Online Appendix A: Supplementary materials

Miguel Angel Carpio and María Eugenia Guerrero

December 2021

Supplementary materials

In this appendix, we present supplementary materials. This document is organized as follows: Section 1 explains the construction of the variables of internal migration; Section 2 presents the robustness exercises we have performed and explains the results; finally, Section 3 presents additional figures and tables.

1 Variables of internal migration

We use the 2007 Peru Census, conducted by the *Instituto Nacional de Estadística e Informática*. This census includes two questions targeting all individuals, which we exploit: "In which district do you currently live?" and "When you were born, in which district did your mother live?". We use these questions to count the number of individuals in the 15-to-60 range who out-migrated from the districts of our sample to the capital Lima.

We also combine the question regarding the district of residence with a second: "Where did you live five years ago?". We identify the individuals that out-migrated by department between 2002 and 2007, as well as the number of individuals that immigrated (an individual currently living in department A that lived in department B five years ago is an immigrant in A and out-migrant in B). We then count immigrants and out-migrants at the district level and obtain rates of current immigration and current out-migration by dividing these district sums into district population size from the 1981 Peru Census. Mathematically, $outmigration_d = outmigrants2007_d/population1981_d * 100$, and $immigration_d = immigrants2007_d/population1981_d * 100$.

2 Robustness tests

2.1 Comparability of the districts near the threshold

Our empirical strategy relies on the assumption that population structure prior to the enactment of *mita* is similar between the subjected region and the control region. Here, we take advantage of the rigorous analysis of this question conducted by Dell (2010). We have already presented Table II in the main document with summary statistics on demographic and fiscal variables. Table A.4 in this online appendix now replicates the test for the continuity of these variables at the threshold. Each of the first three horizontal panel corresponds to one of the specifications of Dell (2010), while the last horizontal panel corresponds to our specification. As was preliminarily shown, the population size before the *mita* was similar and also the structure of population by age, gender and race.

In order to address other concerns related to how comparable the districts inside and outside the *mita* catchment were, we analyze the results of Equation 1 using different samples. In Table A.5 in this online appendix, we report only estimations for the smallest bandwidth (i.e. <50km) using the quadratic polynomial that includes interactions with *mita*. In Panel A, we include the baseline results.

In Panel B, we include the ten districts of Metropolitan Cusco, which we had excluded as a precaution, since their current level of development may be determined by its historically central position as part of the capital of the Inca Empire. In Panel C, we exclude seven Inca states because, in the times of the Inca Empire, these areas served religious purposes and were not used for productive activities.[1] In Panel D, we exclude the districts in which the *mita* boundary coincides with a river, which makes it possible to verify that there is no endogeneity due to geography. In all the cases, the estimates are largely stable, both in terms of sign and size.

2.2 Bandwidth size

We assess the stability of our results with respect to changes in the size of the bandwidth. Instead of focusing on our baseline bandwidths of 50, 75 and 100 kilometers of closeness to the *mita* boundary, Figure A.3 in this online appendix shows how our estimates change when we consider bandwidth sizes ranging from 30 to 100 kilometers in intervals of 5, together with 95 percent confidence intervals. The estimates are largely stable, both in terms of sign and size.

2.3 Specifications of the RD polynomial

We also examine robustness to different specifications of the RD polynomial $f(geographic\ location_d)$. We focus our attention on the selected interacted quadratic polynomial of latitude and longitude. Table A.6 in this online appendix presents three groups of columns, where each group corresponds to the natural logarithm of the three proposed indicators: number of surnames, number of district-exclusive surnames and number of area-exclusive surnames. Within each group, the columns correspond to three possible sub-samples: districts within 100 km, 75 km, and 50 km of the boundary. Each horizontal panel of Table A.6 examines robustness to the following specifications: linear, quadratic (our baseline) and cubic. It also includes an ordinary least squares estimation. Notice that the differences between *mita* and *non-mita* are always statistically significant, whether we use linear, quadratic or cubic polynomials (first, second, and third panels, respectively). The exceptions are two regressions of the linear polynomial in the narrowest bandwidth, probably due to the small number of observations; and four regressions of the cubic polynomial, which is arguably due to the large number of terms. This time the sizes of the coefficients are not fully stable, but interestingly the order of magnitude is always the same: 1) area-exclusive surname, 2) district-exclusive surname, and 3) number of surnames.

We also report, for the sake of completeness, estimates from the three baseline specifications of $f(geographic\ location_d)$ used by Dell (2010). Each of the first three horizontal panels of Table A.7 in this online appendix corresponds to these specifications. Panel A shows the results of a multidimensional RD using a cubic polynomial of longitude and latitude. Most of the coefficients do not have the expected sign and are not statistically significant in any case. The high order of the polynomial is likely the cause of this misleading estimation (see Gelman and Imbens (2014)). Panel B presents the results for a

[1] The excluded districts are: Chinchaypujio, Limatambo, Calca, Lamay, Maras, Ollaytantambo and Yucay. These districts belong to the Cusco department.

traditional single-dimensional RD using a cubic polynomial of the Euclidean distance to the Potosí mine. Controlling for this variable may reduce a potential bias because the distance to Potosí was a criterion for assignment the *mita*. As explained in the main document, the greater the distance to Potosí, the less likely a district would have been deemed eligible for *mita* by the Viceroy Toledo. Moreover, the distance to Potosí was a potential determinant of out-migration. We obtain a negative significant effect in each of the nine point-estimates with this specification. Panel C estimates the results for a single-dimensional RD using a cubic polynomial of the distance to *mita* boundary, similar to traditional one-dimensional RD designs. We obtain a negative difference in nine out of the nine estimations with this specification, but the size is smaller than in the case of the previous specification. In fact, the point estimate is significant in only three out of nine cases. This is likely related to a positive bias in the estimation when we do not control for the distance to Potosí. In order to provide evidence of this bias, we present a fourth specification: Panel D presents the results for cubic polynomials of both distance to *mita* boundary and distance to Potosí. We obtain again a negative significant effect in nine out of the nine point-estimates with this specification.

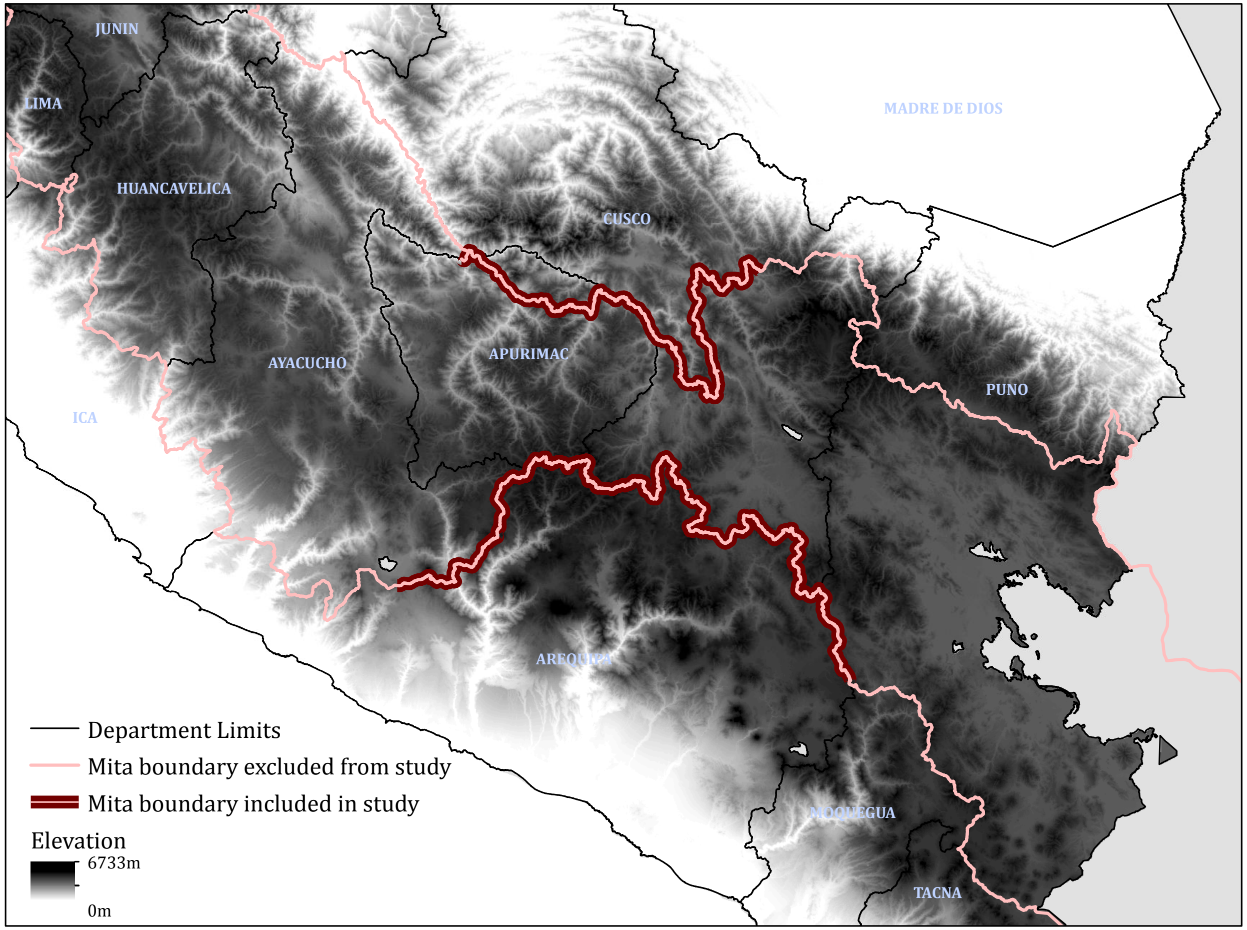
We go deeper in the specification analysis by examining robustness to different polynomials of elevation and slope. This is important because, as explained in the main document, elevation was another criterion used by Viceroy Toledo in assigning *mita*. Panel A of Table A.8 in this online appendix shows our main results as a baseline while Panels B, C and D show the results for quadratic, cubic, and quartic polynomials, respectively. The *mita* effect is markedly stable for these changes. We obtain significance in 36 out of 36 cases, and we also recover the stability of the size of the coefficients.

2.4 Control variables

Table A.9 in this online appendix examines the robustness of our results with respect to the introduction of several control variables. Panel A shows our main results as a baseline. Panels B includes the Euclidean distance to the Potosí mine as a control variable. As explained before, this distance was a criterion for the assignment of *mita* and was a potential determinant of out-migration. Panel C examines the robustness of the preferred specification to the introduction of current population size as a control variable. Panel D includes both the Euclidean distance to the Potosí mine and current population size. In all the cases, the results remain qualitatively the same. They show that, even if we compare districts of the same population size or equally far from the Potosí mine, the coefficients are stable.

3 Additional figures and tables

Figure A.1: Map of the area of study



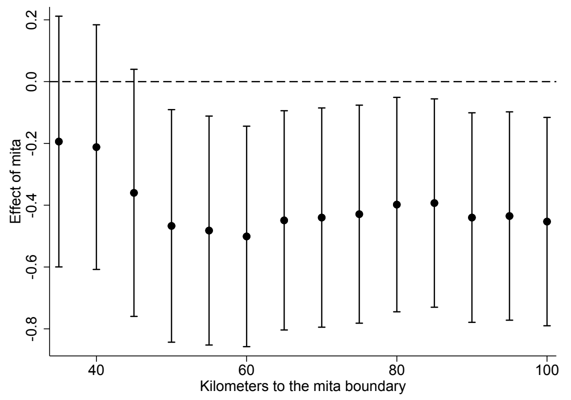
Notes: The map shows the *mita* borders in Peru and the segments focused on in the study. Elevation is shown in the background. Current administrative divisions are in black.

Source: Data from Dell (2010).

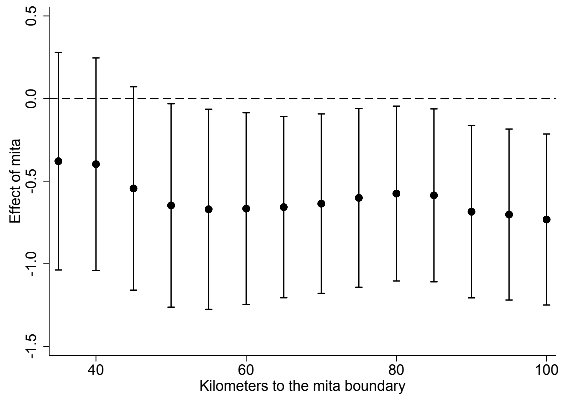
Figure A.2: Autographed manuscript of Guaman Poma de Ayala (1615, fol. 529)

Source: Royal Danish Library, GKS 2232 kvart: Guaman Poma, Nueva corónica y buen gobierno (c. 1615), page [525 [529]]. Permission to use the digital facsimile granted by the Department of Manuscripts & Rare Books. The translation of the title is "The administrator of the royal mines punishes the native lords with great cruelty".

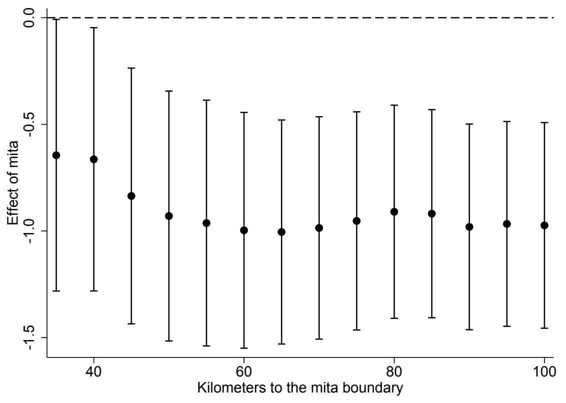
Figure A.3: Robustness to bandwidth choice



Log number of surnames



Log district-exclusive surnames



Log area-exclusive surnames

Notes: Each sub-figure corresponds to one surname indicator. It plots the point estimates of γ in the vertical axis based on Equation 1, for different bandwidth values between 30-100 km in increments of 5 km in the horizontal axis. Lines stemming from the point estimates show 95 percent confidence intervals.

Sources: Peruvian Electoral Roll of 2011 and data from Dell (2010).

Table A.1: Top 20 ER 2011 surnames by classification

Common	Area-exclusive		District-exclusive			
	Mita	*Non-mita*	*Mita*		*Non-mita*	
Surname	Surname	Surname	Surname	Location	Surname	Location
Quispe	Huarhuachi	Olayunca	Chancos	Huancapi, Ayacucho	Berreras	San Jeronimo, Cusco
Mamani	Huachuhuillca	Pfuyo	Yapia	Ongoy, Apurimac	Condorvilca	Coporaque, Arequipa
Huaman	Ccahuay	Antachoque	Vidangos	Puno, Puno	Catcoparco	Alca, Arequipa
Condori	Lago	Curasco	Manchi	Ranracancha, Apurimac	Camilla	Anta, Cusco
Flores	Arohuillca	Yupayccana	Hachiri	Espinar, Cusco	Huaracco	Abancay, Apurimac
Apaza	Sucacahua	Huamanccari	Capajaña	Juliaca, Puno	Sarca	Lari, Arequipa
Huamani	Astoyauri	Loncone	Equiño	Santo Tomas, Cusco	Anayhua	Calca, Cusco
Ramos	Yauyo	Abal	Ilaccaña	Santo Tomas, Cusco	Sega	Urubamba, Cusco
Gutierrez	Marcas	Yarahuaman	Yunca	Puno, Puno	Uchuquicaña	Andagua, Arequipa
Cruz	Licla	Huillcanina	Polloyqueri	Puno, Puno	Mayanasa	Orcopampa, Arequipa
Vargas	Pandia	Victoria	Ulhua	Santo Tomas, Cusco	Torres De La Gala	Cusco, Cusco
Mendoza	Chiquillan	Pañihuara	Titalo	Puno, Puno	Incattito	San Jeronimo, Cusco
Rojas	Pacoricona	Bañares	Larijo	Puno, Puno	Jhuno	Challabamba, Cusco
Puma	Gomel	Auquipuma	Challcha	Puno, Puno	Alcasivincha	Cayarani, Arequipa
Huillca	Nauto	Manga	Catti	Juliaca, Puno	Veredas	San Sebastian, Cusco
Palomino	Chila	Tonccochi	Centellas	Puno, Puno	Quisperroca	San Sebastian, Cusco
Vilca	Secce	Elorrieta	Seje	Juliaca, Puno	Corimaita	Cusco, Cusco
Gonzales	Jahuira	Kancha	Excelmes	Puno, Puno	Cadenillas	Wanchaq, Cusco
Torres	Blancos	Huayaconza	Ccalasani	Quiquijana, Cusco	Corampa	Santiago, Cusco
Huanca	Espillco	Phuyo	Huanso	Coyllurqui, Apurimac	Zamatelo	Huambo, Arequipa

Notes: The table presents the most common 20 surnames of the ER 2011 (departments of Apurimac, Ayacucho, Arequipa, Cusco and Puno) by subset. Each subset represents one of the categories set out in our conceptual framework. In the case of district-exclusive surnames, we include the names of the district and department.
Source: Peruvian Electoral Roll of 2011.

Table A.2: RD estimates of the *mita* effect on surname indicators using bootstrap estimation and population size

	Closeness to the *mita* boundary		
	<100 km	<75 km	<50 km
A. Log number of surnames			
Mita	-0.453**	-0.429**	-0.467**
	(0.196)	(0.183)	(0.228)
B. Log district-exclusive surnames[a]			
Mita	-0.732***	-0.601*	-0.647*
	(0.257)	(0.340)	(0.331)
C. Log area-exclusive surnames			
Mita	-0.974***	-0.953***	-0.930***
	(0.243)	(0.283)	(0.303)
Observations	289	239	185

Notes: The unit of analysis is the district. All regressions are multidimensional RD using a quadratic polynomial of latitude and longitude, including interactions with *mita*. All regressions include geographic controls and boundary segment fixed effects. Results are calculated using nonparametric bootstrap estimation with 100 replications. Robust standard errors are in parentheses. Coefficients that are significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[a]This variable equals ln(number of district-exclusive surnames+1). This transformation in done to preserve the entire sample because some observations are equal to zero (seven obs. in the larger bandwidth, five and two in the subsequent bandwidths).

Sources: Peruvian Electoral Roll of 2011 and data from Dell (2010).

Table A.3: Summary statistics for *post-mita* migration indicators by *mita* status

	Bandwidth 100			Bandwidth 75			Bandwidth 50		
Variable	*Mita*	*Non-mita*	diff	*Mita*	*Non-mita*	diff	*Mita*	*Non-mita*	diff
District with a Japanese surname	0.453	0.384	0.069	0.451	0.416	0.035	0.480	0.387	0.093
	(0.035)	(0.053)	(0.064)	(0.039)	(0.057)	(0.069)	(0.045)	(0.062)	(0.078)
Log number of Japanese surnames[a]	0.431	0.379	0.052	0.414	0.414	0.000	0.422	0.407	0.014
	(0.039)	(0.059)	(0.071)	(0.041)	(0.064)	(0.074)	(0.045)	(0.074)	(0.082)
Log number of out-migrants to Lima[a]	4.435	4.168	0.268	4.387	4.134	0.253	4.351	4.295	0.056
	(0.093)	(0.135)	(0.168)	(0.093)	(0.135)	(0.183)	(0.093)	(0.135)	(0.218)
Current immigration rate	2.634	2.418	0.216	2.815	2.459	0.356	2.884	2.346	0.538
	(0.252)	(0.382)	(0.460)	(0.303)	(0.424)	(0.527)	(0.312)	(0.461)	(0.548)
Current out-migration rate	4.760	4.357	0.403	4.835	4.691	0.144	4.933	5.402	-0.469
	(0.239)	(1.232)	(0.880)	(0.274)	(1.371)	(1.024)	(0.339)	(1.691)	(1.290)
Log population size	7.571	7.370	0.202	7.550	7.472	0.079	7.628	7.624	0.005
	(0.074)	(0.109)	(0.134)	(0.079)	(0.112)	(0.138)	(0.090)	(0.127)	(0.156)
Observations	203	86		162	77		123	62	

Notes: The unit of analysis is the district. The table presents means and differences in means, and tests whether they are statistically different. The first three columns focus on observations that are 100 km from the *mita* boundary, the following columns restricts the boundary to 75 and 50 km. Standard errors for the difference in means are in parentheses. Differences that are significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[a] These variables equal ln(variable+1). This transformation is done to preserve the entire sample, because some observations are equal to zero.

Sources: "Pioneers" project of the Japanese Peruvian Association, 2007 Peru Census, Peruvian Electoral Roll of 2011, and data from Dell (2010).

Table A.4: RD estimates of the *mita* effect on 1572 demographics

	Population 1572	Percent			Log mean tribute	Share of Tribute Revenues			
		Men	Women	Boys		Spanish Nobility	Spanish Priests	Spanish Justices	Indig. Mayors
A. Cubic polynomial of latitude and longitude									
Mita	101.816	-0.002	-0.009	0.011	0.028	-0.010	0.004	0.004	0.003
	(1,574.765)	(0.009)	(0.016)	(0.012)	(0.034)	(0.030)	(0.019)	(0.010)	(0.005)
B. Cubic polynomial of distance to Potosí									
Mita	1,967.153	0.006	-0.011	0.005	0.025	-0.013	0.008	0.006	-0.001
	(1,305.021)	(0.007)	(0.012)	(0.010)	(0.031)	(0.025)	(0.015)	(0.009)	(0.004)
C. Cubic polynomial of distance to *mita* boundary									
Mita	2,077.044*	0.007	-0.008	0.001	0.037	-0.009	0.005	0.003	-0.001
	(1,090.424)	(0.007)	(0.010)	(0.008)	(0.031)	(0.018)	(0.012)	(0.006)	(0.004)
D. Interacted quadratic polynomial of latitude and longitude									
Mita	103.328	0.025*	-0.009	-0.016	0.158***	-0.013	-0.004	0.009	0.011
	(3,224.628)	(0.014)	(0.029)	(0.025)	(0.044)	(0.046)	(0.031)	(0.020)	(0.008)
Observations	65	65	65	65	65	65	65	65	65

Notes: The unit of analysis is the district within 50 km of the *mita* boundary. Panels A, B and C use the cubic polynomials of latitude and longitude, distance to Potosí and distance to the *mita* boundary, respectively. These are exactly the same regressions presented in Table V by Dell (2010). Panel D presents the results using our baseline specification, that is, the quadratic polynomial of latitude and longitude interacted with *mita*. All regressions include geographic controls and boundary segment fixed effects. Regressions for the share of men, women and boys weighted by the square root of the district's total population; and regression for the tribute rate in 1572 weighted by the square root of the district's tributary population. Robust standard errors are in parentheses. Coefficients that are significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Source: Data from Dell (2010).

Table A.5: Specification tests: different samples

	Log number of surnames	Log district-exclusive surnames[a]	Log area-exclusive surnames
A. Baseline			
Mita	-0.467**	-0.647**	-0.930***
	(0.192)	(0.314)	(0.299)
Observations	185	185	185
B. Includes Cusco			
Mita	-0.609***	-0.795**	-1.072***
	(0.199)	(0.311)	(0.298)
Observations	195	195	195
C. Excludes Inca estates			
Mita	-0.549***	-0.793**	-1.040***
	(0.198)	(0.318)	(0.303)
Observations	178	178	178
D. Excludes rivers			
Mita	-0.463**	-0.622*	-0.907***
	(0.195)	(0.316)	(0.300)
Observations	183	183	183

Notes: The unit of analysis is the district within 50 km from the *mita* boundary. All regressions include the quadratic polynomial of latitude and longitude interacted with *mita*, geographic controls and boundary segment fixed effects. Robust standard errors are in parentheses. Coefficients that are significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
[a]This variable equals ln(number of exclusive surnames+1). This transformation is done to preserve the entire sample, because some observations are equal to zero (seven obs. in the larger bandwidth, five and two in the subsequent bandwidths).
Sources: Peruvian Electoral Roll of 2011 and data from Dell (2010).

Table A.6: Specification tests: RD polynomials

	Log number of surnames			Log district-exclusive surnames[a]			Log area-exclusive surnames		
	<100 km	<75 km	<50 km	<100 km	<75 km	<50 km	<100 km	<75 km	<50 km
A. Polynomial of latitude and longitude interacted with *mita*									
Interacted linear polynomial									
Mita	-0.209**	-0.176**	-0.138	-0.305**	-0.315**	-0.230	-0.357***	-0.364***	-0.298**
	(0.089)	(0.087)	(0.096)	(0.128)	(0.126)	(0.142)	(0.130)	(0.131)	(0.151)
Interacted quadratic polynomial - Baseline									
Mita	-0.453***	-0.429**	-0.467**	-0.732***	-0.601**	-0.647**	-0.974***	-0.953***	-0.930***
	(0.172)	(0.180)	(0.192)	(0.264)	(0.276)	(0.314)	(0.246)	(0.261)	(0.299)
Interacted cubic polynomial									
Mita	-0.351*	-0.322	-0.435*	-0.524	-0.324	-0.248	-0.848***	-0.764**	-0.591*
	(0.192)	(0.208)	(0.229)	(0.333)	(0.338)	(0.380)	(0.283)	(0.297)	(0.349)
B. Ordinary Least Squares									
Mita	-0.138	-0.140	-0.161*	-0.168	-0.232*	-0.200	-0.198	-0.286**	-0.312**
	(0.091)	(0.089)	(0.089)	(0.121)	(0.120)	(0.134)	(0.122)	(0.125)	(0.137)
Observations	289	239	185	289	239	185	289	239	185

Notes: The unit of analysis is the district. Panel A presents the results for different specifications of the polynomial of latitude and longitude interacted with *mita*, while Panel B presents the results without polynomials. All regressions include geographic controls and boundary segment fixed effects. Robust standard errors are in parentheses. Coefficients that are significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[a]This variable equals ln(number of exclusive surnames+1). This transformation in done to preserve the entire sample because some observations are equal to zero (seven obs. in the larger bandwidth, five and two in the subsequent bandwidths).

Sources: Peruvian Electoral Roll of 2011 and data from Dell (2010).

Table A.7: Specification tests: results using specifications proposed by Dell (2010)

	Log number of surnames			Log district-exclusive surnames[a]			Log area-exclusive surnames		
	<100 km	<75 km	<50 km	<100 km	<75 km	<50 km	<100 km	<75 km	<50 km
A. Cubic polynomial of latitude and longitude									
Mita	0.057	0.066	0.053	0.006	0.116	0.137	-0.039	-0.037	-0.017
	(0.126)	(0.136)	(0.149)	(0.177)	(0.192)	(0.213)	(0.175)	(0.196)	(0.213)
B. Cubic polynomial of distance to Potosí									
Mita	-0.172*	-0.175*	-0.207**	-0.268**	-0.282**	-0.288**	-0.397***	-0.426***	-0.446***
	(0.096)	(0.097)	(0.095)	(0.135)	(0.134)	(0.138)	(0.137)	(0.141)	(0.143)
C. Cubic polynomial of distance to *mita* boundary									
Mita	-0.101	-0.107	-0.159*	-0.124	-0.184	-0.191	-0.173	-0.252**	-0.323**
	(0.089)	(0.088)	(0.091)	(0.121)	(0.120)	(0.138)	(0.121)	(0.125)	(0.140)
D. Cubic polynomials of distance to Potosí and distance to *mita* boundary									
Mita	-0.185*	-0.168*	-0.198**	-0.268*	-0.267**	-0.271*	-0.403***	-0.412***	-0.434***
	(0.098)	(0.098)	(0.095)	(0.138)	(0.135)	(0.138)	(0.141)	(0.142)	(0.145)
Observations	289	239	185	289	239	185	289	239	185

Notes: The unit of analysis is the district. Panels A, B and C present the results using the same specifications proposed by Dell (2010), that is, cubic polynomials of latitude and longitude, distance to Potosí and distance to the *mita* boundary, respectively. Panel D presents the results using cubic polynomials of distance to Potosí and distance to the *mita* boundary. All regressions include geographic controls and boundary segment fixed effects. Robust standard errors are in parentheses. Coefficients that are significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[a]This variable equals ln(number of district-exclusive surnames+1). This transformation in done to preserve the entire sample because some observations are equal to zero (seven obs. in the larger bandwidth, five and two in the subsequent bandwidths).

Sources: Peruvian Electoral Roll of 2011 and data from Dell (2010).

Table A.8: Specification tests: elevation and slope polynomials

	Log number of surnames			Log district-exclusive surnames[a]			Log area-exclusive surnames		
	<100 km	<75 km	<50 km	<100 km	<75 km	<50 km	<100 km	<75 km	<50 km
A. Linear polynomial of elevation and slope - Baseline									
Mita	-0.453***	-0.429**	-0.467**	-0.732***	-0.601**	-0.647**	-0.974***	-0.953***	-0.930***
	(0.172)	(0.180)	(0.192)	(0.264)	(0.276)	(0.314)	(0.246)	(0.261)	(0.299)
B. Quadratic polynomial of elevation and slope									
Mita	-0.462***	-0.434**	-0.521***	-0.730***	-0.577**	-0.669**	-0.969***	-0.933***	-0.982***
	(0.174)	(0.184)	(0.196)	(0.265)	(0.275)	(0.313)	(0.247)	(0.262)	(0.300)
C. Cubic polynomial of elevation and slope									
Mita	-0.554***	-0.544***	-0.566***	-0.749***	-0.620**	-0.715**	-1.027***	-1.013***	-1.056***
	(0.172)	(0.176)	(0.197)	(0.272)	(0.282)	(0.313)	(0.256)	(0.270)	(0.305)
D. Quartic polynomial of elevation and slope									
Mita	-0.571***	-0.536***	-0.566***	-0.774***	-0.606**	-0.720**	-1.057***	-0.998***	-1.048***
	(0.173)	(0.177)	(0.195)	(0.272)	(0.283)	(0.314)	(0.256)	(0.274)	(0.308)
Observations	289	239	185	289	239	185	289	239	185

Notes: The unit of analysis is the district. All regressions include the quadratic polynomial of latitude and longitude interacted with *mita* and geographic controls and boundary segment fixed effects. Robust standard errors are in parentheses. Coefficients that are significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[a]This variable equals ln(number of exclusive surnames+1). This transformation in done to preserve the entire sample because some observations are equal to zero (seven obs. in the larger bandwidth, five and two in the subsequent bandwidths).

Sources: Peruvian Electoral Roll of 2011 and data from Dell (2010).

Table A.9: Specification tests: additional controls

	Log number of surnames			Log district-exclusive surnames[a]			Log area-exclusive surnames		
	<100 km	<75 km	<50 km	<100 km	<75 km	<50 km	<100 km	<75 km	<50 km
A. Baseline									
Mita	-0.453***	-0.429**	-0.467**	-0.732***	-0.601**	-0.647**	-0.974***	-0.953***	-0.930***
	(0.172)	(0.180)	(0.192)	(0.264)	(0.276)	(0.314)	(0.246)	(0.261)	(0.299)
B. Controlling for the distance to Potosí									
Mita	-0.447**	-0.429**	-0.467**	-0.737***	-0.601**	-0.649**	-0.975***	-0.954***	-0.936***
	(0.172)	(0.181)	(0.193)	(0.264)	(0.276)	(0.314)	(0.246)	(0.261)	(0.298)
C. Controlling for population in 2011									
Mita	-0.377**	-0.323**	-0.337**	-0.618***	-0.463**	-0.476*	-0.871***	-0.814***	-0.757***
	(0.148)	(0.133)	(0.139)	(0.235)	(0.222)	(0.251)	(0.224)	(0.210)	(0.234)
D. Controlling for the distance to Potosí and population in 2011									
Mita	-0.374**	-0.325**	-0.339**	-0.626***	-0.464**	-0.480*	-0.875***	-0.817***	-0.765***
	(0.148)	(0.133)	(0.139)	(0.236)	(0.221)	(0.249)	(0.224)	(0.210)	(0.233)
Observations	289	239	185	289	239	185	289	239	185

Notes: The unit of analysis is the district. All regressions include the quadratic polynomial of latitude and longitude interacted with *mita* and geographic controls and boundary segment fixed effects. Robust standard errors are in parentheses. Coefficients that are significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[a]This variable equals ln(number of exclusive surnames+1). This transformation in done to preserve the entire sample because some observations are equal to zero (seven obs. in the larger bandwidth, five and two in the subsequent bandwidths).

Sources: Peruvian Electoral Roll of 2011 and data from Dell (2010).