599*** (0.0629)		
Battle Index, 2yr lag		0.338* (0.144)			0.164** (0.0628)	
Territorial Control Value, 2yr lag		2.081*** (0.531)			0.586*** (0.0621)	
Battle Index, 5yr lag			0.531* (0.268)			0.250* (0.111)
Territorial Control Value, 5yr lag			1.438*** (0.171)			0.541*** (0.0556)
Year FE	✓	✓	✓	✓	✓	✓
Observations	4482	4233	3486	4482	4233	3486
Log-Likelihood	-438.6	-442.3	-444.4	-809.1	-800.2	-759.3
AIC	951.3	956.7	954.7	1698.3	1678.4	1590.6

Clustered standard errors in parentheses

<sup>+</sup>  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 9: Station Presence, Logit and Ordered Probit, Ethnic Centralization Interaction 1

	(1)	(2)	(3)	(4)	(5)	(6)
	Station Presence	Station Presence	Station Presence	Station Type	Station Type	Station Type
Area	-13.70* (5.779)	-14.22* (5.868)	-15.41* (6.394)	-5.520+ (3.046)	-5.760+ (3.018)	-6.677* (3.076)
Caravan Road Length	1.337+ (0.763)	1.313+ (0.749)	1.341+ (0.727)	0.826* (0.329)	0.811* (0.328)	0.814* (0.330)
Road Length	1.317 (1.025)	1.381 (1.040)	1.514 (1.039)	0.570 (0.399)	0.593 (0.404)	0.628 (0.419)
Disease Environment	0.0304 (0.0642)	0.0280 (0.0640)	0.0192 (0.0637)	0.0132 (0.0327)	0.0137 (0.0329)	0.0150 (0.0340)
Number of Troops, 1889	-0.00425 (0.00511)	-0.00393 (0.00512)	-0.00220 (0.00447)	-0.00167 (0.00146)	-0.00161 (0.00149)	-0.00124 (0.00159)
Ethnic Population	-0.00000225* (0.00000101)	-0.00000233* (0.000000996)	-0.00000238* (0.000000987)	-0.000000763+ (0.000000401)	-0.000000779+ (0.000000405)	-0.000000839* (0.000000425)
Population Density	0.720 (0.553)	0.725 (0.544)	0.705 (0.567)	0.140 (0.267)	0.157 (0.267)	0.183 (0.272)
Mean Elevation	-0.00987** (0.00317)	-0.00976** (0.00310)	-0.00961** (0.00301)	-0.00321** (0.00119)	-0.00326** (0.00118)	-0.00363** (0.00122)
Annual Precipitation	0.000527 (0.00197)	0.000578 (0.00191)	0.000552 (0.00170)	0.000770 (0.000834)	0.000726 (0.000833)	0.000494 (0.000833)
Annual Mean Temp	-0.176** (0.0554)	-0.174*** (0.0528)	-0.170*** (0.0504)	-0.0565* (0.0222)	-0.0576** (0.0220)	-0.0645** (0.0223)
Border Distance	0.0143** (0.00488)	0.0144** (0.00496)	0.0136* (0.00533)	0.00492** (0.00174)	0.00504** (0.00177)	0.00501** (0.00186)
Coastal Distance	0.00797* (0.00327)	0.00819* (0.00320)	0.00857** (0.00314)	0.00234* (0.000980)	0.00237* (0.000989)	0.00260* (0.00102)
Soil Quality	-0.800 (0.535)	-0.925 (0.697)	-1.065 (0.692)	-0.784* (0.344)	-0.747+ (0.434)	-0.735+ (0.404)
Soil Quality × Cotton price, 1yr lag	0.0154+ (0.00918)	0.0183+ (0.00999)	0.0214* (0.00890)	0.0111** (0.00421)	0.0108+ (0.00605)	0.0113* (0.00537)
Rubber Suitability	1.919 (2.557)	1.142 (2.399)	-2.556 (2.665)	0.453 (1.043)	0.153 (1.011)	-1.510 (1.147)
Rubber Suit × Rubber price, 1yr lag	-0.0423+ (0.0231)	-0.0346 (0.0232)	0.00204 (0.0255)	-0.0164+ (0.00911)	-0.0131 (0.00845)	0.00401 (0.00942)
Ethnic Centralization	-2.548* (1.182)	-2.611* (1.138)	-2.641* (1.059)	-0.772* (0.372)	-0.778* (0.375)	-0.788* (0.377)
Battle Index, 1yr lag	0.686* (0.285)			0.360** (0.116)		
Ethnic Central × Battle Index, 1yr lag	-0.259 (0.238)			-0.148+ (0.0812)		
Territorial Control Value, 1yr lag	2.349** (0.829)			0.602*** (0.0630)		
Battle Index, 2yr lag		0.669* (0.297)			0.337* (0.134)	
Ethnic Central × Battle Index, 2yr lag		-0.373 (0.259)			-0.181+ (0.103)	
Territorial Control Value, 2yr lag		2.087** (0.525)			0.592*** (0.0626)	
Battle Index, 5yr lag			1.241** (0.438)			0.599** (0.227)
Ethnic Central × Battle Index, 5yr lag			-0.667* (0.294)			-0.320* (0.149)
Territorial Control Value, 5yr lag			1.464*** (0.168)			0.554*** (0.0579)
Year FE	✓	✓	✓	✓	✓	✓
Observations	4482	4233	3486	4482	4233	3486
Log-Likelihood	-438.1	-441.3	-441.4	-808.0	-798.5	-755.4
AIC	952.3	956.6	950.8	1698.0	1676.9	1584.8

Clustered standard errors in parentheses  
+  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 10: Station Presence, Logit and Ordered Probit, Ethnic Centralization Interaction 2

	(1)	(2)	(3)	(4)	(5)	(6)
	Station Presence	Station Presence	Station Presence	Station Type	Station Type	Station Type
Area	-13.37* (6.518)	-13.82* (6.583)	-14.50* (6.908)	-5.832+ (3.141)	-6.018+ (3.127)	-6.695* (3.222)
Caravan Road Length	1.340+ (0.737)	1.309+ (0.728)	1.333+ (0.713)	0.811* (0.320)	0.798* (0.318)	0.805* (0.319)
Road Length	1.382 (1.078)	1.465 (1.091)	1.642 (1.086)	0.543 (0.435)	0.577 (0.435)	0.651 (0.438)
Disease Environment	0.0279 (0.0645)	0.0247 (0.0642)	0.0101 (0.0637)	0.0157 (0.0350)	0.0154 (0.0350)	0.0140 (0.0359)
Number of Troops, 1889	-0.00408 (0.00528)	-0.00388 (0.00523)	-0.00215 (0.00442)	-0.00190 (0.00142)	-0.00180 (0.00143)	-0.00134 (0.00152)
Ethnic Population	-0.00000196+ (0.00000106)	-0.00000204+ (0.00000106)	-0.00000224* (0.00000105)	-0.000000794* (0.000000397)	-0.000000814* (0.000000404)	-0.000000870* (0.000000417)
Population Density	0.799 (0.577)	0.799 (0.573)	0.700 (0.572)	0.160 (0.252)	0.171 (0.253)	0.178 (0.260)
Mean Elevation	-0.0101** (0.00331)	-0.00989** (0.00320)	-0.00919** (0.00303)	-0.00336* (0.00135)	-0.00336* (0.00133)	-0.00358** (0.00134)
Annual Precipitation	0.000849 (0.00211)	0.000934 (0.00205)	0.000944 (0.00182)	0.000725 (0.000852)	0.000700 (0.000852)	0.000519 (0.000855)
Annual Mean Temp	-0.180** (0.0581)	-0.176** (0.0552)	-0.162*** (0.0490)	-0.0584* (0.0237)	-0.0589* (0.0234)	-0.0637** (0.0234)
Border Distance	0.0145** (0.00485)	0.0146** (0.00493)	0.0133* (0.00519)	0.00505** (0.00167)	0.00514** (0.00169)	0.00503** (0.00178)
Coastal Distance	0.00703* (0.00311)	0.00718* (0.00314)	0.00786* (0.00323)	0.00240* (0.000989)	0.00241* (0.000996)	0.00265** (0.00102)
Soil Quality	-0.710 (0.637)	-0.731 (0.853)	-0.884 (0.813)	-0.736* (0.354)	-0.753+ (0.439)	-0.698+ (0.407)
Soil Quality × Cotton price, 1yr lag	0.0159+ (0.00937)	0.0168 (0.0116)	0.0187+ (0.0102)	0.0105* (0.00441)	0.0111+ (0.00604)	0.0105* (0.00529)
Rubber Suitability	1.417 (2.683)	0.608 (2.562)	-2.320 (2.305)	0.434 (1.033)	0.252 (0.992)	-0.951 (0.989)
Rubber Suitability × Rubber price, 1yr lag	-0.0367 (0.0241)	-0.0291 (0.0235)	-0.000512 (0.0238)	-0.0170+ (0.00919)	-0.0149+ (0.00863)	-0.00224 (0.00865)
Ethnic Centralization	-2.513* (1.002)	-2.517* (1.002)	-2.549* (1.030)	-0.803* (0.379)	-0.808* (0.381)	-0.845* (0.382)
Territorial Control Value, 1yr lag	4.541** (1.562)			0.558*** (0.0897)		
Ethnic Central × Territorial Control Value, 1yr lag	-1.197 (0.729)			0.0438 (0.0717)		
Battle Index, 1yr lag	0.439** (0.169)			0.213** (0.0721)		
Territorial Control Value, 2yr lag		4.136** (1.419)			0.553*** (0.0904)	
Ethnic Central × Territorial Control Value, 2yr lag		-1.106 (0.690)			0.0358 (0.0724)	
Battle Index, 2yr lag		0.323* (0.162)			0.160* (0.0624)	
Territorial Control Value, 5yr lag			2.580** (0.993)			0.538*** (0.0900)
Ethnic Central × Territorial Control Value, 5yr lag			-0.653 (0.507)			0.00392 (0.0722)
Battle Index, 5yr lag			0.557+ (0.289)			0.249* (0.109)
Year FE	✓	✓	✓	✓	✓	✓
Observations	4482	4233	3486	4482	4233	3486
Log-Likelihood	-422.1	-426.8	-436.1	-808.4	-799.7	-759.3
AIC	920.1	927.6	940.2	1698.8	1679.4	1592.6

Clustered standard errors in parentheses  
+  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

the colonial project, when the pace of state expansion was highest. Second, contradictory accounts about ethnic groups in different regions might have also been due to real within-group heterogeneity – which is masked in existing measures like the pre-colonial level of political centralization of ethnic groups. Both scenarios support our argument that “objective” pre-colonial factors, like ethnic political centralization, did not solely shape processes of state-building, most notably because the German administration was very aware of the lack of reliable information and local heterogeneity.



## 7 Christian Missions

Table 11: Station Presence, Logit, Christian Missions

	(1)	(2)	(3)
	Station Presence	Station Presence	Station Presence
Area	-6.138 (5.417)	-6.102 (5.606)	-6.051 (5.681)
Caravan Road Length	0.729 (0.630)	0.688 (0.620)	0.569 (0.605)
Road Length	1.538* (0.671)	1.622* (0.668)	1.784** (0.673)
Disease Environment	-0.00493 (0.0410)	-0.00679 (0.0404)	-0.0131 (0.0392)
Number of Troops, 1889	-0.00487* (0.00246)	-0.00479* (0.00239)	-0.00402+ (0.00221)
Ethnic Population	-0.00000456 (0.00000326)	-0.00000474 (0.00000321)	-0.00000516+ (0.00000310)
Population Density	1.482*** (0.436)	1.444*** (0.436)	1.402** (0.454)
Mean Elevation	-0.00267 (0.00224)	-0.00251 (0.00218)	-0.00222 (0.00203)
Annual Precipitation	0.00142 (0.00107)	0.00136 (0.00107)	0.00112 (0.00110)
Annual Mean Temp	-0.0305 (0.0412)	-0.0269 (0.0398)	-0.0210 (0.0355)
Border Distance	0.00933** (0.00324)	0.00903** (0.00316)	0.00803* (0.00313)
Coastal Distance	0.0000242 (0.00171)	0.000153 (0.00169)	0.000495 (0.00169)
Soil Quality	-0.625 (0.566)	-0.665 (0.687)	-0.780 (0.548)
Cotton price, 1yr lag	-0.0512 (0.0564)	-0.0380 (0.0697)	-0.0110 (0.0575)
Soil Quality × Cotton price, 1yr lag	0.0155+ (0.00923)	0.0163 (0.0117)	0.0182* (0.00869)
Rubber Suitability	-2.133 (1.993)	-3.098 (2.085)	-3.975+ (2.057)
Rubber price, 1yr lag	-0.00638 (0.0149)	-0.0375* (0.0159)	-0.0627* (0.0303)
Rubber Suitability × Rubber price, 1yr lag	0.000267 (0.0191)	0.00988 (0.0198)	0.0196 (0.0187)
Battle Index, 1yr lag	0.623*** (0.169)		
Control Value, 1yr lag	2.298*** (0.582)		
Battle Index, 2yr lag		0.567*** (0.142)	
Control Value, 2yr lag		1.896*** (0.482)	
Battle Index, 5yr lag			0.555** (0.208)
Control Value, 5yr lag			1.039*** (0.151)
Christian Missions	-0.0847 (0.184)	-0.0841 (0.183)	-0.0671 (0.179)
Year FE	✓	✓	✓
Observations	7074	6681	5502
Log-Likelihood	-838.9	-857.0	-882.9
AIC	1751.9	1786.1	1831.9

Clustered standard errors in parentheses  
+  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

## 8 Post-Treatment Bias

Table 12: Station Presence, Logit, Exogenous Variables

	(1)	(2)	(3)
	Station Presence	Station Presence	Station Presence
Area	-5.394 (3.577)	-5.173 (3.624)	-4.975 (3.390)
Caravan Road Length	1.463*** (0.438)	1.447*** (0.435)	1.405** (0.434)
Disease Environment	-0.0101 (0.0391)	-0.0130 (0.0386)	-0.0199 (0.0384)
Mean Elevation	-0.00565** (0.00176)	-0.00541** (0.00172)	-0.00492** (0.00159)
Annual Precipitation	0.000322 (0.000907)	0.000303 (0.000918)	0.000120 (0.000940)
Annual Mean Temp	-0.0908** (0.0343)	-0.0851** (0.0329)	-0.0749** (0.0283)
Border Distance	0.00386 (0.00247)	0.00360 (0.00246)	0.00247 (0.00247)
Coastal Distance	0.00184 (0.00131)	0.00183 (0.00130)	0.00192 (0.00125)
Soil Quality	-0.285 (0.221)	-0.131 (0.241)	-0.0449 (0.242)
Cotton price, 1yr lag	0.0150 (0.0183)	0.0419+ (0.0218)	0.0805** (0.0296)
Soil Quality × Cotton price, 1yr lag	0.00262 (0.00295)	0.000208 (0.00342)	-0.000702 (0.00336)
Rubber Suitability	-0.294 (1.952)	-1.268 (2.034)	-2.255 (2.068)
Rubber price, 1yr lag	0.00332 (0.0123)	-0.0239+ (0.0143)	-0.0440 (0.0281)
Rubber Suitability × Rubber price, 1yr lag	-0.0110 (0.0183)	-0.00179 (0.0191)	0.00908 (0.0185)
Battle Index, 1yr lag	0.574*** (0.171)		
Control Value, 1yr lag	2.333*** (0.567)		
Battle Index, 2yr lag		0.545*** (0.149)	
Control Value, 2yr lag		1.930*** (0.462)	
Battle Index, 5yr lag			0.570** (0.211)
Control Value, 5yr lag			1.092*** (0.147)
Year FE	✓	✓	✓
Observations	8100	7650	6300
Log-Likelihood	-1074.1	-1098.9	-1138.1
AIC	2212.3	2259.8	2332.2

Clustered standard errors in parentheses

+  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

## 9 Battle Index Sub-Components

Table 13: Station Presence, Logit, Battle Index Components

	(1)	(2)	(3)	(4)	(5)	(6)
	Station Presence	Station Presence	Station Presence	Station Presence	Station Presence	Station Presence
Area	-5.847 (5.368)	-5.847 (5.585)	-5.980 (5.638)	-6.330 (5.340)	-6.250 (5.539)	-6.134 (5.615)
Caravan Road Length	0.627 (0.582)	0.592 (0.576)	0.520 (0.556)	0.666 (0.571)	0.626 (0.567)	0.520 (0.567)
Road Length	1.461* (0.685)	1.558* (0.678)	1.758** (0.677)	1.511* (0.674)	1.590* (0.671)	1.754** (0.676)
Disease Environment	-0.00287 (0.0418)	-0.00394 (0.0411)	-0.00811 (0.0391)	-0.00340 (0.0412)	-0.00523 (0.0406)	-0.0121 (0.0394)
Number of Troops, 1889	-0.00467+ (0.00246)	-0.00455+ (0.00239)	-0.00373+ (0.00219)	-0.00470* (0.00239)	-0.00461* (0.00233)	-0.00385+ (0.00216)
Ethnic Population	-0.00000496 (0.000000324)	-0.00000515 (0.000000320)	-0.00000545+ (0.000000310)	-0.00000461 (0.000000322)	-0.00000480 (0.000000317)	-0.00000525+ (0.000000309)
Population Density	1.543*** (0.444)	1.503** (0.444)	1.419** (0.454)	1.471*** (0.439)	1.441** (0.439)	1.413** (0.456)
Mean Elevation	-0.00257 (0.00221)	-0.00250 (0.00217)	-0.00239 (0.00204)	-0.00282 (0.00222)	-0.00263 (0.00217)	-0.00225 (0.00202)
Annual Precipitation	0.00114 (0.00106)	0.00108 (0.00107)	0.000946 (0.00109)	0.00131 (0.00104)	0.00123 (0.00105)	0.000986 (0.00109)
Annual Mean Temp	-0.0284 (0.0405)	-0.0261 (0.0392)	-0.0233 (0.0355)	-0.0321 (0.0409)	-0.0282 (0.0395)	-0.0210 (0.0352)
Border Distance	0.00979** (0.00321)	0.00947** (0.00313)	0.00864** (0.00311)	0.00978** (0.00319)	0.00947** (0.00312)	0.00839** (0.00310)
Coastal Distance	0.000239 (0.00169)	0.000457 (0.00167)	0.000849 (0.00165)	0.000294 (0.00168)	0.000419 (0.00166)	0.000653 (0.00165)
Soil Quality	-0.374 (0.616)	-0.556 (0.700)	-0.547 (0.571)	-0.519 (0.615)	-0.605 (0.723)	-0.713 (0.583)
Soil Quality × Cotton price, 1yr lag	0.0114 (0.00931)	0.0148 (0.0112)	0.0148+ (0.00852)	0.0142 (0.00961)	0.0158 (0.0118)	0.0173+ (0.00889)
Rubber Suitability	-1.798 (1.928)	-1.769 (1.939)	-2.986 (2.060)	-2.429 (2.076)	-3.037 (2.085)	-3.532+ (2.033)
Rubber Suit × Rubber price, 1yr lag	-0.00299 (0.0188)	-0.00307 (0.0185)	0.0102 (0.0187)	0.00411 (0.0199)	0.0101 (0.0198)	0.0158 (0.0185)
Major Battles, 1yr lag	1.204*** (0.263)					
Territorial Control Value, 1yr lag	2.269*** (0.587)			2.289*** (0.588)		
Major Battles, 2yr lag	1.332*** (0.221)					
Territorial Control Value, 2yr lag	1.870*** (0.480)			1.884*** (0.486)		
Major Battles, 5yr lag	1.641*** (0.363)					
Territorial Control Value, 5yr lag	1.028*** (0.150)			1.028*** (0.152)		
Distance Major Battle, 1yr lag	-0.00105** (0.000340)					
Distance Major Battle, 2yr lag	-0.000860** (0.000299)					
Distance Major Battle, 5yr lag	-0.000817+ (0.000437)					
Constant	3.573 (13.36)	7.118 (12.82)	2.413 (11.65)	5.719 (13.26)	6.316 (12.70)	5.391 (11.56)
Year FE	✓	✓	✓	✓	✓	✓
Observations	7074	6681	5502	7074	6681	5502
Log-Likelihood	-835.5	-849.7	-875.7	-843.9	-861.9	-886.1
AIC	1743.0	1769.4	1815.4	1759.8	1793.7	1836.1

Clustered standard errors in parentheses  
+  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

## 10 Fixed Effects

Table 14: Station Presence, Logit and Ordered Probit, Region Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
	Station Presence	Station Presence	Station Presence	Station Type	Station Type	Station Type
Area	-4.357 (5.209)	-4.421 (5.396)	-4.564 (5.545)	-1.855 (2.247)	-2.015 (2.258)	-2.627 (2.283)
Caravan Road Length	0.879 (0.559)	0.831 (0.560)	0.702 (0.569)	0.339 (0.279)	0.340 (0.277)	0.328 (0.272)
Road Length	1.238 <sup>+</sup> (0.747)	1.336 <sup>+</sup> (0.730)	1.598* (0.725)	0.694* (0.337)	0.720* (0.335)	0.787* (0.340)
Disease Environment	-0.0243 (0.0392)	-0.0252 (0.0387)	-0.0303 (0.0381)	-0.0124 (0.0190)	-0.0120 (0.0191)	-0.0111 (0.0197)
Number of Troops, 1889	-0.00665** (0.00246)	-0.00650** (0.00240)	-0.00599** (0.00231)	-0.00265* (0.00103)	-0.00272** (0.00104)	-0.00265* (0.00107)
Ethnic Population	-0.000000540 <sup>+</sup> (0.000000300)	-0.000000556 <sup>+</sup> (0.000000298)	-0.000000593* (0.000000292)	-0.000000356* (0.000000161)	-0.000000356* (0.000000160)	-0.000000360* (0.000000160)
Population Density	1.688** (0.592)	1.640** (0.586)	1.544** (0.573)	0.778** (0.276)	0.789** (0.276)	0.827** (0.282)
Mean Elevation	-0.00236 (0.00241)	-0.00223 (0.00238)	-0.00166 (0.00221)	-0.000885 (0.000890)	-0.000882 (0.000892)	-0.000995 (0.000916)
Annual Precipitation	0.00160 (0.00117)	0.00151 (0.00119)	0.00133 (0.00122)	0.000928 <sup>+</sup> (0.000564)	0.000923 (0.000566)	0.000825 (0.000582)
Annual Mean Temp	-0.0182 (0.0438)	-0.0149 (0.0435)	-0.00456 (0.0399)	-0.00341 (0.0168)	-0.00338 (0.0168)	-0.00630 (0.0171)
Border Distance	0.0106*** (0.00311)	0.0103*** (0.00309)	0.00921** (0.00310)	0.00378** (0.00142)	0.00385** (0.00142)	0.00404** (0.00146)
Coastal Distance	0.0000310 (0.00166)	0.000202 (0.00165)	0.000473 (0.00172)	-0.0000255 (0.000667)	-0.0000116 (0.000672)	0.000101 (0.000694)
Soil Quality	-0.589 (0.659)	-0.700 (0.780)	-0.738 (0.627)	-0.356 (0.358)	-0.388 (0.418)	-0.354 (0.352)
Soil Quality × Cotton price, 1yr lag	0.0143 (0.00997)	0.0163 (0.0127)	0.0168 <sup>+</sup> (0.00982)	0.00712 (0.00433)	0.00799 (0.00578)	0.00792 (0.00489)
Rubber Suitability	-2.193 (2.023)	-2.991 (2.051)	-3.891 <sup>+</sup> (2.025)	-0.780 (0.818)	-0.986 (0.821)	-1.529 <sup>+</sup> (0.817)
Rubber Suitability × Rubber price, 1yr lag	0.00311 (0.0202)	0.0107 (0.0201)	0.0211 (0.0187)	-0.00213 (0.00745)	-0.00000845 (0.00753)	0.00583 (0.00739)
Battle Index, 1yr lag	0.497** (0.165)			0.207*** (0.0570)		
Territorial Control Value, 1yr lag	2.565*** (0.699)			0.512*** (0.0449)		
Battle Index, 2yr lag		0.418** (0.131)			0.178*** (0.0479)	
Territorial Control Value, 2yr lag		2.108*** (0.610)			0.497*** (0.0442)	
Battle Index, 5yr lag			0.299 (0.201)			0.118 (0.0813)
Territorial Control Value, 5yr lag			1.075*** (0.166)			0.433*** (0.0430)
Year FE	✓	✓	✓	✓	✓	✓
Region FE	✓	✓	✓	✓	✓	✓
Observations	7074	6681	5502	7074	6681	5502
Log-Likelihood	-801.4	-820.7	-850.4	-1409.3	-1389.9	-1327.4
AIC	1684.9	1721.5	1774.8	2906.6	2865.9	2734.9

Clustered standard errors in parentheses

<sup>+</sup>  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$



Table 15: Station Presence, Logit and Ordered Probit, Ethnic Group FE

	(1)	(2)	(3)	(4)	(5)	(6)
	Station Presence	Station Presence	Station Presence	Station Type	Station Type	Station Type
Area	-3.217 (12.53)	-5.040 (12.52)	-8.296 (13.26)	-3.326 (4.402)	-3.689 (4.376)	-4.841 (4.464)
Caravan Road Length	0.519 (0.886)	0.529 (0.846)	0.528 (0.786)	0.397 (0.349)	0.404 (0.347)	0.392 (0.351)
Road Length	4.607*** (1.236)	4.413*** (1.179)	3.941*** (1.064)	1.410** (0.476)	1.442** (0.477)	1.566** (0.488)
Disease Environment	-0.130 (0.103)	-0.129 (0.0975)	-0.129 (0.0932)	-0.0605 (0.0372)	-0.0615+ (0.0374)	-0.0668+ (0.0386)
Number of Troops, 1889	-0.00189 (0.0152)	-0.00260 (0.0154)	-0.00190 (0.0170)	0.00552 (0.00459)	0.00554 (0.00467)	0.00563 (0.00455)
Ethnic Population	-0.000000927+ (0.000000483)	-0.000000924+ (0.000000487)	-0.000000971+ (0.000000520)	-0.000000625* (0.000000288)	-0.000000640* (0.000000294)	-0.000000683* (0.000000312)
Population Density	2.615 (2.019)	2.647 (2.032)	2.810 (2.234)	1.212* (0.603)	1.268* (0.595)	1.390* (0.621)
Mean Elevation	0.000992 (0.00667)	0.000993 (0.00666)	0.00103 (0.00731)	0.00312 (0.00275)	0.00316 (0.00280)	0.00336 (0.00295)
Annual Precipitation	0.00932+ (0.00490)	0.00826+ (0.00463)	0.00567 (0.00408)	0.00338** (0.00109)	0.00335** (0.00110)	0.00332** (0.00112)
Annual Mean Temp	0.0786 (0.110)	0.0718 (0.111)	0.0578 (0.123)	0.0690 (0.0468)	0.0696 (0.0477)	0.0730 (0.0505)
Border Distance	-0.0121 (0.0133)	-0.0107 (0.0135)	-0.00665 (0.0142)	-0.00144 (0.00486)	-0.00146 (0.00491)	-0.00181 (0.00515)
Coastal Distance	-0.0147** (0.00482)	-0.0137** (0.00467)	-0.0109* (0.00435)	-0.00585*** (0.00147)	-0.00592*** (0.00145)	-0.00604*** (0.00153)
Soil Quality	0.0375 (1.498)	-0.562 (1.508)	-0.751 (1.515)	-0.440 (0.488)	-0.552 (0.581)	-0.516 (0.527)
Soil Quality × Cotton price, 1yr lag	0.0171 (0.0171)	0.0276+ (0.0167)	0.0293* (0.0134)	0.0108 (0.00743)	0.0134 (0.00952)	0.0142+ (0.00797)
Rubber Suitability	-6.455 (4.503)	-6.988 (4.518)	-7.592+ (4.438)	-3.296+ (1.882)	-3.509+ (1.834)	-4.168* (1.825)
Rubber Suit × Rubber price, 1yr lag	0.0575+ (0.0334)	0.0592+ (0.0355)	0.0557 (0.0356)	0.0166 (0.0163)	0.0185 (0.0158)	0.0241 (0.0150)
Battle Index, 1yr lag	0.276 (0.183)			0.151+ (0.0831)		
Territorial Control Value, 1yr lag	2.702*** (0.665)			0.573*** (0.0680)		
Battle Index, 2yr lag		0.224+ (0.133)			0.105+ (0.0581)	
Territorial Control Value, 2yr lag		2.328*** (0.691)			0.544*** (0.0678)	
Battle Index, 5yr lag			0.0480 (0.198)			0.0485 (0.0843)
Territorial Control Value, 5yr lag			1.211*** (0.242)			0.457*** (0.0675)
Year FE	✓	✓	✓	✓	✓	✓
Ethnic Group FE	✓	✓	✓	✓	✓	✓
Observations	2610	2448	2016	7074	6681	5502
Log-Likelihood	-346.8	-357.6	-383.9	-901.9	-880.1	-823.1
AIC	817.6	837.2	883.9	1883.8	1818.1	1706.2

Clustered standard errors in parentheses

+  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 16: Station Presence, OLS, Grid Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
	Station Presence	Station Presence	Station Presence	Station Type	Station Type	Station Type
Cotton price, 1yr lag	-0.00682* (0.00288)	-0.00536+ (0.00307)	-0.00389 (0.00272)	-0.0251* (0.0116)	-0.0238+ (0.0127)	-0.0186 (0.0116)
Soil Quality × Cotton price, 1yr lag	0.000878+ (0.000467)	0.000853 (0.000526)	0.000770 (0.000491)	0.00349+ (0.00196)	0.00380+ (0.00226)	0.00349+ (0.00212)
Rubber price, 1yr lag	0.00243*** (0.000593)	0.00281*** (0.000822)	0.000616+ (0.000319)	0.00610*** (0.00167)	0.00748*** (0.00217)	0.000950 (0.000976)
Rubber Suitability × Rubber price, 1yr lag	-0.000192 (0.000806)	-0.0000613 (0.000815)	0.000347 (0.000737)	-0.000277 (0.00187)	-0.000290 (0.00185)	0.000239 (0.00169)
Battle Index, 1yr lag	0.0104 (0.00705)			0.00931 (0.0150)		
Control Value, 1yr lag	0.0143* (0.00611)			0.0409* (0.0197)		
Battle Index, 2yr lag		0.00782 (0.00579)			0.00443 (0.0122)	
Control Value, 2yr lag		0.0119* (0.00574)			0.0320+ (0.0187)	
Battle Index, 5yr lag			0.00309 (0.00667)			-0.0162 (0.0172)
Control Value, 5yr lag			0.00317 (0.00457)			0.00979 (0.0153)
Year FE	✓	✓	✓	✓	✓	✓
Grid FE	✓	✓	✓	✓	✓	✓
Observations	7074	6681	5502	7074	6681	5502
Log-Likelihood	5254.4	5045.0	4590.2	-1803.8	-1587.8	-805.7
AIC	-10466.9	-10049.9	-9146.5	3649.6	3215.6	1645.3

Clustered standard errors in parentheses

+  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 17: Station Presence, Correlated RE Model

	(1)	(2)	(3)	(4)	(5)	(6)
	Station Presence	Station Presence	Station Presence	Station Type	Station Type	Station Type
Battle Index, 1yr lag (between)	0.376*** (0.0697)			1.041*** (0.176)		
Battle Index, 1yr lag (within)	0.0104 (0.00706)			0.00931 (0.0150)		
Control Value, 1yr lag (between)	0.161*** (0.00630)			0.508*** (0.0370)		
Control Value, 1yr lag (within)	0.0143* (0.00612)			0.0409* (0.0197)		
Battle Index, 2yr lag (between)		0.405*** (0.0695)			1.160*** (0.179)	
Battle Index, 2yr lag (within)		0.00782 (0.00579)			0.00443 (0.0123)	
Control Value, 2yr lag (between)		0.160*** (0.00651)			0.511*** (0.0372)	
Control Value, 2yr lag (within)		0.0119* (0.00575)			0.0320 <sup>+</sup> (0.0187)	
Battle Index, 5yr lag (between)			0.432*** (0.0709)			1.272*** (0.190)
Battle Index, 5yr lag (within)			0.00309 (0.00668)			-0.0162 (0.0172)
Control Value, 5yr lag (between)			0.154*** (0.00707)			0.497*** (0.0371)
Control Value, 5yr lag (within)			0.00317 (0.00458)			0.00979 (0.0153)
Controls	✓	✓	✓	✓	✓	✓
Grid RE	✓	✓	✓	✓	✓	✓
Observations	7074	6681	5502	7074	6681	5502

Clustered standard errors in parentheses

<sup>+</sup>  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

## 11 Lagged Dependent Variable

Table 18: Station Presence, Ordered Logit

	(1)	(2)	(3)
	Station Type	Station Type	Station Type
Station Type, lag	1.569*** (0.170)	1.630*** (0.163)	1.646*** (0.169)
Area	-0.466 (1.319)	-0.782 (1.380)	-0.884 (1.425)
Caravan Road Length	-0.0898 (0.169)	-0.131 (0.174)	-0.155 (0.182)
Road Length	0.733*** (0.219)	0.782*** (0.221)	0.779*** (0.225)
Disease Environment	-0.00206 (0.0135)	0.000561 (0.0136)	-0.000353 (0.0136)
Number of Troops, 1889	-0.000975 (0.000625)	-0.001000 (0.000623)	-0.000827 (0.000604)
Ethnic Population	-0.000000232* (0.000000106)	-0.000000232* (0.000000112)	-0.000000246* (0.000000116)
Population Density	0.356* (0.144)	0.327* (0.151)	0.302+ (0.155)
Mean Elevation	-0.000691 (0.000562)	-0.000720 (0.000581)	-0.000901 (0.000601)
Annual Precipitation	0.000193 (0.000301)	0.0000938 (0.000316)	-0.0000473 (0.000326)
Annual Mean Temp	-0.00811 (0.0107)	-0.00943 (0.0110)	-0.0137 (0.0116)
Border Distance	0.000834 (0.000989)	0.000791 (0.00102)	0.000667 (0.00103)
Coastal Distance	0.000273 (0.000468)	0.000334 (0.000498)	0.000562 (0.000505)
Soil Quality	0.0679 (0.586)	-0.299 (0.916)	-0.130 (0.913)
Cotton price, 1yr lag	0.0118 (0.0644)	-0.0249 (0.106)	0.00733 (0.104)
Soil Quality × Cotton price, 1yr lag	0.00231 (0.0110)	0.0108 (0.0189)	0.00920 (0.0184)
Rubber Suitability	-2.100* (0.855)	-2.220* (0.921)	-3.037*** (0.819)
Rubber price, 1yr lag	-0.0434* (0.0214)	-0.0596** (0.0215)	-0.0848*** (0.0244)
Rubber Suitability × Rubber price, 1yr lag	0.0173* (0.00878)	0.0191* (0.00948)	0.0280*** (0.00819)
Battle Index, 1yr lag	0.274** (0.101)		
Control Value, 1yr lag	0.0979* (0.0422)		
Battle Index, 2yr lag		0.239** (0.0825)	
Control Value, 2yr lag		0.0949* (0.0446)	
Battle Index, 5yr lag			0.289* (0.119)
Control Value, 5yr lag			0.0631 (0.0424)
Year FE	✓	✓	✓
Observations	7074	6681	5502
Log-Likelihood	-717.6	-676.4	-624.3
AIC	1515.1	1430.8	1320.5

Clustered standard errors in parentheses  
<sup>+</sup>  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

## 12 Instrumental Variables

Table 19: Station Presence, IV Model

	(1) Station Presence
Battle Index, 1yr lag	0.773* (0.302)
Area	-0.642** (0.242)
Caravan Road Length	0.0309 (0.0238)
Road Length	0.00680 (0.0423)
Disease Environment	0.00256 (0.00222)
Number of Troops, 1889	-0.0000812 (0.000153)
Ethnic Population	4.16e-08 (2.83e-08)
Population Density	0.0315 (0.0258)
Mean Elevation	-0.000159 (0.000187)
Annual Precipitation	0.000121 (0.0000804)
Annual Mean Temp	-0.000915 (0.00349)
Border Distance	-0.000431 (0.000346)
Coastal Distance	-0.0000970 (0.000113)
Soil Quality	-0.263* (0.134)
Soil Quality $\times$ Cotton price, 1yr lag	0.00329* (0.00134)
Rubber Suitability	-0.0181 (0.122)
Rubber Suit $\times$ Rubber price, 1yr lag	-0.000610 (0.00126)
Territorial Control Value, 1yr lag	0.119*** (0.0115)
Constant	-0.724 (1.039)
Year FE	✓
Kleibergen-Paap rk LM statistic	16.297
Kleibergen-Paap rk Wald F statistic	8.42
Hansen J statistic	1.997
Observations	4644
Log-Likelihood	-1772.6
AIC	3617.3

Clustered standard errors in parentheses

+  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

## 13 Plantations



Table 20: Number of Plantations, Logit

	(1)	(2)
	Rubber	Cotton
Area	57.43* (23.86)	39.50*** (6.785)
Caravan Road Length	0.0322 (0.317)	0.00956 (0.323)
Road Length	0.0445 (0.496)	1.497** (0.470)
Disease Environment	-0.0126 (0.0334)	0.0201 (0.0352)
Number of Troops, 1889	-0.00188 (0.00605)	0.00380 (0.00241)
Ethnic Population	-0.00000785 <sup>+</sup> (0.00000475)	-0.00000111 (0.00000146)
Population Density	-0.259 (0.337)	0.0398 (0.311)
Mean Elevation	-0.00214 (0.00160)	0.00154 (0.00134)
Annual Precipitation	0.000340 (0.000751)	-0.000346 (0.000861)
Annual Mean Temp	-0.0401 (0.0273)	0.0504* (0.0233)
Border Distance	-0.00409* (0.00198)	-0.00998*** (0.00252)
Coastal Distance	0.00205 (0.00135)	-0.00276* (0.00131)
Rubber Suitability	2.440** (0.784)	
Soil Quality		0.279 (0.576)
Station Presence	-2.136 <sup>+</sup> (1.258)	2.468*** (0.569)
Constant	-2.790 (10.12)	-22.44** (8.290)
Observations	393	393
Log-Likelihood	-165.0	-150.7
AIC	362.0	333.4

Clustered standard errors in parentheses

<sup>+</sup>  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

## 14 Longterm Effects: Nightlights

We draw on nightlight satellite data to measure modern day economic development at the sub-national level. Nightlights are a useful way of measuring economic development and have been used in related studies on the long-term effects of colonial and pre-colonial rule (Michalopoulos and Papaioannou 2013, Chen and Nordhaus 2011). We match nightlight intensity from the Defense Meteorological Satellite Programs Operational Linescan System. We use the annual composite image of nightlights for 2010 and log the intensity score to address extreme skew. We then estimate the effect of historical colonial station presence on nightlight intensity, controlling for all our historical and geographic cross-sectional confounding variables. We find that historical station presence has a positive and statistically significant effect on nightlight intensity, increasing nightlight intensity by nearly two standard deviations.

Table 21: Log(Night Lights), OLS

	(1)	(2)
	Rubber	Cotton
Area	-0.665 (0.587)	-0.452 (0.556)
Caravan Road Length	0.0588* (0.0291)	0.0420 (0.0313)
Road Length	0.0679+ (0.0401)	0.0157 (0.0355)
Disease Environment	-0.00225 (0.00241)	-0.00126 (0.00207)
Number of Troops, 1889	0.0000880 (0.000116)	0.000115 (0.000113)
Ethnic Population	1.11e-08 (2.63e-08)	2.26e-08 (2.33e-08)
Population Density	0.104** (0.0400)	0.0682* (0.0314)
Mean Elevation	0.0000739 (0.000110)	0.0000632 (0.0000933)
Annual Mean Precipitation	0.000113+ (0.0000646)	0.0000472 (0.0000538)
Annual Mean Temp	0.000104 (0.00148)	-0.000579 (0.00130)
Border Distance	0.0000373 (0.000157)	-0.000108 (0.000149)
Coastal Distance	-0.0000987 (0.000109)	-0.0000591 (0.0000920)
Soil Quality	0.0447 (0.0420)	0.0380 (0.0366)
Rubber Suitability	-0.0436 (0.0690)	0.00986 (0.0524)
Station Presence		0.416** (0.150)
Constant	-0.403 (0.530)	-0.150 (0.433)
Observations	393	393
Log-Likelihood	-13.46	8.590
AIC	56.91	14.82

Clustered standard errors in parentheses

+  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

## 15 Motives for state expansion according to the German administration

German officers and station chiefs repeatedly emphasized that this expansion of military presence was essential for the successful pacification of restive regions: “only longer-term presence of the troops, continuous showing of power can lead to continuous success.” According to a similar but more cynical statement, stations were important “to inform the natives about their belonging to the German authority” (Yearly Report 1904/1905: 14). As Graf von Gtzen (1909: 35) explains, these military posts – established in reaction to violent encounters in order to increase territorial control – were maintained to ensure peace, and were subsequently transformed into military districts with increasing administrative functions. Nigmann gives a broader and more systematic overview of the rationale of state expansion by providing information about why 35 of the 81 stations in our dataset were established.

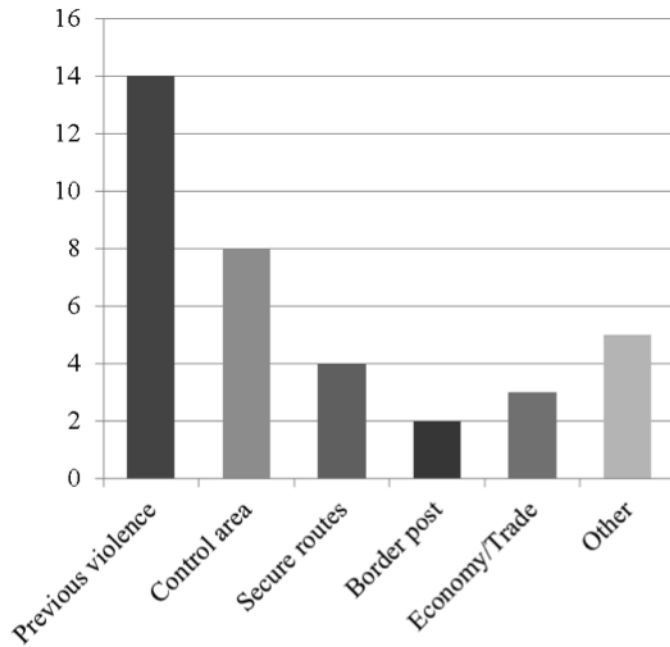


Figure 2: Reasons for the Establishment of German Stations

According to Nigmann, 24 of these 35 stations were established with the explicit aim of pacifying restive regions or securing control over specific areas. Economic reasons feature much less prominently on Nigmann's list. Indeed, caravan routes played an essential role in state expansion, although not in the way that is often emphasized. Caravans were often attacked, and Germans established stations in reaction to these attacks, rather than due to the logistical or economic advantages associated with the routes. This security rationale of a state presence near important trading routes is also often found in German files (R1001/1026: 6–7, 29–35, 49–50), underlining the primary role of perceived threats to stability and order as primary motives for the establishment of German stations.

## 16 Cost considerations in the expansion of the German presence

We have argued that the costs of state expansion cannot explain the spatial patterns of expansion; cost considerations were of secondary concern if investments in additional state presence were considered a high strategic priority, especially in cases of perceived challenges to colonial rule. Colonial files provide ample evidence that the German administration tried to cut costs wherever possible. Suggestions to establish additional stations were repeatedly rejected for financial reasons (R1001/1026: 1–133). However, while costs may have influenced the overall extent of state expansion, we find no evidence to support the assumption that cost considerations shaped its spatial patterns. Costs resulting from unfavorable health conditions occasionally led to relocations of German stations – usually to more favorable areas very close to the original position. The Mikindani station was relocated to Lindi (only around 50 kilometers away) because Mikindani provided “ideal conditions for mosquitos and Malaria” (R1001/215: 19–20). For similar reasons, the Kilo-ssa station was moved 80 kilometers to the west to Morogoro (R1001/224: 2324). In both cases, the original stations were not entirely abandoned, but downgraded to auxiliary stations to maintain a basic state presence. The station of Ulanga was initially established in 1894 in order to suppress the belligerent Wahehe. In 1895 it was moved to nearby Perondo due to unfavorable health conditions. However, as the local population continued to resist, the station was again moved to Iringa, located in the heart of Wahehe territory, just one year later. After successfully dispelling the Wahehe threat, the station was maintained as one of the largest military stations and permanent locales of state presence (Nigmann 1911: 77).

These examples illustrate that relocations due to health conditions did not substantially affect the territorial make-up of the state administration. While they may have led to minor geographical reconfigurations, the moves did not affect the locations of the main German strongholds of state presence. The examples also underline our argument that where there was conflict, strategic considerations related to territorial control overrode conflicting interests related to minimizing administrative or human costs.

## 17 State expansion after violent conflict – reactions to the Maji Maji rebellion

We also investigate our hypotheses in the context of the most severe resistance to the German colonial state. Critics of the colonial administration have repeatedly claimed that the massive Maji Maji rebellion in the south of the colony from 1905-07 was a consequence of insufficient state presence in the region (Götzen 1909; Nigmann 1911). The southern districts were simply not seen as a potential threat by the Germans, who focused more on what they thought to be the largest and most powerful ethnic groups in the territory – the Nyamwezi further north (Götzen 1909; Nigmann 1911; Bühner 2011). Looking at the rebellious districts and comparing state presence before and after the violent uprising, we can see how the colonial administration tried to adapt its strategy of state penetration to strengthen its capacity as a direct response to the uprising. The left-hand side of Figure 1 shows the location of German stations in 1905 before the rebellion, as well as the location of violent events during the rebellion. The latter information stems from a geo-coded violent event dataset collated from weekly issues of the German East African Newspaper for the period under investigation (De Juan 2014). The right-hand side displays the location of German stations in 1907 after the rebellion.



Figure 3: The German Colonial State before and after the Maji Maji Rebellion

We can see that within only two years, nine German stations were either newly established or upgraded in the restive regions. Looking more closely at individual spatial associations, it is evident that stations such as Iraku, Ubena, Mgende, and Kibata were established directly in areas with high levels of violence. Moreover, the stations of Kiberege, Mgenda, and Ubena significantly contributed to filling gaps in German presence in the southern regions. Overall, the establishment of new stations substantially increased German coverage in the region. The total distance from all violent events to their respective nearest German stations was reduced by more than 25%. The minutes from the meeting of the colonial government council (Gouvernementsrat) from 1907 underscore that this was not a coincidence. Members of the council stressed that new stations were essential in the hinterland of Kilwa and Lindi because these districts had experienced intense rebellion. More concretely, according to the council, stations were to be established in areas that constituted a natural center point in their respective areas (R1001/812; 140, 176).