Estimating the Impact of the Marshall Plan on Post-War European Economic Recovery

Bruce Li

Abstract

This paper investigates the impact of the Marshall Plan on the economic recovery of European countries following World War II. The Marshall Plan, officially known as the European Recovery Program (ERP), was a United States initiative that provided over $13 billion in economic aid to help rebuild Western European economies following World War II. Using a Difference-in-Differences (DiD) regression model, this paper estimates whether countries receiving substantial Marshall Plan aid experienced greater growth in GDP per capita than countries receiving relatively little or no aid.

Table of Contents

Estimating the Impact of the Marshall Plan on Post-War European Economic Recovery 1

1. Introduction 3

2. Motivation for the Econometric Model 4

2.1 Debate in the Literature 4

2.2 Why Difference-in-Differences? 5

2.3 Construction of Treatment and Control Groups 6

2.4 Parallel trends 9

3. Data 10

3.1 Sources 10

3.2 Sample 11

3.3 Variables 12

3.4 Descriptive Statistics 13

3.5 Descriptive Patterns 13

4. Estimation Strategy and Empirical Results 15

4.1 Methodology 15

4.2 Regression Results 16

4.3 Robustness checks 20

4.4 Heterogeneity checks 22

5. Conclusion 25

References 28

Appendix 30

1. Introduction

The rapid recovery of Western Europe after World War II is often presented as one of the clearest successes of post-war economic policy. At the centre of this story is the Marshall Plan, officially known as the European Recovery Program, through which the United States provided large-scale financial aid to European countries between 1948 and 1952. The programme is credited with helping to rebuild infrastructure, revive global trade and restore global economic order. However, Marshall aid coincided with the impacts of domestic policy reforms, the return of political stability, and the tendency of poorer economies to catch up with richer ones. As these developments occurred at the same time, it is difficult to determine how much of the recovery can be attributed specifically to Marshall Plan aid.

This paper asks whether countries receiving substantial Marshall Plan aid experienced stronger growth in GDP per capita than countries receiving little or no aid. It contributes to the debate by comparing the economic performance of recipient countries with a broader control group using a Difference-in-Differences framework. The analysis also distinguishes between changes in the level of GDP per capita and proportional growth. This distinction is important because the countries receiving Marshall Plan aid generally began the period with higher income levels than many of the countries in the control group. A larger increase in GDP per capita may therefore not necessarily represent a faster percentage rate of growth.

The empirical analysis uses annual country-level data from 40 countries, producing 960 country-year observations. GDP per capita data are drawn primarily from the Maddison Project Database, while information on Marshall Plan aid is taken from Learn Europe's compilation of OECD, Maddison, and ISTAT data. The pre-treatment period is defined as 1920 to 1935, while the post-treatment period runs from 1949 to 1956. The models use both a binary treatment indicator, based on whether a country received more than $10 of aid per capita, and a continuous measure of aid per capita. This makes it possible to test both whether recipient countries

recovered differently from non-recipients and whether countries receiving more intensive aid experienced larger gains.

The results provide strong evidence that the Marshall Plan was associated with higher absolute recovery than with faster proportional growth. In the binary GDP per capita model, treated countries experienced an estimated additional increase of approximately 1,246 GDP per capita units relative to the control group, and the result is statistically significant at the 1 percent level. The continuous aid model produces a similar pattern: each additional dollar of aid per capita is associated with approximately 15.9 additional GDP per capita units in the post-treatment period. However, when GDP per capita is expressed in logarithmic form, neither the binary treatment coefficient nor the continuous aid coefficient is statistically significant. The evidence therefore suggests that Marshall Plan recipients experienced larger absolute income gains, but it does not show that they grew faster in percentage terms.

2. Motivation for the Econometric Model

2.1 Debate in the Literature

The extent to which the Marshall Plan contributed to Europe's post-war economic recovery remains a subject of considerable debate among economic historians with varying emphases on different areas of economic recovery. One school of thought argues that the Marshall Plan played a decisive role in accelerating the reconstruction of European countries. De Long and Eichengreen (1991), Eichengreen and Uzan (1992), Hogan (1987), and Casella and Eichengreen (1996) contend that the program not only supplied much-needed capital but also promoted economic cooperation and policy reforms that laid the foundations for sustained growth.

In contrast, other scholars argue that Europe's recovery can largely be explained without attributing a central role to the Marshall Plan. Milward (1984), Abelshauser (1983), Vonyó (2008), and Alvarez-Cuadrado and Pintea (2008) emphasise the importance of post-war reconstruction,

arguing that the replacement of wartime destruction naturally generated rapid economic expansion. Similarly, Abramovitz (1986), Barro and Sala-i-Martin (1992), Baum and Barkoulas (2006), and Giannetti (2002) argue that poorer European economies experienced rapid growth through economic convergence, gradually catching up with wealthier countries by adopting existing technologies and production methods. Others, including Cowen (1985), Milward (1984), and Abelshauser (1983), emphasise the role of domestic economic policies and national reforms.

As these explanations are not mutually exclusive and largely occurred simultaneously during Europe's post-war recovery, it is difficult to determine the independent causal effect of the Marshall Plan through historical analysis alone. This motivates the use of an econometric approach capable of isolating the effect of Marshall Plan aid while accounting for broader economic changes affecting all countries during the period.

2.2 Why Difference-in-Differences?

To estimate the causal effect of the Marshall Plan, this paper employs a Difference-in-Differences (DiD) research design. DiD compares the change in GDP over time between countries receiving substantial Marshall Plan aid and a control group receiving little or no aid. By comparing these differences, the method controls for common macroeconomic shocks affecting all European countries during the post-war period.

The identifying assumption underlying the DiD model is that, in the absence of the Marshall Plan, the treatment and control groups would have followed parallel trends in GDP per capita. This assumption is assessed through graphical analysis of pre-treatment trends before estimating the regression model. Under this assumption, the interaction term between the treatment and post-treatment indicators provides an estimate of the average causal effect of substantial Marshall Plan aid on post-war economic recovery.

2.3 Construction of Treatment and Control Groups

Before estimating the Difference-in-Differences model, the treatment and control groups must be constructed in a way where the countries' pre-treatment trends were sufficiently similar.

Time periods

The pre-treatment period is defined as 1920 to 1935. This period provides a sufficiently long baseline for comparing the economic trajectories of the treatment and control groups before the Marshall Plan. Beginning in 1920 allows the analysis to capture the post-World War I recovery, the expansion of the 1920s, and the downturn caused by the Great Depression. Although the Great Depression introduced substantial economic disruption, it affected countries across both the treatment and control groups, making the period useful for assessing whether the two groups moved in broadly similar directions before treatment. The period ends in 1935 to avoid later disruptions such as the Spanish Civil War and World War II, both of which would make pre-treatment comparisons less reliable.

The post-treatment period is defined as 1949 to 1956. Although the Marshall Plan formally began in 1948, 1948 is treated as a transition year because aid was only beginning to be implemented and its economic effects would not necessarily have appeared immediately in GDP per capita. Starting the post-treatment period in 1949 therefore allows the analysis to focus on years when the program was more fully underway. Extending the post-treatment period to 1956 captures both the immediate recovery after the Marshall Plan and possible medium-run effects through reconstruction, investment, industrial production, and trade recovery. The full period from 1920 to 1956 is used for the parallel trends graph, while the regression analysis focuses specifically on the defined pre-treatment and post-treatment windows.

Treatment threshold and aid intensity

Treatment is defined using Marshall Plan aid per capita rather than total aid. Total aid would assign greater importance to larger countries, such as the United Kingdom, France, Italy, and West Germany, because they had larger populations and economies. Aid per capita is more appropriate for this study because it measures the intensity of aid received by the average person and is more consistent with GDP per capita as the dependent variable.

A country is classified as treated if it received more than $10 of Marshall Plan aid per capita. This threshold is chosen because the aid-per-capita data show a natural break between low-aid and higher-aid recipients: Yugoslavia received $1.76 per capita, Portugal received $5.95, and Turkey received $7.04, while the next country, Sweden, received $16.76. Therefore, any threshold between approximately $7 and $17 would produce the same classification. A $10 threshold is a simple and transparent cutoff located within this gap.

Table 1 Country classification based on Marshall Plan aid per capita

Country	Aid per capita (US$)	Classification
United Kingdom	56	Treatment
France	57	Treatment
Italy	28	Treatment
Netherlands	85	Treatment
Austria	81	Treatment
Belgium	61	Treatment
Greece	67	Treatment
Denmark	60	Treatment
Norway	72	Treatment
Turkey	7	Control

Ireland	49	Treatment
Sweden	17	Treatment
Portugal	6	Control
Yugoslavia	2	Control
Iceland	162	Treatment

Source: Adapted from Learn Europe, compiled from OECD, Maddison Project Database and ISTAT data.

Three cases are excluded from the main analysis. Trieste is excluded because it was not a standard independent country during the period and reliable comparable GDP per capita data are not available in the dataset. West Germany and Luxembourg are also excluded from the baseline sample because the dataset lacks complete pre-treatment observations for 1920 to 1935. Since the parallel trends assumption must be assessed before treatment begins, countries without pre-treatment data cannot be used in the baseline DiD comparison.

To increase the size of the control group and improve the precision of the DiD estimates, the study also includes a number of non-recipient countries. These countries were selected because they possess reliable historical GDP per capita data, did not receive Marshall Plan aid, and provide additional observations against which the post-war performance of recipient countries can be compared.

Table 2 Final treatment and control groups

Treatment countries	**Control countries**
United Kingdom, France, Italy, Netherlands, Austria, Belgium, Greece, Denmark, Norway, Sweden, Iceland	Spain, Portugal, Australia, Canada, New Zealand, Switzerland, Finland, Mexico, Chile, Uruguay, Brazil, Argentina, Colombia, Costa Rica, Ecuador, Panama, Peru, Sri Lanka, Bolivia, Cuba, Guatemala, Honduras,

	Venezuela, Nicaragua, El Salvador, Malawi, Zimbabwe, Ghana, Liberia

2.4 Parallel trends

The parallel trends assumption is assessed by comparing the average GDP per capita trajectories of the treatment and control groups before the Marshall Plan. Figure 1 shows a comparison of the average GDP per capita between the treatment and control groups.

Before World War II, the two groups moved in broadly similar directions. Both rose during the 1920s, declined during the Great Depression, and recovered during the late 1930s. The war years show a sharper decline for the treatment group, likely because many treated countries experienced more direct wartime destruction. This makes the parallel trends assumption imperfect, but the pre-war movements still provide enough visual support to proceed with the DiD model.

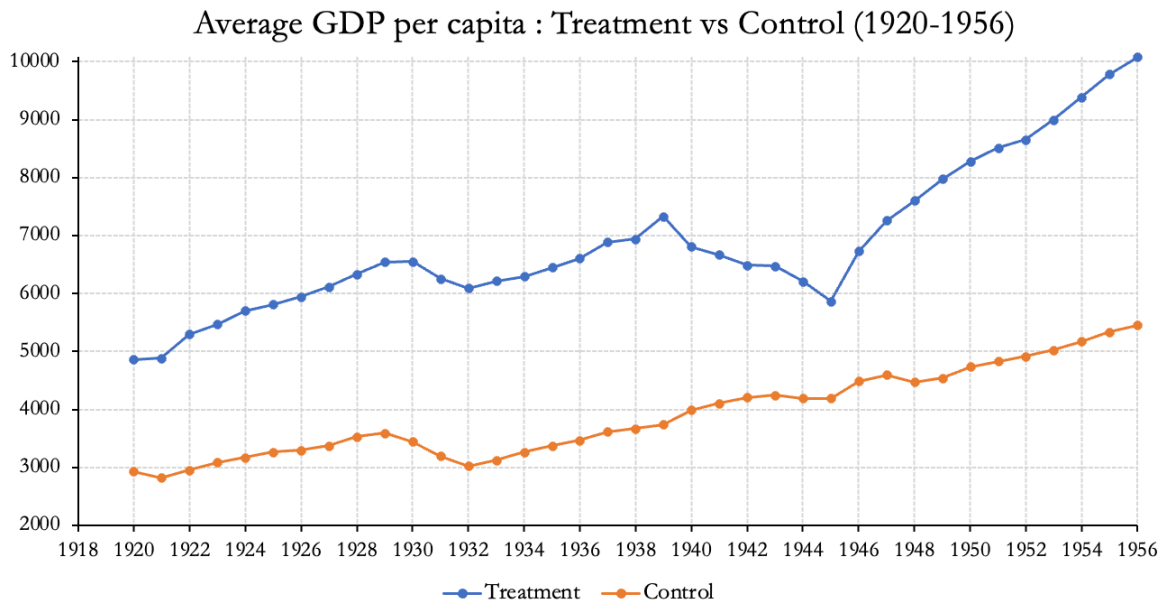


Figure 1 Average GDP per capita of treatment and control groups, 1920-1949.

After the Marshall Plan began in 1948, the treatment group rose more quickly than the control group. This pattern is consistent with the idea that Marshall Plan aid supported faster recovery,

but the graph alone cannot prove causality. The regression analysis is therefore needed to estimate the size and statistical significance of the treatment effect.

3. Data

3.1 Sources

This study combines several historical datasets in order to compare economic performance across countries before and after the Marshall Plan. The main source is the Maddison Project Database, which provides long-run estimates of GDP per capita and population for a large number of countries. Its broad country and year coverage makes it well suited to this study, since the analysis requires comparable observations for both European recipients of Marshall Plan aid and a wider group of non-recipient countries between 1920 and 1956. The Maddison data are used throughout the main analysis and also provide the historical information needed for several robustness and heterogeneity checks, including initial income, wartime GDP per capita decline, and country population. A limitation is that figures from the early twentieth century are reconstructed historical estimates rather than modern national accounts, so they should not be treated as perfectly precise. However, using a single source across the sample improves consistency and makes the figures more comparable than combining separate national series.

Information on Marshall Plan aid is adapted from Learn Europe, which presents country-level aid figures compiled from the Organisation for Economic Co-operation and Development (OECD), the Maddison Project Database, and the Italian National Institute of Statistics (ISTAT). This source provides a consistent summary of the aid received by the participating European economies and allows aid to be compared across countries. It is particularly useful because it reports the information needed to calculate aid intensity on a per-person basis rather than relying only on total aid. The same source is therefore used for both the treatment classification and the continuous measure of Marshall Plan aid.

Two additional historical databases are used for the heterogeneity analysis. Population observations missing from the Maddison data are supplemented using Clio-Infra, a database designed to provide comparable long-run indicators across countries. Clio-Infra is also the source of the historical urbanization data. Its urban population estimates for 1900 and 1950 provide coverage for countries for which annual interwar urbanization figures are unavailable. This makes it possible to include urban development as a country characteristic, although the limited number of historical benchmark years means that the urbanization measure is less precise than the annual GDP per capita series. The paper therefore treats it as an approximate indicator of countries' level of urban development before the Marshall Plan.

Pre-war trade data come from the CEPII TRADHIST database, which contains historical information on merchandise trade and national economic activity. CEPII is used because it provides internationally comparable trade data extending into the interwar period, when coverage from modern sources such as the World Bank is generally unavailable. It allows the study to examine whether countries that were more integrated into international trade before World War II experienced different outcomes after receiving Marshall Plan aid. Historical trade coverage is incomplete for some countries, so this part of the analysis uses a smaller sample than the baseline regressions.

3.2 Sample

The final regression sample contains 40 countries and 960 country-year observations. 11 countries are classified as treated because they received more than $10 of Marshall Plan aid per capita and have complete GDP per capita data for the required years. 29 countries are included in the control group because they received little or no Marshall Plan aid and have complete GDP per capita data

for the required years. The pre-treatment period is 1920 to 1935, while the post-treatment period is 1949 to 1956. This means that each country contributes 24 regression observations.

3.3 Variables

Table 3 Variables used in the regression analysis

Variable	Definition	Role in the model
GDP per capita	Annual GDP per capita for each country	Dependent variable in Models 1 and 3
Treatment	Equals 1 if a country received more than $10 of Marshall Plan aid per capita, and 0 otherwise	Binary treatment group indicator
Post	Equals 1 for 1949 to 1956, and 0 for 1920 to 1935	Post-treatment indicator
Treatment x Post	Equals 1 for treated countries in the post-treatment period	Main DiD variable in Models 1 and 2
Aid per capita	Marshall Plan aid per person, measured in US dollars	Continuous aid intensity variable
Aid per capita x Post	Aid per capita multiplied by the post-treatment indicator	Main DiD variable in Models 3 and 4

The key variable in the binary treatment regressions is Treatment x Post. This variable identifies treated countries after the Marshall Plan began, so its coefficient estimates whether treated countries experienced a larger change in GDP per capita than control countries after treatment. The key variable in the continuous aid intensity regressions is Aid per capita x Post. This tests whether countries receiving more aid per person experienced larger post-treatment gains.

3.4 Descriptive Statistics

Table 4 Descriptive statistics for regression variables

Variable	Mean	Standard deviation	Minimum	Maximum
GDP per capita	5,044	3,081	1,242.00	14,507
Aid per capita	21	37	0.00	162
Treatment	0.33	0.46	0	1
Post	0.33	0.47	0	1
Treatment x Post	0.10	0.30	0	1

The mean of the treatment variable is 0.31 because 11 of the 36 countries in the sample are treated. The mean of the post variable is 0.33 because the regression uses 16 pre-treatment years and 8 post-treatment years. Aid per capita varies substantially across countries, ranging from zero for non-recipient countries to 162.33 dollars for Iceland.

3.5 Descriptive Patterns

Table 3 compares average GDP per capita in the treatment and control groups before and after the Marshall Plan.

Table 5 Average GDP per capita by group and period

Group	Pre-treatment average	Post-treatment average	Change
Treatment group	5,926	8,958	3,032
Control group	3,553	5,528	1,975
Difference	2,373	3,430	1,057

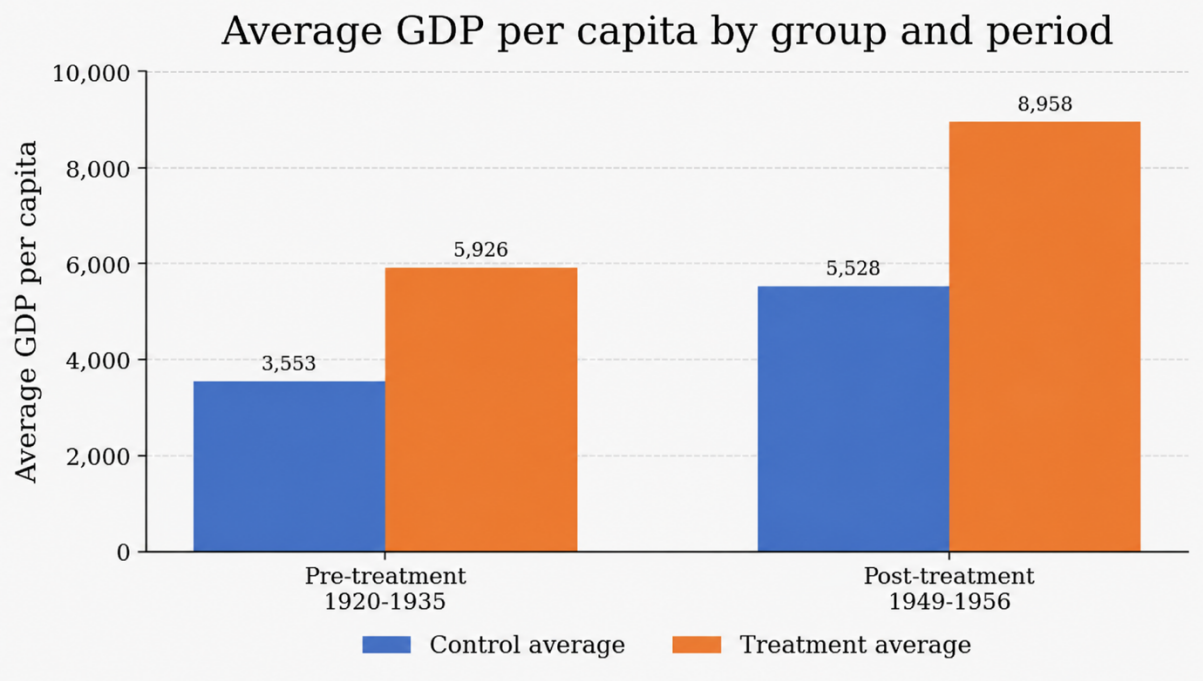


Figure 2 Average GDP per capita by group and period.

The descriptive data show that treated countries had higher GDP per capita than control countries before the Marshall Plan began. This difference in levels does not invalidate the DiD approach, since the model compares changes over time rather than requiring the two groups to begin at the same level.

Both groups experienced growth between the pre-treatment and post-treatment periods. However, the treatment group's average GDP per capita increased by 1,938.13, compared with 1,722.07 for the control group. This produces a raw difference-in-differences value of 216.06. This figure should not be interpreted as a causal estimate on its own, since it does not account for statistical uncertainty or other sources of variation. However, it does show that the descriptive pattern is consistent with the possibility that treated countries recovered somewhat faster after the Marshall Plan began. The next section therefore estimates this relationship using an OLS Difference-in-Differences regression.

4. Estimation Strategy and Empirical Results

4.1 Methodology

This section estimates whether countries that received substantial Marshall Plan aid experienced a larger increase in GDP per capita after 1948 than countries in the control group. The models are estimated using ordinary least squares (OLS). OLS is appropriate here because the dependent variable is continuous, and the regression estimates the average relationship between treatment status, the post-treatment period, and GDP per capita.

This study utilises two different models. First, using a binary treated/control indicator:

$$Y_{it} = \beta_0 + \beta_1 Treatment_i + \beta_2 Post_t + \beta_3 (Treatment_i \times Post_t) + \varepsilon_{it}$$

Y_{it} is defined as being either $GDPpc_{it}$ or $\ln(GDPpc_{it})$, allowing the treatment effect to be interpreted in both raw GDP per capita units and in proportional terms. Results being presented in proportional terms in this study is useful because the treatment and control groups began at different income levels.

Second, the model is also estimated using Marshall Plan aid per capita as a continuous treatment variable rather than a binary treated/control indicator. This tests whether countries receiving more aid per person experienced larger increases in GDP per capita after 1948. The continuous treatment model is:

$$Y_{it} = \beta_0 + \beta_1 Aidpc_i + \beta_2 Post_t + \beta_3 (Aidpc_i \times Post_t) + \varepsilon_{it}$$

Together, these four models test whether the results are sensitive to the way economic recovery is measured and to whether Marshall Plan aid is treated as a binary or continuous variable.

4.2 Regression Results

Model 1: GDP per capita with binary treatment

This model estimates the effect of being classified as a substantial Marshall Plan recipient on raw GDP per capita. The key coefficient is Treatment x Post, which compares the post-1949 change in treated countries with the post-1949 change in the control group.

Results

Coefficient	Standard error	t-statistic	p-value	R-squared
1246	401	3	0.002	0.31

Statistical interpretation

The coefficient on Treatment x Post is 1,246.217, with a standard error of 401.313 and a p-value of 0.0020. The estimate is statistically significant at the 1 percent level, meaning that the probability of observing an effect this large by random sampling variation alone is very small if the true treatment effect were zero. The t-statistic of 3.105 also indicates that the estimate is large relative to its standard error. This provides strong statistical evidence that, in the raw GDP per capita model, the post-treatment change for Marshall Plan recipients was larger than the post-treatment change for the control group.

Economic interpretation

Economically, the coefficient implies that treated countries experienced an estimated 1,246 additional GDP per capita units relative to the control group after 1949. This is a large absolute difference, especially when compared with the pre-treatment control-group average. The result suggests that Marshall Plan recipients recovered more strongly in income-per-person levels than the group of non-recipient controls. However, because this model measures recovery in raw GDP per capita rather than percentage terms, the result is interpreted as evidence of stronger absolute recovery, not faster proportional growth.

Model 2: Log GDP per capita with binary treatment

This model uses the same binary treatment indicator as Model 1, but changes the dependent variable to log GDP per capita. This allows the treatment effect to be interpreted in proportional terms rather than raw GDP per capita units.

Results

Coefficient	Standard error	t-statistic	p-value	R-squared
0.011	0.09	0.12	0.9	0.30

Statistical interpretation

The coefficient on Treatment x Post is 0.0112, with a standard error of 0.0933 and a p-value of 0.9044. This result is not statistically significant at any conventional level. The coefficient is very small relative to its standard error, and the t-statistic of 0.120 shows that the estimate is extremely close to zero in statistical terms. Therefore, once the dependent variable is transformed into log GDP per capita, the regression does not provide evidence that treated countries had a higher proportional growth rate than the control group after 1949.

Economic interpretation

Economically, this result means that the stronger absolute recovery found in Model 1 does not translate into a clearly stronger percentage recovery. The coefficient of 0.0112 corresponds to a very small proportional difference, roughly around 1.1 percent, and this estimate is too imprecise to distinguish from zero. This matters because treated countries generally began at higher income levels, so a large raw GDP per capita gain may still represent a similar or even smaller percentage increase than the gain experienced by poorer control countries. Model 2 therefore weakens the claim that the Marshall Plan accelerated growth in proportional terms.

Model 3: GDP per capita with aid per capita treatment

This model replaces the binary treatment indicator with Marshall Plan aid per capita. Instead of asking only whether recipient countries recovered differently, it tests whether countries receiving more aid per person experienced larger raw GDP per capita gains after 1949.

Results

Coefficient	Standard error	t-statistic	p-value	R-squared
16	5	3	0.003	0.23

Statistical interpretation

The coefficient on Aid per capita x Post is 15.864, with a standard error of 5.310 and a p-value of 0.0029. This result is statistically significant at the 1 percent level. The t-statistic of 2.988 indicates that the positive relationship between aid intensity and raw GDP per capita recovery is not driven merely by noise in the sample. Statistically, this model supports the idea that the intensity of Marshall Plan aid is positively associated with the size of the post-treatment recovery in GDP per capita levels.

Economic interpretation

Economically, the coefficient suggests that each additional dollar of Marshall Plan aid per capita is associated with an estimated 15.864 additional GDP per capita units in the post-treatment period. This dose-response result is important because it moves beyond the simple treated-versus-control comparison and asks whether more intensive aid is connected to larger recovery. The positive and significant coefficient is consistent with the interpretation that Marshall Plan aid helped support reconstruction, investment, and production capacity in recipient countries. At the same time, this result should not be overinterpreted as a pure causal dose effect, because countries receiving more aid may also have differed in wartime destruction, industrial structure, and domestic policy conditions.

Model 4: Log GDP per capita with aid per capita treatment

This model uses aid per capita as the treatment variable and log GDP per capita as the dependent variable. It tests whether higher aid intensity is associated with faster proportional GDP per capita growth after 1949.

Results

Coefficient	Standard error	t-statistic	p-value	R-squared
0.0009	0.01	0.7	0.5	0.21

Statistical interpretation

The coefficient on Aid per capita x Post is 0.000879, with a standard error of 0.001239 and a p-value of 0.4780. The result is positive but not statistically significant. The t-statistic of 0.710 shows that the estimate is not large enough relative to its uncertainty to reject the null hypothesis of no proportional effect. Therefore, the continuous aid model also fails to provide statistically reliable evidence that higher aid intensity produced faster percentage growth in GDP per capita.

Economic interpretation

Economically, this means that the evidence for the Marshall Plan is much weaker when recovery is measured proportionally rather than in raw GDP per capita units. The positive coefficient suggests that higher aid intensity may be associated with slightly higher proportional growth, but the estimate is small and too imprecise to draw a firm conclusion. Taken together with Model 3, this suggests that Marshall Plan aid appears more clearly connected to absolute income recovery than to percentage growth. This distinction is important because richer countries can experience larger raw gains without necessarily growing faster relative to their starting level.

Overall, the sample produces a clear pattern across the four models. The two raw GDP per capita models are positive and statistically significant, while the two log GDP per capita models are not

statistically significant. The results therefore provide stronger evidence that Marshall Plan aid was associated with higher absolute recovery than with faster proportional growth.

4.3 Robustness checks

To test whether the main results depend on the specific model, sample, treatment definition, or time period, a series of robustness checks were conducted. These checks examine whether the main conclusion changes when the analysis is estimated in different ways. The central result remains that the evidence for the Marshall Plan is strongest when recovery is measured in absolute GDP per capita levels. The log models remain much weaker, meaning that the results do not consistently show faster proportional growth among recipient countries. Below is a summary of results of robustness checks. Full data tables of each check can be found in the appendix.

Country and year fixed effects

The level models remain positive and statistically significant. The binary GDP per capita coefficient is 1,246 with $p < 0.001$, and the aid-intensity GDP per capita coefficient is 15.86 with $p < 0.001$. This means that the main level result is not simply driven by permanent country differences or common year shocks.

Event-study and pre-trends test

The formal pre-trends statistic is 1.42 with $p = 0.130$. This means that there is no formal evidence that treatment and control countries were already diverging before the Marshall Plan.

Placebo treatment

Re-estimating Models 1–4 using artificial treatment dates of 1928, 1930, and 1933 produces no placebo interaction that is statistically significant at the 5 percent level[1]. The closest result is Model

[1] The placebo regressions use only observations from 1920 to 1935 and re-estimate Models 1–4 after replacing the actual post-treatment indicator with artificial post-treatment indicators beginning in 1928, 1930, or 1933. The

1 using the 1928 placebo date, which has a coefficient of 626 and a p-value of 0.095. Overall, the placebo tests provide little evidence that the baseline models were detecting treatment effects before the Marshall Plan.

Alternative pre-treatment windows

The level coefficients remain positive for 1920-1929 and 1925-1935, but the 1930-1938 model excluding Spain is weaker and no longer significant at the 5 percent level. The result is generally robust to pre-period choice, but late-1930s models are less supportive.

Alternative post-treatment windows

The level estimates remain positive across shorter and longer post-periods. The 1948-1952 estimate is weaker, while the 1949-1956 estimate is stronger. The effect appears more visible over a medium-run recovery window than during the earliest ERP years alone.

Recipient-only sample

Within aid recipients only, the continuous aid GDP per capita coefficient is positive but not statistically significant, while the log coefficient is positive and significant. The main level result becomes weaker when zero-aid controls are excluded, so the recipient-only evidence is more mixed.

Alternative treatment definitions

Any-aid treatment matches the baseline. The top-tercile aid definition remains positive and significant, while above-median and top-quartile definitions are positive but not significant. The result is not tied only to the $10 cutoff, but it is strongest for broad aid exposure and high-intensity aid definitions.

Alternative aid-intensity measures

reported coefficients correspond to either Treatment × Placebo Post in Models 1 and 2 or Aid per Capita × Placebo Post in Models 3 and 4.

Aid normalized by pre-treatment GDP per capita remains positive and significant in the models. The level result is not limited to measuring aid only in nominal dollars per capita.

Leave one country out

Dropping one country at a time produces binary GDP per capita coefficients from 996 to 1,538, and the estimate remains statistically significant in every run. Therefore, the main binary level result is not driven by one influential country.

Overall, the robustness checks support a cautious version of the main finding. The GDP per capita level results are relatively stable across fixed effects, placebo tests, alternative time windows, alternative aid measures, and leave-one-country-out estimates. The main exception is the recipient-only sample, where the raw GDP per capita result is positive but no longer statistically significant. This suggests that the control group is important for identifying the level effect, and that the evidence should not be interpreted as proving a uniform effect among all Marshall Plan recipients.

The robustness checks also confirm the distinction between absolute recovery and proportional growth as across most checks, the absolute models are more supportive than the log models.

4.4 Heterogeneity checks

The heterogeneity checks examine whether the estimated effect of the Marshall Plan differs across countries with different initial economic conditions and wartime experiences. This is done by extending the baseline DiD model to include a triple interaction between the treatment indicator, the post-treatment period, and the characteristic of interest. The coefficient on the triple interaction therefore measures whether the Marshall Plan treatment effect differs systematically across groups. The discussion below focuses on the most economically meaningful findings.

Wartime Disruption

The strongest heterogeneity results concern wartime disruption. Two alternative measures are considered. The first uses the triple interaction Treatment × Post × High Wartime GDP per Capita Decline, where countries are classified according to whether their decline in GDP per capita between 1939 and 1945 exceeded the sample median. The second uses Treatment × Post × Heavily Affected WWII Status, where countries are classified as heavily affected, occupied, or defeated during the Second World War.

One might expect Marshall Plan aid to have produced the largest benefits in countries that suffered the greatest wartime destruction because these countries faced the largest reconstruction needs. However, the estimates suggest the opposite.

For countries classified as heavily affected during the war, the coefficient on Treatment × Post × Heavily Affected WWII Status is −2,969 ($p < 0.001$) in the GDP per capita model and −0.595 ($p = 0.002$) in the log GDP per capita model. These results indicate that the estimated Marshall Plan treatment effect was significantly smaller for countries experiencing the greatest wartime disruption.

A similar pattern is observed using wartime GDP per capita decline as the measure of disruption. The coefficient on Treatment × Post × High Wartime GDP per Capita Decline is −1,493 ($p = 0.045$) in the GDPpc model, indicating a significantly smaller treatment effect among countries experiencing larger wartime output losses. Although the corresponding log model produces a negative coefficient (−0.212), it is not statistically significant ($p = 0.224$).

One possible explanation is that severe wartime damage created reconstruction bottlenecks that limited the immediate productivity of Marshall Plan aid. Destroyed infrastructure, damaged industrial capacity, housing shortages, administrative disruption, and shortages of complementary

inputs may have delayed the conversion of aid into measurable GDP per capita gains. Rather than suggesting that aid was ineffective, these findings imply that the countries facing the greatest wartime disruption may simply have required more time before reconstruction aid translated into higher economic output.

Initial Income

To examine whether the Marshall Plan had different effects depending on countries' level of economic development, the sample is divided according to median pre-treatment GDP per capita. The results indicate that countries with lower initial income experienced significantly larger post-war gains from Marshall Plan aid. This finding is consistent across both the GDP per capita and log GDP per capita models and supports the idea of catch-up growth from poorer economies following the war.

Country Size

Country size is measured using average pre-treatment population to investigate whether the effectiveness of Marshall Plan aid varied according to the scale of the recipient economy. The GDP per capita model produces a statistically significant interaction term, while the log model does not.

These results suggest some evidence that the impact of Marshall Plan aid differed according to population size, although the relationship is not sufficiently consistent across models to draw strong conclusions.

Trade Openness

Trade openness is measured using the ratio of merchandise exports and imports to GDP prior to World War II. Neither the continuous nor median-split models produce statistically significant interaction terms in either the GDP per capita or log GDP per capita regressions.

This suggests that pre-war integration into international trade did not systematically alter the effectiveness of Marshall Plan aid. Although these regressions are estimated using a smaller

sample because of limited historical trade data, there is little evidence that trade openness was an important source of heterogeneity in the impact of the Marshall Plan.

Urbanization

The final heterogeneity check considers whether the Marshall Plan had different effects depending on countries' level of urbanization before the war. The estimated interaction terms are consistently negative, implying that aid may have generated slightly larger gains in less urbanized economies. However, the coefficients are not statistically significant.

Consequently, there is insufficient evidence to conclude that pre-war urbanization systematically influenced the effectiveness of Marshall Plan aid.

5. Conclusion

This paper has examined the extent to which Marshall Plan aid contributed to post-war economic recovery. Using a Difference-in-Differences framework, it compares changes in GDP per capita among countries receiving substantial Marshall Plan aid with changes among a wider control group. The analysis also tests whether the intensity of aid mattered by replacing the binary treatment indicator with aid per capita. To account for differences in how recovery may be measured, the models use both GDP per capita in levels and log GDP per capita.

The main results show a clear distinction between absolute and proportional recovery. In the level models, Marshall Plan aid is positively and statistically significantly associated with post-war gains in GDP per capita. The binary treatment model estimates that recipient countries experienced an additional increase of approximately 1,246 GDP per capita units relative to the control group after 1949. The continuous aid model produces a similar result, with each additional dollar of aid per capita associated with a larger increase in GDP per capita during the post-treatment period. However, the log models do not produce statistically significant effects.

This means that the evidence is stronger for larger absolute gains in income per person than for faster proportional growth.

The robustness checks generally support this interpretation. The positive level results remain stable when the regressions include country and year fixed effects, use alternative treatment definitions, change the pre-treatment and post-treatment windows, apply placebo treatment dates, use alternative aid-intensity measures, and remove one country at a time. At the same time, the recipient-only model produces weaker evidence in levels, showing that the choice of control group remains important to the results. The robustness analysis therefore strengthens the case that the level findings are not driven by one particular country or one narrow model, but it does not eliminate all uncertainty surrounding the estimated effect.

The heterogeneity analysis further suggests that the impact of the Marshall Plan was not uniform across countries. The strongest differences are linked to wartime disruption. Countries that were more heavily affected by World War II, or that experienced larger wartime declines in GDP per capita, tended to have smaller estimated treatment effects. One possible explanation is that the most severely damaged economies faced reconstruction bottlenecks that delayed the conversion of aid into measurable output gains. Country size also appears to matter, with larger countries experiencing smaller absolute treatment effects, while initial income is associated with stronger gains in the level model. By contrast, the results for trade openness and urbanization are weaker and do not provide clear evidence that either characteristic systematically altered the effect of Marshall Plan aid.

These findings support a balanced interpretation of the Marshall Plan's role in European recovery. The results are consistent with the view that aid helped accelerate reconstruction and raise economic output, particularly in absolute terms. However, they do not suggest that the

Marshall Plan alone caused Europe's post-war expansion. Domestic reforms, the replacement of destroyed capital, renewed trade, technological catch-up, and broader political stabilization were also important. The lack of significant results in the log models is especially important because it indicates that recipient countries did not necessarily grow faster in percentage terms, even if they experienced larger increases in GDP per capita.

Several limitations remain. Marshall Plan aid was not randomly assigned, and treated countries differed from the control group in income levels, geography, wartime experience, and institutional conditions before treatment. The parallel trends assumption is supported only imperfectly, particularly because the treatment group suffered a sharper decline during the war. Historical GDP and aid estimates also contain measurement uncertainty, while some heterogeneity variables rely on incomplete or interpolated data. The trade-openness regressions use a smaller sample because of gaps in historical records, and the wartime-status classification necessarily involves some judgement. These limitations mean that the results should be interpreted as evidence of a strong association rather than definitive proof of a complete causal effect.

Overall, the evidence suggests that the Marshall Plan made a meaningful contribution to post-war economic recovery, but that its effects varied across countries and are most clearly visible in absolute gains in GDP per capita rather than in faster proportional growth.

References

1. Abelshauser, Werner. 1981. "West German Economic Recovery, 1945–1951: A Reassessment." *The Journal of Economic History* 41 (2): 295–312.
2. Abramovitz, Moses. 1986. "Catching Up, Forging Ahead, and Falling Behind." *The Journal of Economic History* 46 (2): 385–406.
3. Alvarez-Cuadrado, Francisco, and Mihaela I. Pintea. 2009. "A Quantitative Exploration of the Golden Age of European Growth." *Journal of Economic Dynamics and Control* 33 (7): 1437–1450.
4. Barro, Robert J., and Xavier Sala-i-Martin. 1992. "Convergence." *Journal of Political Economy* 100 (2): 223–251.
5. Baum, Christopher F., and John Barkoulas. 1998. "A Longitudinal Analysis of Macroeconomic Convergence in the European Union." *Journal of Macroeconomics* 20 (3): 511–525.
6. Casella, Alessandra, and Barry Eichengreen. 1996. "Can Foreign Aid Accelerate Stabilisation?" *The Economic Journal* 106 (436): 605–619.
7. Cowen, Tyler. 1985. "The Marshall Plan: Myths and Realities." In *U.S. Aid to the Developing World: A Free Market Agenda*, edited by Doug Bandow, 61–74. Washington, DC: Heritage Foundation.
8. De Long, J. Bradford, and Barry Eichengreen. 1991. "The Marshall Plan: History's Most Successful Structural Adjustment Program." *NBER Working Paper Series*, No. w3899. National Bureau of Economic Research.
9. Eichengreen, Barry, and Marc Uzan. 1992. "The Marshall Plan: Economic Effects and Implications for Eastern Europe and the Soviet Union." *Economic Policy* 7 (14): 13–75.
10. Giannetti, Mariassunta. 2002. "The Effects of Integration on Regional Disparities: Convergence, Divergence or Both?" *European Economic Review* 46 (3): 539–567.

11. Hogan, Michael J. 1987. *The Marshall Plan: America, Britain, and the Reconstruction of Western Europe, 1947–1952*. Cambridge: Cambridge University Press.
12. Milward, Alan S. 1984. *The Reconstruction of Western Europe, 1945–51*. Berkeley, CA: University of California Press.
13. Vonyó, Tamás. 2008. "The Post-War Growth Miracle: A Reconstruction Myth." *The Economic History Review* 61 (2): 360–384.
14. Fouquin, Michel, and Jules Hugot. 2016. *TRADHIST: Historical Bilateral Trade and Gravity Data, 1827–2014*. Paris: CEPII.
15. Clio-Infra. n.d. *Clio-Infra Database*. Amsterdam: International Institute of Social History.
16. Bolt, Jutta, and Jan Luiten van Zanden. 2024. "Maddison-Style Estimates of the Evolution of the World Economy: A New 2023 Update." *Journal of Economic Surveys*: 1–41.
17. Learn Europe. n.d. *Learn Europe: Educational Resources About the European Union*. University of Lleida. https://learneurope.eu.

Appendix

Appendix Table 1 Main and fixed-effects model results

Check	Coefficient	Standard error	t-stat	p-value	R-squared
Model 1	1,246	401	3.11	0.002	0.31
Model 2	0.011	0.093	0.12	0.904	0.30
Model 3	15.86	5.31	2.99	0.003	0.23
Model 4	0.0009	0.001	0.71	0.478	0.21
Model 1	1,246	143	8.70	<0.001	0.92
Model 2	0.011	0.022	0.50	0.617	0.96
Model 3	15.86	1.79	8.84	<0.001	0.92
Model 4	0.0009	0.0003	3.15	0.002	0.96

Appendix Table 2 Event-study coefficients

Year	Coefficient	Standard error	t-stat	p-value
1920	-1,150	454	-2.53	0.011
1921	-1,019	454	-2.24	0.025
1922	-738	454	-1.63	0.104
1923	-694	454	-1.53	0.126
1924	-550	454	-1.21	0.226
1925	-532	454	-1.17	0.241
1926	-426	454	-0.94	0.349
1927	-344	454	-0.76	0.449
1928	-274	454	-0.60	0.546
1929	-128	454	-0.28	0.779

1930	31.66	454	0.07	0.944
1931	-18.46	454	-0.04	0.968
1932	-14.72	454	-0.03	0.974
1933	6.13	454	0.01	0.989
1934	-45.97	454	-0.10	0.919
1935	0.0000	0.0000	0.00	1.000
1936	60.48	454	0.13	0.894
1937	191	454	0.42	0.673
1938	193	454	0.43	0.671
1939	510	454	1.12	0.262
1940	-72.81	457	-0.16	0.873
1941	-334	457	-0.73	0.465
1942	-566	457	-1.24	0.215
1943	-630	457	-1.38	0.168
1944	-943	455	-2.07	0.038
1945	-1,287	455	-2.83	0.005
1946	-640	457	-1.40	0.161
1947	-213	457	-0.47	0.641
1948	48.03	454	0.11	0.916
1949	350	454	0.77	0.441
1950	466	454	1.03	0.305
1951	605	454	1.33	0.183
1952	660	454	1.45	0.146
1953	898	454	1.98	0.048
1954	1,130	454	2.49	0.013

1955	1,371	454	3.02	0.003
1956	1,542	454	3.40	<0.001

Appendix Table 3 Placebo treatment-date tests

Check		Coefficient	Standard error	t-stat	p-value	R-squared
1928	Model 1	626	374	1.67	0.095	0.25
	Model 2	0.081	0.103	0.79	0.431	0.27
	Model 3	6.90	5.04	1.37	0.171	0.14
	Model 4	0.001	0.001	0.83	0.406	0.16
1930	Model 1	579	388	1.49	0.137	0.25
	Model 2	0.091	0.107	0.85	0.395	0.26
	Model 3	6.45	5.22	1.24	0.217	0.14
	Model 4	0.001	0.001	0.88	0.382	0.16
1933	Model 1	437	482	0.91	0.365	0.25
	Model 2	0.066	0.133	0.49	0.621	0.26
	Model 3	4.71	6.48	0.73	0.468	0.14
	Model 4	0.0009	0.002	0.51	0.607	0.15

Appendix Table 4 Alternative pre-treatment windows

Check		Coefficient	Standard error	t-stat	p-value	R-squared
1920-1929	Model 1	1,463	464	3.15	0.002	0.30
	Model 2	0.045	0.104	0.43	0.664	0.29
	Model 3	18.28	6.11	2.99	0.003	0.23
	Model 4	0.001	0.001	0.98	0.328	0.22
1925-1935	Model 1	1,036	453	2.29	0.022	0.30

	Model 2	-0.013	0.100	-0.13	0.894	0.30
	Model 3	13.62	5.99	2.27	0.023	0.22
	Model 4	0.0005	0.001	0.41	0.685	0.22
1930-1938 (excluding Spain):	Model 1	796	495	1.61	0.108	0.30
	Model 2:	-0.049	0.107	-0.46	0.647	0.30
	Model 3	11.25	6.54	1.72	0.086	0.22
	Model 4	0.0001	0.001	0.08	0.938	0.22

Appendix Table 5 Alternative post-treatment windows

Check		Coefficient	Standard error	t-statistic	p-value	R-squared
1948-1952	Model 1	794	439	1.81	0.071	0.28
	Model 2	-0.028	0.109	-0.26	0.797	0.28
	Model 3	11.58	5.81	1.99	0.047	0.19
	Model 4	0.0005	0.001	0.37	0.715	0.19
1949-1953	Model 1	964	442	2.18	0.029	0.29
	Model 2	-0.012	0.109	-0.11	0.914	0.28
	Model 3	12.67	5.87	2.16	0.031	0.20
	Model 4	0.0006	0.001	0.43	0.671	0.19
1948-1956	Model 1	1,154	389	2.96	0.003	0.30
	Model 2	0.002	0.090	0.02	0.981	0.30
	Model 3	15.20	5.14	2.96	0.003	0.22
	Model 4	0.0008	0.001	0.69	0.492	0.21
1949-1956	Model 1	1,246	401	3.11	0.002	0.31
	Model 2	0.011	0.093	0.12	0.904	0.30
	Model 3	15.86	5.31	2.99	0.003	0.23

	Model 4	0.0009	0.001	0.71	0.478	0.21

Appendix Table 6 Recipient-only continuous aid checks

Check	Coefficient	Standard error	t-stat	p-value	R-squared
Model 3	9.30	7.34	1.27	0.207	0.36
Model 4	0.003	0.001	2.33	0.020	0.31

Appendix Table A7 Alternative treatment definitions

Check		Coefficient	Standard error	t-stat	p-value	R-squared
Any ERP aid	GDPpc	1,246	401	3.11	0.002	0.31
	Log(GDPpc)	0.011	0.093	0.12	0.904	0.30
Above median aid among recipients	GDPpc	787	608	1.30	0.196	0.13
	Log(GDPpc)	0.030	0.142	0.21	0.831	0.11
Top quartile aid among recipients	GDPpc	1,204	760	1.58	0.113	0.14
	Log(GDPpc)	0.109	0.178	0.61	0.539	0.11
Top tercile aid among recipients	GDPpc	1,517	662	2.29	0.022	0.15
	Log(GDPpc)	0.140	0.155	0.90	0.367	0.13

Appendix Table 8 Alternative aid-intensity measures

Check		Coefficient	Standard error	t-stat	p-value	R-squared
Main aid per capita	GDPpc	15.86	5.31	2.99	0.003	0.23
	Log(GDPpc)	0.0009	0.001	0.71	0.478	0.21

Aid per 1,000 units of 1920-1935 avg GDPpc	GDPpc	63.22	23.82	2.65	0.008	0.15
	Log(GDPpc)	0.006	0.006	1.00	0.319	0.14
Aid per 1,000 units of 1935 GDPpc:	GDPpc	69.46	27.33	2.54	0.011	0.16
	Log(GDPpc)	0.006	0.006	0.87	0.387	0.15

Appendix Table 9 Leave-one-country-out summary ranges

Model	Minimum coefficient	Maximum coefficient	Country at minimum	Country at maximum	Minimum p-value	Maximum p-value
Model 1	996	1,538	Iceland	Greece	<0.001	0.017
Model 2	-0.042	0.049	Iceland	Greece	0.610	1.000
Model 3	10.24	18.75	Iceland	Greece	<0.001	0.130
Model 4	-0.001	0.001	Iceland	Greece	0.306	0.585

Appendix Table 10 Initial income heterogeneity check

Check	Coefficient	Standard error	t-stat	p-value	R-squared
GDPpc	623	139	4.49	<0.001	0.67
Log (GDPpc)	0.006	0.031	0.18	0.857	0.69

Notes: The reported coefficient is Treatment x Post x High Initial Income. High Initial Income equals 1 when a country's average GDP per capita during 1920-1935 is at or above the sample median. GDP per capita data are from the Maddison Project Database. Both regressions use the full 40-country, 960-observation sample and are estimated by OLS with conventional standard errors.

Appendix Table 11 Wartime GDP per capita decline heterogeneity check

Check	Coefficient	Standard error	t-stat	p-value	R-squared
GDPpc	-1,493	745	-2.00	0.045	0.31
Log (GDPpc)	-0.212	0.174	-1.22	0.224	0.30

Notes: The reported coefficient is Treatment x Post x High Wartime GDP per Capita Decline. The indicator equals 1 when a country's decline in GDP per capita from 1939 to 1945 exceeds the sample median. GDP per capita data are from the Maddison Project Database. Both regressions use the full 40-country, 960-observation sample and are estimated by OLS with conventional standard errors.

Appendix Table 12 Wartime status heterogeneity check

Check	Coefficient	Standard error	t-stat	p-value	R-squared
GDPpc	-2,969	806	-3.68	<0.001	0.32
Log (GDPpc)	-0.595	0.187	-3.18	0.002	0.31

Notes: The reported coefficient is Treatment x Post x Heavily Affected, Occupied, or Defeated WWII Status. The indicator is author-coded and equals 1 for countries classified as heavily affected, occupied, or defeated during World War II. Both regressions use the full 40-country, 960-observation sample and are estimated by OLS with conventional standard errors.

Appendix Table 13 Country size heterogeneity checks

Check	Outcome	Coefficient	Standard error	t-stat	p-value	R-squared
Continuous (1 SD)	GDPpc	-902	370	-2.44	0.015	0.34
Continuous (1 SD)	Log (GDPpc)	-0.141	0.086	-1.65	0.099	0.34
Above sample median	GDPpc	-2,385	837	-2.85	0.004	0.37
Above sample median	Log (GDPpc)	-0.296	0.191	-1.55	0.122	0.39

Notes: The reported coefficient is the triple interaction between Treatment, Post, and country size. Country size is the natural logarithm of average population from 1920-1935. The continuous measure is standardized, while the binary measure equals 1 above the sample median. Population data are from the Maddison Project Database, with Clio-Infra used where Maddison lacks prewar values. All regressions use 40 countries and 960 observations and are estimated by OLS with conventional standard errors.

Appendix Table 14 Trade openness heterogeneity checks

Check		Coefficient	Standard error	t-stat	p-value	R-squared
Continuous (1 SD)	GDPpc	-516	540	-0.96	0.339	0.27
	Log (GDPpc)	-0.086	0.113	-0.76	0.446	0.26

Above sample median	GDPpc	228	989	0.23	0.818	0.30
	Log (GDPpc)	-0.0002	0.211	-0.00	0.999	0.26

Notes: The reported coefficient is the triple interaction between Treatment, Post, and trade openness. Trade openness is the average of 100 x (merchandise exports + merchandise imports) / nominal GDP over usable years from 1920-1935. The continuous measure is standardized, while the binary measure equals 1 above the sample median. Data are from CEPII TRADHIST. Countries require at least eight usable prewar years, leaving 27 countries and 648 observations. Regressions are estimated by OLS with conventional standard errors.

Appendix Table 15 Urbanization heterogeneity checks

Check	Outcome	Coefficient	Standard error	t-stat	p-value	R-squared
Continuous (1 SD)	GDPpc	-863	462	-1.87	0.062	0.44
	Log (GDPpc)	-0.101	0.097	-1.05	0.296	0.53
Above sample median	GDPpc	-1,473	784	-1.88	0.061	0.35
	Log (GDPpc)	-0.213	0.177	-1.21	0.228	0.38

Notes: The reported coefficient is the triple interaction between Treatment, Post, and urbanization. Urbanization is the estimated urban population share in 1930, constructed as 0.4 times the 1900 value plus 0.6 times the 1950 value from Clio-Infra. The continuous measure is standardized, while the binary measure equals 1 above the sample median. All regressions use 40 countries and 960 observations and are estimated by OLS with conventional standard errors.