

Nguyen, Duc A.; Brakman, Steven; Garretsen, Harry; Kohl, Tristan

Working Paper

Mean Reversion of the German City System After the WWII Bombing of Cities: What Is the Mean?

CESifo Working Paper, No. 11423

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Nguyen, Duc A.; Brakman, Steven; Garretsen, Harry; Kohl, Tristan (2024) : Mean Reversion of the German City System After the WWII Bombing of Cities: What Is the Mean?, CESifo Working Paper, No. 11423, CESifo GmbH, Munich

This Version is available at:

<https://hdl.handle.net/10419/307353>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Mean Reversion of the German City System After the WWII Bombing of Cities: What Is the Mean?

Duc A. Nguyen, Steven Brakman, Harry Garretsen, Tristan Kohl

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: <https://www.cesifo.org/en/wp>

Mean Reversion of the German City System After the WWII Bombing of Cities: What Is the Mean?

Abstract

We study the post-war effects of the bombing of German cities during WWII on urban growth and use the synthetic control method (SCM) to construct comparison units for affected West-German cities. The reason to use SCM is that cities might experience structural changes that have nothing to do with the bombing of cities. Ignoring these structural changes could incorrectly attribute the decline of cities to the WWII bombing shock, while other factors are at work. The SCM takes these structural changes onboard. The synthetic units are used as counterfactuals to assess the long-run impact following the WWII bombing on the size distribution of 53 West-German cities. Our main contribution is that we do not only study whether bombed cities are mean-reverting, but also use the counterfactual to determine whether individual cities experienced a positive or negative impact. In general, we find mean reversion for 50-70% of cities, as well as a roughly balanced ratio of positively to negatively impacted cities.

JEL-Codes: R120, C930, B400.

Keywords: urban growth, synthetic control method, WWII shock, city size distribution.

Duc A. Nguyen
Ruhr University Bochum / Germany
ducanhnguyen91@googlemail.com

*Steven Brakman**
University of Groningen / The Netherlands
s.brakman@rug.nl

Harry Garretsen
University of Groningen / The Netherlands
J.H.Garretsen@rug.nl

Tristan Kohl
University of Groningen / The Netherlands
t.kohl@rug.nl

*corresponding author

October 17, 2024

The authors thank Donald Davis, Herman de Jong, Dimitrios Soudis and Marcel Timmer for helpful comments and suggestions, and Maarten Bosker for sharing the German city population data.

1 Introduction

Large but temporary shocks can have a substantial impact on urban growth. For example, [Davis & Weinstein \(2002\)](#) explored the impact of World War II (WWII) bombings on the growth of Japanese cities, [Brakman et al. \(2004\)](#) and [Bosker et al. \(2008\)](#) did so for Germany, and [Ahlfeldt et al. \(2015\)](#) for the effects of the fall of the German wall for the inner structure of Berlin. In these as well as this study, one of the central questions is whether such shocks fundamentally alter post-shock urban growth, or whether after some time these cities revert back to their mean population trajectory.

Most of these studies rely on time series/unit root methods that tests whether the population trajectory reverts back to its mean pre-shock growth trajectory, or whether it exhibits a unit root and follows a random growth path. However, a fundamental question is: To what mean is the city exactly supposed to be reverting after the temporary shock? Similarly, does the supposed population trajectory estimated by these methods reflect the one the city would have taken without the temporary shock? In the study of these temporary shocks, these questions are important since we cannot easily distinguish whether the post-WWII decline of the city population is due to the temporary (bombing) shock and/or due to other trends that simultaneously occurred during the reconstruction period.

For Germany in particular, the *Strukturwandel* is a case in point and it refers to the period of structural change through the decline of the coal industry in the Ruhr area around the 1960s. As [Glaeser \(2005\)](#) argues, there is no reason to remain in a city close to a productive natural resource once the demand for this particular natural resource declines. Such a period of structural change could very well induce a change in underlying urban growth trends. If we do not account for this potential effect, then we would incorrectly attribute the decline of cities fully to the WWII bombing shock and not (sufficiently) to the decline of the coal industry. In a similar manner, the development away from a manufacturing to a service economy such as the decline of the United States rust belt or sub-urbanization of cities ([Glaeser & Kahn, 2004](#)) come to mind as underlying developments which could induce a change in the trend.

Although basic time series methods employed in previous papers ([Davis & Weinstein, 2002](#); [Brakman et al., 2004](#); [Bosker et al., 2008](#)) can, to an extent, account for it by fitting trend terms that encompass both the pre- and post-shock period, they alone do not distinguish between the population effects induced by the temporary shock and these possible other changing trends.

In order to address this, we use the synthetic control method (SCM) ([Abadie et al., 2010](#); [Abadie, 2021](#)) to construct counterfactual cities that are absent of the temporary shock but can take into account these other changing trends. SCM does this by constructing the counterfactual city, the *synthetic control* itself, from a weighted sum/convex combination of cities chosen from a set of ‘untreated’ donor cities. Then, similar to the growing literature following [Harvey & Thiele \(2021\)](#), we combine the usual time series methods that can test for mean reversion with SCM, where the counterfactual allows us to abstract the temporary shock from other changing trends. We apply this method to the WWII bombings of German cities to evaluate whether German cities revert to their ‘synthetic mean’.

This combination of methods also keeps us as close as possible within the time series framework widely employed within the literature, which in turn allows us to better reconcile the previous, contradicting results that concludes post-WWII mean reversion of German cities ([Brakman et al., 2004](#)) against the opposite conclusion of no mean reversion ([Bosker et al., 2008](#)) and a mostly negative impact on German cities.¹

¹With simultaneous non-linear trends, it may lead to an overacceptance of the unit root null in models that consider a

This study mainly contributes to the strand of empirical research on the effect of a large, temporary shock on a system of cities (Davis & Weinstein, 2002; Brakman et al., 2004; Bosker et al., 2008; Miguel & Roland, 2011), with the aim of providing a stronger empirical method to distinguish the large, temporary shock from any other parallel development happening. Note that the results of this particular strand of research also has implications on the relevance of some competing fundamental theories that explain the location of economic activity. Davis & Weinstein (2002) define those theories as locational fundamentals theory or ‘first-nature geography’ (e.g., Rappaport & Sachs, 2003), random growth theory (e.g., Gabaix, 1999) and increasing returns theory (following Krugman, 1991).

This study is also related to those on multiple equilibria (Bosker et al., 2007; Davis & Weinstein, 2008; Redding et al., 2011; Bleakley & Lin, 2012), although the existence of multiple equilibria is not our focus. To a lesser extent, our research is also related to the disaster literature (duPont IV & Noy, 2015; Siodla, 2015), with the difference that they look at natural disasters typically affecting a single city and not a city system.²

Our results are as follows. First, we find that the majority of (West-)German cities recover to their supposed, synthetic mean in size and city share. We also find that the synthetic city exhibits an inverse-U shaped population trajectory during the 20th century. For those cities where we find a negative impact after WWII, we find that they are on average smaller and situated in less agglomerated areas, which hints at a better ability of larger cities to recover. The remainder of this study is structured as follows: Section 2 discusses the construction of our synthetic controls. Section 3 describes our baseline empirical strategy, Section 4 reports the results, Section 5 discusses the implications of the new findings with existing theories, and Section 6 concludes.

2 Synthetic Cities

2.1 Methodology

The construction of synthetic controls is well-established and we follow the procedure and notation of Abadie et al. (2015). To summarize this method: In the first step, we first choose covariates that we think are good predictors of the outcome variable (city population), such as past population or coal worker population, where SCM then gives more weight to covariates that have a stronger predictive power on the outcome variable.³ In

linear trend (Perron, 1989; Taylor et al., 2001). Note that Bosker et al. (2008) used a sharp structural break while detrending through linear trends i.e. allowing for a kink, which they found to be around WWII. Initially, they failed to reject the unit root hypothesis for all German cities with the population linear-trend stationary model, but once allowing for a structural break, they reject the unit root hypothesis for about 75 percent of cities.

²This distinction could be important in that shocks on the city system may induce relatively little migration given the magnitude of the shock if most cities simultaneously experienced some form of destruction. This gives room for the conjecture that as long as the large, temporary shock on the city system is even, it does not induce migration whereas a shock on a single city in a city does lead to migration. This notion could explain why we overall find a permanent impact in the disaster literature concerning a single city, while mean reversion if the disasters affect the whole country.

³In choosing covariates, the literature follows the principle that it should be a good predictor of the variable of interest. Hence, the pre-treatment lagged outcome variables are included as covariates. However, in our case, we should not include all pre-treatment lags of the outcome as covariates for two reasons. First, the SCM chooses V such that the synthetic and actual population in the pre-treatment years are as close as possible, rendering other covariates irrelevant in the matching process (Kaul et al., 2015). As will become clear later, adding all pre-treatment lags will ignore the “specialized” cities that we try to match with their synthetic specialized counterpart(s) through the sea, river and coal covariates. Second, we also want the covariates closer to the WWII treatment period, as covariates further away may have lower predictive power on the post-WWII period, including the lagged population covariates.

the second step, out of a donor pool of untreated/comparison cities, SCM constructs a large number synthetic cities for the treated city, which are each simply a different convex combinations of untreated/comparison cities. Out of these, SCM chooses the ‘best’ one with lowest pre-treatment mean squared prediction error, which then constitutes our synthetic city for the given treated city.⁴

Formally, the synthetic German city, indexed by i , is represented by a convex combination of donor pool cities as indexed by j . Therefore, we have a sample of $I + J$ individual cities, with I number of ‘treated unit’ German cities, and J ‘comparison units’ donor cities in the ‘donor pool’. Hence, the donor pool is $j = I + 1, \dots, I + J + 1$.

For simplicity, we omit the individual city subscript i for each variable. The synthetic control for the individual city is a convex combination of cities in the donor pool, i.e., it can be represented by the $(J \times 1)$ vector of weights $W = (w_{I+1}, \dots, w_{I+J+1})$ for each city, where $0 \leq w_j \leq 1$ and $w_{I+1} + \dots + w_{I+J+1} = 1$. Let X_1 be a $(K \times 1)$ vector containing the pre-treatment covariate of a treated city and let X_0 be a $(K \times J)$ matrix with the same pre-treatment covariate from the donor pool. In the following two-step procedure, we minimize the discrepancy between X_1 and X_0W ,

$$\|X_1 - X_0W\| = \sqrt{(X_1 - X_0W)'V(X_1 - X_0W)} \quad (1)$$

Conditional on V , the synthetic control $W^*(V) = (w_{I+1}^*, \dots, w_{I+J+1}^*)$ is selected to minimize Equation 1 subject to $w_{I+1}^* + \dots + w_{I+J+1}^* = 1$ and $0 \leq w_j^* \leq 1 \forall j$. V is a $(k \times k)$ diagonal matrix with non-negative entries, where its values weight the relative importance of each pre-treatment covariate.

Equivalently, let X_{0jm} be the value for the m^{th} covariate for donor pool unit j and X_{1m} be the value for the treated unit covariate, where $m = 1, \dots, K$. We choose W^* to minimize

$$\sum_{m=1}^K v_m \left(X_{1m} - \sum_{j=I+1}^{I+J+1} w_j X_{0jm} \right)^2 \quad (2)$$

where v_m are the diagonal entries of V , i.e. $V = \text{diag}(v_1, \dots, v_K)$. Following [Abadie & Gardeazabal \(2003\)](#) and [Abadie et al. \(2010, 2015\)](#), we choose V based on minimizing the mean squared prediction error (MSPE) of the city population outcome variable S given the chosen W^* :

$$\sum_{t=1}^{T_0} \left(S_{1t} - \sum_{j=I+1}^{I+J+1} w_j^*(V) S_{jt} \right)^2 \quad (3)$$

where T_0 is the treatment period. Summarizing, we first calculate $W^*(V)$ for any V in the first step, and then choose the V that gives us the $W^*(V)$ with the lowest MSPE, which constitutes the synthetic city. From these weights, we can draw the synthetic population $S_{i,t}^{synth}$, which we will then combine within a unit root framework in Section 3.2.

⁴See [Abadie \(2021\)](#) for a comprehensive guide and discussion on the use of the synthetic control method. SCM has recently also been applied in, for example, regional science ([Adjei et al., 2023](#)), comparative economic history ([Gilchrist et al., 2023](#)) and criminology ([Rydberg et al., 2018](#)).

2.2 Dataset

For our outcome variable of interest, city population, we build on the original German city dataset of [Brakman et al. \(2004\)](#) and [Bosker et al. \(2008\)](#) covering cities with at least 50,000 inhabitants at some point in the 1880-1990 time period. Our final sample of treated, i.e. bombed cities, includes 53 West-German cities. The cities in the donor pool must be comparable to German cities, but should not have been affected (i.e., damaged or destroyed) by WWII. For this reason, the cities in our donor pool are from those European countries that remained neutral during WWII: Switzerland, Sweden, and Portugal. We also include Denmark because its cities came out unscathed during the “Six Hour War”.⁵ Importantly, these neutral countries lack coal fields, making it unlikely that their cities would reflect the German *Strukturwandel* decline of the coal industry in the 1960s. In contrast, Great Britain has large coal fields in its industrialized regions mainly around the Midland regions. Given data availability limitations, we add English and Welsh cities.

Imposing the requirement that a donor city must have at least 50,000 inhabitants at some point between 1880-1990, the final sample of 144 donor cities includes 16 Swiss cities ([Bundesamt für Statistik, Schweiz, 2016](#)), 38 Swedish cities ([Swärd, 2016](#)), 30 Portuguese cities ([Instituto Nacional de Estatística, 2011](#)), 10 Danish cities ([Danmarks Statistik, 1985](#); [Danmarks Statistik, 2011](#)), and 50 cities in the United Kingdom ([GB Historical GIS / University of Portsmouth, 2021](#)). All our city data are observed between 1880-1990 at ten-year intervals.⁶

Given that specialized cities could have developed differently from other cities irrespective of the WWII shock, we also collect data to construct a coal covariate. This is largely motivated by the decline of the coal industry in the Ruhr area and the UK, which started around the 1960s. For Germany, we obtain the number of workers in the coal industry from [Fischer \(1989, 1995\)](#) on a *Regierungsbezirk* (administration area) level in Germany. Similarly, we have historic county-level data on the number of coal workers employed for both England and Wales from UK commerce reports [Dye \(1921\)](#). There was no significant coal deposit in any other country, or large industrial areas for that matter. There is a large iron industry in Sweden, but it is located in the remote and sparsely populated Kiruna. The iron industry in that area did not decline over the course of the 20th century.

As we will highlight in the next section, most synthetic cities exhibit an inverse U-shaped trajectory. This development may be attributable to post-war suburbanization. Furthermore, we have vastly different population trajectories of coal vs non-coal cities, as the often rapidly industrializing coal cities had much higher growth rates in the pre-WWII period than non-coal cities. The purpose of the coal covariate should be to capture the ‘extra’ decline of the population due to that decline in coal, apart from any other occurring declines. Second, the main industrial areas are found within or nearby coal regions, so that a coal covariate should also cover the decline of manufacturing cities or areas of deindustrialization. Indeed, the industrial heartland of both Germany and the United Kingdom, the Ruhr and the Midland area respectively, are close to or within major coal fields.

Moreover, the population trajectory may already capture the coal covariate, as industrializing cities

⁵We do not include Belgium, France, Ireland, the Netherlands, or Norway due to lack of census data at 10-year intervals in the 19th century.

⁶Due to possible changes in municipal borders over time, we opted for the constant border population estimates for European cities by [DG REGIO \(2021\)](#), available for 1960 onwards. We have also looked up whether there were any municipal border changes before 1960 for each city in our dataset and adjusted these population numbers if possible to allow constant borders throughout.

around the beginning of the 20th century attracted notably more workers, that the lagged population covariates would capture as well. However, we may have a city in the donor pool that exhibits a very similar pre-war trajectory as these coal cities but are not subject to that decline. Thus, if the SCM would consider these cities to match coal cities, we may underestimate the counterfactual post-war decline. The idea of the coal covariate is that the SCM will give ‘preference’, given equally well-fitting pre-war population trajectories, towards matching cities within coal regions with coal region synthetic cities specifically, and vice versa cities within non-coal regions with non-coal region synthetic cities. For that reason, we consider the coal covariate important to include as it may capture the ‘extra’ population decline, apart from the ones induced by suburbanization in general.

For similar reasons as the coal covariate, we also include a river and sea-access dummy variable to account for elements of first-nature geography may be conducive to adjustments of international trade over time. We consider a city to be near a river or sea if today’s border is 1 kilometer away from either. These may capture the ‘extra’ urban population increase if trade became more important in the post-WWII era.

Finally, we construct a simple measure of 1940 city-level urban potential (within national borders), where $\tau_{i,j}$ is the kilometer distance between city i and j , and $S_{i,1940}$ is the population of city i in year 1940:

$$UP_i = \sum_j^n \frac{S_{i,1940}}{\tau_{i,j}}. \quad (4)$$

2.3 Descriptive Statistics

Table 1 visualizes how synthetic cities are constructed with three arbitrarily chosen examples for the cities of Aachen, Lübeck and Düsseldorf. For Aachen, the “Treatment” column shows the observed values for 1900, 1920, 1930 and 1940 population, that it is within 1 kilometer from a river or sea, and the 1920 value of coal workers and its 1940 urban potential measure. The “Synthetic Aachen” column shows that it is constructed based on data from six donor cities and their respective weights. The discrepancy between the actual and synthetic population levels is very low, but the discrepancy for the coal worker covariate is relatively high. As discussed in the previous subsection, this may be due to county-level coal worker counts between regions not being directly comparable, or because we have cities that are far away from coal fields included within coal counties.

For Lübeck, the synthetic values are very much consistent with the observed values. A similar conclusion holds for Düsseldorf. Here, note that the synthetic control method does not match the city with any coal region city. Indeed, we would not think of Düsseldorf as a typical city that is characterized as a coal area city, as it is close but not within the Ruhr coal area. Coal workers assigned at a *Regierungsbezirk* level may then wrongly include cities such as Düsseldorf as a coal area. However, the fact that Düsseldorf is not matched with any coal area counterfactual cities is reassuring, as it indicates that most weight in the matching process is given towards the lagged population covariates. This may, as said, be due to the different population trajectories of coal and non-coal cities, as the often rapidly industrializing coal cities had much higher growth rates in the pre-WWII period.

More generally, Table 2 presents summary statistics on the 53 German sample cities, the summary population statistics for 53 synthetic cities matched on covariates, and summary statistics for the 143 cities in the donor pool.

3 Empirical Strategy

3.1 Examining Non-Linearities

To motivate our approach, we first determine whether synthetic city population growth displays evidence of non-linearity. Consider the behavior of the synthetically estimated population growth of German cities over time in the following quadratic model:

$$S_{i,t}^{synth} = \alpha_i + \beta_{1,i}T + \beta_{2,i}T^2 + \epsilon_{i,t} \quad (5)$$

with time trend T and where $S_{i,t}^{synth}$ is the synthetically estimated city population for the corresponding German city i at time t . We construct synthetic cities based on the following covariates: 1900 and 1920-1940 lagged population, the 1920 number of coal workers, the dummies for sea or river access, and the simple 1940 urban potential measure.

We find evidence of an inverted-U shape of population growth for 42 out of 53 synthetic cities, i.e., $\beta_{1,i} > 0$ and $\beta_{2,i} < 0$. Out of these 42 cases, 36 have statistically significant parameter estimates at the 5 percent level. The average maximum of these 36 cities is around the year 1979.⁷ These results indicate that there are underlying systemic non-random developments at play which influence the majority of city growth patterns in our sample, and that these developments are not initiated by a large, temporary WWII bombing shock. This result further justifies our approach to accommodate for a possible non-linear counterfactual.

3.2 Unit Root Tests

We now turn to our main empirical specification. To evaluate whether the bombing shock exerted any permanent impact on German cities' population growth after WWII, or whether these cities eventually reverted to their mean population trajectory. To that end, we test for mean reversion of the city population with an augmented [Dickey & Fuller \(1979\)](#) unit root test:

$$\Delta S_{i,t} = c_i + \delta_i t + \rho_i S_{i,t-1} + \sum_{k=1}^p \beta_{i,k} \Delta S_{i,t-k} + \epsilon_{i,t} \quad (6)$$

for the city size $S_{i,t}$ for city i at time t .⁸ Finding $\rho_i = 0$ yields a unit root, which implies that the city population trajectory follows a random walk. However, if $\rho_i < 0$, we will find a trend-stationary process so that the city population growth is linear trend mean-reverting after the bombing shock.

We obtain the synthetically corrected (SC) form as the difference of the actual individual city population $S_{i,t}$ with its corresponding synthetic city population $S_{i,t}^{synth}$, thus “detrending” it as follows:

$$S_{i,t}^{SC} \equiv S_{i,t} - S_{i,t}^{synth}. \quad (7)$$

Note that Equation (7) is equivalent to the cointegration relation of $S_{i,t} = \alpha_i + \gamma_i S_{i,t}^{synth} + u_{i,t}$ by imposing $\alpha_i = 0$ and $\gamma_i = 1$, so that $S_{i,t}^{SC} = u_{i,t}$. In this case of known coefficients, the usual Dickey-Fuller

⁷If we consider 1920 to 1990, we find that 46 out of 53 cities fit the same pattern, out of which 32 have parameter estimates significant at the 5 percent level. The average maximum of these 32 cities is around the year 1977.

⁸Note the close resemblance of this specification to the main empirical specifications in [Davis & Weinstein \(2002\)](#); [Brakman et al. \(2004\)](#); [Bosker et al. \(2008\)](#) and [Miguel & Roland \(2011\)](#).

critical values apply. However, we are only interested in whether the difference between the actual ($S_{i,t}$) and synthetic ($S_{i,t}^{synth}$) series is mean-reverting, and not whether there is a α_i and γ_i such that two I(1) series are cointegrated. Hence, we are not interested in the order of integration. Then, equivalent to Equation (6), we write:

$$\Delta S_{i,t}^{SC} = \rho_i S_{i,t-1}^{SC} + \sum_{k=1}^p \beta_{i,k} \Delta S_{i,t-k}^{SC} + \epsilon_{i,t}^{SC}. \quad (8)$$

That is, as we detrend through the corresponding synthetic city for each German city, we drop the linear-trend term $\delta_i t$. By construction, the values for $S_{i,t}^{SC}$ should be close to 0 before the war and are allowed to diverge from 0 after the war. Therefore, we also drop the constant c_i in Equation (6) and consider the Dickey-Fuller critical values without a constant/drift in the following to be appropriate. The interpretation is similar as above, with the difference that we estimate whether cities return to their supposed mean based on their synthetic counterpart. Finding that $\rho_i = 0$ implies a unit root in the population trajectory after detrending with the synthetic city. In contrast, $\rho_i < 0$ implies reversion to the synthetic mean population growth trajectory after a negative divergence, i.e., no permanent effect of the bombing shock on urban growth.

4 Results

4.1 Impact on City Size

We start with a visual assessment of how the bombing shock affects the gap between observed city-level population and the synthetic control city population, i.e., $S_{i,t}^{SC}$ from Equation (7). This result is plotted in Figure 1. As expected, the gap is close to 0 for almost all cities between 1920-1940 by construction, due to our choice of lagged covariates. The orange dashed linear fit from 1900-1940 is close to 0, which reaffirms our exclusion of the trend and constant terms in Equation (8). Note that the yellow dashed linear fit for 1960-1990 is also close to 0.⁹ From Figure 1 we conclude that on average, there is no systematic difference between German city population and its synthetic counterpart in the decades before or after WWII.

Figure 2 depicts a placebo test, where we construct synthetic counterparts for the 56 cities in the donor pool that have a weight of at least 0.05 as a donor city in Figure 1. We show the gap of these individual comparison cities with their synthetic control, where the donor pool is simply the same donor pool as before with the exception of the comparison city itself. We exclude Birmingham because there are no city comparable in size that we can use as a counterfactual. That the linear fit is a horizontal line around 0 is reassuring because this is what we would expect. After all, the donor pool cities did not experience a large, temporary shock on their city system as did Germany, so the population gap measure should (on average) be 0.

Table 3, panel A, shows the results of the city-level Dickey-Fuller tests on the detrended population as in Equation (8). For the cutoff year 1990, we reject the unit root null for 24.5% of the cities at the 5% level. We reject a similar share of cities at the 5% level for the cutoff years 1980 and 1970. But, for the cutoff year 1960, we reject the unit root null for 54.7% of cities at the 5% level, and even 66% at the 10% level. The clear pattern is that an earlier cutoff year leads to a higher rejection of the unit root null hypothesis. A

⁹Following Davis & Weinstein (2002) and Brakman et al. (2004), we ignore the WWII-period and set the cutoff at 1960.

reason for this may be that the synthetic counterfactual is less accurate, the further away we move from the treatment period. Nevertheless, the share of cities with rejected unit roots remains remarkably stable over time: the majority of cities revert to their mean population trajectory after the bombing shock.

The lower part of Table 3, Panel A shows that around half of the cities with a unit root also have a larger population than suggested by their synthetic counterpart. This may indicate that the share of rejected unit roots is underestimated if the counterfactual is less accurate over time. This stands in contrast with the, mostly negative, structural breaks around WWII in Bosker et al. (2008).¹⁰ They also found a unit root for all cities when not considering a structural break, which is consistent with the conventional linear detrended approach results in Table 3, Panel A. Here, we reject the unit root null for only 5.7% of the cases at the 5% level with a cutoff year of 1990 or 1980. Overall, we consider this as strong support for the notion of overacceptance of the unit root null in the conventional linear detrended approach due to simultaneous non-linear adjustments.¹¹

For the sake of completeness, Table 3, Panel B also includes some panel unit root tests similar to Bosker et al. (2008) if one is concerned about the low power of individual unit root tests with few observations. Yet, as Perron (1991) notes, a short time span would be problematic and not the frequency of the observations as such. Nevertheless, we include the Levin et al. (2002) test, with the null of all panels containing the unit root against the alternative of all panels being stationary with the same autoregressive parameters. For both the synthetic approach (panel B.a) and linear detrended approach (panel B.b), we reject the null of all panels containing a unit root.

A somewhat less restrictive test that allows for different autoregressive parameters between cities is Im et al. (2003), which tests the null of all panels having a unit root against the alternative of some panels being stationary and some having a unit root. This test however does not allow for suppressing the constant/drift term. Hence, we decided not to test the synthetically detrended approach with a constant, as it would artificially inflate the critical values to overaccept the null hypothesis for the synthetically detrended approach. This is because the constant is by construction set to 0 in the pre-war period. For the linear detrended approach, we fail to reject the null of all panels having a unit root until 1980, which is more in line with the corresponding findings for individual cities. Hence, it seems that allowing for different autoregressive parameters matters. However, the conclusions drawn seem to be sensitive to the cutoff year.

As a robustness exercise, we may think of taking a log transformation of $S_{i,t}$ and $S_{i,t}^{synth}$ in Equation (6) and Equation (7), similar to Bosker et al. (2008). Taking the log for synthetic controls would not be needed as the deterministic mean is already taken away. Instead, a log transformation would de-emphasize the divergences of the gap of actual and synthetic population that happens after the treatment, as seen in Figure 1. In Table 4, Panel A, we report the same (log-transformed) specification as in Table 3, Panel A. As compared to Table 3, Panel A, we generally reject the unit root null more often by around 15 percentage points for all specifications, consistent with the idea that such a transformation would decrease the fanning out after the treatment in Figure 1. Whether this is the preferred specification is not obvious, but it may reduce the likely inaccuracy of the synthetic control the further we move away from the treatment period.¹²

¹⁰Out of 62 cities, they find 17 cities with a structural break around WWII, out of which 15 have a negative structural break. For the other cities, they find a unit root.

¹¹We find slightly lower rejection rates of the unit root null when repeating the exercise of Table 3, Panel B on the synthetic counterpart series instead of the actual population. These results are available from the authors upon request.

¹²The results presented in this subsection are robust to reducing the threshold of deaths per capita during the Blitz to 0.001: doing so further raises the proportion of cities with mean reversion. Results are available from the authors upon request.

Finally, [Bosker et al. \(2008\)](#) also looked at the evolution of the city size distribution. They found a more even city size distribution i.e. relatively more middle sized cities. However, they were only able to compare the post-war distribution to the pre-war distribution. We think that we are now in a better position for this comparison, since we have an actual counterfactual in the synthetic control, so we can compare the actual city size distribution with the synthetic city size distribution over time. For that, we will perform the two-sample Kolmogorov-Smirnov test on the actual and the synthetic city size for each year. We report the results in Table 5. For each year, we fail to reject the null that both distributions are the same. Note that we more strongly reject the null between 1920 and 1940, which is by construction due to the synthetic control lag covariates we included in those years. Also note that the p-value declines after the war over time, which may indicate that the effect of the bombings on the city size distribution did not fully dissipate yet and that we may reject the null if we extend the cutoff year. This would go against the very idea of having a cutoff year. Furthermore, given 33 cities out of 53 in our sample were above to the synthetic city population already in 1960, and heavily based on our estimate of this section, it seems unlikely that it would take another 50 years for the effect of WWII to fully dissipate.

To conclude this section, despite the widespread bombings on German cities, the city size distribution did not diverge from its supposed counterfactual. Instead, the distribution change results of [Bosker et al. \(2008\)](#) appear to be driven by underlying developments that are not captured by their method and not necessarily by WWII, as well as an overacceptance of the unit root null. This does not suggest that there was no permanent effect of the WWII bombings, but it did not induce a change in the city-size distribution.

4.2 Impact on Relative City Size

[Davis & Weinstein \(2002\)](#) also looked at the impact of WWII bombings on the *relative* city size of Japanese cities on total population, which is also done in [Brakman et al. \(2004\)](#) and [Bosker et al. \(2008\)](#) for the impact of WWII on the relative city share of German cities among the total population. This comparison is relevant as it allows us to specifically look at whether the relative population of a city permanently changed its position within the city size distribution.

In their analysis, [Davis & Weinstein \(2002\)](#) concluded that after the WWII bombings, the relative city size in Japan reverted to the mean (cutoff year: 1963). [Brakman et al. \(2004\)](#) confirm this result for West-Germany (cutoff year: 1965). In contrast, [Bosker et al. \(2008\)](#) concluded that there was no reversion to the mean for the relative city size (cutoff year: 1999). That the WWII shock did not affect relative city size two decades after WWII, but only later after five decades, seems to be a peculiar result. Yet, we can reconcile these findings by allowing for non-linear mean reversion, especially as specialized cities may have developed differently to others that could have driven the relative city size distribution irrespective of the WWII shock.

We follow [Brakman et al. \(2004\)](#) by assuming that bombed cities were (eventually) targeted at random, irrespective of their size, military importance or industrial capacity.¹³ With the total West-German population¹⁴ $S_{i,t}^{total}$, the (synthetic) city share is:

¹³While targets were initially assigned based on their military relevance, this strategy proved unsuccessful in limiting war production. Area bombing was subsequently adopted as the new strategy, targeting both industrialized areas and also easier to hit historical city centers, destroying the housing stock and disrupting overall production by dislocating workers and lowering morale. Another criterion was the visibility from the air (weather conditions, landmarks), susceptibility to incendiary bombs ([Friedrich, 2002](#)), and distance to Britain as earlier aircraft could not reach more Southern areas.

¹⁴We consider the territory of 1957-90 and keep it constant for the total population between 1880-1990.

$$s_{i,t} \equiv \frac{S_{i,t}}{S_{i,t}^{total}}, \quad (9)$$

$$s_{i,t}^{synth} \equiv \frac{S_{i,t}^{synth}}{S_{i,t}^{total}}. \quad (10)$$

Note that we consider the synthetic city share to be the ratio of the same synthetic control as in the previous section, and the actual total West-German population $s_{i,t}^{total}$. In line with our earlier approach, the synthetically detrended city share is:

$$s_{i,t}^{SC} \equiv s_{i,t} - s_{i,t}^{synth} \quad (11)$$

Then, equivalent to Equation (8), we obtain:

$$\Delta s_{i,t}^{SC} = \rho s_{i,t-1}^{SC} + \sum_{k=1}^p \beta_{i,k} \Delta s_{i,t-k}^{SC} + \nu_{i,t}^{SC} \quad (12)$$

Table 6 reports the results of the relative city size. Compared to the absolute size effect in section 4.1, we find a higher rejection rate of the unit root null for the relative share of the population. Second, we again find that the percentage of cities with a unit root and for which we find the population to be above the corresponding synthetic population to be slightly below 50% for all cutoff years. Third, depending on the cutoff year, we find a higher to much higher rejection rate of the unit root null compared to the conventional linear detrended approach. For example, we find mean-reversion for 71.7% of cases at a 5% level in cities with a cutoff of 1960, and 69.8% of cases for 1970. In contrast, we only find mean reversion for 13.2% and 15.1%, respectively, for the corresponding linear detrended approach.

For our preferred cutoff year 1980, which we can consider to be late enough to register any permanent effect after WWII, we still find at a 5% level that 49.1% of the cities are mean-reverting to the synthetic counterfactual as compared to 22.6% with the linear trend method. Given the balanced positive to negative impact within unit root cities and because the synthetic counterfactual is likely to be less accurate for later cutoff years, our results suggest that the city share reverted to a non-linear mean for most cities. In Table 7, we show some of the the pre-war characteristics of cities for which we find either a stationary series (below a 5% significance level), or otherwise a positive or a negative impact unit root of the city share with a cutoff year of 1980. Additionally, Figure 3 shows the spatial distribution of these cities.

On average, cities for which we find mean reversion or a positive impact in the city share were larger before WWII than those negative impact unit root cities. These negative impact cities seem to have been more remote from other larger cities than mean-reverting or positive impact cities, as judged by the urban potential measures in 1939. Figure 3 shows that remote cities have been negatively impacted. Interestingly, cities in the Ruhr area industrial heartland of Germany (mid-western part of the map), which is also the largest agglomeration of cities, mostly recovered or were positively impacted in 1980 (except Herne). The same holds for most of the Rhine-Main/Neckar region (to the south-west around Frankfurt down to Mannheim).

Furthermore, positive impact cities have experienced on average significantly less wartime destruction, as measured by the rate of housing lost of total housing, as well as the cubic meter rubble per 1939 capita.¹⁵

¹⁵Both data are the same as in Brakman et al. (2004), taken from Kästner (1949).

However, there is no significant difference in wartime destruction between stationary and negative impact cities; the difference is that stationary cities had a higher rate of refugees in 1960 as a percentage of population compared to negatively impacted cities. This may also reflect refugees preferring larger cities or agglomerations to settle in.¹⁶ There also are no significant differences between being located in a coal region when it comes to the differential impacts.

Overall, it seems that there is little difference in wartime destruction between cities that were stationary, and those for which we find a permanent negative impact. The discrepancy in whether a city recovers or not seems to be driven by other factors, such as the larger population size or (being part or near to) agglomerations. However, the positive impact cities also experienced less wartime destruction, which is in line with expectations.

What can we conclude here? We find mean-reversion for the majority of city shares throughout all cutoff years, which is very much in line with previous findings that considered a shorter cutoff year in the 1960s (Davis & Weinstein, 2002; Brakman et al., 2004). We can reconcile these findings with the results of a unit root/negative structural breaks concluded by Bosker et al. (2008), who considered a later cutoff year in 1999. With later cutoff years, it opens up the possibility that non-linearities could have driven the results that lead to overacceptance of the unit root null. Secondly, trying to estimate the deterministic mean through non-linear methods would make interpretation of the results difficult to attribute to either non-linearities or the WWII shock. Arguably, our approach reconciles the literature by distinguishing between non-linear developments and the long-term impact of the WWII bombings.

5 Implications on Principal Theories

We can shed some more light about the links between the evidence and the three principal theories as posited by Davis & Weinstein (2002), namely locational fundamentals theory that considers fundamentals of a location, such as river or sea access, as the main driver; increasing returns theory following Krugman (1991), and finally random growth theories that can generate a Zipf distribution (e.g., Gabaix, 1999). With random growth, urban growth is scale-independent so that the city size distribution does not change, even if the rank of individual cities within this distribution may change. However, the usual version of testing random growth on the German dataset with more cities and a higher frequency of observation has been done already by Bosker et al. (2008), so that we will not repeat that exercise again. They find that evidence for independence of size and growth.¹⁷

¹⁶Note that initially after WWII, the French occupied zone restricted the number of refugees. They only received 60,000 at the end of 1947, around 1% of the total population, compared to the total of 4.379 million in the Soviet zone (around 24.3% of the total at that time), 2.957 million in the US zone (around 17.7%) and 3.320 million in the British zone (around 14.5%) occupied zones (Volkman, 1995), which more than compensated for the total German deaths during WWII. Yet, even though refugees were hindered initially to settle in the French occupied zone, we still find that the average rate is 11% in 1960 for the 6 French occupied cities in our sample, Freiburg, Koblenz, Kaiserslautern, Ludwigshafen, Mainz and Trier. In fact, the other city not making it below 12.5% are Freiburg and Kaiserslautern with 12.8% and 12.6%. It seems not likely that differences in impact is driven by refugees not being allowed in one zone, but in others at least in the long-term.

¹⁷They also test for Gibrat's law, but reject this law based on their findings of a change in the city size distribution; We do not find a significant difference in the city size distribution compared to the synthetic distribution. Furthermore, the inverse-U shaped development of synthetic cities would suggest an over acceptance of the unit root null, and thus, of Gibrat's law if the expected mean is not allowed to change accordingly. There is no natural explanation within random growth theory that would explain this non-linearity. Gabaix & Ioannides (2004) argued for a weakened version of Gibrat's law that would allow for possible non-linearity in the mean reverting component.

The inverse-U shaped development of synthetic cities suggest underlying non-random developments possibly explaining this non-linearity. These patterns are consistent with locational fundamentals changing over time, for example through technological change, see also [Michaels & Rauch \(2018\)](#) and [Bleakley & Lin \(2012\)](#) for a discussion. As an example, motorization would change how people interact with the existing road and street network that made living in the city center relatively less attractive by reducing commuting costs (see [Baum-Snow, 2007](#); [Baum-Snow et al., 2017](#)); this is similar to the impact of steamships that made living close to ports attractive. So, changing locational fundamentals can be compatible with a reversion to the (synthetic) mean even if the fundamentals themselves are not altered by the bombing.

Finally, increasing returns theory explain agglomeration by the interaction and proximity of people. Theories in this manner leave the possibility for a permanent effect from a large, temporary shock, a shift in equilibria, and a change in the shape of the city size distribution open. However, that the cities for which we find a permanent negative impact were smaller and more remote on average in 1939 and vice versa, positively impacted cities were larger and less remote can be explained by increasing returns, where the impact of WWII could have acted as an accelerator of agglomeration. This is not consistent with the independence of growth, while changing locational fundamentals do not offer a plausible explanation for this. Hence, we tend to consider increasing returns as the more convincing explanation.

Where does this leave us? Similar to [Davis & Weinstein \(2002\)](#), varying locational fundamentals seem to fit the synthetic city development. However, for the negative impact on smaller cities and the positive impact on large cities in the post-WWII period, increasing returns is the more plausible driver.

6 Conclusion

We began by outlining the potential problems that could arise from not separating the WWII shock from other underlying developments, such as motorization, suburbanization, and the decline of coal in the 1960s-70s that could explain the non-linearity of the population trajectory of those not affected by WWII bombings. Not accounting for these non-linearities may lead to an overacceptance of the unit root, as well as potential misattribution of an effect between the WWII shock and these underlying developments. By creating a synthetic control city for each of the 53 German cities in our sample, we obtain a benchmark with which we can compare the corresponding German cities after WWII, thus potentially controlling for these underlying developments.

Our main findings are as follows: We find that the synthetic city population is typically characterized by an inverse U-shaped development, explainable by underlying non-random developments. We find that the majority of German cities recover to their supposed, synthetic mean in size and city share. We also find a roughly balanced ratio of positively to negatively impacted (unit root) cities, which further supports mean reversion. Those cities where we find a negative impact after WWII were on average smaller and situated in less agglomerated areas, which hints at the better ability of larger cities to recover. This can be rationalized by WWII accelerating migration away from these remote places that may have happened in any case. However, more research would be needed within this line of thought.

Our findings reconcile earlier opposing results in the literature ([Davis & Weinstein, 2002](#); [Brakman et al., 2004](#); [Bosker et al., 2008](#)), together suggesting a permanent effect being likely driven by non-linearities combined with larger timespans (or cutoff date) considered, making it more likely that these non-linearities

become significant. This explanation is consistent with the cutoff year of earlier studies in the 1960s and 1999, while the average peak of the non-linearity of the synthetic city was in 1979. The results also tend to show the relevance of changing locational fundamentals that explain the non-linear counterfactual population trajectory in the 20th century, with increasing returns effects better explaining the migration out of smaller and remote cities after the war.

References

- Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59, 391–425.
- Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California’s tobacco control program. *Journal of the American Statistical Association*, 105(490), 493–505.
- Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2), 495–510.
- Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, 93(1), 113–132.
- Adjei, E. K., Eriksson, R., & Lundberg, J. (2023). The effects of a large industrial investment on employment in a remote and sparsely populated area using a synthetic control approach. *Regional Science Policy & Practice*, 15(7), 1553–1576.
- Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., & Wolf, N. (2015). The economics of density: Evidence from the Berlin Wall. *Econometrica*, 83(6), 2127–2189.
- Baum-Snow, N. (2007). Did highways cause suburbanization? *The Quarterly Journal of Economics*, 122(2), 775–805.
- Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., & Zhang, Q. (2017). Roads, railroads, and decentralization of Chinese cities. *Review of Economics and Statistics*, 99(3), 435–448.
- Bleakley, H., & Lin, J. (2012). Portage and path dependence. *The Quarterly Journal of Economics*, 127(2), 587–644.
- Bosker, M., Brakman, S., Garretsen, H., & Schramm, M. (2007). Looking for multiple equilibria when geography matters: German city growth and the WWII shock. *Journal of Urban Economics*, 61(1), 152–169.
- Bosker, M., Brakman, S., Garretsen, H., & Schramm, M. (2008). A century of shocks: the evolution of the German city size distribution 1925–1999. *Regional Science and Urban Economics*, 38(4), 330–347.
- Brakman, S., Garretsen, H., & Schramm, M. (2004). The strategic bombing of German cities during World War II and its impact on city growth. *Journal of Economic Geography*, 4(2), 201–218.

- Bundesamt für Statistik, Schweiz. (2016). Bevölkerungsentwicklung der Gemeinden 1850-2000.
- Danmarks Statistik. (1985). Danske byers folketal 1801-1981.
- Danmarks Statistik. (2011). Statbank danmark folketal.
- Davis, D. R., & Weinstein, D. E. (2002). Bones, bombs, and break points: The geography of economic activity. *American Economic Review*, 92(5), 1269–1289.
- Davis, D. R., & Weinstein, D. E. (2008). A search for multiple equilibria in urban industrial structure. *Journal of Regional Science*, 48(1), 29–65.
- Defence Analytical Services and Advice. (2011). Total strength, uk service personnel. *Institute for fiscal studies*.
- DG REGIO. (2021). Population data collection for european local administrative units from 1960 onward. *Directorate-General for Regional and Urban Policy, European Commission*.
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366), 427–431.
- duPont IV, W., & Noy, I. (2015). What happened to Kobe? a reassessment of the impact of the 1995 earthquake in Japan. *Economic Development and Cultural Change*, 63(4), 777–812.
- Dye, A. V. (1921). Coal production in the United Kingdom. *US Commerce Reports, Volume 1, Issue 3*.
- Fischer, W. (1989). Statistik der Bergbauproduktion Deutschlands 1850 – 1914. *GESIS Datenarchiv, Köln. histat. Studiennummer 8448, Datenfile Version 1.0.0*.
- Fischer, W. (1995). Statistik der Montanproduktion Deutschlands 1915 – 1985. *GESIS Datenarchiv, Köln. histat. Studiennummer 8400, Datenfile Version 1.0.0*.
- Friedrich, J. (2002). Der Brand; Deutschland im Bombenkrieg 1940-1945. *Munich, Propylaen*.
- Gabaix, X. (1999). Zipf’s law for cities: an explanation. *The Quarterly Journal of Economics*, 114(3), 739–767.
- Gabaix, X., & Ioannides, Y. M. (2004). The evolution of city size distributions. In *Handbook of regional and urban economics* (Vol. 4, pp. 2341–2378). Elsevier.
- GB Historical GIS / University of Portsmouth. (2021). Population Statistics — Total Population, A Vision of Britain through Time.
- Gilchrist, D., Emery, T., Garoupa, N., & Spruk, R. (2023). Synthetic control method: A tool for comparative case studies in economic history. *Journal of Economic Surveys*, 37(2), 409–445.
- Glaeser, E. L. (2005). Reinventing Boston: 1630–2003. *Journal of Economic Geography*, 5(2), 119–153.
- Glaeser, E. L., & Kahn, M. E. (2004). Sprawl and urban growth. In *Handbook of regional and urban economics* (Vol. 4, pp. 2481–2527). Elsevier.

- Harvey, A., & Thiele, S. (2021). Cointegration and control: Assessing the impact of events using time series data. *Journal of Applied Econometrics*, 36(1), 71–85.
- Heinzelmann, M. (2003). Göttingen im Luftkrieg. *Die Werkstatt, Göttingen*.
- Im, K. S., Pesaran, M. H., & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, 115(1), 53–74.
- Instituto Nacional de Estadística. (2011). Censos 1878-2011.
- Kaul, A., Klößner, S., Pfeifer, G., & Schieler, M. (2015). Synthetic control methods: Never use all pre-intervention outcomes together with covariates. *MPRA Paper 83790, University Library of Munich, Germany*.
- Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy*, 99(3), 483–499.
- Kästner, F. (1949). Kriegsschäden (Trümmermengen, Wohnungsverluste, Grundsteuerausfall und Vermögensteuerausfall. *Statistisches Jahrbuch Deutscher Gemeinden*, 37: 361–391.
- Lauer, J. (2020). Das "Bombenloch" blieb noch Jahrzehnte. *Rhein-Neckar Zeitung*.
- Levin, A., Lin, C.-F., & Chu, C.-S. J. (2002). Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics*, 108(1), 1–24.
- Michaels, G., & Rauch, F. (2018). Resetting the urban network: 117–2012. *The Economic Journal*, 128(608), 378–412.
- Miguel, E., & Roland, G. (2011). The long-run impact of bombing Vietnam. *Journal of Development Economics*, 96(1), 1–15.
- National Archives UK. (2021). Ministry of Home Security: Daily Intelligence Reports.
- Ng, S., & Perron, P. (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association*, 90(429), 268–281.
- Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica: Journal of the Econometric Society*, 1361–1401.
- Perron, P. (1991). Test consistency with varying sampling frequency. *Econometric Theory*, 7(3), 341–368.
- Rappaport, J., & Sachs, J. D. (2003). The United States as a coastal nation. *Journal of Economic Growth*, 8(1), 5–46.
- Redding, S. J., Sturm, D. M., & Wolf, N. (2011). History and industry location: evidence from German airports. *Review of Economics and Statistics*, 93(3), 814–831.
- Rydberg, J., McGarrell, E., & Norris, A. (2018). A quasi-experimental synthetic control evaluation of a place-based police-directed patrol intervention on violent crime. *Journal of Experimental Criminology*, 14, 83–109.

- Siodla, J. (2015). Razing San Francisco: The 1906 disaster as a natural experiment in urban redevelopment. *Journal of Urban Economics*, 89, 48–61.
- Swärd, C. (2016). Folknet Database 1810-1990.
- Taylor, M. P., Peel, D. A., & Sarno, L. (2001). Nonlinear mean-reversion in real exchange rates: toward a solution to the purchasing power parity puzzles. *International Economic Review*, 42(4), 1015–1042.
- Tucker, S. C. (2003). *The Second World War*. Red Globe Press.
- Volkman, H.-E. (1995). Ende des Dritten Reiches, Ende des Zweiten Weltkriegs: Eine Perspektivische Rückschau. *Piper*.

7 Data Appendix

7.1 Notes on cities in Germany

Table 8 shows the West-German cities in our sample. As compared to the 62 cities in [Bosker et al. \(2008\)](#), we exclude 7 cities due to incomplete population data at 10 year intervals between 1871-1920. The excluded cities are: Flensburg, Fürth, Kiel, Oldenburg, Wanne-Eickel, Wattenscheid, and Wuppertal. We have excluded Berlin and Hamburg due to the lack of donor cities that are as large as them before WWII. City amalgamations, predominantly in the 1920s and to a lesser extent in the 1930s, are adjusted for following the procedure in [Bosker et al. \(2008\)](#).

7.2 Notes on cities in England and Wales

We do not have a census for England and Wales in 1941 due to the war. However, there was a register during the war in 1939 to issue identity cards in order to coordinate rations. These do not include military personnel, so that we may especially undercount the population in cities with unusually large military presence, such as Portsmouth as a Navy port. To address this, we used the 1951 number of military personnel in a city and added them to the 1939 register with a factor 1.54, which reflects the ratio of the number of total military personnel (1.27 million, see [Tucker, 2003](#)) in September 1939 when the register took place and the personnel (0.8238 million, see [Defence Analytical Services and Advice, 2011](#)) in 1951.¹⁸ For the cities that are not naval bases, this adjustment would on average increase the population figures by 1.3 percent, so that this adjustment changes relatively little for most cities. For the naval bases Plymouth and Portsmouth, this increase is 11.3 and 13.9 percent respectively.

Cities in the United Kingdom were also bombed during what is commonly known as the Blitz, but overall less severely and with the main focus on London and air fields in particular, as opposed to the widely used strategy of 'area bombing' employed by commander Arthur Harris during the war that targeted German city centers more severely without a particular focus on military importance ([Brakman et al., 2004](#)). The

¹⁸Note that the high number of military personnel, even in 1951, is partially due to the National Service Act of 1948 that kept the number of military personnel high for a few years after WWII. On the other hand, the relatively low number of personnel conscripted is reflected in September 1939 with 1.27 million, compared to the 3.29 in April 1941 or 4.91 million in April 1945. The comparable personnel numbers for 1939 and 1951 helps us avoid over-extrapolation through a low factor of 1.54.

UK accurately registered deaths and injuries for each bombing raid sortie, which was documented twice a day for the Ministry of War Security and were made available in the National Archives.¹⁹ With these data, we construct for each U.K. city a measure of deaths per capita caused by the Blitz to determine which cities we can consider to be “unscathed” by the war and thus suitable donor cities for the SCM.

The lowest estimated deaths per 1939 capita of German cities for which we have casualty data are Heidelberg, Göttingen, and Regensburg with 0.0008, 0.0017, and 0.0021 respectively, and which are considered to have come out of the war unscathed, leaving the city largely undamaged (Lauer, 2020; Heinzelmann, 2003; Friedrich, 2002).²⁰ We have in total census data for 75 cities in England and Wales, taken from GB Historical GIS / University of Portsmouth (2021) at the district or unitary authority level at a decennial interval from 1881 to 1991. Taking these figures as a point of orientation on which cities can be considered as largely undamaged out of the English and Welsh cities, we have 13 that are above an “unscathed” threshold of 0.002, and 20 that are above 0.001 deaths per 1940 capita.

For comparison, at least for the cities for which we have estimated death figures by using the source of Friedrich (2002), the average rate for German cities is a magnitude higher at around 0.02 against the average of 0.0008 during the Blitz. Only 2 out of 42 German cities for which we have estimations on death figures caused by the WWII bombings would be below the 0.002 threshold, which stands in contrast to the cities of 75 Great Britain in our dataset with at least 50,000 inhabitants at some point between 1880 to 1990, where 62 English and Welsh cities are below the 0.002 threshold, and 55 are below the 0.001 threshold. Hence, although most cities in Great Britain were raided, most can be considered as undamaged or only slightly damaged. From the 62 “unscathed” cities, we discard 12 due to inconsistent borders over time that we were not able to fix. This leaves us with 50 English and Welsh cities in total.

¹⁹The dataset was cleaned and made available at www.warstateandsociety.com/Bombing-Britain from the original source (National Archives UK, 2021).

²⁰Respectively, the total deaths were 72 (Lauer, 2020), 107 (Heinzelmann, 2003) and 203 (Friedrich, 2002).

Table 1: Individual City Examples

covariates	Treatment Aachen	Synthetic Aachen	Treatment Lübeck	Synthetic Lübeck	Treatment Düsseldorf	Synthetic Düsseldorf
population						
- (1900)	142427	142164.9	82098	82168.29	281492	293331.3
- (1920)	151731	151319.6	118709	118850.4	453026	435615.5
- (1930)	154682	154856	129842	130004.7	476300	482981.9
- (1940)	162164	161832.1	154811	155007.7	541410	533579.4
river	0	0.041	1	1	1	1
sea	0	0.252	1	1	0	1
coal worker	15426	43509.6	0	270.8	134065	0
- (1920)						
urb. pot.	265614.3	268480	245207.3	249135.5	745703.6	621672.5
- (1940)						
	Donor city	Weight	Donor city	Weight	Donor city	Weight
	Barrow, UK	2.9%	Aarhus, DK	32.1%	Kopenhagen, DK	57.9%
	Bradford, UK	10.9%	Blackpool, UK	0.2%	Lisboa, PT	1.7%
	Braga, PT	71.8%	Brighton, UK	28.3%	Southend, UK	40.5%
	Brighton, UK	1.2%	Kolding, DK	13.9%		
	Oldham, UK	5%	Malmö, SWE	7.3%		
	Sheffield, UK	8.1%	Southend, UK	18.3%		

Table 2: Descriptive Statistics

	mean (std. dev.)	min / max	mean (std. dev.)	min / max	mean (std. dev.)	min / max
Covariates	German cities		Synthetic cities		Donor pool	
population						
- (1900)	134480 (112831)	11704 / 534175	137582 (117275)	16198 / 533926	101383 (129882)	1138 / 842277
- (1920)	194007 (170816)	46990 / 699513	188816 (161679)	47003 / 698638	120747 (151217)	1628 / 943941
- (1930)	209416 (184136)	54500 / 759381	207006 (180159)	52197 / 767957	129505 (164122)	2716 / 1030956
- (1940)	220939 (193467)	58713 / 840188	222419 (195616)	57125 / 835473	135698 (168584)	3584 / 1046228
river	0.774 (0.423)	0 / 1			0.917 (0.277)	0 / 1
sea	0.019 (0.137)	0 / 1			0.319 (0.468)	0 / 1
coal worker						
- (1920)	62518 (82687)	0 / 219758			39346 (69607)	0 / 255000
urb. pot.	381356 (214026)	136102 / 972784			233538 (235975)	35326 / 1298276
- (1940)						
Non-covariates						
population						
- (1950)	211312 (175908)	54100 / 816924	248277 (219625)	62761 / 918915	149520 (188550)	7269 / 1148872
- (1960)	263290 (226693)	68400 / 1061359	255478 (212027)	70829 / 884986	159883 (197952)	13688 / 1249621
- (1970)	268313 (241572)	68700 / 1290071	251027 (190815)	78391 / 784843	166579 (201427)	26743 / 1485589
- (1980)	252265 (227949)	65154 / 1277230	240350 (176452)	79449 / 778944	168212 (219135)	40193 / 1961031
- (1990)	249724 (219536)	63525 / 1208484	236850 (171675)	80939 / 780176	168226 (219198)	40462 / 2010080
Sample size	53		53		143	

Table 3: Unit Root Test on City Size

A. Individual City (Augmented) Dickey-Fuller Test					
Period	1900-1990	1900-1980	1900-1970	1900-1960	1900-2000
a. Synthetic Detrended ($c_i = 0$)					
Significance level	% Unit Root Rejected				
1%	7.5	11.3	11.3	28.3	
5%	24.5	26.4	34.0	54.7	
10%	32.1	35.8	49.1	66.0	
	% Unit Root with Positive Impact ($S_{i,t} > S_{i,t}^{synth}$)				
1%	53.1	48.9	53.2	42.1	
5%	47.5	46.2	48.6	45.8	
10%	47.2	41.2	51.9	50.0	
b. Conventional Method with Linear Trend ($\delta_i t = 0$ or $\delta_i t \neq 0$)					
Significance level	% Unit Root Rejected				
1%	0	0	0	3.8	0
5%	5.7	5.7	7.5	9.4	3.8
10%	7.5	9.4	13.2	13.2	5.7
B. Panel Unit Root Test					
a. Synthetic Detrended ($c_i = 0$)					
	t-stat (p-val)	t-stat (p-val)	t-stat (p-val)	t-stat (p-val)	t-stat (p-val)
Levin-Lin-Chu	-3.425 (0.000)	-4.992 (0.000)	-7.415 (0.000)	-11.807 (0.000)	
b. Conventional Method with Linear Trend (i.e., with Trend Term $\delta_i t$)					
Levin-Lin-Chu	-11.856 (0.000)	-13.078 (0.000)	-18.242 (0.000)	-18.530 (0.000)	-12.179 (0.000)
Im-Pesaran-Shin	-0.071 (0.472)	-0.600 (0.274)	-3.833 (0.000)	-4.811 (0.000)	-0.732 (0.232)

Note: The null hypothesis is a unit root in city size detrended by the synthetic city size. The total number of cities is 53 for all specifications, and 143 in the donor pool which had below 0.002 deaths per capita during the Blitz. The synthetic detrended Dickey-Fuller test in panel (A.a), as in equation (8), is estimated by using as covariates the 1900 and 1920-1940 lagged population, the 1920 number of coal workers, the dummies for sea or river access, and the 1940 urban potential measure. The MSPE was set between 1880-1940. The linear detrended Dickey-Fuller test in panel (A.b) follows equation (6) with trend term $\delta_i t$. In Panel B, the individual city unit root number of lags was chosen following [Ng & Perron \(1995\)](#), where the maximum lag length started at 2. Incidentally, the lag is chosen to be zero for all cities. $c_i = 0$ indicates a suppressed constant in the regression. $\delta_i t = 0$ or $\delta_i t \neq 0$ indicates an added trend in the regression. Given the individual city optimal number of lags, we set the lag to 0 for the panel unit root tests. Panel statistics are calculated with adjustment for cross-sectional dependence.

Table 4: Unit Root Test on Log City Size (Donor Pool: <0.002 bombing deaths p.c.)

A. Individual City (augmented) Dickey-Fuller test					
Period	1900-1990	1900-1980	1900-1970	1900-1960	1900-2000
a. Synthetic detrended ($c_i = 0$)					
Significance level	% Unit Root Rejected				
1%	22.6	26.4	30.2	47.2	
5%	41.5	50.9	69.8	75.5	
10%	54.7	67.9	79.2	75.5	
Out of which % Unit Root with Positive Impact ($S_{i,t} > S_{i,t}^{synth}$)					
1%	48.8	48.7	48.6	35.7	
5%	48.4	42.3	37.5	46.2	
10%	45.8	35.3	36.3	46.2	
b. Conventional Method with Linear Trend (i.e. with trend term $\delta_i t$)					
Significance level	% Unit Root Rejected				
1%	5.7	3.8	7.5	3.8	7.5
5%	13.2	17.0	18.9	11.3	20.8
10%	28.3	26.4	34.0	22.6	32.1
B. Panel Unit root test					
a. Synthetic detrended ($c_i = 0$)					
	t-stat (p-val)	t-stat (p-val)	t-stat (p-val)	t-stat (p-val)	t-stat (p-val)
Levin-Lin-Chu	-10.934 (0.000)	-13.474 (0.000)	-16.709 (0.000)	-18.563 (0.000)	
b. Conventional Method with Linear Trend (i.e. with trend term $\delta_i t$)					
Levin-Lin-Chu	-26.995 (0.000)	-28.216 (0.000)	-24.732 (0.000)	-24.112 (0.000)	-26.995 (0.000)
Im-Pesaran-Shin	-7.066 (0.000)	-5.990 (0.000)	-4.828 (0.000)	-5.632 (0.000)	-7.066 (0.000)

Note: The null hypothesis is a unit root in log city size. The total number of German cities is 53 for all specifications, and 143 in the donor pool which had below 0.002 deaths per capita during the Blitz. The synthetic detrend Dickey-Fuller test in 3a, as in equation (8), is estimated through the first specification with the 1900 and 1920 to 1940 lagged population, 1920 number of coal workers, dummies for sea or river access, and the 1940 urban potential measure. The MSPE was set between 1880 and 1940. The linear detrended Dickey-Fuller test in 3b follows equation (6) with trend term $\delta_i t$. The individual city unit root number of lags are chosen following Ng & Perron (1995), where the maximum lag length started at 2 for this procedure. Incidentally, the lag is chosen to be zero for all cities. $c_i = 0$ indicates a suppressed constant in the regression. $\delta_i t = 0$ or $\delta_i t \neq 0$ indicates an added trend in the regression. Given the individual city optimal number of lags, we set the lag to 0 for the panel unit root tests. Panel statistics are calculated with adjusting for cross-sectional dependence.

Table 5: Actual and synthetic city equality of distributions

Kolmogorov-Smirnov			
Year	p-value	Year	p-value
1880	0.273	1940	0.958
1890	0.529	1950	0.837
1900	0.905	1960	0.987
1910	0.744	1970	0.920
1920	0.784	1980	0.476
1930	1.000	1990	0.617

Note: The null hypothesis is the equality of both actual and synthetic distribution. The synthetic cities are the first specification (<0.002 bombing deaths p.c.) with the 1900 and 1920 to 1940 lagged population, 1920 number of coal workers, dummies for sea or river access, and the 1940 urban potential measure. The MSPE was set between 1880 and 1940. We report the exact p-values. The same exercise with the second specification (<0.001 bombing deaths p.c.) yields the same conclusion of no significant p-value for any time.

Table 6: Unit Root Tests on Relative City Size

A. Individual City (Augmented) Dickey-Fuller Test					
Period	1900-1990	1900-1980	1900-1970	1900-1960	1900-2000
a. Synthetic Detrended ($c_i = 0$)					
Significance level	% Unit Root Rejected				
1%	17.0	22.6	24.5	35.8	
5%	43.4	49.1	69.8	71.7	
10%	58.5	66.0	77.4	75.5	
% Unit Root with Positive Impact ($S_{i,t} > S_{i,t}^{synth}$)					
1%	50.0	46.3	45.0	38.2	
5%	46.7	37.0	29.4	46.7	
10%	36.3	22.2	33.3	46.2	
b. Conventional Method with Linear Trend (i.e., with Trend Term $\delta_i t$)					
Significance level	% Unit Root Rejected				
1%	9.4	5.7	5.7	3.8	13.2
5%	26.4	22.6	15.1	13.2	28.3
10%	37.7	22.6	30.2	32.1	45.3
B. Panel Unit Root Test					
a. Synthetic Detrended ($c_i = 0$)					
	t-stat (p-val)	t-stat (p-val)	t-stat (p-val)	t-stat (p-val)	t-stat (p-val)
Levin-Lin-Chu	-10.713 (0.000)	-12.723 (0.000)	-15.242 (0.000)	-16.843 (0.000)	
b. Conventional Method with Linear Trend (i.e. with Trend Term $\delta_i t$)					
Levin-Lin-Chu	-15.619 (0.000)	-14.536 (0.000)	-14.717 (0.000)	-17.233 (0.000)	-16.074 (0.000)
Im-Pesaran-Shin	-4.800 (0.000)	-3.202 (0.001)	-3.012 (0.001)	-3.159 (0.001)	-6.068 (0.000)

Note: The null hypothesis is a unit root in the relative city size detrended by the synthetic relative city size. The total number of German cities is 53 for all specifications, and 143 in the donor pool which had below 0.002 deaths per capita during the Blitz. The synthetic cities are obtained with as covariates the 1900 and 1920-1940 lagged population, the 1920 number of coal workers, the dummies for sea or river access, and the 1940 urban potential measure. The MSPE was set between 1880-1940. The individual city unit root number of lags is chosen following [Ng & Perron \(1995\)](#), where the maximum lag length starts at 2. Incidentally, the lag is chosen to be 0 for all cities. $c_i = 0$ indicates a suppressed constant in the regression. $\delta_i t = 0$ or $\delta_i t \neq 0$ indicates an added trend in the regression. Given the individual city optimal number of lags, we set the lag to 0 for the panel unit root tests.

Table 7: City Characteristics Conditional on Impact (Cutoff 1980)

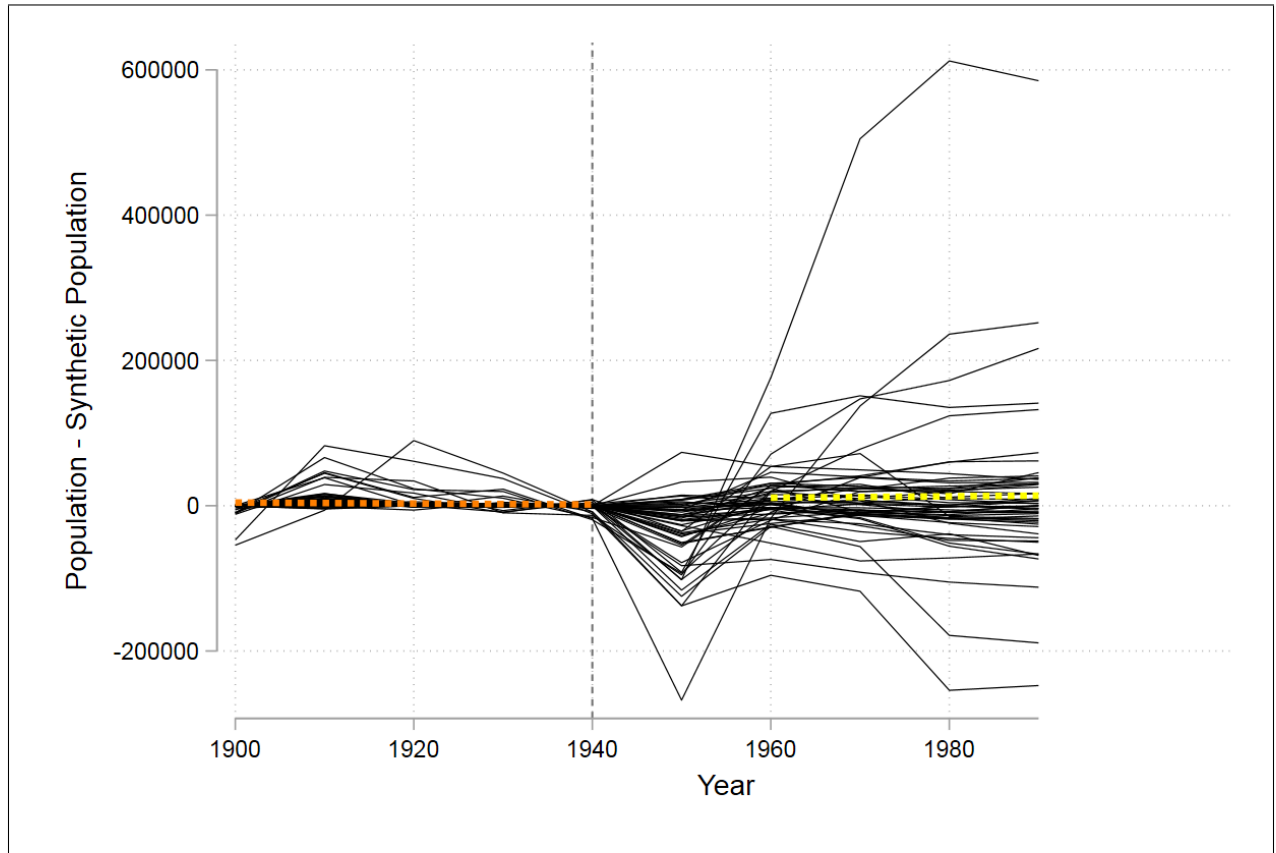
	Impact			t-test		
	Unit Root (pos.)	Stationary	Unit Root (neg.)			
	i.e. $S_{i,t} > S_{i,t}^{synth}$		i.e. $S_{i,t} < S_{i,t}^{synth}$	pos.-stat.	pos.-neg.	stat.-neg.
mean Pop (1939)	302525	228327	162651	0.8948	1.6302	1.1894
std. dev.	(275038)	(188814)	(140687)			
mean Urb. Pot. (1939)	484187	402582	284059	0.9259	2.2326	1.9209
std. dev.	(268524)	(208565)	(163805)			
mean Urb Pot. - Pop (1939)	181661	174256	121408	0.1966	1.5820	1.8440
std. dev.	(95761)	(94402)	(82263)			
mean Coal Worker (1920)	60986	72740	42893	-0.3360	0.5402	1.1451
std. dev.	(88225)	(87829)	(69903)			
mean refugee rate (1960)	0.182	0.165	0.161	0.9555	1.0617	0.3742
std. dev.	(0.062)	(0.039)	(0.036)			
mean rate of housing lost	0.392	0.428	0.392	-0.5641	0.0053	0.6674
std. dev.	(0.214)	(0.141)	(0.219)			
mean rubble p.c.	11.913	12.657	14.180	-0.2591	-0.4739	-0.5961
std. dev.	(10.612)	(6.124)	(11.085)			
Number of Cities	8	30	15			

Note: We consider 'stat.' a city where we find stationarity at a 5% significance level and a unit root city if the significance level is above 5%, while 'pos.' and 'neg.' respectively denote a positive and negative impact on the city population with a cutoff year of 1980, conditional on finding a unit root at a 5% significance level. 'Rubble p.c.' denotes the post-war cubic volume of rubble per 1939 capita in the given city.

Table 8: List of West-German cities

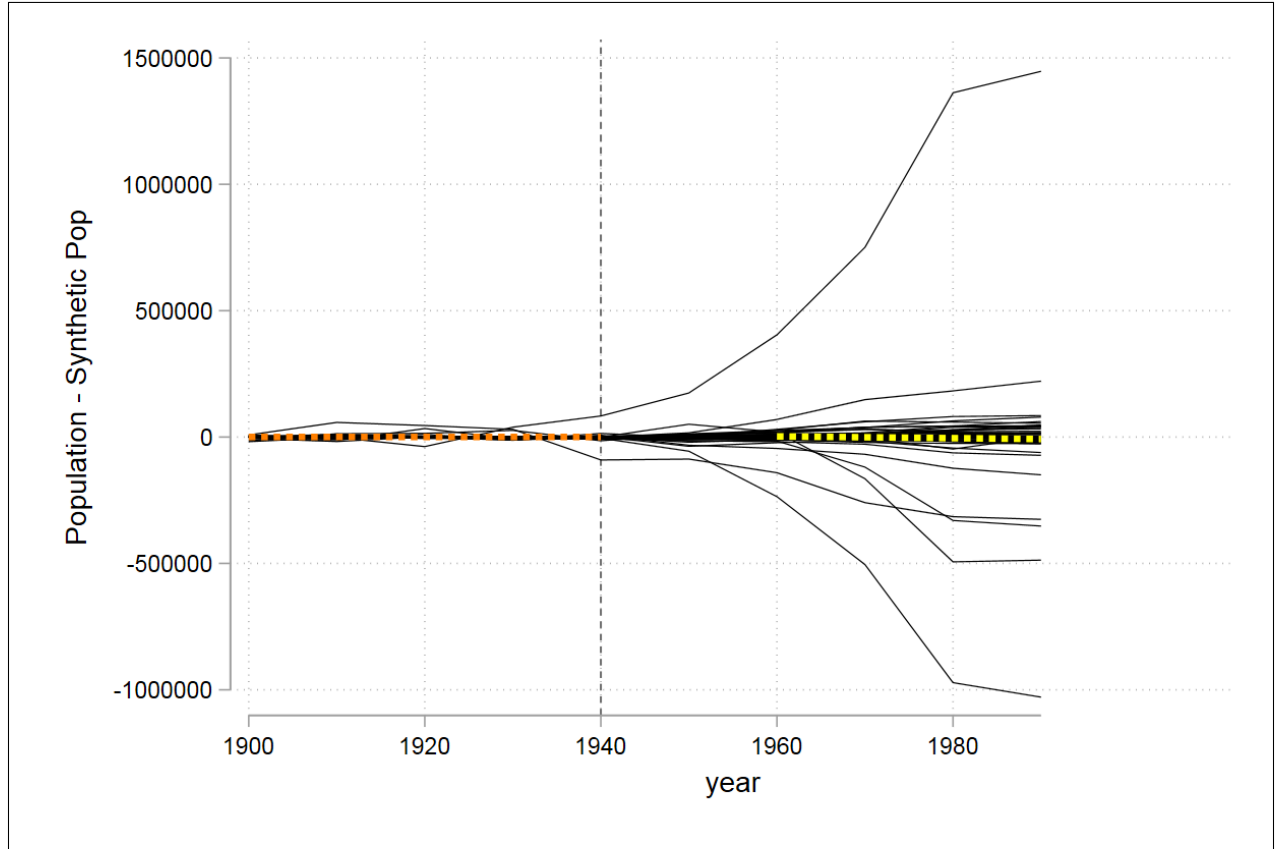
Aachen	Essen	Kassel	Oberhausen
Augsburg	Frankfurt am Main	Koblenz	Offenbach am Main
Bamberg	Freiburg im Breisgau	Köln	Osnabrück
Bielefeld	Gelsenkirchen	Krefeld	Pforzheim
Bochum	Gladbeck	Ludwigshafen am Rhein	Recklinghausen
Bonn	Hagen	Lübeck	Regensburg
Bottrop	Hamm	Mainz	Remscheid
Braunschweig	Hannover	Mannheim	Solingen
Bremen	Heidelberg	Mönchengladbach	Stuttgart
Darmstadt	Herne	Mülheim an der Ruhr	Trier
Dortmund	Hildesheim	München	Ulm
Duisburg	Kaiserslautern	Münster	Wiesbaden
Düsseldorf	Karlsruhe	Nürnberg	Witten
			Würzburg

Figure 1: Individual City Gap of Population and Synthetic Population



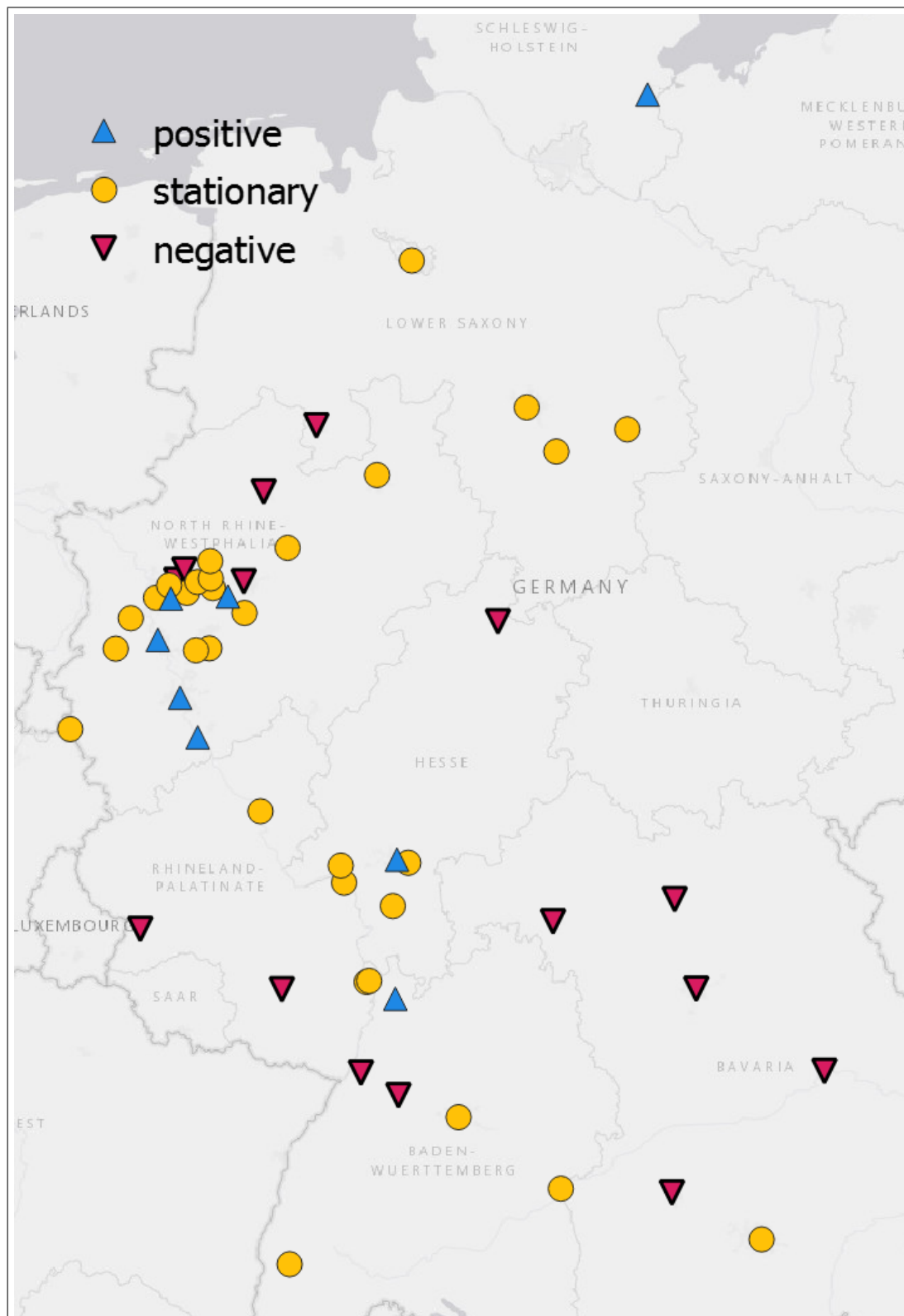
Note: The total number of cities in this Figure is 53 and the synthetic cities are drawn from the donor pool of 143 cities with less than 0.002 deaths per capita during the Blitz. Synthetic population is based on the specification with as covariates the 1900 and 1920-1940 lagged population, the 1920 number of coal workers, the dummies for sea or river access, and the 1940 urban potential measure. The MSPE was set between 1880-1940. The orange (yellow) dashed line shows the linear fit before 1940 (after 1960) to exclude the WWII shock. The city reaching a gap of over 500,000 in 1980 is Munich.

Figure 2: ≥ 0.05 Weight Comparison City Gap



Note: The total number of cities is 56 and corresponds to the cities in the donor pool that had at least a weight of 0.05 in the creation of any of the 53 synthetic cities of Figure 1; also see Table 3 for the synthetic control creation specifications. Birmingham is excluded because there is no city in the donor pool that matches its pre-WWII population. The orange (yellow) dashed line shows the linear fit before 1940 (after 1960).

Figure 3: City Location by Relative City Size Impact (Cutoff 1980)



Note: The total number of cities is 53, and 143 in the donor pool which had below 0.002 deaths per capita during the Blitz.