Online Appendix (Not for Publication)

Stunted firms: The long-term impacts of colonial taxation, by Gabriel Natividad

A. Data sources

1. *The 2008 Business Census of Peru.* Conducted by the National Institute of Statistics and Informatics. The census covers firms of all sizes and all industries except agriculture and financial services as of December 2007 in a cross-sectional fashion. Information is collected personally by census officers visiting business establishments; consistency checks for the information collected are performed by census officers. Firms' locations are identified at the granular level of each census block. The census information used in the paper is only from the 299 districts of the zone of study detailed in Figure 1 of this Online Appendix.

2. *Mita boundaries.* The mita geography raw files were obtained from `http://scholar.harvard.edu/dell/publications/persistent-effects-perus-mining-mita`, Melissa Dell's website, last accessed on March 22, 2017. The mita boundaries coincide with district boundaries among the 299 districts of the zone of study.

3. *Other digital maps.* Digital maps at the census-block level were produced by the National Institute of Statistics and Informatics and were provided by the Ministry of Production in `.shp` format. These maps are required to create the distance-to-boundary variables for the regression discontinuity setup. Maps for Peru's districts were developed by the National Institute of Statistics and Informatics, and were

obtained from the Poverty Map of Peru: http://iinei.inei.gob.pe/iinei/srienaho/descarga/DocumentosZIP/2013-134/Informacioncartografica.zip, last accessed on April 12, 2018. Maps of rivers were developed by the National Authority for Water in Peru and obtained from GeoGPSPeru: http://www.geogpsperu.com/2014/02/descargar-cuencas-hidrograficas-del.html, last accessed on April 12, 2018.

4. *NASA's Shuttle Radar Topography Mission.* This source collects satellite measurements that permit the modeling of two geospatial characteristics: elevation and slope. Information was obtained from OpenTopography: http://opentopo.sdsc.edu/raster?opentopoID=OTSRTM.082016.4326.1, last accessed on April 12, 2018. The geographic variables were created for each census block of the digital maps described above.

5. *The Census of 1572 provided by Cook, Málaga, and Bouysse (1975).* This is a document originally published in 1583, an exact copy of whose content was published in book format in 1975. A .pdf version of this book was used to more efficiently search for the information required to create the variables of interest.

6. *Credit registry of regulated financial institutions in Peru.* This credit registry, provided by the Superintendency of Banking, Insurance, and Pension Fund Administrators in Peru, contains information on all the loans made in 2007 to firms by any supervised financial institution operating in Peru. The location of client firms was geocoded through a proprietary algorithm contracted from Business Analytics, a commercial data services provider in Peru.

7. *The 2007 Population Census of Peru.* Conducted by the National Institute of Statistics and Informatics, and provided by the Ministry of Production, this census includes demographic information as well as census block identifiers. The census was taken months before the business census. Concurrently, a housing census was also taken.

8. *TePaske and Klein's (1982) registry of transactions of the royal treasuries of the Spanish crown in Peru.* Available on Richard Garner's website: `http://www.insidemydesk.com/hdd.html`, last accessed on March 22, 2017. The matching of locations in this registry with communities in the zone of interest was based on Cook, Málaga, and Bouysse (1975).

9. *The 2017 National Survey of Households (ENAHO).* Conducted by the National Institute of Statistics and Informatics, it surveys a sample of households that is representative at the department level (i.e., more aggregate than districts). Locations of respondents are pooled at the level of each "conglomerate," defined as a set of approximately 100-140 houses depending on whether the region is rural or urban. These locations were geocoded and matched using the digital maps described above.

Table A.1: The fixed assets of those firms that hold them

Observations are at the firm level for firms covered by the census in the study sample. The models estimate equation (1) introducing all controls and fixed effects. t-statistics based on standard errors clustered by district are in parentheses.

	Dependent Variable:		
	Log of Fixed Assets (for Firms with Fixed Assets)		
Sample:	< 100 km of Boundary	< 50 km of Boundary	Border district
Panel A. Cubic Polynomial in Distance to Mita Boundary			
Mita	0.027	0.089	0.003
	(0.12)	(0.46)	(0.02)
R^2	0.33	0.40	0.42
Panel B. Cubic Polynomial in Distance to Potosi			
Mita	0.219	0.051	0.015
	(0.89)	(0.25)	(0.08)
R^2	0.35	0.40	0.39
Panel C. Cubic Polynomial in Latitude and Longitude			
Mita	−0.142	0.007	−0.050
	(−0.60)	(0.03)	(−0.19)
R^2	0.37	0.41	0.41
N. clusters (districts)	268	169	60
Sample size	26593	9614	3681

Table A.2: Robustness to alternative specifications

This table presents the coefficient on the mita variable for all robustness checks reported in Section 5.5 of the paper. Each entry is from a different regression. Each line in the table, as indicated on the left-hand side labels, uses a different dependent variable and specification with cubic polynomials of running variable A (distance to the mita boundary), B (distance to Potosi), or C (flexible function of latitude and longitude) as in all models of the paper using border districts only. Ten specifications are reported in the ten columns. First, the cubic polynomial specifications for function $g(\cdot)$ in equation (1) are changed to linear polynomials. Second, the sample of firms close to the mita boundary is further narrowed to 10 kilometers. Third, the granularity of the number of segments of the mita boundary for fixed effects is varied to 60 segments. Fourth, the number of these segments is reduced to 20. Fifth, the industry fixed effects are dropped. Sixth, the firm age fixed effects are dropped. Seventh, the geographic control variables and the industry and age fixed effects are all dropped. Eighth, metropolitan Cusco is included in the estimation sample. Ninth, a definition of Cusco that excludes the Lucre and Oropesa districts is used to exclude Cusco from the tests. Tenth, areas that are close to a mita boundary segment that includes a river are excluded. t-statistics based on standard errors clustered by district are in parentheses. ***, **, * stand for significance at the 1%, 5% and 10% level, respectively.

Dep.Var, Run.Var.:	Linear Poly.	Narrow 10km	60 Segm.	20 Segm.	No Ind. FE	No Age FE	No Geo., No Ind. FE, No Age FE	With Cusco	No Alt. Cusco	No Rivers
Log Sales, A	−0.478***	−0.307	−0.393***	−0.179	−0.487***	−0.474***	−0.561***	−0.437***	−0.416***	−0.520***
	(−2.90)	(−1.33)	(−3.95)	(−1.13)	(−3.07)	(−3.13)	(−2.66)	(−3.09)	(−2.75)	(−3.82)
Log Sales, B	−0.482***	−0.115	−0.470***	−0.231	−0.391***	−0.390***	−0.405**	−0.372***	−0.344***	−0.462***
	(−3.09)	(−0.63)	(−4.00)	(−1.36)	(−3.04)	(−3.15)	(−2.58)	(−3.54)	(−2.90)	(−3.95)
Log Sales, C	−0.400**	0.024	−0.391*	−0.706***	−0.085	−0.164	−0.056	−0.141	−0.025	−0.108
	(−2.38)	(0.11)	(−1.79)	(−4.18)	(−0.39)	(−0.82)	(−0.21)	(−0.90)	(−0.15)	(−0.59)
No F.Assets, A	0.126**	0.249***	0.108**	0.147***	0.118**	0.119**	0.131***	0.117**	0.128***	0.124**
	(2.49)	(3.93)	(2.04)	(3.68)	(2.27)	(2.30)	(3.20)	(2.52)	(2.74)	(2.36)
No F.Assets, B	0.127**	0.231***	0.092	0.142***	0.123**	0.134***	0.123**	0.125**	0.140***	0.145**
	(2.32)	(3.44)	(1.63)	(3.47)	(2.36)	(2.72)	(2.38)	(2.64)	(3.06)	(2.48)
No F.Assets, C	0.108**	0.172**	0.260***	0.188***	0.268***	0.279***	0.254***	0.206***	0.235***	0.320***
	(2.09)	(2.60)	(3.37)	(2.70)	(3.83)	(4.51)	(3.37)	(3.67)	(4.68)	(4.57)
C.Name, A	−0.173***	−0.187***	−0.149***	−0.141***	−0.174***	−0.173***	−0.179***	−0.167***	−0.169***	−0.186***
	(−3.68)	(−4.11)	(−3.43)	(−2.95)	(−3.19)	(−3.65)	(−3.64)	(−3.63)	(−3.61)	(−3.99)
C.Name, B	−0.172***	−0.164***	−0.132***	−0.149***	−0.133***	−0.138***	−0.136***	−0.143***	−0.145***	−0.146***
	(−3.85)	(−4.67)	(−3.92)	(−3.79)	(−3.41)	(−4.66)	(−2.67)	(−4.45)	(−4.45)	(−4.91)
C.Name, C	−0.168***	−0.307***	−0.139***	−0.229***	−0.167**	−0.165***	−0.178**	−0.180***	−0.176***	−0.170***
	(−4.15)	(−4.91)	(−3.24)	(−5.50)	(−2.55)	(−3.84)	(−2.37)	(−5.19)	(−4.65)	(−3.90)
Tax ID, A	−0.114***	−0.096*	−0.129***	−0.076**	−0.122**	−0.111***	−0.145**	−0.113***	−0.108***	−0.119***
	(−3.08)	(−1.95)	(−4.16)	(−2.32)	(−2.56)	(−3.12)	(−2.31)	(−3.17)	(−2.99)	(−3.39)
Tax ID, B	−0.111***	−0.095*	−0.108***	−0.087**	−0.085**	−0.086***	−0.092*	−0.099***	−0.093***	−0.093***
	(−3.03)	(−1.68)	(−3.28)	(−2.45)	(−2.02)	(−2.92)	(−1.82)	(−3.56)	(−3.30)	(−3.10)
Tax ID, C	−0.100***	−0.200**	−0.125*	−0.169***	−0.034	−0.016	−0.037	−0.069	−0.048	−0.012
	(−3.00)	(−2.09)	(−1.88)	(−3.72)	(−0.37)	(−0.25)	(−0.35)	(−1.27)	(−0.92)	(−0.18)

Table A.3: Placebo tests

Observations are at the firm level for all firms in "border districts" but for artificially created mita borders. The models shift the mita borders artificially to perform placebo tests. The first placebo (top panel) artificially moves the boundaries a couple of districts to the north of the true boundary and run narrow-window ten kilometer tests with polynomials in distance to the placebo boundaries. The second placebo (bottom panel) alternatively shifts the boundaries a couple of districts to the south of the real boundary limits with the same specification as the first placebo test. The usual control variables and fixed effects are included. t-statistics based on standard errors clustered by district are in parentheses.

	Dependent Variable:			
	Log of Sales	**No Fixed Assets**	**Has a Commercial Name**	**Has a Business Tax ID Number**
Placebo I: Artificially shifting all mita boundaries two districts to the North				
Mita	0.132	−0.034	−0.042	−0.028
	(1.57)	(−1.21)	(−1.28)	(−1.20)
R^2	0.46	0.33	0.22	0.31
N. clusters (districts)	80	80	80	80
Sample size	4584	4584	4584	4584
Placebo II: Artificially shifting all mita boundaries two districts to the South				
Mita	0.215	−0.022	0.013	0.025
	(1.23)	(−0.57)	(0.44)	(0.78)
R^2	0.48	0.34	0.24	0.27
N. clusters (districts)	62	62	62	62
Sample size	2579	2579	2579	2579

Table A.4: Relating firm outcomes to childhood stunting, *haciendas* and road density using all specifications

Observations are at the firm level for firms covered by the census in the study sample. Only models using the narrow sample of border districts are reported for brevity. The models estimate equation (1) for the usual firm-level dependent variables, controls, and fixed effects. The models include the following additional controls when indicated: the childhood stunting ratio for 2005 at the district level, the number of *haciendas* per district in 1689, and the density of regional road networks at the district level. *t*-statistics based on standard errors clustered by district are in parentheses. ***, **, * stand for significance at the 1%, 5% and 10% level, respectively.

	Dependent Variable:											
	Log of Sales			**No Fixed Assets**			**Commercial Name**			**Has a Business Tax ID Number**		
Control:	Stunted	Hac.	Roads	Stunted	Hac.	Roads	Stunted	Hac.	Roads	Stunted	Hac.	Roads
Panel A. Cubic Polynomial in Distance to Mita Boundary												
Mita	-0.438^{***}	-0.261	-0.429^{***}	0.135^{***}	0.131^{**}	0.137^{**}	-0.172^{***}	-0.115^{*}	-0.145^{***}	-0.113^{***}	-0.076^{*}	-0.099^{***}
	(-2.94)	(-1.46)	(-3.05)	(2.93)	(2.06)	(2.42)	(-3.67)	(-1.91)	(-4.18)	(-3.00)	(-1.86)	(-2.91)
R^2	0.44	0.44	0.44	0.26	0.26	0.26	0.24	0.24	0.24	0.33	0.33	0.33
Panel B. Cubic Polynomial in Distance to Potosi												
Mita	-0.375^{***}	-0.342^{**}	-0.372^{***}	0.146^{***}	0.140^{**}	0.147^{***}	-0.144^{***}	-0.121^{***}	-0.129^{***}	-0.093^{***}	-0.076^{*}	-0.083^{***}
	(-3.19)	(-2.09)	(-3.07)	(3.02)	(2.18)	(2.75)	(-4.78)	(-2.76)	(-4.84)	(-2.93)	(-1.86)	(-2.96)
R^2	0.44	0.44	0.44	0.26	0.26	0.26	0.24	0.24	0.24	0.33	0.33	0.33
Panel C. Cubic Polynomial in Latitude and Longitude												
Mita	-0.162	0.051	-0.145	0.297^{***}	0.285^{***}	0.315^{***}	-0.173^{***}	-0.142^{***}	-0.157^{***}	-0.024	-0.012	-0.005
	(-0.88)	(0.26)	(-0.73)	(4.50)	(3.42)	(5.18)	(-4.08)	(-2.97)	(-3.49)	(-0.37)	(-0.17)	(-0.09)
R^2	0.45	0.45	0.45	0.27	0.27	0.27	0.25	0.25	0.25	0.33	0.33	0.33
N. clusters (districts)	64	64	64	64	64	64	64	64	64	64	64	64
Sample size	5895	5895	5895	5895	5895	5895	5895	5895	5895	5895	5895	5895

Table A.5: Home ownership and house value index in local populations

The data are from the population census. Observations are at the household level for all border districts covered by the business census. The models estimate an equation analogous to (1). House index is the first, most important component of a principal components analysis of 22 home-specific variables. t-statistics based on standard errors clustered by district are in parentheses. ***, **, * stand for significance at the 1%, 5% and 10% level, respectively.

	Dependent Variable:	
	Owns House (1/0)	**House Index**
Panel A. Cubic Polynomial in Distance to Mita Boundary		
Mita	0.067*	−0.161
	(1.91)	(−1.06)
R^2	0.09	0.36
Panel B. Cubic Polynomial in Distance to Potosi		
Mita	0.071*	−0.170
	(1.97)	(−1.24)
R^2	0.09	0.37
Panel C. Cubic Polynomial in Latitude and Longitude		
Mita	0.102*	−0.209
	(1.93)	(−1.48)
R^2	0.09	0.38
N. clusters (districts)	64	64
Sample size	88682	88682

Table A.6: Exploring performance differences between firms in border districts and firms in the central zone considering only inside sections of mita region

Observations are at the firm level exclusively for firms inside the mita region. Only the set of firms located in border districts and the set of firms in the "central zone" are included in the analysis: border districts are those bordering with the mita boundary, whereas the central zone is defined in three alternative ways. First, whatever is not a border district is considered to be a district in the central zone (left-hand side panel of the table). Second, all territories at least 75 kilometers away from the mita border are considered the central zone (middle panel of the table). Third, all territories at least 90 kilometers away from the mita border are labeled the central zone (right-hand side panel of the table). The explanatory variable of interest is a dummy for border districts. Young firm is defined as a dummy equal to one for firms of an age less than or equal to the median age of firms. t-statistics based on standard errors clustered by district are in parentheses. ***, **, * stand for significance at the 1%, 5% and 10% level, respectively.

Dependent Variable: **Log of Sales**									
Central zone defined as:	Wherever not a border district			75 km away from border			90 km away from border		
Border district	−0.145	−0.133	−0.123	−0.564	−0.566	−0.573	−0.102	−0.104	−0.111
	(−0.75)	(−0.69)	(−0.69)	(−1.31)	(−1.31)	(−1.36)	(−0.35)	(−0.35)	(−0.37)
Border × Young			−0.021			0.033			0.001
			(−0.32)			(0.61)			(0.01)
Young firm			−0.151***			−0.187***			−0.164*
			(−5.53)			(−13.73)			(−1.97)
Geo.controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Segment F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age F.E.	Yes	No	No	Yes	No	No	Yes	No	No
R^2	0.37	0.36	0.36	0.39	0.37	0.38	0.41	0.39	0.40
N. clusters...									
...(districts)	203	203	203	75	75	75	48	48	48
Sample size	31138	31138	31138	21367	21367	21367	3677	3677	3677

Table A.7: Performance implications of the mita when considering only old firms

Observations are at the firm level for firms in border districts. The models are exactly as those in Table 2 but keeping in the sample only old firms. Old firms are defined alternatively as those older than the median age (first column) or those older than 20 years of age (second column). t-statistics based on standard errors clustered by district are in parentheses. ***, **, * stand for significance at the 1%, 5% and 10% level, respectively.

	Dependent Variable:	
	Log of Sales	
Sample:	Firms older than the median age	Firms older than 20 years of age
Panel A. Cubic Polynomial in Distance to Mita Boundary		
Mita	-0.552^{***}	-1.075^{**}
	(-3.26)	(-2.17)
R^2	0.46	0.44
Panel B. Cubic Polynomial in Distance to Potosi		
Mita	-0.425^{***}	-0.513^{**}
	(-2.74)	(-2.30)
R^2	0.46	0.47
Panel C. Cubic Polynomial in Latitude and Longitude		
Mita	-0.275	-1.047
	(-1.20)	(-1.32)
R^2	0.47	0.49
N. clusters (districts)	62	34
Sample size	2769	256

Table A.8: Testing for differential performance implications of the mita for old firms

Observations are at the firm level for firms in border districts. The models are as in Table 2 but extended to include an interaction for an old firm dummy. Old firms are defined alternatively as those older than the median age (columns 1 and 2) or those older than 20 years of age (columns 3 and 4). t-statistics based on standard errors clustered by district are in parentheses. ***, **, * stand for significance at the 1%, 5% and 10% level, respectively.

	Dependent Variable:			
	Log of Sales			
Old firm defined as:	Older than median age		Older than 20 years	
Age fixed effects:	Not included	Included	Not included	Included
	Panel A. Cubic Polynomial in Distance to Mita Boundary			
Mita	-0.439^{***}	-0.448^{***}	-0.469^{***}	-0.474^{***}
	(-2.86)	(-2.99)	(-3.11)	(-3.19)
Mita × Old firm	-0.083	-0.068	-0.104	-0.095
	(-1.43)	(-1.44)	(-0.77)	(-0.79)
Old firm	0.214^{***}		-0.017	
	(7.69)		(-0.19)	
R^2	0.43	0.44	0.42	0.44
	Panel B. Cubic Polynomial in Distance to Potosi			
Mita	-0.364^{***}	-0.376^{***}	-0.385^{***}	-0.393^{***}
	(-2.83)	(-2.99)	(-3.10)	(-3.23)
Mita × Old firm	-0.062	-0.047	-0.104	-0.097
	(-1.07)	(-0.98)	(-0.76)	(-0.76)
Old firm	0.211^{***}		-0.012	
	(7.94)		(-0.12)	
R^2	0.43	0.44	0.42	0.44
	Panel C. Cubic Polynomial in Latitude and Longitude			
Mita	-0.136	-0.171	-0.161	-0.190
	(-0.68)	(-0.85)	(-0.81)	(-0.96)
Mita × Old firm	-0.066	-0.051	-0.092	-0.086
	(-1.13)	(-1.07)	(-0.69)	(-0.70)
Old firm	0.209^{***}		-0.022	
	(7.69)		(-0.25)	
R^2	0.43	0.44	0.43	0.44
N. clusters (districts)	64	64	64	64
Sample size	5895	5895	5895	5895

Table A.9: Testing for the differential financial performance of tax-registered firms with banking relationships

The models are exactly as those in Table 7 but extended to include an interaction term for those mita firms that had a business tax ID at the moment their financial performance is measured. This interaction coefficient is reported. All models additionally include as explanatory variables whether the firm is in a mita region, whether the firm had a business tax ID that month, as well as the usual controls and fixed effects. t-statistics based on standard errors clustered by district are in parentheses.

	Dependent Variable:		
	Worst Classification	**W.Average Classification**	**Fraction of Troubled Debt**
Panel A. Cubic Polynomial in Distance to Mita Boundary			
Mita× Has Business Tax ID	0.189	0.015	0.045
	(1.43)	(1.33)	(1.40)
Variables in levels	Yes	Yes	Yes
R^2	0.08	0.07	0.07
Panel B. Cubic Polynomial in Distance to Potosi			
Mita× Has Business Tax ID	0.198	0.016	0.048
	(1.47)	(1.37)	(1.46)
Variables in levels	Yes	Yes	Yes
R^2	0.08	0.06	0.07
Panel C. Cubic Polynomial in Latitude and Longitude			
Mita× Has Business Tax ID	0.181	0.014	0.044
	(1.32)	(1.23)	(1.30)
Variables in levels	Yes	Yes	Yes
R^2	0.08	0.07	0.07
N. clusters (firms)	1720	1720	1720
N. clusters (districts)	39	39	39
Sample size	17699	17699	17699

Table A.10: Additional results on trust in institutions

Observations are at the respondent level for all respondents with locations within 50 kilometers of the mita border. The models follow exactly those in the top section of Table 10 with alternative dependent variables. t-statistics based on standard errors clustered by district are in parentheses.

Dependent Variables:	**Trust in Institutions...**	
	Political Parties	**Ministry of Education**
Panel A. Cubic Polynomial in Distance to Mita Boundary		
Mita	0.004	0.014
	(0.42)	(0.43)
R^2	0.06	0.08
Panel B. Cubic Polynomial in Distance to Potosi		
Mita	−0.009	0.041
	(−0.90)	(1.11)
R^2	0.07	0.08
Panel C. Cubic Polynomial in Latitude and Longitude		
Mita	−0.000	0.008
	(−0.02)	(0.14)
R^2	0.08	0.09
N. clusters (districts)	116	116
Sample size	1313	1469