INSTITUTE for REPLICATION

No. 82
I4R DISCUSSION PAPER SERIES

The Origin of the State: Land Productivity or Appropriability?: Replication

Nikolai Cook

Thibaut Duprey

Anthony Heyes

Martino Pelli

December 2023

I4R DISCUSSION PAPER SERIES

I4R DP No. 82

The Origin of State: Land Productivity or Appropriability?: Replication

Nikolai Cook[1], Thibaut Duprey[2], Anthony Heyes[3], Martino Pelli[4]

[1]Wilfrid Laurier University, Waterloo/Canada
[2]Bank of Canada, Ottawa/Canada
[3]University of Birmingham/Great Britain
[4]Université de Sherbrooke/Canada

DECEMBER 2023

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

ISSN: 2752-1931

The Origin of the State: Land Productivity or Appropriability?: Replication*

Nikolai Cook
Wilfrid Laurier University

Thibaut Duprey
Bank of Canada

Anthony Heyes
University of Birmingham

Martino Pelli
Université de Sherbrooke

Updated: December 2, 2023

1 Introduction

This is a replication of Mayshar et al. (2022) (henceforth MMP).[1] The article posits that the state (defined as societal hierarchy such as tax-levying elites) originated from cultivation of appropriable cereal grains, contrary to the conventional theory that the state originated from increased land productivity following the adoption of agriculture. The article uses multiple datasets to demonstrate a causal effect of cereal cultivation on hierarchy (Claim 1) without finding a similar effect for land productivity (Claim 2), and that societies based on roots or tubers display levels of hierarchy similar to nonfarming societies (Claim 3).

The results of our replication in brief are:

1. The data and code provided by MMP closely reproduce the main results presented in their Table 1 (see our Table 1).

2. Concurrently testing the cereal cultivation and land productivity claims leads to slightly less statistical significance, on average, than the published article (Table 2).

*This replication is part of the 2023-05-05 Ottawa Replication Games held by the Institute for Replication and hosted by the University of Ottawa. The choice of article was made by the replicators from a list of studies given to them by the Institute. We thank other participants of the Games for helpful conversations. Errors are ours. Does not represent the views of the Bank of Canada.

[1]Published by the Journal of Political Economy in April 2022, and at the time of writing has an Altmetric Score of 571, placing it in the 99th percentile for research attention compared to all research from the JPE.

3. Removing the inherited 1-5 scale of the dependent variable (hierarchy) finds that cereal production is not as effective at moving across all levels of hierarchy compared to the more general claim (Table 3 and 4).

4. Using the same procedures with an aim to confirm the conventional hypothesis (land productivity leads to increased hierarchy conditional on cereal growth) offers statistically significant evidence both for and against Claims 1 and 2 and against Claim 3 (Table 6).

5. The statistical significance of Claim 1 is sensitive to the removal of the top 3% of observations outliers (Table 7).

6. Correction of mis-classified 'none or none specified' crop societies alters the interpretation of coefficients behind Claim 3. Societies that rely more on agriculture among farming societies experience more complex hierarchies, irrespective of being primarily cereal producing or tubers growing (Table 8 and 9).

7. Correction of sample restrictions (which keeps only areas where cereals are the dominant crops) results in less statistically significant evidence for Claim 1 (Table 10).

8. Combining the corrections leads to further reduced statistical evidence for Claim 1, and presents statistically significant evidence against Claim 3 (Table 11).

9. Adding new controls for the type of climate experienced by each society reduces statistical evidence for Claim 1 and points to some evidence against Claim 2 (Table 12).

The remainder of this replication is structured as follows. Section 2 discusses the computational reproducibility of our replication. Section 3 details the scope of the replication and our results. Section 4 concludes.

2 Computational Reproducibility

We first confirm the computational reproducibility[2] of MMP. In other words, using the provided Supplemental Material almost reproduces the estimates of the published article

[2]Following the definitions provided by the Institute for Replication, where **Computational Reproducibility**: The ability to duplicate the results of a prior study using the same data and procedures as were used by the original investigator. Reproducibility is done using the same computer code (possibly rebuilt from scratch), but can be achieved using a different software package. **Robustness Replicability**: The ability to duplicate the results of a prior study using the same data but different procedures as were used by the original investigator. Robustness replicability can be done using the raw, intermediate or final data sets used by the original authors. **Direct Replicability**: The ability to duplicate the results of a prior study using new data but the same procedures as were used by the original investigator.

(see Table 1). Any discrepancies we have found are presented in red text; these discrepancies are small and do not change the qualitative interpretation of the results.

3 Scope of Robustness Replicability

The remainder of this replication is dedicated to examining the robustness replicability of Table 1 of MMP. In other words, answering the question "Does using the same data provided by the authors but reasonable alternative procedures produce the same (or at least similar) estimates?"

MMP section II contains their empirical analysis. Section IIA contains data descriptions. MMP Section IIB uses Murdock's (1967) *Ethnographic Atlas* for a measure of hierarchical complexity (the 1-5 integer scaled variable "Jurisdictional Hierarchy Beyond Local Community").[3] MMP use ordinary least squares (OLS) to regress that main outcome variable against a dummy that indicates societies that use grains as their main crop. As choice of crop might depend on hierarchy, MMP employ two-stage least squares (2SLS) and use the caloric advantage of cereals over tubers in a society as an instrumental variable (calculated using land-suitability data provided by the Food and Agriculture Organization). MMP claim (page 1096):

1. a "considerable positive effect" of cultivating cereals on hierarchical complexity
2. no positive effect of land productivity on hierarchy
3. societies based on roots or tubers display levels of hierarchy similar to nonfarming societies

We consider these to be MMP's main results,[4] and focus our investigation exclusively on investigating the robustness of these results as presented in section IIB. The remaining subsections of MMP consider the same research questions using different datasets to find similar results.[5]

Conceptual Replicability: The ability to duplicate the results of a prior study using new data and different procedures as were used by the original investigator.

[3]The *Ethnographic Atlas* provides cultural, institutional, and economic features on more than 1,200 precolonial societies around the world. These societies are not a random sample, but a cross-section biased towards relatively isolated societies and mainly pertain to the eighteenth and nineteenth centuries.

[4]Although Claim 3 is arguably less central to their appropriability theory than Claims 1 and 2.

[5]MMP Section IIC uses Borcan, Olsson, and Putterman (2018) for a measure of hierarchical complexity, using borders of 159 countries and institutional information every half-century. This analysis offers reduced-form only, as data on the prevalent crop for countries is unavailable. MMP Section IID uses data from various sources on the location of ancient cities and archaeological sites. MMP Section IIE uses data from the *Archaeological Atlas of the World* (Whitehouse and Whitehouse 1975).

3.1 Econometric Model

MMP regress the following equation:

$$Y_i = \alpha_1 CerMain_i + \alpha_2 LandProd_i + X_i^{'}\beta + \mu_i \tag{1}$$

"where Y_i is a measure of hierarchy in society i, $CerMain_i$ is a dummy variable that identifies societies that rely mainly on cereals for their subsistence; $LandProd_i$ is a measure of land productivity, and $X_i^{'}$ is a vector of control variables."

For their 2SLS estimates, the first stage is

$$CerMain_i = \beta_1 CerAdv_i + \beta_2 LandProd_i + X_i^{'}\beta + \varepsilon_i \tag{2}$$

where definitions are the same as above and $CerAdv_i$ is "the difference between the maximum potential caloric yield of cereals and that of roots or tubers under a rain-fed subsistence agriculture."

3.2 Robustness Results

Table 1 is the result of our computational reproduction of MMP. The results we estimate using the same data and code are almost the same as the published version. Any discrepancies we have found are presented in red text; these discrepancies are small and do not change the qualitative interpretation of the results. Column 1 presents the OLS estimate[6] of Claim 1 (the coefficient of $CerMain$ is positive and statistically significant, reflecting the fact that cultivating cereals is associated with an increase of 0.70 levels in the 1-5 integer social hierarchy measure). The remaining columns are 2SLS estimates which apply Equation 2 as the first stage. Column 2 presents the 2SLS estimate of Claim 1 (cultivating cereals is associated with an increase of hierarchy of more than 1 level - equivalent to "moving from a tribe to a small chiefdom, or from a large chiefdom to a state.") Column 3 adds continent fixed effects. Column 4 introduces an estimate for Claim 2 by adding land productivity as a control variable. That the coefficient on $LandProd$ (the standardized value of the maximum caloric yield possible per hectare of land for the society) is not statistically significant and small in magnitude is interpreted (conditional on cultivating cereals) as land productivity has no significant effect on hierarchy. Column 5 introduces an estimate for Claim 3 by adding a control for dependence on agriculture. That the coefficient on $Dep.Agri.$ is not statistically significant is interpreted that "societies that practice agriculture are not characterized by more complex hierarchies unless they cultivate cereals." Column 6 uses a post-double-selection methodology (which uses a square-root lasso) to identify relevant controls. Panel B displays first-stage estimates,

[6] Applying Equation 1.

where an increase of 1 standard deviation in the productivity advantage of cereals over tubers increases the probability of growing cereals as a main crop by about 20%.

In Table 2, we present estimates which always include $CerMain$ and $LandProd$ in order to better examine Claim 1 and Claim 2. In MMP Table 1 (our Table 1), only column 4 directly includes both variables required to answer the research question posed in MMP "is it cereal cultivation and not land productivity that leads to social hierarchy?". Table 2 column 1 presents OLS estimates. Consistent with MMP's hypothesis, the coefficient of $CerMain$ is positive and statistically significant ($p < 0.000$) while the coefficient of $LandProd$ is 18% the magnitude and not statistically significant ($p = 0.226$). Column 2 presents 2SLS estimates. The coefficient of $CerMain$ is positive but now statistically insignificant ($p = 0.215$) and the coefficient of $LandProd$ is 9% the magnitude and not statistically significant ($p = 0.658$). In column 3 (which introduces continent fixed effects and coincides exactly with Table 1 column 4), the coefficient of $CerMain$ returns to statistical significance. Claim 3 is tested in column 4, where the coefficient of $Dep.Agri.$ remains statistically insignificant ($p = 0.810$). In this table, we also present F-statistics for the first stage; we find the instrument is considerably weaker when Claim 1 and Claim 2 are tested concurrently - a necessity we believe in order to show positive evidence for Claim 1 and insignificant evidence for Claim 2.

In Table 3, we present estimates which remove a point of concern: Murdock's (1967) *Ethnographic Atlas* provides a measure of hierarchical complexity "Jurisdictional Hierarchy Beyond Local Community" as an integer between 1 and 5 that both Equation 1 and Equation 2 apply without alteration. While we remain silent about the interpretability of the regression coefficients when using this scale, the fact that the dependent variable is arguably arbitrarily scaled and entered into a linear regression may be cause for concern. An alternative would be to create an indicator that takes the value zero for tribal societies and one for societies that are more complex (chiefdoms and states). The resulting coefficient on $CerMain$ could then be interpreted as the effect of primarily cereal cultivation on moving from less hierarchical to more hierarchical societies. Column 1's dependent variable is 0 for tribal societies (the lowest hierarchy in the *Atlas*) and 1 for all others. Column 2's dependent variable is 0 for tribal and small chiefdom societies and 1 for all others (and so on). Columns 1-4 use OLS and columns 5-8 use 2SLS. In columns 6 and 7, we find that cereal cultivation has a statistically significant effect only when organizing beyond a small or large chiefdom.

In Table 4, we present estimates similar to those in Table 3 but each column now compares only two hierarchy levels. For example, column 1 focuses on the sample of tribal and small chiefdoms only. Column 2 presents estimates on the sample of small chiefdoms and large chiefdoms only. By estimating on these subsamples, we note a reduction in statistical power as compared to those in Table 3 and Table 4. We note that, with the 2SLS estimation in columns 5 to 8, the coefficient of $CerMain$ is positive but far from statistical significance at traditional levels. In column 5 that compares small chiefdoms to tribal societies, the coefficient on $LanProd$ becomes significant. As in Table 3, OLS regressions suggest that the dependence on agriculture is often associated with more hierarchy (inconsistent with Claim 3), but when accounting for possible endogeneity with 2SLS, Claim 3 holds.

In Table 5, our structure is exactly the same as in Table 1 however we have re-coded the dependent variable from a 1-5 integer scale into standardized values. Reassuringly, while the coefficients on $CerMain$ change to reflect the interpretation difference (movements along a standardized scale rather than 'levels' of hierarchy) the statistical significance does not change.

In Table 6, we present estimates of what we believe is an equally reasonable exercise as that presented in Table 1 to test Claim 1, 2, and 3. While the dependent variable remains the same 1-5 integer valued hierarchical measure, the primary independent variable is now $LandProd$ (instead of $CerMain$). In this manner, the analysis would be designed to support the conventional theory (that land productivity led to the origination of the state). Column 1 presents OLS estimates - a one standard deviation increase in land productivity is associated with an increase in hierarchy of 0.239 levels. In column 2, which presents 2SLS estimates, we note the instrument's first stage is much more powerful than in Table 1, with an F-statistic 6 times that of the largest in Table 1. Column 4 includes both land productivity and cereal cultivation, to find that cereal cultivation has a positive and statistically significant effect on hierarchy (consistent with Claim 1) while land productivity, conditional on cereal cultivation, does not have a statistically significant effect on hierarchy (consistent with Claim 2). However, column 6 is inconsistent with both claims - once additional controls are selected by MMP's post-double-selection methodology (which uses a square-root lasso to identify relevant controls), it is land productivity not cereal cultivation which is associated with greater hierarchical organization in society.

In Table 7, we present estimates following winsorization of the instrumental variable—how much better cereals are than tubers in a society. We present estimates where the instrument is winsorized from above at 3% (winsorization at the 4% and 5% level offer even starker results). After winsorization, what we would consider the main estimate of MMP (column 4, coefficient of $CerMain$) is not statistically significant ($p = 0.191$) and half of the original published estimate. In fact, even column 3 (which introduces continent fixed effects and is not conditional on $LandProd$) presents a statistically insignificant coefficient on $CerMain$ ($p = 0.130$).

In Table 8 and 9, we present evidence primarily complementing Claim 3. MMP use Murdock's primary crop to identify cereal and tuber growing societies, while non-farming societies are those with "None or none specified" major crop. Non-farming societies are thus included in their control (non-cereal) group. MMP find that societies that practice agriculture has no statistically significant effect, unless they cultivate cereals, and conclude that the main difference is between cereals or not, instead of farming or not (claim 3). In Table 8 we remove the non-farming observations. In Table 9 we keep only those societies that use either cereals or tubers as their main crop (following the narrative in MMP). The results in both tables are similar. We find that in column 3, the coefficient on $CerMain$ is no longer statistically significant ($p = 0.311$ and $p = 0.379$). In column 5, $CerMain$ is again statistically insignificant ($p = 0.332$ and $p = 0.408$) and, most importantly, the coefficient on $Dep.Agri.$ is large and statistically significant ($p = 0.027$ and $p = 0.010$). We conclude that societies that rely more on agriculture among those farming societies do experience more complex hierarchies, irrespective of being primarily cereal producing or tubers growing.

In Table 10, we remove a sample restriction present throughout MMP's analysis. Table 1 and others report estimates from a sample of societies that could grow both tubers and cereals. However, whenever tubers and cereals are both able to be grown, cereals are *always* the dominant crop (in terms of expected caloric yield per hectare). After removing the restriction (thereby adding societies that could only grow tubers to the 'control group') the statistical significance of $CerMain$ becomes less robust and loses statistical significance in column 3 (barely as $p = 0.103$) and in columns 5 and 6 ($p = 0.385$ and $p = 0.406$, respectively).

In Table 11, we present a table that is identical in structure to Table 1, but incorporates a collection of reasonable alternative research decisions. First, there is no restriction that societies must be able to grow both tubers and cereals in order to be included in the regressions. Second, we do not include societies that are marked as having 'none or none specified' when it comes to their major crop. Third, we replace the integer 1-5 scale of the dependent variable with standard deviations (as employed elsewhere by MMP). We note that statistical significance for Claim 1 is retained only in columns 1 and 4. As in Tables 8 and 9, column 5 points to the relevance of agriculture intensity (cereal or tubers) for farming societies instead of cereal cultivation, altering the scope of Claim 3.

In Table 12, we present a table that is also similar in structure to Table 1, but add climate fixed effects. We used the latitude and longitude of the original MMP dataset that we mapped into the raster cells of climate classification using the R code of Kottek et al. (2006). We obtained 26 out of 31 possible climate classifications. To avoid adding too many new controls, we introduced indicators for the 5 main climate groups (equatorial, arid, warm, snow, polar). The polar group is dropped while the dummy variables for the other groups are usually significant. When controlling for climate fixed effects without land productivity or the dependence on agriculture in columns 1 to 3, the cereal advantage of Claim 1 remains with similar magnitudes and significance levels. When controlling for climate fixed effects together with land productivity or the dependence on agriculture in columns 4 and 5, the cereal advantage of Claim 1 is not significant. Adding the climate fixed effects to the post-double-selection methodology does not substantially alter the list of selected controls, and thus the main results of MMP remain. If instead the post-double-selection methodology is performed while always keeping the climate fixed effects in column 7, then Claims 1 and 2 are no longer supported: now it is land productivity instead of cereal cultivation that is associated with greater hierarchical organization in society. We note that compared to the MMP results replicated in Table 1, adding the climate fixed effects makes the instrument weaker in columns 4, 6 and 7 but stronger in columns 2 and 5.

4 Conclusion

The computational reproduction of Mayshar et al. (2022) confirms that their data and code closely reproduce the main results presented in their Table 1. Our robustness replication with a range of plausible alternative specifications points to additional nuance and sometimes eroded support for the main results.

Tables

Table 1: Cereals and Hierarchy: OLS and 2SLS Computational Reproduction: Same Data and Code as MMP

	DV: Jurisdictional hierarchy beyond local community (1-5 scale)					
	OLS	2SLS	2SLS	2SLS	2SLS	2SLS PDS
	(1)	(2)	(3)	(4)	(5)	(6)
			A. Second stage			
CerMain	0.707***	1.170***	0.892**	1.064*	0.830	0.797**
	(0.114)	(0.352)	(0.447)	(0.447)	(0.554)	(0.378)
	[0.000]	[0.001]	[0.046]	[0.056]	[0.134]	[0.035]
LandProd				-0.037		0.020
				(0.086)		(0.065)
				[0.670]		[0.765]
DepAgri					0.259	
					(0.544)	
					[0.635]	
Continent FE			Y	Y	Y	...
Obs.	952	952	952	952	952	877
F-statistic		52.15	33.13	13.06	20.38	16.10
R2	0.113					
			B. First stage			
CerAdv		0.209***	0.155***	0.258***	0.130***	0.256***
		(0.029)	(0.027)	(0.071)	(0.029)	(0.063)
		[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

Reproduction of MMP Table 1 using the same data and code. Red text identifies discrepancies between what the replication package produces and the published version. In Panel A, column 6, land productivity is included in the regression and is not reported in the published table (see their footnote 28). Panel A, column 6 has an associated F-statistic of 16.10 (16.11 is reported). Panel B, the column 5 coefficient is missing its decimal place in the published version. The same has a standard error of 0.029 instead of the published version's reported 0.068. The dependent variable is a 1-5 integer scaled variable with 1 representing a tribal society and 5 representing a large state. *CerMain* is an indicator variable that cereals are the primary crop for the society. *LandProd* is a measure of the maximum caloric yield per hectare (either cereals or tubers). *DepAgri* is the average percent of calories derived from agriculture. Each observation is a society in Murdock's (1967) *Ethnoatlas*. Standard errors, clustered at the country level, in parentheses. *p*-values in square brackets. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table 2: Examining Claim 1 and Claim 2 concurrently

	DV: Hierarchy (1-5 scale)			
	(1)	(2)	(3)	(4)
	OLS	2SLS	2SLS	2SLS
CerMain	0.646***	0.863	1.064*	1.082**
	(0.000)	(0.215)	(0.056)	(0.031)
LandProd	0.121	0.081	-0.037	-0.046
	(0.226)	(0.658)	(0.670)	(0.609)
DepAgri				0.109
				(0.810)
Continent FE			Y	Y
Obs.	952	952	952	952
F-statistic		6.02	13.06	18.11

The dependent variable is a 1-5 integer scaled variable with 1 representing a tribal society and 5 representing a large state. *CerMain* is an indicator variable that cereals are the primary crop for the society. *LandProd* is a measure of the maximum caloric yield per hectare (either cereals or tubers). *DepAgri* is the average percent of calories derived from agriculture. Each observation is a society in Murdock's (1967) *Ethnoatlas*. p-values representing standard errors clustered at the country level in parentheses. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table 3: Removing the arbitrary integer scale of dependent variable

	DV: Indicator for above hierarchy level...							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Tribe	S. Chief.	L. Chief	S. State	Tribe	S. Chief.	L. Chief	S. State
CerMain	0.134***	0.016	0.019	0.002	0.251	0.413**	0.287*	0.131
	(0.002)	(0.716)	(0.506)	(0.880)	(0.108)	(0.037)	(0.059)	(0.124)
LandProd	0.052	-0.020	-0.011	-0.004	0.044	-0.048	-0.030*	-0.013
	(0.335)	(0.524)	(0.487)	(0.573)	(0.377)	(0.148)	(0.093)	(0.119)
DepAgri	0.170*	0.289***	0.188***	0.088***	0.090	0.016	0.004	-0.001
	(0.050)	(0.000)	(0.002)	(0.007)	(0.558)	(0.924)	(0.975)	(0.985)
Continent FE	Y	Y	Y	Y	Y	Y	Y	Y
Obs.	952	952	952	952	952	952	952	952
F-statistic					18.11	18.11	18.11	18.11

Each column has a differently coded indicator dependent variable. Column 1's dependent variable is 0 for tribal societies and 1 for all others. Column 2's dependent variable is 0 for tribal and small chiefdom societies and 1 for all others (and so on). $CerMain$ is an indicator variable that cereals are the primary crop for the society. *LandProd* is a measure of the maximum caloric yield per hectare (either cereals or tubers). *DepAgri* is the average percent of calories derived from agriculture. Each observation is a society in Murdock's (1967) *Ethnoatlas*. p-values representing standard errors clustered at the country level in parentheses. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table 4: Removing the arbitrary integer scale of dependent variable (repeated coding)

	DV: Indicator for one above level...							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Tribe	S. Chief.	L. Chief	S. State	Tribe	S. Chief.	L. Chief	S. State
CerMain	0.156***	-0.057	0.056	0.006	0.064	0.296	0.270	1.213
	(0.001)	(0.351)	(0.458)	(0.955)	(0.718)	(0.122)	(0.307)	(0.586)
LandProd	0.075	-0.039	-0.012	-0.019	0.085*	-0.066*	-0.006	-0.063
	(0.165)	(0.311)	(0.841)	(0.645)	(0.064)	(0.098)	(0.918)	(0.529)
DepAgri	-0.016	0.314**	0.283	0.526**	0.055	0.178	0.300	0.847
	(0.859)	(0.013)	(0.168)	(0.013)	(0.737)	(0.213)	(0.156)	(0.221)
Continent FE	Y	Y	Y	Y	Y	Y	Y	Y
Obs.	725	422	204	93	725	422	204	93
F-statistic					20.27	16.49	14.44	0.47

Each column compares adjacent levels of the integer scaled 1-5 dependent variable. Column 1's dependent variable is 0 for tribal societies and 1 for small chiefdoms. Column 2's dependent variable is 0 for small chiefdoms and 1 for large chiefdoms. $CerMain$ is an indicator variable that cereals are the primary crop for the society. *LandProd* is a measure of the maximum caloric yield per hectare (either cereals or tubers). *DepAgri* is the average percent of calories derived from agriculture. Each observation is a society in Murdock's (1967) *Ethnoatlas*. p-values representing standard errors clustered at the country level in parentheses. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table 5: Removing the arbitrary integer scale of dependent variable (standardization)

DV: Jurisdictional hierarchy beyond local community (standardized)						
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	2SLS	2SLS	2SLS	2SLS	2SLS PDS
CerMain	0.678***	1.121***	0.854**	1.019*	0.795	0.764**
	(0.000)	(0.001)	(0.046)	(0.056)	(0.134)	(0.035)
LandProd				-0.035		0.019
				(0.670)		(0.765)
DepAgri					0.248	
					(0.635)	
Continent FE			Y	Y	Y	...
Obs.	952	952	952	952	952	877
F-statistic		52.15	33.13	13.06	20.38	16.10

The dependent variable is a 1-5 integer scaled variable that is standardized. *CerMain* is an indicator variable that cereals are the primary crop for the society. *LandProd* is a measure of the maximum caloric yield per hectare (either cereals or tubers). *DepAgri* is the average percent of calories derived from agriculture. Each observation is a society in Murdock's (1967) *Ethnoatlas*. p-values representing standard errors clustered at the country level in parentheses. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table 6: Land productivity and hierarchy

DV: Jurisdictional hierarchy beyond local community (1-5 scale)						
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	2SLS	2SLS	2SLS	2SLS	2SLS PDS
LandProd	0.239**	0.309***	0.191**	0.135	0.152*	0.152*
	(0.023)	(0.000)	(0.023)	(0.103)	(0.062)	(0.066)
CerMain				0.259**		0.125
				(0.012)		(0.187)
DepAgri					0.749***	
					(0.000)	
Continent FE			Y	Y	Y	...
Obs.	952	952	952	952	952	877
F-statistic		307.77	361.75	460.92	356.52	365.91

Columns are the same as in Table 1, with *LandProd* now taken to be the primary independent variable. In columns 2-6, *LandProd* is instrumented by the caloric-advantage cereals have over tubers. The dependent variable is a 1-5 integer scaled variable with 1 representing a tribal society and 5 representing a large state. *CerMain* is an indicator variable that cereals are the primary crop for the society. *LandProd* is a measure of the maximum caloric yield per hectare (either cereals or tubers). *DepAgri* is the average percent of calories derived from agriculture. Each observation is a society in Murdock's (1967) *Ethnoatlas*. p-values representing standard errors clustered at the country level in parentheses. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table 7: Winsorization of instrumental variable

DV: Jurisdictional hierarchy beyond local community (1-5 scale)						
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	2SLS	2SLS	2SLS	2SLS	2SLS PDS
CerMain	0.707***	1.083***	0.691	0.627	0.595	0.368
	(0.000)	(0.002)	(0.109)	(0.147)	(0.243)	(0.291)
LandProd				0.014		0.054
				(0.874)		(0.485)
DepAgri					0.433	
					(0.365)	
Continent FE			Y	Y	Y	···
Obs.	952	952	952	952	952	877
F-statistic		63.67	32.58	14.37	22.48	14.40

Columns are the same as in Table 1. In columns 2-6, *LandProd* is instrumented by the caloric-advantage cereals have over tubers, with the top 3% (31 total) of cereal-advantaged observations winsorized. The dependent variable is a 1-5 integer scaled variable with 1 representing a tribal society and 5 representing a large state. *CerMain* is an indicator variable that cereals are the primary crop for the society. *LandProd* is a measure of the maximum caloric yield per hectare (either cereals or tubers). *DepAgri* is the average percent of calories derived from agriculture. Each observation is a society in Murdock's (1967) *Ethnoatlas*. p-values representing standard errors clustered at the country level in parentheses. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table 8: Removing 'no crop specified'

DV: Jurisdictional hierarchy beyond local community (1-5 scale)						
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	2SLS	2SLS	2SLS	2SLS	2SLS PDS
CerMain	0.478***	1.071**	0.974	1.003*	0.935	0.376
	(0.006)	(0.048)	(0.311)	(0.073)	(0.332)	(0.574)
LandProd				-0.004		-0.027
				(0.967)		(0.715)
DepAgri					0.523**	
					(0.027)	
Continent FE			Y	Y	Y	...
Obs.	764	764	764	764	764	715
F-statistic		11.81	6.03	13.78	6.03	7.49

MMP included societies which Murdock's *Atlas* classified the main crop as 'None or None Specified' as the control group; we remove those here. Columns are the same as in Table 1. The dependent variable is a 1-5 integer scaled variable with 1 representing a tribal society and 5 representing a large state. $CerMain$ is an indicator variable that cereals are the primary crop for the society. *LandProd* is a measure of the maximum caloric yield per hectare (either cereals or tubers). *DepAgri* is the average percent of calories derived from agriculture. Each observation is a society in Murdock's (1967) *Ethnoatlas*. p-values representing standard errors clustered at the country level in parentheses. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table 9: Keeping only tubers and cereals

DV: Jurisdictional hierarchy beyond local community (1-5 scale)						
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	2SLS	2SLS	2SLS	2SLS	2SLS PDS
CerMain	0.535***	1.246*	1.281	1.263*	1.190	.878
	(0.006)	(0.059)	(0.379)	(0.098)	(0.408)	(0.409)
LandProd				0.002		-0.071
				(0.987)		(0.395)
DepAgri					0.683**	
					(0.010)	
Continent FE			Y	Y	Y	...
Obs.	704	704	704	704	704	662
F-statistic		9.18	4.09	10.52	4.12	4.61

MMP included societies which Murdock's *Atlas* classified as "None or none specified", "Non food crops only", "Vegetables", and "Tree fruits" as the roots and tubers control group. We remove those here. Columns are the same as in Table 1. The dependent variable is a 1-5 integer scaled variable with 1 representing a tribal society and 5 representing a large state. $CerMain$ is an indicator variable that cereals are the primary crop for the society. *LandProd* is a measure of the maximum caloric yield per hectare (either cereals or tubers). *DepAgri* is the average percent of calories derived from agriculture. Each observation is a society in Murdock's (1967) *Ethnoatlas*. p-values representing standard errors clustered at the country level in parentheses. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table 10: Full sample

DV: Jurisdictional hierarchy beyond local community (1-5 scale)						
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	2SLS	2SLS	2SLS	2SLS	2SLS PDS
CerMain	0.712***	1.200***	0.839	1.180**	0.722	0.318
	(0.000)	(0.005)	(0.103)	(0.034)	(0.385)	(0.406)
LandProd				-0.049		0.089
				(0.463)		(0.120)
DepAgri					0.238	
					(0.735)	
Continent FE			Y	Y	Y	...
Obs.	1,059	1,059	1,059	1,059	1,059	932
F-statistic		40.64	36.02	13.93	10.67	12.44

MMP restricted all analyses to societies which could (potentially) grow tubers and cereals. However, whenever both types of crops can be grown, cereals are *always* the highest calorie-per-hectare option. In this table, we remove the restriction and therefore include societies that could not grow cereals into the control group. Columns are the same as in Table 1. The dependent variable is a 1-5 integer scaled variable with 1 representing a tribal society and 5 representing a large state. $CerMain$ is an indicator variable that cereals are the primary crop for the society. *LandProd* is a measure of the maximum caloric yield per hectare (either cereals or tubers). *DepAgri* is the average percent of calories derived from agriculture. Each observation is a society in Murdock's (1967) *Ethnoatlas*. p-values representing standard errors clustered at the country level in parentheses. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table 11: Cereals and hierarchy: OLS and 2SLS robustness reproduction: same data, different codes

DV: Jurisdictional hierarchy beyond local community (standardized)						
	OLS	2SLS	2SLS	2SLS	2SLS	2SLS PDS
	(1)	(2)	(3)	(4)	(5)	(6)
CerMain	0.505***	1.608	1.584	1.269*	0.994	.840
	(0.009)	(0.230)	(0.530)	(0.084)	(0.620)	(0.437)
LandProd				0.013		-0.034
				(0.871)		(0.551)
DepAgri					0.609*	
					(0.078)	
Continent FE			Y	Y	Y	···
Obs.	752	752	752	752	752	696
F-statistic		4.49	2.20	10.66	2.74	4.25

Columns are the same as in Table 1. The dependent variable is a 1-5 integer scaled variable, now standardized. MMP restricted all analyses to societies which could (potentially) grow tubers and cereals. However, whenever both types of crops can be grown, cereals are *always* the highest calorie-per-hectare option. In this table, we remove the restriction and therefore include societies that could not grow cereals into the control group. MMP included societies which Murdock's *Atlas* classified as "None or none specified", "Non food crops only", "Vegetables", and "Tree fruits" as the roots and tubers control group. We remove those here. $CerMain$ is an indicator variable that cereals are the primary crop for the society. *LandProd* is a measure of the maximum caloric yield per hectare (either cereals or tubers). *DepAgri* is the average percent of calories derived from agriculture. Each observation is a society in Murdock's (1967) *Ethnoatlas*. p-values representing standard errors clustered at the country level in parentheses. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table 12: Cereals and hierarchy: OLS and 2SLS robustness reproduction: additional data, same codes

	DV: Jurisdictional hierarchy beyond local community (1-5 scale)						
	OLS	2SLS	2SLS	2SLS	2SLS	2SLS PDS	2SLS PDS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CerMain	0.731***	1.243***	0.891**	0.421	0.816	0.855**	-0.039
	(0.000)	(0.000)	(0.038)	(0.580)	(0.138)	(0.046)	(0.955)
LandProd				0.105		0.053	0.154*
				(0.430)		(0.394)	(0.084)
DepAgri					.314		
					(.617)		
Continent FE			Y	Y	Y	...	...
Climate FE	Y	Y	Y	Y	Y	...	Y
Obs.	952	952	952	952	952	877	877
F-statistic		56.29	32.91	6.89	23.82	11.08	7.50

Columns are the same as in Table 1, with the addition of climate fixed effects (equatorial, arid, warm, snow, polar) from Kottek et al. (2006). The dependent variable is a 1-5 integer scaled variable with 1 representing a tribal society and 5 representing a large state. *CerMain* is an indicator variable that cereals are the primary crop for the society. *LandProd* is a measure of the maximum caloric yield per hectare (either cereals or tubers). *DepAgri* is the average percent of calories derived from agriculture. Each observation is a society in Murdock's (1967) *Ethnoatlas*. p-values representing standard errors clustered at the country level in parentheses. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

References

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F. (2006). World map of the köppen-geiger climate classification updated. *Meteorologische Zeitschrift*, 15(3):259–263.

Mayshar, J., Moav, O., and Pascali, L. (2022). The origin of the state: Land productivity or appropriability? *Journal of Political Economy*, 130(4):1091–1144.