Appendix A

Instrumental Variables Results

- Figure A1 shows that the logic of our instrument holds. Here it uses 20th century data on temperature and two measures of the ability to grow wheat to demonstrate the robust relationship between the two. We employ two data sets from the FAO. The first, the "Agro-climatically attainable yield for rain fed wheat", is from the Global Agro-ecological Assessment for Agriculture in the 21st century. It captures the ability of land to produce wheat absent of modern irrigation techniques. We estimate the optimal climate to grow wheat (at around 10.5 °C). A clear a parabolic relationship between temperature and this FAO measure is observed simply by plotting it against average annual temperature between 1960 and 2000.

- In Table A1 we regress the FAO measure of wheat suitability on the absolute deviation from 10.5 degrees showing that there is a negative relationship between the two. The results from this regression are summarized in the first column of Table A1. The effect of a one degree deviation from the optimal temperature is substantial, decreasing the FAO measure by .61 units. This is a particularly large effect since the FAO measure is on a fourteen point scale. Moreover, a large amount of the variation in the FAO wheat suitability measure is explained by deviation from this optimal growing temperature, the R^2 statistic is calculated to be .55. In addition, regressing average annual wheat yields between 1960 and 1990 on deviation from the optimal growing temperature again shows a similarly robust relationship. A one degree deviation from the optimal temperature has a large effect on average annual wheat yields – approximately 1600 hectograms per hectare.

- We follow Conley et al (2012) and conduct a sensitivity analysis using their union of confidence intervals method and allow for a direct effect of the instrument on the outcome over a range of symmetric bounds. Formally, we estimate via a 2SLS-like procedure $Y = X\beta + Z\gamma + \epsilon$, where Z is our instrument and X is the variable of interest and allow the parameter γ - capturing the direct effect of the instrument on the outcome - to take on a range of values $\gamma_0 \in [-\gamma, \gamma]$.

  In Table A2 we provide results from this procedure, giving values of γ where β - our parameter of interest - is null. We conduct this exercise, treating as the outcome of interest the number of textile centers, iron production centers, and their sum, respectively. Across each specification it would take a substantively large direct effect of the instrument in order to make the effect of early urban density null. Comparing our estimates from this procedure to the reduced-form regression of the presence of proto-industry (each of our measures) on our instrument, the size of the violation required to make our estimates null is between two thirds and just over three quarters of this reduced form relationship between the instrument and our outcomes.

- Table A2 estimates the effect of early urbanization (measured in the year 1200) on the number of proto-industrial centers existing before 1500 on a given 225 x 225 km grid-unit with an instrumental variables approach. It uses the optimal growing temperature for cereals like wheat as an instrument for the size of the urban population in existence on a particular grid-unit and via two-stage least squares we estimate the effect of urban density in 1200 on the presence of proto-industry (textile centers, iron forges, and their sum) before 1500. As in the main text, across specification we find a positive and statistically significant relationship between the size of urban populations living on a grid-square in 1200 and the number of proto-industrial centers before 1500.

Robustness Tests

- Table A3 reproduces the results from Table 1 in the main text. However, instead of using 225 km $\times$ 225 km units of observation, it employs units approximately half that size. The results remain nearly identical and indicate that urban population is driven by a non-stationary, exploding, process.
- Table A4 reproduces Table 3 from the main text. However, instead of using 225 km $\times$ 225 km units of observation, it relies on units approximately half that size. The results remain nearly identical and indicate that urban population caused the emergence of a proto-industrial core.
- Table A5 reproduces Table 3 from the main text now accounting for the fact that not all grid-squares have the same land-mass by dividing each unit by its total land area. The results are substantively identical. Here it is important to note that point estimates though on a different scale than those in the main text - a result of scaling each unit by its territory - the results yielded are close in magnitude to those presented in the main text.
- Tables A6 & A7 reproduce Tables 5 and 6 from the main text now treating the outcome as the natural log of parliamentary frequency. The results remain substantively identical.
- Table A8 examines the effect of proto-industrial activity before 1500 on the existence of a large city in 1500 and 1800 after conditioning on the level of urban density in 1200, defining as large cities as those with populations greater than 20,000 in 1500 and larger 100,000 in 1800. We dichotomize the independent variable to reflect the presence or absence of proto-industry. The table presents the marginal effect of adding a proto-industrial center: textile production, iron production, or either. These estimates range from a 45.4 % to a 52.8% percent increase in the presence of large city.
- Table A9 reproduces Table 1 from the main text. However, instead of using urban population as the main independent variable, it uses units estimates of urban population density (urban population/total population) derived by HYDE project. The results remain nearly identical and indicate that urban population is driven by a non-stationary, exploding, process.
- Table A10 reproduces Table 5 from the main text. However, instead of using urban population as the main independent variable, it uses units estimates of urban population density (urban population/total population) derived by HYDE project. The results remain nearly identical and indicate that urban population density caused the emergence of a proto-industrial core.
- Table A11 reproduces Table 2 from the main text. However, instead of using urban population as the main independent variable, it uses units estimates of urban population density (urban population/total population) derived by HYDE project. The results remain nearly identical

and indicate that urban population density is correlated with nineteenth and contemporary levels of per capita income.

- Table A12 estimates the relationship between early urban population (measured in 1200) and the presence of porto-industrial activity after accounting for differences in ethnolinguistic characteristics of groups present on each grid-square. We account for this by constructing dummies for each linguistic group that exist on each grid-square as derived from the World Language Mapping System (2015). The results again show that there is a persistent relationship between early urban development and the presence of proto-industrial activity even after accounting for ethnolinguistic differences.

Table A13 Reproduces Table 7. from the main text. However, instead of using the distance weighted measure of conflict, we take the count of conflicts that occurred within the boundaries of a given unit as our main measure of exposure to conflict. The results are qualitatively similar to those presented in main text.

Tables A14-15 Reproduces Table 5 from the main text and explores the relationship between our measures of conflict, loss of sovereignty, and parliamentary frequency and future urban density.

Table A16 Investigates the sensitivity of our main results to the presence of spatial interference. We model this as follows: $\mu_{i,t} = \beta\mu_{it-1} + \lambda(I \otimes W_n)\mu_{i't} + \epsilon_{i,t}$ Where W_n is an NxN spatial weights, I the identity matrix, and λ the spatial autoregressive parameter. Furthermore, we model the disturbance $\epsilon_{i,t} = \rho(I \otimes W_n)\epsilon_{i,t} + \eta_{it}$ and assume $\epsilon_{i,t} \sim IID(0, \sigma^2_\epsilon)$ and $\eta_{i,t} \sim IID(0, \sigma^2_\eta)$. That is, we account for spatial interference through the random component ρ and directly through the spatial lag λ. We find that our estimates of the inter-temporal relationship between levels of urban density are qualitatively unchanged when we account for spatial interference.

Optimal Growing Temperature and Wheat Propensity/Yeilds (1960-2000)

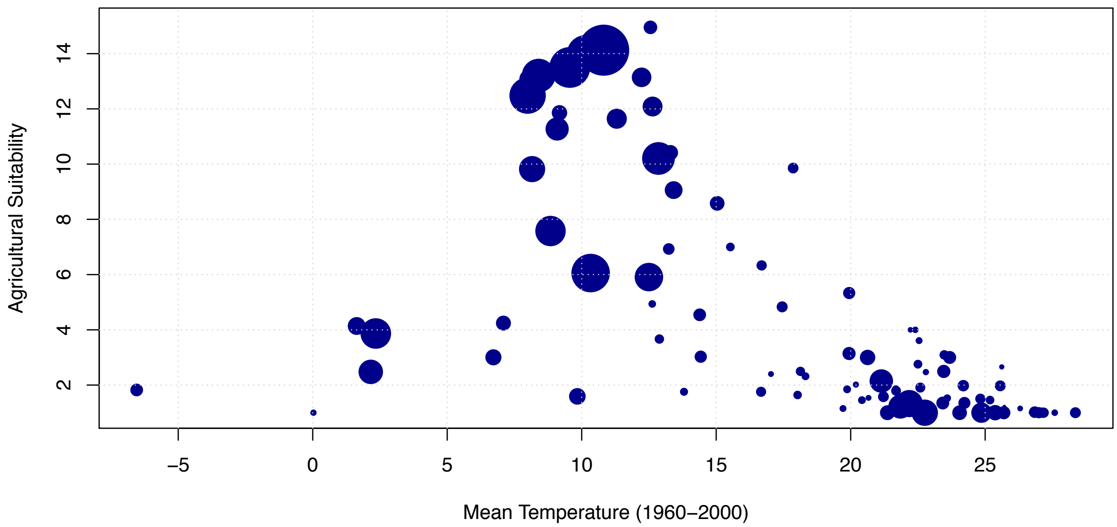
	FAO Wheat Suitability	Avg Wheat Yield
Intercept	10.20***	36372***
	(.49)	(2538)
|Temp - Optimal Growing Temp |	-.64***	-1574.0***
	(.05)	(242.7)
R^2	.57	.30
N	125	100

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 1: Wheat Suitability Measures

The relationship between temperature and the suitability to produce wheat. The first column regresses the FAO wheat suitability index against the absolute average distance from 10.5 °C between 1960 and 2000. The second column regresses the average wheat yield on this measure.

Figure 1: The FAO Wheat Suitability Index



The FAO wheat suitability index is plotted on the y-axis against average annual temperature on the x-axis. The FAO measure is the "Agro-climatically attainable yield for rain fed wheat" is from the Global Agro-ecological Assessment for Agriculture in the 21st century. It captures the ability of land to produce wheat absent of modern irrigation techniques. A clear parabolic relationship with a peak at approximately 10.5 degrees Celsius is observed. The radius of each circle is proportional to the average wheat yield between 1960 and 2000.

The Relationship Between Urban Density and Incomes Proto-Industry by 1500 - 2SLS Results

	Textiles	*Iron*	*Total Proto-Industry*	*Textiles*	*Iron*	*Total Proto-Industry*
Urban Density_{1200}	0.0735***	0.1277***	0.2011***			
	(0.0146)	(0.0285)	(0.0379)			
log(Urban Density_{1200})				0.3116***	0.5417***	0.8533***
				(0.0566)	(0.0914)	(0.1195)
Conley et al (2012) Violation : s.t. $\beta = 0$	-0.051	-0.095	-0.158	-0.050	-0.095	-0.156
R^2	-0.2746	-0.7560	-0.8045	0.1587	-0.0312	0.0517
N	444	444	444	444	444	444

	Textiles	*Iron*	*Total Proto-Industry*	Urban Density_{1200}	Urban Density_{1200}
| Temp - Optimal Growing Temp |	-0.0784***	-0.1364***	-0.2148***	-1.0680***	-0.2517***
	(0.0172)	(0.0242)	(0.0352)	(0.2534)	(0.0338)
R^2	0.0556	0.0546	0.0760	0.0339	0.0849
N	444	444	444	444	444
Cragg-Donald F Stat				17.7636	55.3544

Table 2: The top panel of this table presents instrumental variables results of the effect of early urban density (measured in the year 1200) on the number of proto-industrial centers on a given geographic unit before 1500. In each specification it uses the deviation from the optimal growing temperature as an instrument for urban density. For each model we give the exclusion restriction violation (the magnitude of the direct effect of the instrument on each outcome of interest) as derived from Conley et al. (2012)s union of confidence intervals method of sensitivity analysis that would yield a null result. In the first three columns of the lower panel we regress each of our measures of proto-industry on the instrument. In the last two we present the first stage results from the instrumental variables analysis. Heteroskedasticity robust standard errors in parentheses.

Autoregressive Models of Urban Growth

	Outcome: Urban Pop_t							Outcome: log(Urban Pop)$_t$					
	OLS			*GMM*				*OLS*			*GMM*		
Urban Pop_{t-1}	1.26***	1.11***	1.23***	1.27***	0.91***	0.84***	log(Urban Pop_{t-1})	0.88***	0.62***	0.62***	0.65***	0.69***	0.62***
	(0.06)	(0.17)	(0.19)	(0.08)	(0.11)	(0.19)		(0.01)	(0.02)	(0.03)	(0.04)	(0.03)	(0.03)
Urban Pop_{t-2}		0.24	0.14		0.59***	0.49***	log(Urban Pop_{t-2})		0.36***	0.28***		0.32***	0.36***
		(0.16)	(0.12)		(0.10)	(0.11)			(0.02)	(0.03)		(0.03)	(0.05)
Urban Pop_{t-3}			-0.00			0.34***	log(Urban Pop_{t-3})			0.13***			0.15***
			(0.12)			(0.08)				(0.02)			(0.03)
N	10110	8425	6740	10110	8425	6740		10110	8425	6740	10110	8425	6740
R^2	0.63	0.70	0.71					0.78	0.80	0.82			
$m2$				0.0160	0.0009	0.2653					0.9170	0.0975	0.3700
	Breitung Test (H_0: All Panels Contain Unit Roots)			Hadri LM Test (H_0:All Panels Are Stationary)				Breitung Test (H_0: All Panels Contain Unit Roots)			Hadri LM Test (H_0:All Panels Are Stationary)		
	8.25	50.14	60.47	79.14***	78.77***	30.35***		1.36	33.0971	23.75	43.06***	35.25***	4.83***
	(0.99)	(0.99)	(0.99)	(0.00)	(0.00)	(0.00)		(0.99)	(0.99)	(0.99)	(0.00)	(0.00)	0.00)

Table 3: This table presents the estimates of the autoregressive relationship between past and present urban development. The unit of observation is the 100 km x 100 km grid-square. The top panel measures total urban population and the lower takes the logarithm of this number. Heteroskedaticity robust standard errors clustered by unit in parentheses. p-value for the Arellano-Bond test of second order serial correlation in the errors denoted as $m2$.

The Effect of Early Urbanization on the Development of Protoindustry by 1500

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
	Textile Production				**Iron Production**				**Total Protoindustry**			
Urban Density$_{1200}$	0.0229***	0.0196***	0.0998***	0.0326***	0.0057***	0.0029*	0.0346***	0.0066	0.0286***	0.0225***	0.0779***	0.0229***
	(0.0039)	(0.0035)	(0.0170)	(0.0078)	(0.0017)	(0.0015)	(0.0072)	(0.0044)	(0.0050)	(0.0043)	(0.0127)	(0.0047)
R^2	0.20	0.26			0.023	0.12			0.17	0.28		
θ			5.31	1.39			2.24	0.50			3.74	0.97
Effect at $1.3 \times R^2$ of Controlled Regression		0.016				0.0018				0.017		
log(Urban Density$_{1200}$)	0.0937***	0.0796***	0.3252***	0.1769***	0.0327***	0.0193***	0.1892***	0.0586**	0.1264***	0.0989***	0.2649***	0.1257***
	(0.0140)	(0.0133)	(0.0212)	(0.0214)	(0.0061)	(0.0066)	(0.0194)	(0.0234)	(0.0164)	(0.0159)	(0.0159)	(0.0162)
R^2	0.23	0.28			0.06	0.13			0.23	0.31		
θ			1.5892	0.8488			0.9198	0.4374			1.6923	0.7140
Effect at $1.3 \times R^2$ of Controlled Regression		0.05				-.026				0.07		
Model:	OLS	OLS	NegBin	NegBin	OLS	OLS	NegBin	NegBin	OLS	OLS	NegBin	NegBin
Controls	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
N	1693	1693	1693	1693	1693	1693	1693	1693	1693	1693	1693	1693

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$

Table 4: This table presents estimates of the effect of early urban development (in the year 1200) on the number of proto-industrial centers in existence on a given 100 km × 100 km unit. Heteroscedasticity robust standard errors in parentheses. Controls are: terrain ruggedness, agricultural suitability, distance to coast, river length, coast length, latitude, and longitude. Following Oster (2013) we provide sensitivity estimates of the effects under the hypothetical condition when unobservables account for 1.3 × the R^2 from the controlled regressions.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
	Textile Production				**Iron Production**				**Total Protoindustry**			
Urban Density_{1200}	408.6814*	323.9035*	1262.6269***	293.3361**	396.3106	250.2302	871.2381***	360.1849	804.9920*	574.1338*	1078.0977***	324.8560**
	(208.7774)	(167.4876)	(468.1530)	(132.6728)	(258.3946)	(186.6418)	(277.1392)	(228.1648)	(456.5578)	(340.2046)	(323.3648)	(157.4880)
R^2	0.11	0.30			0.04	0.33			0.08	0.42		
θ			4.93	0.84			16.50	2.90			9.25	1.46
Effect at $1.3 \times R^2$ of Controlled Regression		282.85				201.24				442.17		
log(Urban Density_{1200})	1.9059***	1.5188***	3.0630***	1.1778***	2.1157***	1.2870**	2.3985***	1.0895**	4.0216***	2.8058***	2.7679***	1.0600***
	(0.4851)	(0.4171)	(0.5641)	(0.2148)	(0.6995)	(0.6432)	(0.5308)	(0.4294)	(1.0852)	(0.9567)	(0.4864)	(0.2850)
R^2	0.24	0.36			0.10	0.35			0.19	0.46		
θ			1.59	0.85			0.92	0.44			1.69	0.71
Effect at $1.3 \times R^2$ of Controlled Regression		1.91				1.82				2.18		
Model:	OLS	OLS	NegBin	NegBin	OLS	OLS	NegBin	NegBin	OLS	OLS	NegBin	NegBin
Controls	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
N	444	444	444	444	444	444	444	444	444	444	444	444

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table 5: This table presents estimates of the effect of early urban development (in the year 1200) on the number of proto-industrial centers in existence on a given 225 km × 225 km unit. Heteroscedasticity robust standard errors in parentheses. To account for the observations with coastline, we divide each grid-square by the size of territory in square kilometers within it. First stage results are presented in the lower panel.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
	Outcome: log(Frequency of Parliaments)				Outcome: log(Urban Density$_{t+1}$)					
log(Urban Density$_t$)	0.461***	0.487***	0.032	0.000			0.718***			0.299**
	(0.05)	(0.05)	(0.05)	(0.04)			(0.05)			(0.15)
log(Frequency of Parliaments)					0.221***	0.230***	0.085***	-0.038	-0.023	-0.002
					(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.04)
Region Effects	N	Y	N	N	N	Y	Y	N	N	N
Year Effects	N	Y	Y	Y	N	Y	Y	Y	Y	Y
Country Effects	N	N	Y	Y	N	N	N	Y	Y	Y
System GMM	N	N	N	N	N	N	N	N	N	Y
R^2	0.093	0.117	0.001	0.057	0.094	0.174	0.539	0.003	0.141	
N × T	1790	1790	1790	1790	1793	1793	1793	1793	1793	1793
N = 309, T=6										
$m2$										0.04

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 6: This table estimates Table 3 in the main text now taking the natural logarithm of the parliamentary frequency index. The first four columns of this table presets the estimated effect of urban density in period t on the frequency with which parliaments met in periods t to t+1. Columns 5-10 presets results of the relationship between the frequency of past parliamentary meetings and urban density in the subsequent century. Heteroscedasticity robust standard errors clustered by semi-sovereign unit in parentheses.

Initial Urban Conditions and Parliamentary Life Across Time

	1.	2.	3.	4.	5.	6.	7.	8.	9.
	1300-1400	*1400-1500*	*1500-1600*	*1600-1700*	*1700-1800*	*1400-1500*	*1500-1600*	*1600-1700*	*1700-1800*
log(Urban Density$_{1200}$)	0.046***	0.042****	0.040***	0.038***	0.038***	0.053***	0.052***	0.053***	0.050***
	(0.007)	(0.007)	(0.005)	(0.005)	(0.006)	(0.007)	(0.006)	(0.007)	(0.009)
Δ_{t-1200}log(Urban Density)	0.022**	-0.003	0.000	0.001	-0.001	-0.032***	0.033**	-0.037***	0.032***
	(0.007)	(0.005)	(0.002)	(0.001)	(0.001)	(0.008)	(0.007)	(0.008)	(0.007)
Δ_{t-1300}log(Urban Density)						0.062***	-0.098***	0.147***	-0.150***
						(0.013)	(0.021)	(0.028)	(0.038)
Δ_{t-1400}log(Urban Density)							0.079***	-0.204***	0.263**
							(0.021)	(0.041)	(0.073)
Δ_{t-1500}log(Urban Density)								0.110***	-0.210***
								(0.026)	(0.070)
Δ_{t-1600}log(Urban Density)									0.055
									(0.032)
Region Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	290	289	286	287	288	289	286	287	288
R^2	0.100	0.085	0.100	0.090	0.090.	0.145	0.160	0.172	0.147

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 7: This table provides estimates of the relationship between initial urban density and parliamentary meeting frequency. Each column regresses the fraction of years in a given century on the logged value of urban density in the year 1200. We account for the overall change between any set of periods, such that Δ_t represents the change in urban density over t centuries. Hetereskedacisity robust standard errors clustered by semi-sovereign unit in parentheses.

Proto-Industrialization and the Presence of a Large Cities

A.)

	City > 20,000 in 1500			**City > 100,000 in 1800**		
Textile Center	0.454			0.464		
	(0.078,0.742)			(0.059,0.744)		
Iron Production		0.526			0.528	
		(0.326,0.678)			(0.337,0.696)	
Either Proto-Industry			0.519			0.519
			(0.059,0.744)			(0.346,0.683)
Log Likelihood	-134.74	-124.17	-121.01	-132.19	-125.06	-118.63
N	444	444	444	444	444	444

B.)

	Textiles Center		**Iron Production**		**Either Proto-Industry**	
City > 5,000 in 1200	0.438		0.421		0.555	
	(0.311,0.572)		(0.299,0.552)		(0.419,0.678)	
City > 20,000 in 1500		0.293		0.388		0.494
		(0.197,0.402)		(0.286,0.498)		(0.389,0.603)
Log Likelihood	-44.06	-59.37	-107.17	-98.16	-117.86	-117.86
N	444	444	444	444	444	444

Table 8: Panel A presents the marginal effect of the presence of proto-industry before 1500 on the presence of a large city in 1500 and 1800. All estimates are derived from a logistic regression of a binary indicator of large city presence on a binary indicator of the presence of proto-industry, controlling for the total urban population in 1200. Ninety-five percent heteroskedacity robust confidence intervals derived from quasi-Bayesian simulation in parentheses.

Autoregressive Models of Urban Population Density

	Outcome: Urban Pop Density$_t$							**Outcome: log(Urban Pop Density)$_t$**					
	OLS			*GMM*				*OLS*			*GMM*		
Urban Pop$_{t-1}$	1.45***	1.83***	3.43***	1.64***	1.05***	2.00***	log(Urban Pop$_{t-1}$)	0.90***	0.98***	0.77***	0.88***	0.75***	0.38***
	(0.04)	(0.20)	(0.24)	(0.07)	(0.10)	(0.25)		(0.00)	(0.01)	(0.02)	(0.00)	(0.02)	(0.01)
Urban Pop$_{t-2}$		-0.45*	-1.33***		0.92***	1.19***	log(Urban Pop$_{t-2}$)		-0.08***	0.22***		0.09***	0.34****
		(0.22)	(0.19)		(0.10)	(0.18)			(0.01)	(0.02)		(0.02)	(0.01)
Urban Pop$_{t-3}$			-1.22***			-1.63***	log(Urban Pop$_{t-3}$)			-0.15****			0.02
			(0.15)			(0.31)				(0.02)			(0.02)
N	2664	2220	1776	2664	2220	1776		2664	2220	1776	2664	2220	
R^2	0.84	0.80	0.87					0.98	0.98	0.98			
$m2$				-3.24	6.42	5.40					15.00	7.75	-9.29
	Breitung Test (H_0: All Panels Contain Unit Roots)			Hadri LM Test (H_0:All Panels Are Stationary)				Breitung Test (H_0: All Panels Contain Unit Roots)			Hadri LM Test (H_0:All Panels Are Stationary)		
	30.61	27.61	38.63	35.70***	34.51***	15.36***		29.04	17.06	7.827	60.25***	58.02***	8.73***
	(.99)	(0.99)	(0.99)	(0.00)	(0.00)	(0.00)		(0.99)	(0.99)	(0.99)	(0.00)	(0.00)	0.00)

Table 9: This table presents the estimates of the autoregressive relationship between past and present urban population density as measured by the HYDE project. The unit of observation is the 225 km x 225km grid-square. The top panel measures total urban population and the lower takes the logarithm of this number. Heteroskedaticity robust standard errors clustered by unit in parentheses. p-value for the Arellano-Bond test of second order serial correlation in the errors denoted as $m2$.

Early Urban Population Density and The Development of Protoindustry

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
	Textile Production				**Iron Production**				**Total Protoindustry**			
Urban Pop Density$_{1200}$	0.67***	0.56**	1.87***	0.67**	1.40***	1.17**	2.58***	1.03***	2.07***	1.73**	2.42***	0.68**
	(0.18)	(0.20)	(0.31)	(0.22)	(0.35)	(0.38)	(0.35)	(0.30)	(0.48)	(0.53)	(0.28)	(0.23)
R^2	0.10	0.27			0.14	0.37			0.17	0.44		
θ			4.52	0.80			9.79	2.48			6.00	1.37
Effect at $1.3 \times R^2$ of Controlled Regression		0.503				1.065				1.56311		
log(Urban Pop Density$_{1200}$)	0.08***	0.07*	0.50***	0.26*	0.16***	0.20***	0.74***	0.74***	0.24***	0.26***	0.60***	0.42***
	(0.02)	(0.03)	(0.07)	(0.10)	(0.03)	(0.05)	(0.09)	(0.16)	(0.04)	(0.07)	(0.05)	(0.10)
R^2	0.07	0.24			0.09	0.33			0.12	0.39		
θ			2.99	0.95			6.43	2.30			4.02	1.35
Effect at $1.3 \times R^2$ of Controlled Regression		0.063				0.209				0.273		
Model:	OLS	OLS	NegBin	NegBin	OLS	OLS	NegBin	NegBin	OLS	OLS	NegBin	NegBin
Controls	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
N	444	444	444	444	444	444	444	444	444	444	444	444

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table 10: This table presents estimates of the effect of early urban population density (in the year 1200) derived from the HYDE project's estimates on the number of proto-industrial centers in existence on a given 225 km × 225 km unit. Heteroscedasticity robust standard errors in parentheses.

The Relationship Between Urban Population Density in 1800 and Per Capita Income in the 19th and 20th Centuries

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
	GDP per capita 1870						**GDP per capita 1900**					
	GDP per capita			log*(GDP per capita)*			*GDP per capita*			log*(GDP per capita)*		
log(Urban Density$_{1800}$)	495.24	121.95	117.58	0.26	0.08	0.08	559.18	110.99	230.33	0.21	0.07	0.11
	(228.39)	(165.62)	(117.44)	(0.12)	(0.10)	(0.07)	(318.13)	(204.44)	(130.67)	(0.13)	(0.09)	(0.07)
Constant	917.51	1408.83***	-3170.14*	6.95***	7.23***	4.61***	1483.32*	2244.81***	-5160.84**	7.36***	7.69***	4.64***
	(427.97)	(90.13)	(938.93)	(0.24)	(0.05)	(0.58)	(604.26)	(111.27)	(1012.69)	(0.27)	(0.05)	(0.67)
Controls	N	N	Y	N	N	Y	N	N	Y	N	N	Y
Country Fixed Effects	N	Y	Y	N	Y	Y	N	Y	Y	N	Y	Y
N	99	99	99	99	99	99	100	100	100	100	100	100
R^2	0.21	0.67	0.73	0.23	0.66	0.73		0.15	0.59	0.65	0.15	0.62
0.70												

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
	GDP per Capita in 2008											
	All Europe						**Western Europe**					
	GDP per capita			log*(GDP per capita)*			*GDP per capita*			log*(GDP per capita)*		
log(Urban Pop Density$_{1800}$)	5319.78***	4481.22**	3683.41*	0.36***	0.22***	0.18*	-1641.78	6242.94***	4330.50*	-0.04	0.25***	0.18*
	(1385.37)	(1324.55)	(1765.26)	(0.08)	(0.05)	(0.08)	(2128.58)	(1181.12)	(1749.26)	(0.07)	(0.04)	(0.07)
Constant	14872.82***	64755.11***	-8491.55	9.32***	10.66***	8.77***	31223.24***	29469.43***	-3562.10	10.29***	10.23***	8.91***
	(3169.80)	(3441.97)	(13752.80)	(0.19)	(0.12)	(0.63)	(5307.54)	(1052.65)	(20537.00)	(0.19)	(0.04)	(0.93)
Controls	N	N	Y	N	N	Y	N	N	Y	N	N	Y
Country Fixed Effects	N	Y	Y	N	Y	Y	N	Y	Y	N	Y	Y
N	245	245	245	245	245	245	156	156	156	156	156	156
R^2	0.25	0.80	0.81	0.42	0.88	0.89 0.02	0.31	0.40	0.01	0.32	0.43	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 11: The top panel describes the relationship between urban population density (derived from the HYDE project's estimates) in 1800 and per capita income in 1870 and 1900, respectively. The lower of panel describes the relationship between urban density in 1800 and per capita income in 2008. All observations are at the NUTS-2 level. In the lower panel the first six columns use all NUTS-2 regions for which there is income data. The last six columns use only those in Western Europe. Hetereoskedacticty robust standard errors clustered by country are in parentheses. Controls are: terrain ruggedness, agricultural suitability, distance to coast, coast length, latitude, and longitude. Following Oster (2013) we provide sensitivity estimates of the effects under the hypothetical condition when unobservables account for 1.3 × the R^2 from the controlled regressions.

The Relationship Between Urban Density and Incomes Proto-Industry by 1500

	Iron		*Textiles*		*Total Proto-Industry*	
log(Urban Density_{1200})	0.05**	0.03	0.03*	0.03	0.08**	0.06*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Controls	N	Y	N	Y	N	Y
Language Group Dummies	Y	Y	Y	Y	Y	Y
R^2	0.91	0.92	0.93	0.93	0.93	0.94
N	444	444	444	444	444	444

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 12: This table presents OLS estimates of the effect of early urban density (measured in the year 1200) on the number of proto-industrial centers on a given geographic unit before 1500. We include in each model dummies for each ethno-linguistic group present on each unit (225 x 225 km grid-squres). Heteroskedasticity robust standard errors in parentheses. Data on linguistic group are taken from the World Language Mapping System (2015).

The Effect of Conflict on Parliamentary Meeting Frequency

	(1.)	(2.)	(3.)	(4.)
	OLS		*GMM*	
$\log(\text{Conflict}_{i,t} + 1)$	0.048*	0.064**	-0.028	-0.006
	(0.024)	(0.023)	(0.032)	(0.026)
Sovereignty $\text{Index}_{i,t}$		0.258***		0.125**
		(0.033)		(0.039)
Parliamentary $\text{Index}_{i,t-1}$			0.938***	0.926***
			(0.044)	(0.043)
N	1790	1785	1496	1491
R^2	0.046	0.191		
$m2$			-2.45	-2.25

Table 13: The unit of observation is the sovereign/semi-sovereign political unit. This table presets the estimated effect of conflict (derived from Dincecco and Onorato, 2013) and the loss of sovereignty, respectively, in period t on the frequency with which parliaments met in periods t to t+1. Heteroscedasticity robust standard errors clustered by semi-sovereign unit in parentheses.

The Effect of Conflict and Parliamentary Meeting Frequency on Urban Development

	(1.)	(2.)	(3.)	(4.)	(5.)	(6.)
		OLS			*GMM*	
log(Conflict$_{i,t-1}+1$)	0.919		0.936	0.769		0.761
	(0.845)		(0.883)	(0.561)		(0.573)
Sovereignty Index$_{i,t-1}$		-1.389	-1.327		-0.183	-0.193
		(1.887)	(1.911)		(4.901)	(4.899)
Parliamentary Index$_{i,t-1}$	-7.728*	-6.795	-7.066*	-2.981	-2.449	-2.674
	(3.876)	(3.606)	(3.574)	(13.287)	(10.692)	(10.758)
Urban Density$_{i,t-1}$				-.0282	-.0284	-.0285
				(0.0285)	(0.0285)	(0.0284)
N	1783	1778	1778	1783	1778	1778
R^2	0.022	0.022	0.022			
m_2				-.976	-.978	-.978

Table 14: The unit of observation is the sovereign/semi-sovereign political unit. This table presets the estimated effect of conflict (derived from Dincecco and Onorato, 2013), the loss of sovereignty, respectively, in period $t-1$ and the frequency with which parliaments met between periods $t-1$ and t, on urban density in period t. Heteroscedasticity robust standard errors clustered by semi-sovereign unit in parentheses.

The Effect of Conflict and Parliamentary Meeting Frequency on Urban Development (Logged)

	(1.)	(2.)	(3.)	(4.)	(5.)	(6.)
		OLS			*GMM*	
$\log(\text{Conflict}_{i,t-1} + 1)$	0.160		0.174*	0.382**		0.389**
	(0.086)		(0.087)	(0.135)		(0.139)
Sovereignty $\text{Index}_{i,t-1}$		0.267	0.279		0.332	0.318
		(0.180)	(0.180)		(0.306)	(0.296)
Parliamentary $\text{Index}_{i,t-1}$	-0.202	-0.319	-0.369	0.270	0.180	0.166
	(0.349)	(0.363)	(0.367)	(0.637)	(0.734)	(0.697)
$\log(\text{Urban Density}_{i,t-1})$				0.377***	0.353***	0.365***
				(0.104)	(0.105)	(0.105)
N	1783	1778	1778	1783	1778	1778
R^2	0.120	0.120	0.121			
m_2				2.578	2.559	2.555

Table 15: The unit of observation is the sovereign/semi-sovereign political unit. This table presets the estimated effect of conflict (derived from Dincecco and Onorato, 2013), the loss of sovereignty, respectively, in period $t-1$ and the frequency with which parliaments met between periods $t-1$ and t, on the log of urban density in period t. Heteroscedasticity robust standard errors clustered by semi-sovereign unit in parentheses.

The Pre-Industrial Structure of Urban Growth (Autoregressive Models) Accounting for Spatial Interference

	(1.)	(2.)	(3.)	(4.)	(5.)	(6.)
Urban Density$_{t-1}$	1.22***	1.23***	1.099***	1.098***	1.208***	1.212***
	(0.0111)	(0.0111)	(0.0272)	(0.0272)	(0.0326)	(0.0326)
Urban Density$_{t-2}$			0.177***	0.187***	0.185***	
			(0.0347)	(0.0347)	(0.048)	(0.048)
Urban Density$_{t-3}$					-0.176***	-0.171***
					(0.050)	(0.050)
ρ	0.623***	0.519***	0.630***	.534***	0.626***	0.496***
	(0.020)	(0.030)	(0.022)	(0.033)	(0.025)	(0.037)
λ		0.108***		.094***		0.124***
		(0.021)		(0.022)		(0.022)

Table 16: The unit of observation is the 225 km x 225km grid-square. This table presents the autoregressive models of Urban Density accounting for spatial interference across units. ρ gives the spatial autoregressive parameter from the interference in errors component. λ gives the direct spatial autoregressive parameter.

Appendix B

Sources of Regional Income Data

Belgium Buyst, Erik. 2009. “Reversal of Fortune in a Small, Open Economy: Regional GDP in Belgium, 1896-2000.” Vives discussion paper 8. Electronic copy available at:

`http://ssrn.com/abstract=1586762`

France Combes, Pierre-Philippe, Miren Lafourcade, Jacques-Francois Thisse, and Jean-Claude Toutain. 2011. “The rise and fall of spatial inequalities in France: A long-run perspective.” Explorations in Economic History 48 (2): 243-271.

Germany Frank, Harald. 1994. Regionale Entwicklungsdisparitaten im deutschen Industrialisierungsprozess 1849-1939 : eine empirisch-analytische Untersuchung. MŁnster. Hoffmann, Walther G., and Josef Heinz MŁller. 1959. Das deutsche Volkseinkommen 1851-1957. Mohr.

Great Britain Crafts, Nicholas F. R. 2004. “Regional GDP in Britain, 1871-1911: Some Estimates.” Working Paper No. 03/04. Department of Economic History. London School of Economics.

Italy Felice, Emanuele. 2009. ”Regional value added in Italy (1891-2001): estimates, elaborations”. Working paper.

`http://e-archivo.uc3m.es/bitstream/10016/5332/5/wp-09-08.pdf`

Felice, Emanuele. 2011. “Regional value added in Italy, 1891-2001, and the foundation of a long-term picture.” The Economic History Review 64 (3): 929-950.

Spain Martinez Galarraga, Julio. 2007. “New Estimates of Regional GDP in Spain, 1860-1930.” Documents de treball (Facultat d’Economia i Empresa. Espai de Recerca en Economia), 2007, E07/177. Tirado, Daniel A., Joan R. Roses and Julio Martinez-Galarraga. 2010. “The Long-Term Patterns of Regional Income Inequality in Spain (1860-2000).” Working Papers in Economic History, Univ. Carlos III, no. 10-08.

`http://www.uc3m.es/uc3m/dpto/HISEC/working_papers/working_papers_general.html`

Sweden Enflo, Kerstin, Martin Henning and Lennart Schon. 2010. ”Swedish regional GDP 1855-2000. Estimations and general trends in the Swedish regional system.” Working Papers in Economic History, Univ Carlos III, no. 10-03.

`http://www.uc3m.es/uc3m/dpto/HISEC/working_papers/working_papers_general.html`

Henning, Martin, Kerstin Enflo and Fredrik NG Andersson. 2010. “Trends and cycles in regional economic growth. How spatial differences formed the Swedish growth experience 1860-2009.” Working Papers in Economic History, Univ Carlos III, no. WP 10-10.

`http://www.uc3m.es/uc3m/dpto/HISEC/working_papers/working_papers_general.html`

Appendix C- Coal

In addition (or independently) of the technological and institutional factors examined so far, easy access to cheap and abundant energy sources (coal) has been seen as a crucial condition behind the growth of Europe and, particularly, the location of the industrial revolution (Landes 1965, Pomeranz 2002). To test this hypothesis, following Findlay and O'Rourke (2014) we have digitized the "Les Houillères Européennes" map from Chatel and Dollfus (1931), which records the location of 124 major 19th-century coal fields in Europe, and measure the distance for each of our units - either arbitrary grid squares or regions - to the closest coal field.

To explore the relationship between access to coal and urban development in 1800 we estimate the following model:

$$\mu_{i,t} = \alpha + \beta_1 \mu_{i,t-1} + \beta_2 D_i + \beta_3 D_i \times \mu_{i,t-1} + \epsilon_i \tag{1}$$

where $\mu_{i,t}$ is the logged urban population of quadrant i in the year 1800, $\mu_{i,t-1}$ is logged urban population in the same quadrant in an earlier period - either 1500 or 1700, D_i is the logged distance of quadrant i to the nearest coal field, and ϵ_i is a mean zero random disturbance. The parameter β_1 tells us the direct effect of past urban density, β_2 the direct effect of distance to coal, and β_3 the effect of urban density as it varies by distance to coal. The marginal effect of distance to coal is given by $\beta_2 + \beta_3 \mu_{i,t-1}$.

Estimates of these effects are given in the top panel of Table C1. Both the effect of past urban density and distance to coal are significant and in the expected direction in models where we exclude the interaction term (Columns 1 and 3). The effect of urban population is substantively larger than that associated with distance to coal. According to model 1, for example, a 100% change in distance to coal reduced urban population in 1800 by 23% whereas the effect of a 100% change in urban population in 1500 is associated with 53% change. Once we interact past values of urban population with the distance to the nearest coal field (Columns 2 and 4 of Table 1), we find, however, that the negative effect of distance to coal declines in past levels of urban agglomeration.

Table C2 examine both urban density in 1800 and distance to coal on regional per capita incomes

The Relationship Between Coal and Urban Density, and Incomes

	1.	2.	3.	4. .
	Outcome: $\log(UrbanDensity_{1800})$			
$\log(UrbanDensity_{1500})$	0.53***	0.33***		
	(0.04)	(0.06)		
$\log(UrbanDensity_{1700})$			0.63***	0.43***
			(0.05)	(0.06)
Distance to Coal	−0.23***	−0.13***	−0.18***	−0.09***
	(0.02)	(0.02)	(0.02)	(0.01)
Distance to Coal × $\log(UrbanDensity_{1500})$		0.02***		
		(0.00)		
Distance to Coal × $\log(UrbanDensity_{1700})$				0.02***
				(0.00)
Constant	5.43***	4.31***	4.29***	3.31***
N	444	444	444	444
	0.65	0.65	0.71	0.71

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 1: This table provides estimates of the effects of coal access and historical urban population on urban population in 1800. The unit of observation is the 225 km × 225 km grid-square. Heteroskedasticty robust standard errors in parentheses.

in 1870 and 1900, respectively. Corroborating the results from Table 2 in the main text, a 100% change in urban density is associated with an increase in per capita income of between \$346 and \$442 in 1870. The relationship between a 100% change in distance to coal mines and income is 1870 is between -\$297 and -\$407. However, once we include country fixed effects, the relationship between coal and incomes becomes statistically indistinguishable from zero. Moreover, when we interact urban density in 1800 with distance to 19th century coal mines, the effect of distance to coal is not statistically significant in any model with control variables.

Table C3 reproduces these results now treating income in 2008 as the outcome. Instead of a negative relationship between distance to coal we find the opposite: the farther from 19th-century coal mines a NUTS 2 region was, the wealthier it is predicted to be in 2008. However, after including country fixed effects this relationship becomes statistically null. In the lower panel of Table C3 we reproduce this result, subsetting our data just to Western Europe.

The Relationship Between Coal and Incomes

	Outcome: GDP Per Capita 1870							
$\log(UrbanDensity_{1800})$	346.94**	442.20***	350.65***	369.47***	660.00***	416.33***	430.30***	389.44***
	(111.94)	(37.14)	(60.27)	(49.49)	(121.64)	(71.05)	(81.90)	(67.95)
Distance to Coal	-406.55*	-296.95*	-94.10	7.19	-1937.71***	-148.56	-552.42	-113.32
	(186.82)	(94.61)	(140.40)	(102.17)	(360.08)	(380.02)	(456.40)	(381.81)
Distance to Coal × $\log(UrbanDensity_{1800})$					-298.82**	28.52	-89.12	-23.13
					(73.08)	(74.22)	(70.58)	(65.59)
Constant	4006.70***	-16.94	3585.00***	27.81	5564.33***	-257.40	3975.43***	181.29
	(624.81)	(633.76)	(283.92)	(393.12)	(491.31)	(1057.09)	(469.40)	(798.32)
N	104	104	104	104	104	104	1104	104
R^2	0.43	0.78	0.80	0.85	0.53	0.78	0.80	0.85
Controls	N	Y	N	Y	N	Y	N	Y
Country Effects	N	N	Y	Y	N	N	Y	Y

	Outcome: GDP Per Capita 1900							
$\log(UrbanDensity_{1800})$	508.85**	630.61***	533.55***	585.60***	995.70***	695.07***	651.97***	634.59***
	(157.83)	(55.97)	(61.53)	(50.65)	(131.36)	(90.07)	(117.30)	(61.71)
Distance to Coal	-489.37	-312.66*	-112.90	-40.28	-2874.70***	-660.18	-779.37	-319.29
	(299.18)	(112.49)	(204.42)	(157.64)	(485.52)	(403.06)	(676.60)	(520.07)
Distance to Coal × $\log(UrbanDensity_{1800})$					-465.54**	-66.63	-129.97	-53.68
					(107.20)	(76.53)	(114.73)	(89.18)
Constant	5553.75***	286.83	5083.33***	124.62	7975.29***	862.33	5664.10***	484.20
	(871.14)	(504.29)	(318.83)	(825.50)	(487.68)	(903.21)	(643.68)	(1028.51)
N	106	106	106	106	106	106	106	106
R^2	0.45	0.80	0.79	0.85	0.58	0.80	0.80	0.85
Controls	N	Y	N	Y	N	Y	N	Y
Country Effects	N	N	Y	Y	N	N	Y	Y

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2: This table provide estimates of the effect of urban density in 1800 and coal extraction in the nineteenth century at the regional level on incomes in the nineteenth century. The unit of analysis is the NUTS2 region. Heteroskedasticty robust standard errors clustered by country in parentheses. Controls are: terrain ruggedness, agricultural suitability, distance to coast, coast length, latitude, and longitude.

The Relationship Between Coal and Urban Density, and Incomes

	1.	2.	3.	4.	5.	6.	7.	8.
	Outcome: GDP Per Capita 2008							
$\log(UrbanDensity_{1800})$	2570.98*	3920.06***	3067.96***	3425.86***	2465.47*	3873.22***	3078.98***	3443.17***
	(969.72)	(364.72)	(476.44)	(346.69)	(979.28)	(377.77)	(482.96)	(358.06)
Distance to Coal	-1250.76*	-996.76**	-270.99	-491.67*	-4821.27**	-2426.56	121.35	-19.40
	(497.35)	(304.34)	(163.63)	(215.79)	(1328.49)	(1409.37)	(412.55)	(542.00)
Distance to Coal ×				1037.16**	412.72	-112.53	-136.76	
$\log(UrbanDensity_{1800})$					(322.81)	(351.13)	(106.42)	(147.49)
Constant	15748.66**	-45949.78***	6786.29***	839.57	16121.05**	-45825.76***	6774.38***	1235.56
	(4553.72)	(9174.20)	(514.61)	(26927.86)	(4590.82)	(9157.66)	(521.65)	(27196.10)
N	254	254	254	254	254	254	254	254
R^2	0.13	0.50	0.76	0.79	0.14	0.50	0.76	0.79
Controls	N	Y	N	Y	N	Y	N	Y
Country Effects	N	N	Y	Y	N	N	Y	Y

	Outcome: GDP Per Capita 2008 Western Europe							
$\log(UrbanDensity_{1800})$	1842.17	2703.46**	3041.57**	3590.76***	1828.47	2695.39**	3039.39**	3586.56***
	(868.14)	(536.88)	(616.73)	(446.50)	(872.90)	(538.29)	(624.79)	(448.63)
Distance to Coal	-284.65	-324.11	-288.28	-497.04	-4875.25	-3477.95	-846.73	-2122.17
	(205.00)	(320.83)	(276.74)	(359.76)	(5625.10)	(6661.73)	(6211.82)	(5180.60)
Distance to Coal ×					1044.76	717.51	126.94	368.13
$\log(UrbanDensity_{1800})$					(1236.99)	(1437.10)	(1349.14)	(1101.46)
Constant	21053.49***	-27422.08	32215.18***	-11586.39	21103.85***	-27331.55	32220.21***	-11492.00
	(3994.48)	(12887.80)	(1420.34)	(19528.59)	(4019.20)	(12887.49)	(1438.90)	(19710.91)
N	163	163	163	163	163	163	163	163
R^2	0.10	0.31	0.36	0.46	0.10	0.31	0.36	0.46
Controls	N	Y	N	Y	N	Y	N	Y
Country Effects	N	N	Y	Y	N	N	Y	Y

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3: The top panel provides estimates of the effect of urban density in 1800 and coal extraction in the nineteenth century at the regional level on incomes in 2008. The unit of analysis is the NUTS2 region. To make the results completely comparable to those in the main text, the lower panel excludes countries for which we do not have data from the 19th century incomes. Heteroskedasticty robust standard errors clustered by country in parentheses. Controls are: terrain ruggedness, agricultural suitability, distance to coast, coast length, latitude, and longitude.