http://www.diva-portal.org

This is the published version of a paper published in .

Citation for the original published paper (version of record):

Hammar, O. (2026)
Income and Inequality in the Aztec Empire: A Replication of Alfani and Carballo (Nature Human Behaviour, 2023)
Journal of Comments and Replications in Economics, 5: 2026-14
https://doi.org/10.18718/81781.65

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-150296

Income and Inequality in the Aztec Empire.

A Replication of Alfani and Carballo (*Nature Human Behaviour*, 2023)

Olle Hammar *

Journal of Comments and Replications in Economics, Volume 5, 2026-14, DOI:10.18718/81781.65

JEL: C18, D31, N01, N36
Keywords: Aztec Empire, Economic history, Income inequality, Replication, Reproducibility, Robustness
Data Availability: The Stata code and the data to reproduce the results in this replication can be downloaded at JCRE's data archive (DOI: 10.15456/j1.2026065.1258856488).

Please cite as: Hammar, O. (2026). Income and Inequality in the Aztec Empire. A Replication of Alfani and Carballo (*Nature Human Behaviour*, 2023). *Journal of Comments and Replications in Economics*, Vol. 5 (2026-14). DOI: 10.18718/81781.65

Abstract

Alfani and Carballo (2023) estimate income inequality in the Aztec Empire around 1492. Their preferred estimate gives a Gini index of 50.4. They conclude that inequality was high even before the Spanish conquest. First, I confirm that the main outcomes are computationally reproducible from the analysis data provided in the replication package. Second, I examine two inconsistencies. Third, I test the robustness reproducibility of the relationship between population density and per capita income through eight different robustness tests, confirming a positive and statistically significant relationship but with considerable variation in the point estimates. Finally, I use these estimates as inputs for robustness measures of the main outcomes. The results support the qualitative conclusion that inequality and extraction were substantial before the Spanish conquest. Their exact magnitudes, however, depend on modelling choices and should be interpreted as approximate historical estimates rather than precise measurements.

*Linnaeus University and Stockholm School of Economics. olle.hammar@lnu.se
I thank the editor W. Robert Reed and three anonymous reviewers for comments that substantially improved the paper. I am grateful to the original authors Guido Alfani and Alfonso Carballo for providing additional data and code. I also thank Abel Brodeur and the Institute for Replication for organizing the replication project from which this study originated. All remaining errors are my own.

Received March 6, 2026; Revised August 11, 2026; Accepted August 12, 2026; Published October 21, 2026.
©Author(s) 2026. Licensed under the Creative Commons License - Attribution 4.0 International (CC BY 4.0).

1 Introduction

Alfani and Carballo (2023) estimate the levels of income inequality and imperial extraction in the Aztec Empire around 1492, that is, before the Spanish conquest. Their analysis consists of two steps. First, they estimate the relationship between population density and per capita income in preindustrial societies. Second, they use these regression results as inputs for calculations of their main outcomes, namely the levels of income and inequality in the Aztec Empire. Their main estimate finds per capita income levels of 690 USD (1990 PPP), corresponding to 1.72 times the subsistence minimum. In terms of income inequality, they find that the richest 1% earned 41.8% of total income, the richest 5% earned 50.8% of total income, and the poorest 50% earned 23.3% of total income. Their estimates suggest that the imperial ruling class extracted 4.54% of total income, and that the Gini index was 50.4. They conclude that income inequality in the Aztec Empire was high even before the Spanish conquest, questioning to what extent today's high levels of economic inequality in Mexico can be explained by the Spanish conquest and extractive institutions imposed by the colonizers.

In this replication study, I investigate whether their analytical results are computationally reproducible and further test their replicability and robustness to: (1) including country and year fixed effects, (2) including all observations, (3) using the full Ashraf and Galor (2011) sample, and (4) using population density from Ashraf and Galor (2011). When possible, the robustness analyses are done for both models used in the original paper, that is, ordinary least squares (OLS) and instrumental variables (IV). As such, in total, I conduct eight robustness reproducibility analyses. Finally, I compare these robustness reproduction results to those in the original paper, including their estimated income distribution in the Aztec Empire, and those of modern Mexico and the United States.

Why is a careful reassessment of these estimates important? They matter because they affect how we interpret the history of inequality in Mexico and Latin America. If the Aztec Empire was already highly unequal before the Spanish conquest, then colonial institutions are unlikely to be the whole explanation for later inequality. At the same time, the estimates rest on limited evidence and several strong assumptions, for example, about the relationship between population density and income in preindustrial societies. The purpose of this replication is to assess how much the results change when those assumptions are varied, and to what extent the original findings are sensitive to such changes.

Overall, the qualitative conclusions in the original paper are supported, although the variation across specifications indicates considerable uncertainty about the exact magnitudes.

2 Computational Reproducibility

For the computational reproducibility, I used the original authors' replication package available here: https://www.openicpsr.org/openicpsr/project/186521/version/V2/view.

The replication package included two parts. In the first part, the analysis data and codes for the econometric regression analyses were provided. In this part, the per capita income levels were estimated. These estimates were then used as inputs for the second part. The first part was

conducted in Stata.

In the second part, the income estimates from the first part, together with other input data, were used to estimate the levels of income inequality. These analyses were done in Microsoft Excel.

Neither raw data nor cleaning codes were provided (see Table 1). In the original replication package, analysis code was also missing for the Kolmogorov-Smirnov tests comparing the estimated income distribution of the Aztec Empire to those of modern Mexico and the USA (original paper's Figure 2 and p. 1266); but this was later provided separately by the original authors.

All main outcomes were computationally reproducible from the analysis data. The main implication of this is that the paper was reproducible from analysis data, but not from raw data. I could verify the econometric and spreadsheet calculations once the final analysis files were taken as given, but I could not assess whether choices made in collecting or cleaning the raw inputs affected the results. While this is not central to the robustness exercises below, which use the authors' analysis data as the starting point, it limits the scope of the reproduction.

Table 1: Results of computational reproduction

	Fully	**Partial**	**No**
Raw data provided			x
Cleaning code provided			x
Analysis data provided	x		
Analysis code provided		x	
Reproducible from raw data			x
Reproducible from analysis data	x		

As an alternative to the original authors' Excel calculations, I also computationally reproduced the main Gini index (for the Aztec Empire as a whole) using the Stata command *ineqdeco* (Jenkins 1999). This reproduced the original Gini estimate of 50.4.

In the first version of this replication study (Hammar 2025), I was not able to reproduce the p-values reported for the Kolmogorov-Smirnov tests in the note to the original paper's Figure 2. After the first round of review, the original authors provided the Stata code and data used for these tests. Using these files, I reproduce their reported results: the comparison between the Aztec Empire and Mexico gives $D = 0.55$ and an exact p-value of 0.004, while the comparison between the Aztec Empire and the United States gives $D = 0.65$ and an exact p-value below 0.001.

The discrepancy in my initial reproduction came from the way the income distributions were grouped. The authors' test used ventiles. My initial implementation used the income groups reported in the Excel replication file, namely deciles with the top 5% reported separately. This gives very similar D-statistics, but larger exact p-values because the distributions are represented by fewer groups. I therefore treat the original Kolmogorov-Smirnov results as computationally reproduced. The alternative grouping is retained below only as a robustness check of the distributional comparison (Table 2), not as a failure to reproduce the original test.

Table 2: Computational reproduction results for Kolmogorov-Smirnov tests

Test for equality of income distributions	Reported in original paper (Figure 2)	Reproduction using *ksmirnov* (alternative aggregation)
Aztec Empire 1492 versus Mexico 2018	D = 0.55*** (p-value = 0.004)	D = 0.55* (exact p-value = 0.075)
Aztec Empire 1492 versus USA 2018	D = 0.65*** (p-value = 0.000)	D = 0.64** (exact p-value = 0.021)

Notes: Exact p-values from Stata's *ksmirnov* command. The authors' test uses ventiles. The alternative 11-group aggregation uses deciles with the top 5% reported separately. Significant at the ***[1%] **[5%] *[10%] level.

A coding error was found in the main IV analysis (Model 2), which used the *reg* and *predict* commands in Stata. In this analysis, the restriction of the natural logarithm of population density having to be greater than or equal to zero was only imposed in the first stage and not in the second stage. Moreover, estimating the IV model in two steps, instead of using a single command such as *ivregress*, resulted in different standard errors. The Hausman specification test for choosing between the OLS Model 1 and the two-stage least squares (2SLS) Model 2 (see discussion in the original paper p. 1269), however, used the *ivreg* command (but without robust standard errors). The different results are presented in Table 3 (Columns 1–3).

Table 3: Computational reproduction results for original IV analysis

Model 2 **(IV)**	Original used in main analysis (1)	Original using same samples in both stages (2)	Original used in Hausman test using *ivreg* (3)	Reproduction original corrected using *ivregress* (4)
(Ln) Population density	0.198*** (0.048) [0.000]	0.157*** (0.048) [0.002]	0.157 (0.095) [0.104]	0.157*** (0.054) [0.005]
Constant	5.829*** (0.151) [0.000]	5.807*** (0.234) [0.000]	5.807*** (0.304) [0.000]	5.807*** (0.313) [0.000]
Controls	Yes	Yes	Yes	Yes
Observations	66	57	57	57
R-squared	0.086	0.048	0.275	0.275

Notes: IV regressions using 2SLS. Number of domesticable species of animals used as instrumental variable. The dependent variable is the natural logarithm of income per capita. The main independent variable is the natural logarithm of population density. Control variables include natural logarithm of land productivity, natural logarithm of absolute latitude, mean distance to nearest coast or river, and percentage of land within 100 km of coast or river. Population density from Maddison (2003). Small-sample degrees-of-freedom adjustments, except for (1) and (2). Adjusted R-squared, except for (1) and (2). Robust standard errors are in parentheses for (1), (2) and (4), and non-robust standard errors are in parentheses for (3). P-values are reported in brackets. Significant at the ***[1%] **[5%] *[10%] level.

Table 3 (Column 4) also reports the corresponding results using the *ivregress* command (with robust standard errors and small-sample degrees-of-freedom adjustments), which is what I have used in the remainder of this reproduction. As shown in Table 3, the coefficient estimates were the same for the corrected reproduction (Column 4) as for the one used in the Hausman test in the original paper (Column 3), that is, 0.157, but smaller than the original estimate used in the main analysis and reported in the original paper (Column 1),

that is, 0.198. While the original results used in the Hausman test without robust standard errors were not statistically significant (Column 3), the corrected results with robust standard errors were found statistically significant at the 1% level (Column 4).

3 Robustness Reproduction

3.1 Robustness reproduction of the population density and income relationship

I conducted four pre-specified robustness reproductions of the analysis estimating the relationship between population density and per capita income in preindustrial societies (see pre-re-analysis plan in the Appendix):[1] (1) including country and year fixed effects (FE) instead of other control variables, which asks whether the estimated relationship between population density and income is sensitive to controlling for unobserved country characteristics and common year shocks instead of the original set of covariates, (2) including all observations including those where the logarithmic population density variable is smaller than zero, which tests the importance of the original sample restriction, (3) estimating the relationship in levels instead of natural logarithms for population density and per capita income, which is a diagnostic, functional-form check, and (4) using the full sample of observations provided by Ashraf and Galor (2011), which tests whether the results depend on the subset of observations used in the original specification. Three of these four pre-specified checks are used for the main outcomes below (the levels specification was estimated as planned but yielded implausibly high estimates and is therefore excluded from the reported results).[2] The robustness reproduction analyses were pre-specified to focus on the OLS model (Model 1). When possible, I have also conducted them for the IV model (Model 2), which was the main model used in the original paper.

In addition to the pre-specified robustness reproductions, I have also conducted an exploratory (that is, non-pre-specified) robustness reproduction, based on reading Ashraf and Galor's (2011) paper and looking at their data: (5) using log population density from Ashraf and Galor (2011) instead of Maddison (2003), which is a measurement check.[3]

These robustness checks are not meant to replace the preferred specification in the original paper. Instead, they are used to assess how much the first-step income estimates depend on choices about controls, sample restrictions, and data source. The results should therefore be interpreted relative to the original specification: they show how the estimates move when one assumption is changed, not which specification is correct.

The robustness reproducibility results for Model 1 (OLS) are presented in Table 4, and the robustness reproducibility results for Model 2 (IV) are presented in Table 5. The sample sizes differ across specifications for three reasons. First, the IV models require non-missing instrument data and therefore use fewer observations than the corresponding OLS models. Second, the "All observations" specifications remove the original log-population-density restriction but still use the original replication data. Third, the "Full AG2011 sample" specifications instead use the full Ashraf and Galor (2011) source sample, which is slightly larger.

[1] An earlier version of this replication study is available as part of the Institute for Replication's (I4R) replication project with *Nature Human Behaviour* (Brodeur et al. 2024); see Hammar (2025).

[2] The levels specification was pre-specified in the pre-re-analysis plan (see point 2c in the Appendix) and estimated as planned. The relationship between population density and per capita income was positive and statistically significant. Used as inputs for the income and inequality calculations, however, these OLS and IV estimates implied per capita incomes of 907 and 998 USD (1990 PPP), which appear historically implausible. The levels specification is therefore excluded from the rest of the analysis; the code remains in the replication package, and the estimates are available in Hammar (2025).

[3] Another exploratory robustness reproduction, using "plants" instead of "animals" as the instrumental variable, was also estimated as a check of instrument choice. However, the first-stage F-statistic of 1.621 indicated a very weak first stage, and the specification is therefore not reported. This specification was included in the earlier version of this replication study (Hammar 2025).

Table 4: Robustness reproduction results for original paper Model 1 (OLS)

Model 1 (OLS)	Original (1)	Country and year fixed effects (2)	All observations (3)	Full AG2011 sample (4)	Population density from AG2011 (5)
(Ln) Population density	0.189*** (0.040) [0.000]	0.335*** (0.044) [0.000]	0.083*** (0.012) [0.000]	0.084*** (0.012) [0.000]	0.080*** (0.011) [0.000]
Constant	5.728*** (0.387) [0.000]	5.456*** (0.094) [0.000]	5.995*** (0.165) [0.000]	5.999*** (0.164) [0.000]	6.062*** (0.162) [0.000]
Controls	Yes	No	Yes	Yes	Yes
Fixed effects	No	Yes	No	No	No
Observations	66	66	88	89	86
R-squared	0.317	0.847	0.321	0.321	0.315

Notes: The dependent variable is the natural logarithm of income per capita. The main independent variable is the natural logarithm of population density. Control variables include natural logarithm of land productivity, natural logarithm of absolute latitude, mean distance to nearest coast or river, and percentage of land within 100 km of coast or river. Population density from Maddison (2003), except for (5) where it is from Ashraf and Galor (2011) [AG2011]. While (3) removes the original restriction on log population density using the original replication data, (4) includes one additional country-year observation as it uses the full AG2011 source sample. Robust standard errors are in parentheses. P-values are reported in brackets. Significant at the ***[1%] **[5%] *[10%] level.

Table 5: Robustness reproduction results for original paper Model 2 (IV)

Model 2 (IV)	Original (1)	Original corrected (2)	All observations (3)	Full AG2011 sample (4)	Population density from AG2011 (5)
(Ln) Population density	0.198*** (0.048) [0.000]	0.157*** (0.054) [0.005]	0.118*** (0.032) [0.000]	0.121*** (0.032) [0.000]	0.107*** (0.026) [0.000]
Constant	5.829*** (0.151) [0.000]	5.807*** (0.313) [0.000]	5.909*** (0.242) [0.000]	5.916*** (0.243) [0.000]	6.033*** (0.197) [0.000]
Controls	Yes	Yes	Yes	Yes	Yes
Observations	66	57	66	67	65
R-squared	0.086	0.275	0.253	0.254	0.269
First-stage F	.	13.097	10.996	11.058	16.258

Notes: IV regressions using 2SLS. Number of domesticable species of animals used as instrumental variable. The dependent variable is the natural logarithm of income per capita. The main independent variable is the natural logarithm of population density. Control variables include natural logarithm of land productivity, natural logarithm of absolute latitude, mean distance to nearest coast or river, and percentage of land within 100 km of coast or river. Population density from Maddison (2003), except for (5) where it is from Ashraf and Galor (2011) [AG2011]. Small-sample degrees-of-freedom adjustments, except for (1). Adjusted R-squared, except for (1). Columns (1) and (2) same as columns (1) and (4) in Table 3, respectively. Robust F-statistics reported from first-stage regressions. Robust standard errors are in parentheses. P-values are reported in brackets. Significant at the ***[1%] **[5%] *[10%] level.

3.1.1 Country and year fixed effects

First, the results could be affected by the choice of control variables included and excluded in the empirical specification estimating the relationship between population density and per capita income in preindustrial societies. As such, in a first reproducibility robustness check, I included country and year fixed effects instead of other control variables. These results are presented in Table 4 (Column 2). The relationship was still found positive and statistically significant, but with a larger point estimate (0.335) than in the original OLS specification (0.189).

3.1.2 All observations

Second, in the original analysis the sample was restricted to the observations where the logarithmic population density was greater than zero. In a second robustness check, I included all observations including those where the logarithmic population density variable was smaller than zero. This increased the sample size from 66 to 88 country-years in the OLS specification (additional countries in the larger sample were Algeria, Australia, Brazil, Canada, Finland, Libya, New Zealand, Norway, and the United States), and from 57 to 66 observations in the IV specification (additional countries Brazil, Finland, and Norway). The IV sample is smaller because it also requires non-missing values for the instrument and the first-stage variables. These results are presented in Table 4 (Column 3) for the OLS results, and in Table 5 (Column 3) for the IV results. The relationship was found positive and statistically significant, but with smaller point estimates (0.083 for OLS and 0.118 for IV) than in the original specifications.

3.1.3 Full Ashraf and Galor (2011) sample

Third, according to the original paper, the analysis was conducted using "a sample of 66 pre-industrial societies, which included 22 distinct world areas all observed in three different years (1, 1000 and 1500 AD)" (p. 1269), based on data from Ashraf and Galor (2011). This statement, however, was not fully correct. The original OLS analysis included 66 observations, but not from a balanced panel of countries and years. Instead, it included 24 countries, of which 21 were observed in all three years, while 3 countries were only observed in one year. The first stage in the original IV analysis included 21 countries, of which 18 were observed in all three years. As a third robustness check, I used the full sample of observations provided by Ashraf and Galor (2011). This increased the sample size to 89 country-years in the OLS specification and to 67 country-years in the IV specification (the additional country relative to the "All observations" robustness specification in Column 3 was Tunisia, observed in year 1). These larger-sample results are presented in Table 4 (Column 4) for OLS and in Table 5 (Column 4) for IV. The relationship between population density and per capita income was still found positive and statistically significant, but with smaller coefficients (0.084 for OLS and 0.121 for IV) than found in the original specifications.

3.1.4 Population density from Ashraf and Galor (2011)

Fourth, and finally, the data from Ashraf and Galor (2011) included two different estimates of the log population density: one compiled from various historical sources (Ashraf and Galor 2011), and one taken from Maddison (2003). In the original paper, only the latter variable, that is, based on Maddison (2003), was used. Hence, as a fourth (non-pre-specified) robustness check, I instead used Ashraf and Galor's (2011) main log population density variable (that is, the one based on various sources). The results are presented in Table 4 (Column 5) for OLS, and in Table 5 (Column 5) for IV. The relationship was still found positive and statistically significant, but with smaller point estimates (0.080 for OLS and 0.107 for IV) than in the original paper.

3.2 Robustness reproduction of the main outcomes: income and inequality

Overall, the robustness reproduction results in Section 3.1 showed that the positive and statistically significant relationship between population density and per capita income was robust and reproducible. This result, however, was not a main finding of the original paper. Instead, the coefficients from these regressions were used as input variables for estimating the main outcomes: income levels, income inequality, and imperial inequality in the Aztec Empire. Next, I have thus used the coefficient estimates from Section 3.1 as inputs to analyze the robustness reproducibility of the following main outcomes: absolute per capita income, relative per capita income, Gini index, richest 1% income share, richest 5% income share, poorest 50% income share, and extraction imperial ruling class. The calculations of these outcomes were conducted in the same way as in the original paper, but using the coefficient estimates from the robustness regressions as inputs. These robustness exercises, however, should not be interpreted as a set of equally preferred alternatives to the original specification, but rather their aim is to vary different assumptions in order to show which parts of the results depend on those modelling choices. As such, the robustness results should be read mainly as evidence on uncertainty, where the magnitudes may depend on assumptions made in the first-step income regression. The results are presented in Table 6.

Table 6: Robustness reproduction results for main outcomes

		Per capita income		Income inequality	Income share (%)			Imperial inequality (%)
Model	Specification	Absolute (1990 PPP)	Relative (sub. min.)	Gini index	Richest 1%	Richest 5%	Poorest 50%	Extraction ruling class
IV.1	Original	690	1.72	50.4	41.8	50.8	23.3	4.54
OLS.1	Original	611	1.53	44.2	36.2	44.4	26.3	5.12
OLS.2	Country and year FE	771	1.93	55.6	47.5	56.0	20.9	4.06
OLS.3	All observations	557	1.39	38.8	30.7	39.1	28.9	5.61
OLS.4	Full AG2011 sample	560	1.40	39.1	30.9	39.3	28.8	5.59
OLS.5	Pop. density AG2011	586	1.46	41.7	33.2	42.0	27.5	5.34
IV.2	Original corrected	593	1.48	42.5	34.5	42.7	27.1	5.27
IV.3	All observations	575	1.44	40.6	32.5	40.9	28.0	5.45
IV.4	Full AG2011 sample	583	1.46	41.4	33.3	41.7	27.6	5.37
IV.5	Pop. density AG2011	621	1.55	45.0	36.3	45.3	25.9	5.04
	Mean (robustness)	*606*	*1.51*	*43.1*	*34.9*	*43.4*	*26.8*	*5.22*

Notes: Comparison between estimates of income levels, income inequality, and imperial inequality in the original paper and robustness reproductions. Mean (robustness) measured as average values of all robustness specifications, that is, OLS.2–5 and IV.2–5, which should be interpreted as a descriptive average and not an average over equally preferred estimates. For description of how regression results are used to calculate income and inequality levels, see the original paper.

3.2.1 Income levels

In the original paper, the absolute per capita income for the Aztec Empire as a whole was estimated to 690 USD (1990 PPP) using the original IV specification, and to 611 USD using the original OLS specification.

The average value for my robustness reproduction estimates was 606 USD, that is, similar to the original paper's OLS estimate. At the same time, there is relatively large variation in the values estimated in the different robustness specifications, ranging from 557 USD to 771 USD, suggesting a high degree of uncertainty about the true value. For relative per capita income, in the original paper, this was estimated to 1.72 times the subsistence minimum in the IV specification, and to 1.53 subsistence minimums in the OLS specification. In my different robustness specifications, this value ranged between 1.39 and 1.93, with an average value of 1.51 subsistence minimums. Note, however, that all estimates should not be read as equally plausible historical estimates of Aztec income, but rather as a range of possible alternatives. It should also be noted that the original paper emphasizes that the income estimates are approximations, based on strong assumptions. It also includes a related robustness exercise, reducing each provincial GDP estimate by 10%, which shows that the estimated level of inequality seems to be relatively sensitive to the income level, while the conclusion of high extraction appears less sensitive.

3.2.2 Income inequality

For income inequality, in the original paper, the Gini index was estimated to 50.4 using Model 2 (IV) and to 44.2 using Model 1 (OLS). In my robustness estimations, the average value for the Gini index was 43.1, with estimates ranging from 38.8 to 55.6. The income share of the richest 1% was estimated to 41.8% of total income (36.2% for OLS) in the original paper. The average value in my robustness analyses was 34.9%, with estimates ranging between 30.7% and 47.5%. For the income share of the richest 5%, the original paper's estimate was 50.8% (44.4% for OLS). The average robustness reproduction estimate was 43.4%, with estimated values between 39.1% and 56.0%. The income share of the poorest 50% was estimated to 23.3% (26.3% for OLS) in the original paper. My robustness estimates ranged from 20.9% to 28.9%, with the average estimated income share of the bottom half being 26.8% of total income. As with income levels, the original paper's income inequality estimate thus seems a plausible estimate, albeit with large uncertainty.

3.2.3 Imperial inequality

The original paper also estimated the level of imperial extraction across the Aztec Empire. The imperial ruling class extraction was estimated to 4.54% of total income in Model 2 (IV), and 5.12% in Model 1 (OLS). The average estimate in my robustness analyses was 5.22%, with estimates ranging between 4.06% and 5.61% of total income.

3.3 Testing the statistical difference between the income distributions

Finally, in the original paper a Kolmogorov-Smirnov test was conducted to assess if the estimated income distribution of the Aztec Empire in 1492 was statistically different from those of modern Mexico and the United States in 2018. It should be noted, however, that these comparisons were used mainly to illustrate the magnitude of the estimated Aztec income distribution, not as the original paper's central finding. Nevertheless, the analysis in the original paper found that the distribution was statistically different (with a higher level of inequality in the Aztec Empire than in modern Mexico and United States). While I was initially not able to fully reproduce this result using only the data and code provided in the original replication package, the Kolmogorov-Smirnov results were later fully reproduced using the ventile data and code supplied separately by the original authors.

In a final analysis, I have used the Aztec Empire income distributions estimated from the different robustness reproductions and tested if they were statistically different from 1) the estimated Aztec Empire distribution in the original paper, 2) the income distribution in modern Mexico, and 3) the income distribution in modern United States. The comparisons were conducted through a Kolmogorov-Smirnov test and based on estimated income shares of deciles and the top 5% income group (that is, the

comparisons below use the 11-group aggregation described in Section 2). The results are presented in Table 7.

Table 7: Kolmogorov-Smirnov tests comparing income distributions

Model	Specification	Original paper Aztec Empire 1492 (1)	Mexico 2018 (2)	USA 2018 (3)
IV.1	Original	.	0.55* [0.075]	0.64** [0.021]
OLS.1	Original	0.45 [0.211]	0.55* [0.075]	0.55* [0.075]
OLS.2	Country and year fixed effects	0.45 [0.211]	0.64** [0.021]	0.64** [0.021]
OLS.3	All observations	0.55* [0.075]	0.45 [0.211]	0.55* [0.075]
OLS.4	Full AG2011 sample	0.55* [0.075]	0.45 [0.211]	0.55* [0.075]
OLS.5	Population density from AG2011	0.45 [0.211]	0.45 [0.211]	0.55* [0.075]
IV.2	Original corrected	0.45 [0.211]	0.45 [0.211]	0.55* [0.075]
IV.3	All observations	0.45 [0.211]	0.45 [0.211]	0.55* [0.075]
IV.4	Full AG2011 sample	0.45 [0.211]	0.45 [0.211]	0.55* [0.075]
IV.5	Population density from AG2011	0.45 [0.211]	0.55* [0.075]	0.55* [0.075]

Notes: Results from Kolmogorov-Smirnov tests comparing the estimated income distributions for the Aztec Empire from the different reproduction robustness analyses with (1) the distribution estimated in the original paper, (2) the distribution in Mexico 2018, and (3) the distribution in the United States 2018, using the *ksmirnov* command in Stata. Estimations based on income distributions in deciles and the top 5% as a separate group. Combined Kolmogorov–Smirnov D. Exact p-values are reported in brackets. Difference between the distributions significant at the ***[1%] **[5%] *[10%] level.

The results from the Kolmogorov-Smirnov tests reported in Table 7 show that out of the eight different robustness reproduction analyses, six yielded income distributions that were not statistically different from the original paper. However, two were statistically different at the 10% significance level.

Comparing the estimated distributions from the robustness reproductions with the income distribution in modern Mexico, those were not statistically different in six specifications, statistically different at the 10% significance level in one specification, and at the 5% level in one specification. The comparison with the income distribution in modern United States showed statistically different distributions in all cases (at the 10% significance level in seven specifications, and at the 5% level in one specification). In other words, the original paper's result that incomes in the Aztec Empire were more unequally distributed than those in modern United States seems to be a robust finding. The original paper's conclusion that "[t]he income distribution of the Aztec Empire differs significantly from those of modern-day Mexico" (p. 1266), however, does not appear to be robust. In other words, the comparison with modern Mexico is more sensitive to the specification than the comparison with the United States. This suggests that fine comparisons with modern

distributions should be interpreted cautiously.

Finally, a graphical comparison of the different Gini index estimates is illustrated in Figure 1.

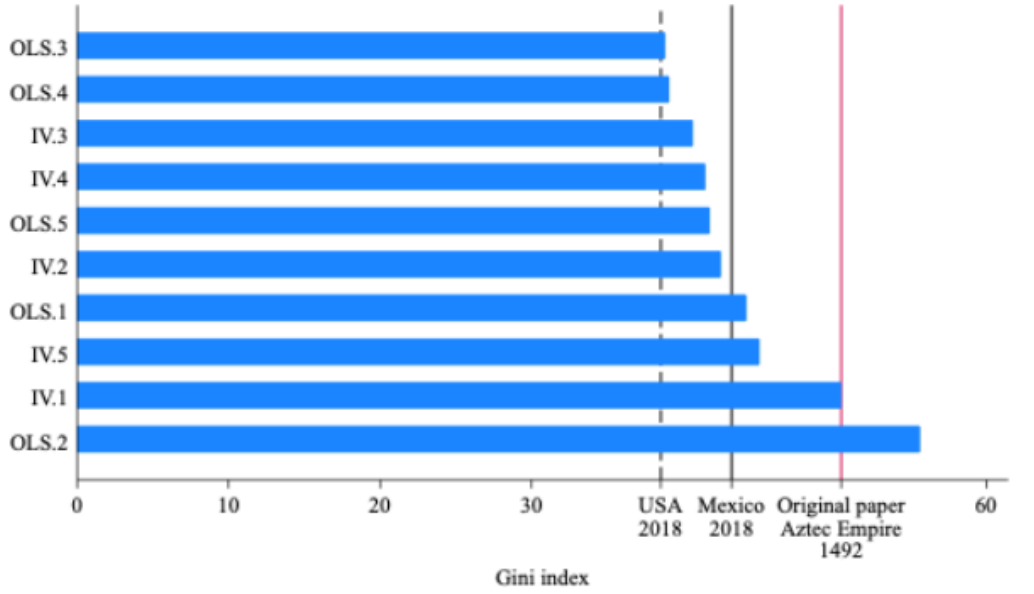

Figure 1: Gini index comparison between original paper and robustness reproductions
Notes: Gini indices for the Aztec Empire computed based on the coefficient estimates from the different robustness reproduction regressions, as well as the original paper's Gini index for the Aztec Empire (red solid line), modern Mexico (black solid line), and the United States (black dashed line). For description of the different specifications, see main text.

Figure 1 shows the relatively large variation in Gini indices estimated for the Aztec Empire in the different robustness reproductions. While, on average, the robustness estimates were relatively similar to the one in the original paper, all except one were lower and the exact value appears very uncertain. All robustness estimations resulted in higher Gini coefficients than the one of modern United States, supporting the original paper's result that income inequality was higher in the Aztec Empire than in contemporary United States. Comparing the income distribution in the Aztec Empire with the one in current Mexico, however, six of the eight alternative estimations resulted in lower Gini coefficients and two in higher. Again, the original paper's finding that income inequality was higher in the Aztec Empire than in modern Mexico does not appear to be a robustly reproducible result. Yet, the qualitative conclusions that the Aztec Empire had high income inequality and high extraction are stable across the exercises.

4 Conclusion

The results of this replication study support the main qualitative conclusion in Alfani and Carballo (2023). The Aztec Empire appears to have been a highly unequal society before the Spanish conquest, and the evidence for high imperial extraction is especially stable. The original results are computationally reproducible from the analysis data, and the robustness checks do not overturn the broad interpretation of the original paper.

The main lesson from the robustness exercises, however, is about uncertainty. In particular, this replication challenges how the point estimates should be interpreted, where the exact level of income and inequality is more uncertain than the preferred estimate alone would suggest. This is not surprising given the limited evidence and the strong assumptions needed to estimate income distributions for the Aztec Empire before the Spanish conquest. Still, the Gini estimate moves enough that fine comparisons with modern income distributions should be interpreted cautiously. Also, in these robustness analyses, I have only varied the specifications for the regression analysis estimating the relationship between population density and per capita income. Further variations in, for example, the raw input data or the model for estimating the levels of income and inequality would likely yield even more uncertainty.

Hence, the main takeaway from this replication is a distinction between the direction of the result and its precision. The evidence supports the view that inequality and extraction were high in the Aztec Empire before Spanish rule. It is less informative about the exact income and inequality levels, and about whether Aztec inequality should be ranked above or below a particular modern country. Readers should therefore treat the original estimates as historically informative approximations, not as precise measurements.

For the broader literature, the results suggest that colonial institutions should not be treated as the only possible source of high inequality in Mexico and Latin America, as the estimates remain consistent with substantial inequality before the Spanish conquest. At the same time, the exercise shows the limits of precision in this setting. Quantitative estimates for historical societies can be informative for assessing broad patterns, but they are sensitive to assumptions that are difficult to verify.

References

Alfani, G., and Carballo, A. (2023). *Income and Inequality in the Aztec Empire on the Eve of the Spanish Conquest*. Nature Human Behaviour, **7**, 1265–1274. DOI: 10.1038/s41562-023-01636-3.

Ashraf, Q., and Galor, O. (2011). *Dynamics and Stagnation in the Malthusian Epoch*. American Economic Review, **101**, 2003–2041.DOI: 10.1257/aer.101.5.2003.

Brodeur, A., Dreber, A., Hoces de la Guardia, F., and Miguel, E. (2024). *Reproduction and Replication at Scale*. Nature Human Behaviour, **8**, 2–3. DOI: 10.1038/s41562-023-01807-2.

Hammar, O. (2025). *A Comment on "Income and Inequality in the Aztec Empire on the Eve of the Spanish Conquest"*. I4R Discussion Paper, **221**. Institute for Replication.PDF: https://www.econstor.eu/bitstream/10419/315048/1/I4R-DP221.pdf.

Jenkins, S. P. (1999). *INEQDECO: Stata Module to Calculate Inequality Indices with Decomposition by Subgroup*. Statistical Software Components, **S366002**. Boston College Department of Economics.

Maddison, A. (2003). *The World Economy: Historical Statistics*. Development Centre Studies. Paris: OECD. URL: https://www.oecd.org/en/publications/2003/09/the-world-economy_g1gh38a7.html.

Appendix

I4R NHB

Pre-re-analysis plan

Name of replicator: Olle Hammar

Title of paper: Income and inequality in the Aztec Empire on the eve of the Spanish conquest

Original authors: Guido Alfani & Alfonso Carballo

I will conduct a computational and robustness reproduction of Alfani and Carballo (2023) "Income and inequality in the Aztec Empire on the eve of the Spanish conquest" (*Nature Human Behaviour* 7: 1265–1274), including the following steps:

1. The analysis in the original paper is conducted in Stata and Excel. Since codes are only provided for parts of the analysis, a first check will be to recode the main analysis in Stata based on the information provided in the original article and its supplementary materials.
2. Next, I will check the robustness of the estimation results analyzing the relationship between population density and per capita income for preindustrial societies, focusing on the OLS model (Model 1). The robustness checks will consist of four alterations:
    a. The results could be affected by the choice of control variables included/excluded in the specification. As such, in a first robustness check, I will include country and year fixed effects instead of other control variables.
    b. In the original analysis, the sample is restricted to the observations where the logarithmic population density is greater than zero. In a second robustness check, I will include all observations including those where the logarithmic population density variable is smaller than zero.
    c. As shown in the original paper's Supplementary Figure 1, there is a positive correlation between population density and per capita income both in absolute values and for their natural logarithmic transformation. Hence, as a third robustness check, I will use absolute values instead of natural logarithms for population density and per capita income.
    d. The original analysis is conducted using "a sample of 66 pre-industrial societies, which included 22 distinct world areas all observed in three different years (1, 1000 and 1500 AD)", based on data from Ashraf and Galor (2011) "Dynamics and stagnation in the Malthusian epoch" (*American Economic Review* 101: 2003–2041). As a final robustness check, I will use the full sample of observations provided by Ashraf and Galor (2011).
3. Finally, I will check to what extent these robustness analyses change the following main outcomes (measured for the Aztec Empire as a whole):
    a. Income levels
        i. Per capita income, absolute: **US$690** (1990 PPP)
        ii. Per capita income, relative: **1.72** times subsistence minimum
    b. Income inequality
        i. Gini index: **50.4**
        ii. Richest 1% income share: **41.8%** of total income
        iii. Richest 5% income share: **50.8%** of total income
        iv. Poorest 50% income share: **23.3%** of total income
    c. Imperial inequality
        i. Extraction imperial ruling class: **4.54%** of total income

To test whether the income distribution becomes different from that reported in the paper, I will use the Kolmogorov–Smirnov test.