Analyzing the Impact of a New Battery Factory on Local Housing Prices: Evidence from Skellefteå, Sweden

Marcus Moerup, Petter Hallqvist, Vladislav Virtonen
Minerva University
14 Mint Plaza, 3 Floor, San Francisco, CA 94103

Submitted: April 21, 2023

Abstract

In October 2017, Northvolt AB, a Swedish lithium-ion battery manufacturer, announced the construction of Northvolt Ett, Europe's largest lithium-ion battery factory, in the northern Swedish municipality of Skellefteå. The estimated investment until 2023 was up to €4 billion, providing 3,000 new jobs to the local town. Many articles have been written about the subsequent increase in housing prices in Skellefteå, focusing both on economic revitalization and the subsequent increase in housing prices. This paper uses synthetic control and synthetic difference-in-differences (DiD) methods to investigate the effect of the announcement of factory establishment on local housing prices. To estimate a treatment effect, both models produce a synthetic counterfactual of housing price development in Skellefteå using a donor pool of municipalities selected for their similarity based on socioeconomic factors. The counterfactual was constructed and the analysis was completed using the Synth (Abadie et al. 2011) and Synthdid (Arkhangelsky et al. 2021) packages in R. Our Synthetic Control models estimate that as of 2021, Northvolt Ett increased apartment valuations by 6,822 SEK/m^2, and house K/T (price/taxable value) ratios by 0.42, both estimates considered statistically significant. This allows us to reject our null hypothesis. Our synthetic DiD model produces positive estimates without statistical significance. We note that due to lacking median income data for 2022, our synthetic control models only range until 2021. Given the trend in housing prices in Skellefteå in comparison to Synthetic Skellefteå, we consider it a high likelihood that treatment estimates would be higher for 2022.

1 Introduction

In October 2017, Northvolt AB, a Swedish lithium-ion battery manufacturer, announced its plan to build Northvolt Ett, Europe's largest lithium-ion battery factory. It was estimated to create upwards of 3,000 new jobs and would be built in the Northern Swedish town of Skellefteå, where the total population at the time was around 72,000 (SCB 2023).

It is clear that the total estimated investment of €4 billion by 2023 presents a massive economic opportunity to Skellefteå, serving the booming electric vehicle and energy storage markets (Nilsen 2017). However, locals have complained about the new housing shortage in Skellefteå, and some have attributed it to the newly built factory. As the European EV supply chain grows, the economic impacts of corporate investment on rural housing markets will be a highly relevant relationship to understand. That is what this paper aims to do, aiming to inform public policy and corporations of local impacts resulting from their decisions going forward.

Among the existing literature on the topic, several studies have investigated the potential adverse effect of industrial sites on housing prices through negative externalities such as congestion, obstruction of view, and traffic noise disturbance. De Vor & De Groot (2011) found that proximity to industrial sites in the regions of Randstad and Noord-Brabant in the Netherlands reduced residential property values. However, it is equally important to consider the potential positive effects caused by an influx of new residents who increase the demand for housing, shifting the equilibrium and increasing the local housing prices. Qian & Tan (2021) analyzed 391 high-skilled firm entries in the U.S. from 1990 to 2010 to estimate the impact on incumbent residents and found a significant effect on housing prices, adversely affecting low-income residents living near the entry sites.

These "big fish, small pond"-scenarios have become increasingly common since Elon Musk coined the term "gigafactory" in 2013. Tesla has since established five of these large-scale facilities, with the most recent being situated outside of Berlin, Germany. Even the smallest of the gigafactories now employs over 1,400 individuals (Crider 2021), and in several locations, reports of soaring housing prices have emerged in the wake of these factories opening their doors.

This study's primary purpose is to investigate whether the increase in housing prices can be directly attributed to the announcement of the new firm entry. Our research question is whether a corporation's announcement of building a new factory in a small town causes an increase in local housing prices.

Many news agencies cite studies of housing price increases. For example, Norrland Living (2022) reported an increase in housing prices in the city by 18% and apartment prices by 25% upon the completion of Northvolt 1. We note the lack of rigorous academic causal study investigating the effect of the newly built factory on housing prices in Skellefteå. This is what we set out to investigate in this paper.

We use a set of panel data representing the socioeconomic trends of Skellefteå to model a synthetic counterfactual - what would have happened to house prices if there were no factory? For this model, we use the following variables: housing price, house price/value ratio, unemployment, average age, median income, and population density. We also use difference-in-differences (DiD) for this analysis.

This paper contributes to the literature by carrying out the first causal study about the impact of the announcement of a new factory building on housing prices in small municipalities in Sweden. It also complements existing empirical evidence of the Keynesian economic theory of animal spirits, showing that the expectation of a new Northvolt factory in Skellefteå contributed to driving up the prices of the housing market.

2 Hypothesis

The main causal question of this paper is: How did the 2017 establishment of Northvolt Ett (Northvolts EV battery factory) in Skellefteå influence apartment and house prices in Skellefteå? The central hypothesis of this study is that Northvolt's announcement of a new battery factory in Skellefteå will have a significant positive impact on local average housing prices. The units of analysis are municipalities. Thus, the null hypothesis is that Nortvolt's establishment in Skellefteå has no impact on the local average housing prices. The units of the outcome variable are 1) the average Swedish kronor per square meter (SEK/m^2) and 2) the average price per taxable value ratio (K/T) for apartments and houses respectively, given the form of publically available data. Our hypothesis is based on several reasons, including:

1) The establishment of the battery factory is expected to attract a large number of new residents to Skellefteå, creating a surge in demand for housing,

2) The presence of a new industrial plant will signal economic growth and development in the area, which may attract outside investors and further drive up housing prices,

3) The increased economic activity generated by the battery factory will create new employment opportunities, leading to an influx of highly skilled workers who will have greater purchasing power and thus demand more expensive housing,

4) The location of the battery factory in Skellefteå is expected to stimulate the local economy, leading to increased demand for housing not only from employees of the factory but also from people who want to participate in the growing economy,

5) The potential constraints on the supply of housing in Skellefteå, as the town may face challenges in rapidly expanding its housing stock to accommodate the expected increase in demand.

To investigate this hypothesis, the study will analyze data on local housing prices in Skellefteå both before and after the announcement of the battery factory, as well as compare these results to those of similar municipalities in Sweden that did not experience similar treatment.

3 Literature Review

Much existing literature has attempted to investigate the impact of firm entries on local economic factors, including housing prices. Some argue that large firms bring negative externalities such as congestion, traffic noise disturbance, and obstruction of view, which could adversely affect local property values. De Vor & De Groot (2011) studied industrial sites and their local impacts in the regions of Randstad and Noord-Brabant in the Netherlands. They found that the presence of an industrial site caused a significant value reduction in nearby residential properties due to an increase in negative externalities.

Others argue that firm entries bring an influx of new residents seeking housing in the local area, which drives up demand and causes an increase in housing and renting costs, often resulting in gentrification because the original residents no longer can afford to live in the area and thus are replaced by wealthier and more educated individuals. Sable (2007) investigated the "impact of the biotechnology industry on local economic development in the Boston and San Diego metropolitan areas." The paper is meant to contrast policymakers' generally favorable perception of biotech companies and their influence on local development by highlighting the negative impacts on several different factors, including gentrification. The paper finds that: "the downside of the gene boom for longtime residents of Cambridge [Boston] is that wealthy entrepreneurs and well-paid employees of pharmaceutical companies are helping to drive up rents and house prices " (Sable 2007).

Qian & Tan (2021) analyzed 391 high-skilled firm entries in the U.S. from 1990 to 2010 to estimate the impact on incumbent residents. For years, small local governments in the U.S. have been fighting to attract big businesses using tools like subsidies and tax incentives, often without knowing the implications of such entries on the local residents. The study finds that the overall welfare benefits for incumbent residents caused by high-skilled firm entries are minimal yet positive. However, the welfare is unequally distributed, with high-income homeowners benefitting the most, while low-income renters who live close to the entry point often have to move as a result of increasing rent costs.

However, the results found in Sable (2007) and Qian & Tan (2021) might have limited generalizability to European countries and/or smaller towns given the specific contextual circumstances of American metropolitan areas. A study that was more applicable to a Swedish context was conducted by Ismail & Wilhelmsson (2022). They analyzed the impact of construction projects on the socio-economic backgrounds of local residents (gentrification) and the affordability of housing. They used a combination of propensity score weighting and difference-in-differences to study construction projects completed between 2009 and 2014 in Stockholm, Sweden. While they found a limited effect of the construction projects on residents' socio-economic background, they found a significant positive effect on housing affordability. This effect is not directly applicable to firm entries in small towns, because 1) new construction projects increase housing supply which likely offsets some of the price increase caused by higher demand from new incoming residents, and 2) the impact of firm entries on local housing prices is likely to differ between metropolitan areas such as Stockholm and rural towns, partly but not only due to the impact of increased demand from new residents relative to total population size.

None of the papers above used the synthetic control method, and to the best of our knowledge, it has not been used to investigate the impacts of firm entries on housing prices in the past. The closest example might be Gautier et al. (2009), who used synthetic control and difference-in-differences to estimate the effect of the Theo Van Gogh murder on housing prices in Amsterdam. Moreover, one of the most well-known papers using the synthetic control method was a case study of Sweden by Andersson (2019), investigating the effect of a new carbon tax on emissions.

4 Data

All outcome and covariate data comes from Statistics Sweden (SCB), which is the government agency operating under the Ministry of Finance, tasked with collecting and presenting official Swedish statistics for political decision-making, debate, and research (SCB 2023). Statistics on housing prices come from the Swedish Realtor Statistical Organization, which collects and presents data on housing prices, which is controlled and verified by SCB (Svensk Mäklarstatistik 2023). Treatment data is defined as all time periods for Skelleftå in the Synthetic Control models, and Skellefteå post-2017 in the synthetic DiD Models.

The units of analysis in this study are Swedish municipalities. This was chosen because it is the smallest level of high-quality data across housing prices and covariates that captures local effects of economic shocks like factory construction. Because many Swedish municipalities are large (including Skellefteå), many have economic cores around which industry is situated, and large areas are very sparsely populated areas of low economic activity and population density. To be able to use data on the municipality level we need to control for population density (as seen in the methodology section) in order to make the comparison across units of similar geographic and economic structures.

Outcome data for housing comes in two kinds: data for apartments sold and data for houses sold. In this case, the word "apartments" means legally available for private purchase on the market ("Bostadsrättsföreningar", as opposed to "Hyresrätter," which means apartments legally available for hire, and translates to apartments owned by private people). Houses are defined as single or double-unit houses owned by private people.

As seen in the methodology section, this paper only includes municipal housing markets of similar structures (apartment-to-villa ratio) for comparability. Data for apartments is measured in price per square meter (SEK/m^2) and data for villas is measured in the K/T ratio. The K/T ratio represents data for the price paid divided by the taxable value of houses. According to the SCB, this is a better valuation measure for houses because it represents how the willingness of consumers to pay for underlying value. Furthermore, it smooths out data across time, as the price per square meter ratio for villas in particular in small housing markets is highly dependent on the location, age, and quality of housing units sold that particular year - meaning that price per square meter is more prone to reacting strongly to new development and other local endogenous factors (Statistiska Centralbyrån 2023).

For our purposes, trying to measure the impacts of shocks on the local economy, which is a clear drawback. However, apartment prices across municipalities are available from 2005 to 2022, and K/T ratios for houses are available from 1996 to 2022, allowing for more data to create more accurate synthetic controls and better pre-treatment matching for the DiD method.[1]

[1] Mäklarstatistik and SCB were contacted by email to ask for price per square meter data for houses sold, and asked follow-up questions about why the K/T ratio was their preferred measure. The reply was that the data was in theory available, but they implied that they did not have time to give it to us. They said that the K/T measurement was

Limiting temporal data availability of the datasets used are:

1) Unemployment (easily accessible, most recent dataset): 2006-2022, - Median income (across municipalities): 1996-2021[2]

2) K/T Ratio: 1996-2022,

3) Apartment price per square meter: 2005-2022.[3]

Data for the selection of donors was also from the SCB official statistics and Google News. The collection and filtering process is described in the following methodology section. While we use the percentage of villas as a selection criterion for the donor pool, we do not include them as a confounder controlled for in our models, as it is not related to both treatment and outcomes, removing it makes the model more interpretable, and it makes the model more conservative as fewer degrees of freedom makes a type 1 error less likely.

The list of synthetic control donor units is (they can be also seen on the map in Fig 25):

1. Karlskrona
2. Norrtälje
3. Falun
4. Lidköping
5. Hudiksvall
6. Västervik
7. Mark
8. Värnamo
9. Katrineholm
10. Falköping
11. Nässjö

The summary statistic table for the entire period available in the dataset, including pre- and post-treatment, is shown in Fig 1.

5 Methodology

We chose synthetic control and synthetic difference-in-differences for this investigation. The primary reason for only using synthetic control methods was that there is one treated unit compared to many control units, which was suitable for our causal question. Using two different methods also allows us to investigate two different aspects of our causal question. Synthetic control calculates the average treatment effect for the treated (ATT) as the post-treatment difference between outcomes

preferable because it was smoothed out and hence suitable for their aims of providing statistics for policy-making and trend measurement across time. K/T data is considered to be the preferred measure of Swedish housing prices, even to median house prices by the SCB

[2] We reached out to the SCB for 2022 median income data because it would enable us to add another post-treatment datapoint, but they said that median income data for 2022 is published on a 13-month trailing basis

[3] Data for Apartment price per square meter for Falkenberg municipality was missing, so that control unit was removed from the data pool for our apartment pricing models.

	av_age	houses_price_value_ratio	median_income	normalized_unemployment_perc
Mean	42.94	1.18	237.35	3.79
Std.Dev	1.26	0.31	31.12	1.12
Min	41.10	0.69	183.30	1.27
Q1	42.10	0.95	211.30	2.89
Median	42.55	1.06	232.85	3.84
Q3	43.35	1.43	262.50	4.68
Max	46.70	2.28	310.90	6.42
MAD	0.82	0.27	36.77	1.32
IQR	1.22	0.48	51.00	1.77
CV	0.03	0.27	0.13	0.30
Skewness	1.12	0.82	0.32	0.03
SE.Skewness	0.18	0.18	0.18	0.18
Kurtosis	0.46	-0.08	-0.92	-0.77
N.Valid	192.00	192.00	192.00	192.00
Pct.Valid	100.00	100.00	100.00	100.00

	pop_km2	population	price_sek_m2	unemployment_num	year
Mean	31.61	44024.85	11588.07	1654.44	2013.50
Std.Dev	14.31	14207.62	6552.58	715.27	4.62
Min	10.40	29339.00	3138.00	418.00	2006.00
Q1	23.20	33269.00	6383.50	1212.50	2009.50
Median	30.00	36765.50	10475.00	1489.00	2013.50
Q3	35.10	56870.50	15036.50	1918.50	2017.50
Max	64.00	73393.00	29844.00	4482.00	2021.00
MAD	7.56	6279.55	6421.88	526.32	5.93
IQR	9.65	23588.25	8593.50	702.00	7.50
CV	0.45	0.32	0.57	0.43	0.00
Skewness	0.75	0.83	0.92	1.20	0.00
SE.Skewness	0.18	0.18	0.18	0.18	0.18
Kurtosis	-0.02	-0.93	0.10	1.49	-1.23
N.Valid	192.00	192.00	176.00	192.00	192.00
Pct.Valid	100.00	100.00	91.67	100.00	100.00

Fig. 1. Descriptive statistics table for the entire period of 2006-2021. The data can be found in the appendix

of treatment and synthetic treatment units, and synthetic DiD calculates ATT as the difference in pre- and post-treatment differences between treated and synthetic control units.

We ran synthetic control and synthetic DiD models for both apartments and houses. Firstly, because the data is available in different forms (price per square meter and K/T ratio) and making models based on both could indicate biases introduced by these data forms. Secondly, apartment data on the municipal level is only available from 2005-2022 and housing data is available from 1996-2022. This means that housing data allows us to run a synthetic DiD model with a significantly longer pre-treatment period which may improve the model itself.

Synthetic control has two main benefits: that weights are not model dependent and that the weights for each contributor to the synthetic control unit can be made explicit. This also allows for rerunning the model with the largest synthetic control donor removed as a robustness check. This benefit is also true for the difference-in-differences method. Synthetic DiD has the advantage of only requiring data for the outcome variable, as long as the parallel trends assumption is upheld. It also allows for unit-fixed effects, which enables a larger pool of donor units for the synthetic control. Given the previously mentioned range of available apartment data and the limitations of covariate data mentioned in the Data section, we were able to run the synthetic control models from 2006-2021. By including a synthetic DiD approach we were able to extend at least one model (housing prices) from 1996 to 2022. This is because the DiD requires only outcome data for the whole period (checking the pre-treatment parallel trends assumption).

The covariates we chose for our models were population per square meter (to control for the degree of urbanization); normalized unemployment (in relation to population), the median income (to control for local economic health); and average age (as a proxy for structural-demographic factors). These factors were chosen because they would influence both housing prices and the likelihood of treatment assignment (local economic factors and demographics of local populations are highly related to where one chooses to place a factory). This would make them confounders that need to be controlled for in the synthetic control model.

As seen in the DAG in Fig 2, there likely are endogenous factors influencing both the treatment and outcome. For example, the reputation of different municipalities and cities influences both housing prices and which units were given treatment. We chose our covariates because we thought among those publically available at SCB, these had the highest signal to cost (in terms of degrees of freedom and interpretability), controlling both for long- and short-term structural economic factors. Given the multitude and rigor of the methods and robustness checks used, we take the ignorability assumption to hold true for our purposes.

Synthetic control requires data for all covariates and outcomes across the whole time period, so both synthetic control models for apartments and houses range from 2006 to 2021. The DiD requires only outcome data for the whole period (checking the pre-treatment parallel trends assumption), and is therefore constructed from 2005 to 2022 for apartments and 1998 to 2022 for houses (a longer pre-treatment period means higher confidence in DiD treatment effect). The DiD Synthetic Control models are constructed only using outcome data.[4]

Two robustness checks were performed on the synthetic control models, both placebo tests and removing the highest weighted control and re-running the model (Fig 10, 12, 14, 16, 18, and 20). Placebo tests (Fig 22, 24) were also performed for the synthetic control DiD. For the synthetic control apartment model, Hudiksvall had a weight of 0.999, so it was removed before the treatment effect calculation. For the robustness check, Västervik was weighted the most (0.374), so it was removed, and the model was recreated (Fig 11, 12, 13). Falköping did not have apartment data for municipalities, so it was removed from that model. For the synthetic control house model, the distribution of weights was even, so Hudiksvall was kept in the model. Mark had the largest weight among the municipalities (0.207) so for robustness check it was removed (Fig 17, 18, 19).

The most challenging part of the methodology was choosing representative units for the donor pool. The "Synth" package in R tends to be good at choosing donors for the synthetic control, but we believe that some factors are required to be satisfied through pre-filtering before letting the "dataprep" function in the Synth package assign weights to each respective control unit donor. Firstly, the housing markets need to have a similar distribution between villas and apartments. Secondly, the chosen units must be of a similar level of economic scale so that a given treatment can be assumed to have an effect of a similar scale. Thirdly, control units cannot have received any similar treatment. The third criterion is linked to the second, as a large factory investment (or similar treatment) is rare in rural and small municipalities, and very common in larger, urban, and more densely populated municipalities.

[4] The Synthetic Difference in Differences Method was used to calculate the discrepancy of trends in outcomes between Skellefteå and the synthetic control before and after treatment. It is hard to find the exact specifications of methods used in the DiD estimate function, but given our results, it seems the model picks a data point in time that is 11 years after the beginning of the dataset, and one datapoint two years after treatment. Between those lines, two trendlines are established for the control and treatment respectively, and compared, their difference showing the treatment effect. We have already established our limitation of pre-treatment data, and recognize that it likely is required for the model to establish a robust pre-treatment trend, but we did not find a way to move the post-treatment datapoint for trend calculation either. If this parameter is truly rigid it means that Synthetic DiD would only be able to capture short-term treatment effects.

Based on the pre-filtering requirements, the criteria below were applied in the following order: Stages 1-2 are related to requiring a similar economic structure, stage 3 is related to a similar housing market, and stage 4 controls for the non-treatment criterion. Each stage removed units, but as expected the highest percentage of units were removed in the fourth stage as most units had other forms of treatment.

Criteria for the donor pool were:

1. Municipality population between 30 000 and 150 000

2. Rural/Non-Rural (population density under 100 people per square kilometer)

3. The percentage of people living in privately owned houses in 2022 was larger than 40 (rounded up to the nearest percent)

4. No similar treatment (no new huge business expansion / new business move-in to town, removing units with economic shocks larger than 1 billion SEK, ca €100 million)

Data to remove units that had received similar treatment in the period 2005-2021 was taken from Google News. This data was the hardest to compile, relying on our judgment. For each municipality that made it to this stage of the donor selection process, three Google News searches were made with a customized timespan of Jan 1st, 2005 - Jan 1st, 2022: 1) "[Name of Municipality/Main town] new factory", 2) "[Name of Municipality] investment", 3) "[Name of Municipality] public funding". The three search prompts were meant to capture a range of public and private investments in these areas. The results for these searches were then looked through, and any leads of large-scale economic shocks disqualifying the municipalities were followed up on (we define large economic shocks as investments larger than 1 billion SEK). Furthermore, if the municipality was within a one-hour drive of a large economic center (100 000+ population city, port, or other population centers), the process was repeated for that center too, as we assumed that commuters and other economic spillover effects would disqualify nearby municipalities by influencing their housing prices too. This criterion was the most stringent, disqualifying all urban municipalities and leaving a few rural municipalities with small populations for the donor pool. While we tried to create a donor pool criterion that was effective in ensuring that donors had not received similar treatment, applied consistently across units, and was replicable, we recognize that this part of the data collection process was the most dependent on our judgment. Economic activity is inherently interconnected, calling the stable unit treatment value assumption (SUTVA) into question as spillover effects of economic intervention happen on both the local and national scales. However, given available time and resources, and the large number of units removed from the donor pool based on our treatment criterion we believe our method upholds SUTVA and leads to representative results. Limitations and potential improvements to this method will be discussed in the limitations section.

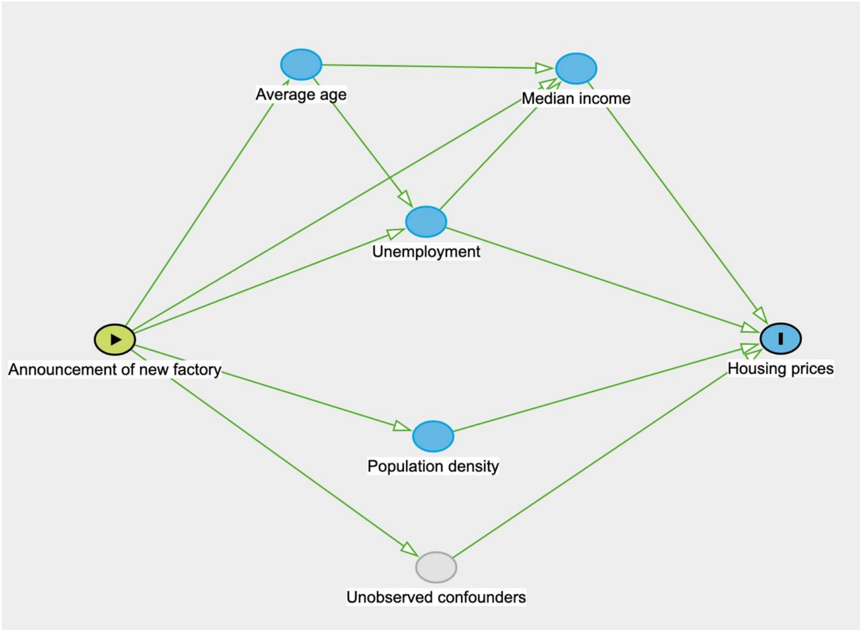

Fig. 2. The directed acyclic graph of our causal design. The yellow node represents the treatment variable, and the blue node on the right represents the outcome variable. The grey node at the bottom represents unobserved confounders, such as the perception of socially desirable housing.

6 Analysis & Results

The treatment effects discussed in this section are the causal effect of the announcement of the building of a new Northvolt factory in Skelleftea municipality on the average house and apartment prices. The outcome measure for each is the average price in SEK per square meter, and the average K/T ratio respectively.

6.1 Synthetic Control

6.1.1 Synthetic Control for Apartments

The treatment effect in 2021 is 6821.9 SEK/m^2. This effect represents the difference between the outcome data for the treated and the synthetic control data in SEK per square meter for 2021. It is also useful to mention that the difference between the groups increased drastically from 2019 to 2021. This indicates a lagged treatment effect, which can be expected given the realization of plans announced in 2017. There is a clear increase in differences between the groups going from 166 SEK/m^2 in 2019 to 3575 SEK/m^2 in 2020 and 6822 SEK/m^2 in 2021. The difference between the treated and control groups for 2021 (6821 SEK/m^2) is about 5-10 times larger than previous differences across the time period 2014-2017. It is clear that there both seem to be a treatment effect and that it is increasing over time. It would be interesting to extrapolate the trend further to 2022, but there was a limitation in median income data until 2021 (data for 2022 comes out Feb 2024). If we could find a proxy for the data or another dataset such as average income we would be able to model the treatment effect more accurately and get a better understanding of how it develops over time. In this model, we calculate the treatment effect as the difference between Skellefteå and its Synthetic control in 2021 of 6821 SEK/m^2.

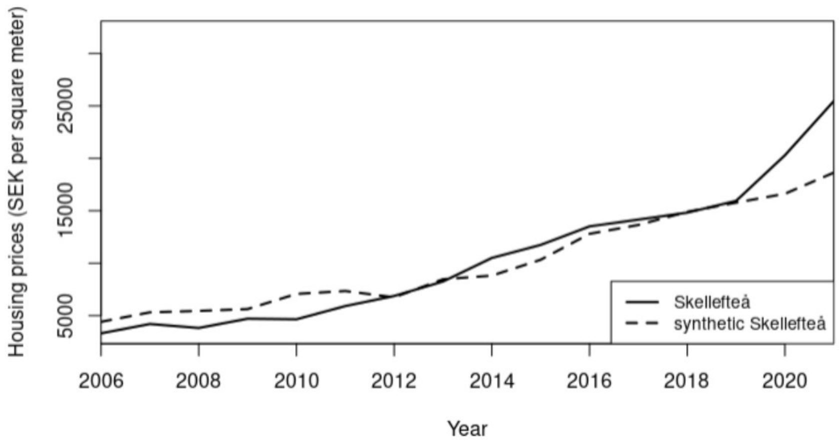

Fig. 3. This is the synthetic control graph for apartments. The treatment effect is the difference between the continuous black line and the dotted line. The weights for Synthetic Skellefteå are reported in Fig 11 in Appendix A.2.

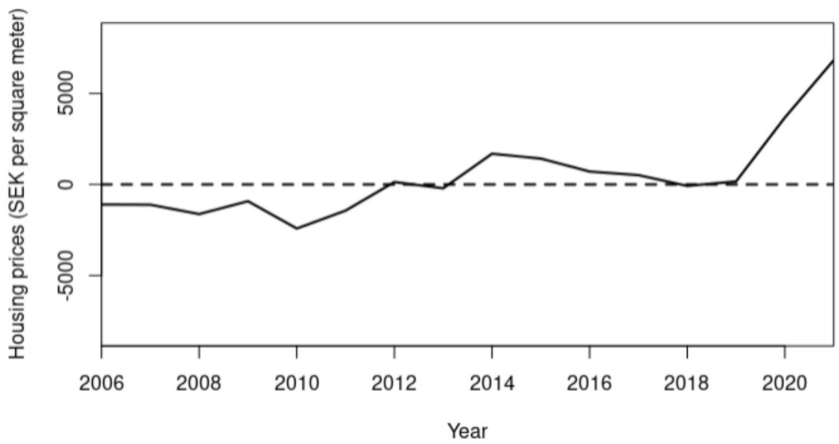

Fig. 4. This is the treatment effect graph for apartments. The treatment effect is the difference between the continuous black line and the dotted line. The table representing treatment effects in each year can be found in Fig 9 in Appendix A.2.

We perform two robustness checks on the Synthetic Control model for apartments, the placebo test across space, and recreating the synthetic control without the highest weighted unit in the model (Västervik). If the trendline and treatment effects of the new synthetic model hold, we find evidence that our estimated treatment effects and not model dependent and hence robust. It is very similar to our previous synthetic control model so we take this to be true (see Fig 12 in Appendix A.2). The placebo test across space presents the outcome trendline of the top weighted units selected to create the synthetic control. Skellefteå shows a clear divergence from the overall trend in the post-2017 period of the data (see Appendix A.2). In relation to the output of the placebo test, we calculate the Mean Standard Prediction Error of the Synthetic control model, as suggested by Abadie et al. (2010) as the best method of calculating the significance of results. We find the post-treatment MSPE for Skellefteå in the Apartment Synthetic Model around 6.8 times larger than the pre-treatment MSPE, which is the highest among control units (see Fig. 14 Appendix A.2). This indicates that our results are statistically significant and that this model allows us to reject the null hypothesis with regard to apartments.

6.1.2 Synthetic Control for Houses

Similar to the apartment model the valuation of houses (as seen by the difference between Skellefteå and Synthetic Skellefteå for each year) see a clear, but lagged treatment effect, that shows up in 2020 and gains in strength in 2021. In 2018 the K/T ratio difference is around 0.068, increasing to 0.216 in 2020 and 0.421 in 2021. There is both an increase in difference and an acceleration of change for every year from 2017 to 2021. The difference between the K/T ratio in 2021 to the time period of 2014-2017 is around 5 times larger and increasing. We have the same limitation on data post-2021 due to no collection of 2022 data on median income which would allow us to extend our model another year and see if the trend of increasing differences in comparison to control holds. However, similar to the Synthetic Apartment model, we see a clear positive post-treatment effect in the difference between Skellefteå and the Synthetic Control. In the case of the Synthetic Control for houses, we estimate a treatment effect of 0.421 K/T ratio in 2021.

We perform two robustness checks on the Synthetic Control model for houses, the placebo test across space, and recreating the synthetic control without the highest weighted unit in the model (Mark). In the recreation without Mark we very limited change in comparison to the original model, indicating that these results are not model dependent (See Fig 18 in Appendix A.3). In our House Synthetic Control placebo model we see that Skellefteå stands out as clearly changing its own trend and diverging from the general trend around the time of treatment (see Fig 16 in Appendix A.3). We find the post-treatment MSPE for Skellefteå in the Housing Synthetic Model around 5.4 times larger than the pre-treatment MSPE, which is the highest among control units (see Fig 20 in Appendix A.3). This indicates that the results for the Synthetic Control House models are statistically significant, and allows us to reject the null hypothesis.

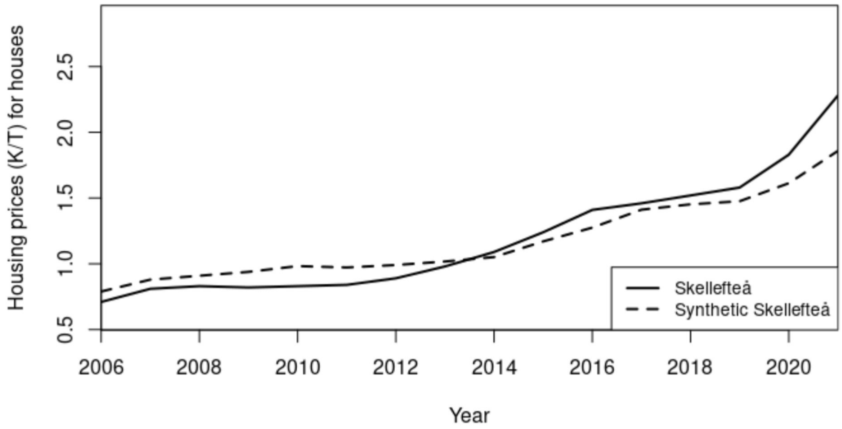

Fig. 5. This is the synthetic control graph for houses. The treatment effect is the difference between the continuous black line and the dotted line. The weights for Synthetic Skellefteå are reported in Fig 17 in Appendix A.3.

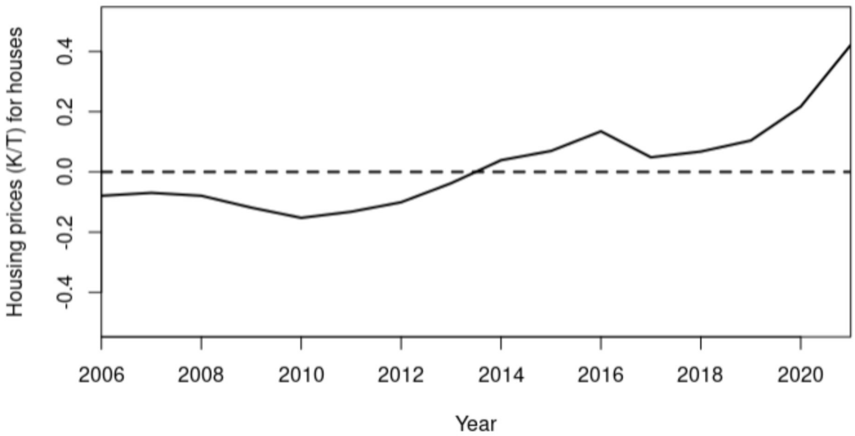

Fig. 6. This is the treatment effect graph for houses. The treatment effect is the difference between the continuous black line and the dotted line. The table representing treatment effects in each year can be found in Fig 15 in Appendix A.3.

6.2 Synthetic Difference-in-Differences

6.2.1 Synthetic DiD for Houses

In the Synthetic DiD for houses, the estimate function measured the outcome trend for Skellefteå and Synthetic Skellefteå from 2007 to 2019. After comparing the differences between the trends the model showed us a treatment estimate of 0.20 in the K/T ratio between the groups over that time. It is notable that after 2019, the trend in the treated unit increases more rapidly than the Synthetic Control unit - clearly increasing the difference between the units. However, given the previously noted rigidity in the estimate and synthdid_plot function, we are only able to calculate the treatment effect for the given period.

We perform a robustness check on the Synthetic Control DiD house model, the placebo test across space, outputting a spaghetti plot showing similar divergence between trends as already presented in the original model. We also calculate the 95% Confidence Interval for this model and find the bounds to be between 0.10 and 0.30. This produces significant doubt over our confidence in the estimated treatment effect as we can only say with 95% confidence that our estimate is within a 50% range of itself. We take this to mean that this model does not provide sufficient evidence to allow us to reject the null hypothesis.

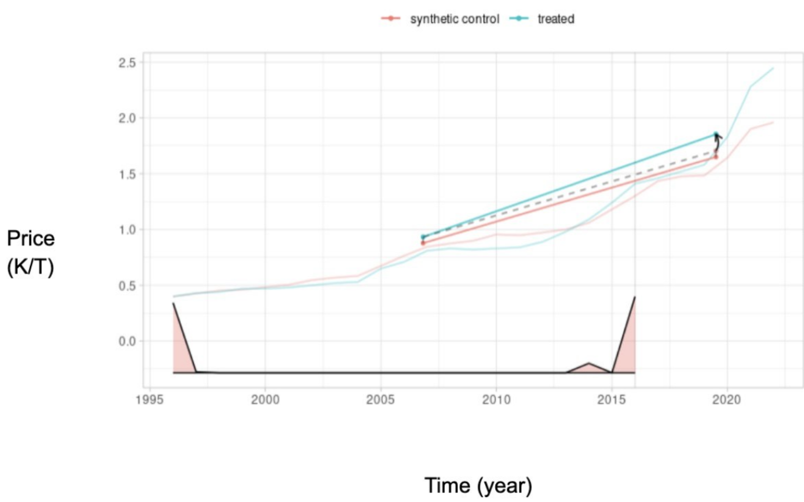

Fig. 7. This is the synthetic DiD plot for houses. It shows a positive treatment effect, calculated as the difference between the blue line and the red line.

6.2.2 Synthetic DiD for Apartments

The Synthetic DiD for apartments model measures the difference in trends for Skellefteå and Synthetic Skellefteå between 2016 and 2019. The estimated treatment effect is $2851.26 SEK/m^2$. Similar to the Synthetic DiD model for houses that after 2019 the trend for the treated outcomes increases at a much higher pace than the trend for control, which is not captured in the model.

We perform robustness checks through placebos and find some differences between treatment and control groups, though these are limited. We calculate the 95% Confidence Interval for this model and find it to be between -1858 SEK/m^2 and 7561 SEK/m^2. This means that the estimated treatment effect is clearly not statistically significant (Appendix A.5).

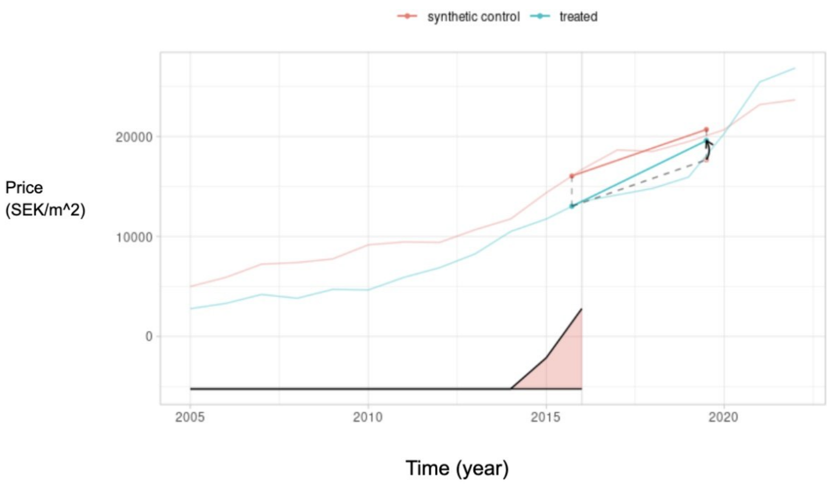

Fig. 8. This is the synthetic DiD plot for apartments. It shows a positive treatment effect, calculated as the difference between the blue line and the red line.

6.3 Comparison Between Synthetic Control Models

There is a clear similarity of magnitude between the 2021 data in comparison to 2014-2017 data across both synthetic control models. The similarities between the models indicate that our treatment effect is accurate at least on an order of magnitude scale. It is encouraging that they are similar, indicating a general effect across the Skellefteå housing market as a result of the announcement and buildout of Northvolt 1. The treatment effects increase from 2017 to 2021 for housing increase in a smoother manner in comparison to the apartment model, where there appears to be more of a clear shock. This can likely be attributed to the nature of the data itself, where the K/T value is less sensitive to year-over-year short-term shocks as it is smoothed out by dividing price by taxable value. The models are comparable in terms of directionality and relative magnitude because both of them represent measures of valuation. However, they are not comparable in terms of their specific numbers.

6.4 Comparison Between Synthetic DiD Models

Neither of the Synthetic DiD model estimates is statistically significant. This could be for a range of reasons. It is likely due to lacking availability of pre-treatment data. While the Synthetic House DiD has poor Confidence intervals for our purposes, they are much better than that of the Apartment model. The House model has 9 more years of pre-treatment data than the Apartment model, meaning that it is probably able to produce a pre-treatment trend with a higher degree of certainty.

Even in the hypothetical case in which the DiD models would produce statistically significant estimates, the fact that it uses 2019 data to estimate a treatment effect that clearly increases significantly post-2019 would make them unreliable in a practical sense.

6.5 Comparison Between Synthetic Control and Synthetic DiD

An interesting difference fact that we only created the Synthetic DiD based on the outcome variable, but in the Synthetic Control Model the covariates of also control units determine the shape of Synthetic Skellefteå. It is possible to add covariates to a Synthetic DiD model, but we did not do that in this case.

A learning that we can take from this comparison is that the Synthetic Control Method can produce estimates with a much higher degree of fidelity from a much shorter timespan of data than the DiD method can. Of course, this relies on the availability of covariates necessary to create the control unit. It is also clear that a clear limitation of the Difference in Differences methods is the requirement of a large amount of pre-treatment data.

7 Limitations

In terms of the data, the primary limitation is that the data on apartment prices are only available from 2005, and data were available only for certain municipalities in Sweden. Moreover, the datasets used only contained measures of certain variables, which limited our ability to check the influence of potential confounders. Thus, this sort of analysis is always prone to confounding bias caused by the unobserved confounders.

The second significant limitation of our analysis is the selection of the initial donor pool which is prone to all sorts of subjective bias. In order to ensure the minimization of those biases, we devised a robust list of criteria that the donor pool observations had to meet. Yet, there are still many regards in which Skellefteå (the treatment unit) differs from the donor pool units, including municipalities' political leadership, primary economic focus, etc.

The third limitation comes from our method of ensuring that the donor pool units were not treated during the study period. To ensure the donor pool units satisfied this condition, we conducted a brief analysis of top headlines during the treatment period. However, due to the vague nature of our treatment (the establishment of a new factory) and its practical similarity to other phenomena like local business expansion or the opening of any other significant institution (e.g., school), our methodology of screening donor pool unit candidates could be improved by a more rigorous data analysis during the treatment period (e.g., checking local employment statistics).

Another limitation is the lack of data on unemployment rates, calculated as the number of unemployed people divided by the total number in the labor force, multiplied by 100. As only the number of unemployed people per municipality was available, we normalized it by dividing it by the total population in the municipality. A more rigorous measure of unemployment would be calculated as a share of the total population eligible to work.

8 Conclusion

We found evidence supporting our hypothesis that Northvolt's announcement of a new battery factory in Skellefteå will have a significant positive impact on local housing prices. In conclusion, our synthetic control model estimates that the establishment of Northvolt Ett in Skellefteå has increased apartment prices by 6,822 SEK/m^2, and house prices by K/T ratios by 0.42. Based on these outcomes we reject the Null Hypothesis.

Given an especially wide confidence interval, the synthetic difference-in-differences treatment estimate for average apartment prices is inconclusive. Our synthetic difference-in-differences result indicates that K/T ratios for housing price valuation increased by 0.2. The Confidence interval for our housing model is also sufficiently wide to prevent us from rejecting the Null Hypothesis with regard to the Difference-in-Differences model.

This study shows the promise of the synthetic control method to be applied in other similar settings to understand developments in economic factors such as housing markets as a result of corporate investment. Models could be made even more accurate by using monthly data and potentially collaborating with local municipalities to measure economic activity at an even more granular level.

The main area in which this study can be improved is donor selection for synthetic control. We apply rigorous criteria given our resources and time at hand, but determining a criterion that can be applied at low technical and time cost (such as a measure of corporate and public investment on a municipal level). This would allow for a more algorithmic selection of donor units, potentially increasing the accuracy of the synthetic control unit, and hence producing a treatment estimate with higher fidelity.

Future studies aiming to investigate the treatment effect of the establishment of a new factory in a municipality using synthetic DiD should utilize datasets with a greater range or higher frequency of pre-treatment data. DiD methods rely on a lot of pre-treatment data to establish statistically significant results. However, we note the practical limitations of using synthetic DiD, as to our knowledge there seems to be a limitation in calculating lagging treatment effects due to the lack of flexibility in the model.[5][6]

A Statement of contributions of the authors

Marcus Moerup: expanded on the initial draft of 'Introduction,' wrote 'Literature Review,' wrote 'Hypothesis', created the DAG, ran the paper through Grammarly, and was responsible for citations and 'References', including formatting in LaTex.

Petter Halqvist: focused on data collection and cleaning, selection and iteration

[5] #casestudy: We clearly defined the bounds of our study. We conducted a thorough analysis of the opening of the new EV factory in Skellefteå, including the media analysis, pre-treatment period analysis, the treatment effect, and the post-treatment period.

[6] #modeling: We accurately produced, presented, and interpreted synthetic control and synthetic DiD models.

of the donor pool, the initial draft of the 'Data' and the 'Methodology' sections, and the code for both synthetic control and synthetic difference-in-differences.

Vladislav Virtonen: focused on descriptive stats, grammar correction, the 'Limitations' section, communication with the professor and TA, and the code for both synthetic control and synthetic difference-in-differences.

A AI use

We used AI primarily to help us resolve unfamiliar code errors in R. We used AI-powered Grammarly to enhance grammar and clarity.

References

Abadie, A., Diamond, A. & Hainmueller, J. (2010), 'Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program'.
URL: *https://www.tandfonline.com/doi/epdf/10.1198/jasa.2009.ap08746?needAccess=true*

Abadie, A., Diamond, A. & Hainmueller, J. (2011), 'Synth: Anrpackage for synthetic control methods in comparative case studies'.
URL: *https://www.jstatsoft.org/article/view/v042i13*

Andersson, J. J. (2019), 'Carbon taxes and co 2 emissions: Sweden as a case study', *American Economic Journal: Economic Policy* **11**(4), 1–30.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W. & Wager, S. (2021), 'Synthetic difference-in-differences'.

Crider, J.-n. (2021), 'Tesla reaches job target for buffalo gigafactory, avoids 41.2 million penalty - cleantechnica'.
URL: *https://cleantechnica.com/2021/12/29/tesla-reaches-job-target-for-buffalo-gigafactory-avoids-41-2-million-penalty/*

De Vor, F. & De Groot, H. L. (2011), 'The impact of industrial sites on residential property values: A hedon-ic pricing analysis from the netherlands', *Regional Studies* **45**(5), 609–623.

Gautier, P. A., Siegmann, A. & Van Vuuren, A. (2009), 'Terrorism and attitudes towards minorities: The effect of the theo van gogh murder on house prices in amsterdam', *Journal of Urban Economics* **65**(2), 113–126.

Ismail, M. & Wilhelmsson, M. (2022), 'New housing investments' effects on gentrification and affordability in stockholm, sweden'.

Nilsen, T. (2017), 'Northern swedish city gets europe's largest battery factory'.
URL: *https://thebarentsobserver.com/en/industry-and-energy/2017/10/northern-swedish-city-gets-europes-largest-battery-factory*

Norrland Living, N. (2022), 'Norrland news - may 2022'.
URL: *https://norrlandliving.com/2022/04/norrland-news/*

Qian, F. & Tan, R. (2021), *The effects of high-skilled firm entry on incumbent residents*, Stanford Institute for Economic Policy Research (SIEPR).

Sable, M. (2007), 'The impact of the biotechnology industry on local economic development in the boston and san diego metropolitan areas', *Technological Forecasting and Social Change* **74**(1), 36–60.

SCB (2023).
URL: *https://www.scb.se/en/*

Statistiska Centralbyrån, . (2023), 'Hur mäter man prisutvecklingen på småhus?'.
URL: *https://www.scb.se/hitta-statistik/statistik-efter-amne/boende-byggande-och-bebyggelse/fastighetspriser-och-lagfarter/fastighetspriser-och-lagfarter/produktrelaterat/Fordjupad-information/hur-mater-man-prisutvecklingen-pa-smahus/*

Svensk Mäklarstatistik, . (2023).
URL: *https://www.maklarstatistik.se/om-oss/*

A Appendix

A.1 Links

The link to the synthetic control data used in this paper can be found here.

The link to the synthetic DiD data can be found here: here.

The Rmd file with code can be found here: here.[7] [8]

A.2 Synthetic Control Figures for Apartments

2006	-1099.7172
2007	-1112.1796
2008	-1626.6534
2009	-918.7978
2010	-2418.8964
2011	-1435.8020
2012	132.6959
2013	-209.9154
2014	1686.8611
2015	1417.8292
2016	709.5622
2017	511.8933
2018	-79.0215
2019	166.2228
2020	3674.4073
2021	6821.9081

Fig. 9. This is a table of synthetic control numerical estimates for apartments. The left column is years and the right column is treatment effects in SEK/m^2

.
.
.
.

[7] #dataviz: We proceed and presented numerous plots, histograms, graphs, and tables in a professional manner, ensuring the axes have titles, and every figure has a caption. Moreover, every figure is accurately referenced in the text.

[8] #algorithms: We professionally worked with R code and included the Rmd file with access to the professor and TA. We showed a logical and easy-to-read series of steps for finding treatment effects and outputting graphs for synthetic control and synthetic DiD.

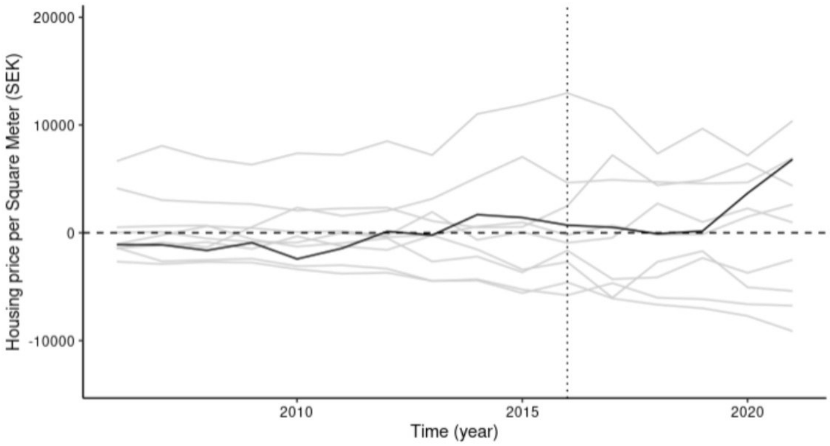

Fig. 10. This spaghetti plot shows the placebo test for the synthetic control for apartments.

Unit	**Weight**
Karlskrona	0.023
Norrtälje	0.016
Falun	0.097
Lindköping	0.020
Västervik	0.374
Mark	0.029
Värnamo	0.061
Katrineholm	0.310
Nässljö	0.071

Fig. 11. This table shows the distribution of weights of the donor for synthetic control performed for apartments.

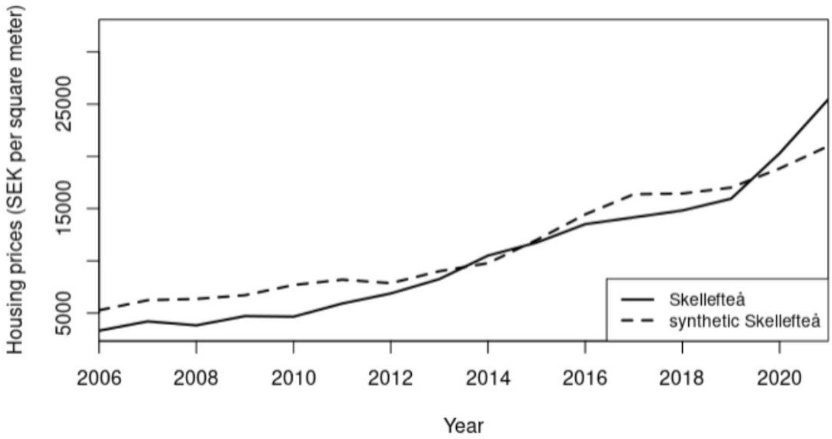

Fig. 12. Synthetic control graph for apartments. It is the result of a robustness check that was performed by removing the donor with the highest weight from the pool: Västervik. The treatment effect is the difference between the continuous black line and the dotted line.

Unit	**Weight**
Karlskrona	0.038
Norrtälje	0.072
Falun	0.122
Lindköping	0.051
Mark	0.071
Värnamo	0.110
Katrineholm	0.400
Nässljö	0.135

Fig. 13. This table shows the distribution of weights of the donor pool of the robustness check for synthetic control performed for apartments, without the observation that had the highest weight in the original donor pool: Västervik.

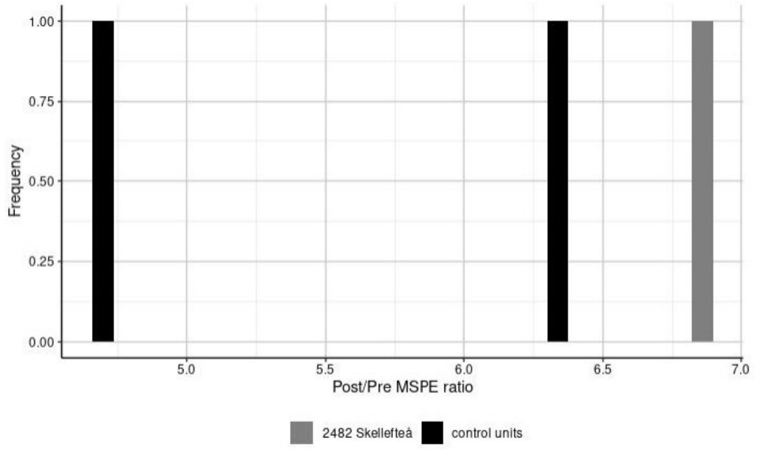

Fig. 14. This histogram shows the MSPE ratio for Apartments, which is another robustness check for the synthetic control for apartments.

A.3 Synthetic Control Figures for Houses

2006	-0.07918768
2007	-0.06958667
2008	-0.07917235
2009	-0.11860747
2010	-0.15254883
2011	-0.13221082
2012	-0.10070806
2013	-0.03755267
2014	0.03891467
2015	0.06965608
2016	0.13478321
2017	0.04857121
2018	0.06767668
2019	0.10417393
2020	0.21706223
2021	0.42148372

Fig. 15. This table shows synthetic control numerical estimates for houses. The left column is years and the right column is treatment effects in K/T, which is the price-to-taxable value ratio, on average per municipality.

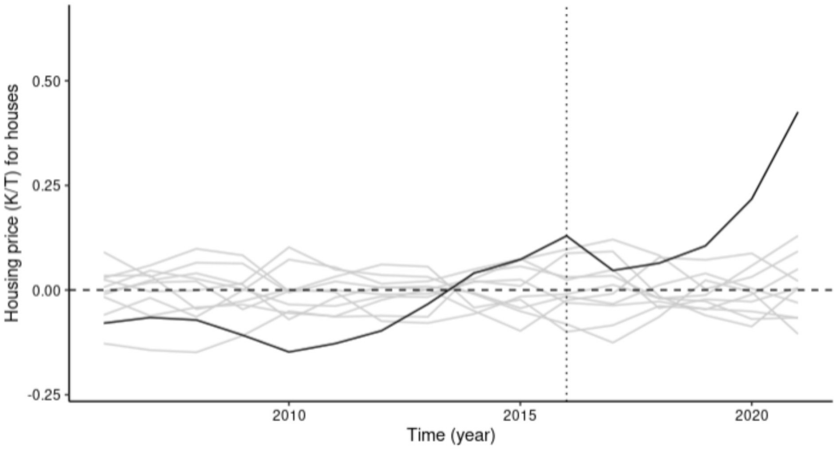

Fig. 16. This spaghetti plot shows the placebo test for the synthetic control for houses.

Unit	**Weight**
Karlskrona	0.014
Norrtälje	0.065
Falun	0.170
Lindköping	0.084
Västervik	0.053
Mark	0.207
Värnamo	0.063
Katrineholm	0.067
Hudiksvall	0.178
Falköping	0.026

Fig. 17. This table shows the distribution of weights of the donor pool for synthetic control for houses.

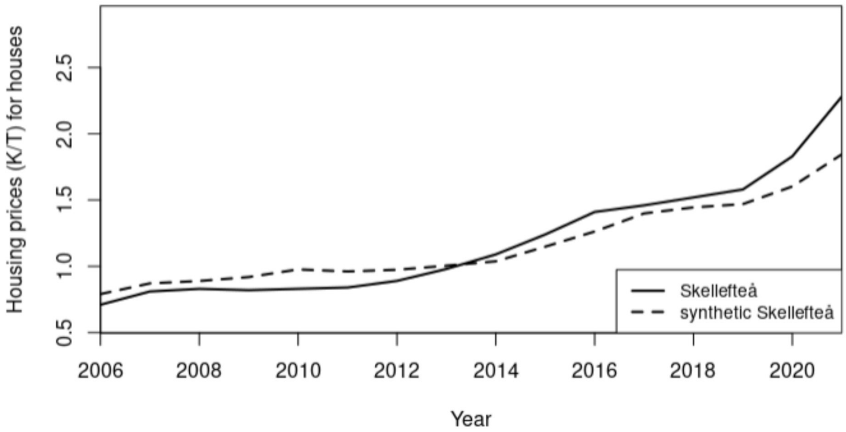

Fig. 18. This is the synthetic control graph for houses. It is the result of a robustness check that was performed by removing the donor with the highest weight from the pool: Mark. The treatment effect is the difference between the continuous black line and the dotted line.

Unit	**Weight**
Karlskrona	0.004
Norrtälje	0.099
Falun	0.157
Lindköping	0.100
Västervik	0.164
Värnamo	0.084
Katrineholm	0.102
Hudiksvall	0.031
Falköping	0.156

Fig. 19. This table shows the distribution of weights of the donor pool of the robustness check for synthetic control for houses, without the observation that had the highest weight in the original donor pool: Mark.

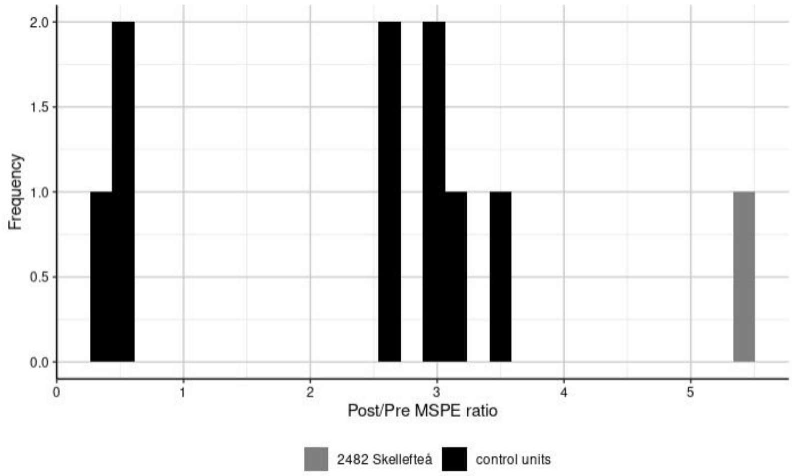

Fig. 20. This histogram shows the MSPE ratio for houses, which is another robustness check for the synthetic control for houses.

A.4 Synthetic DiD Figures for Apartments

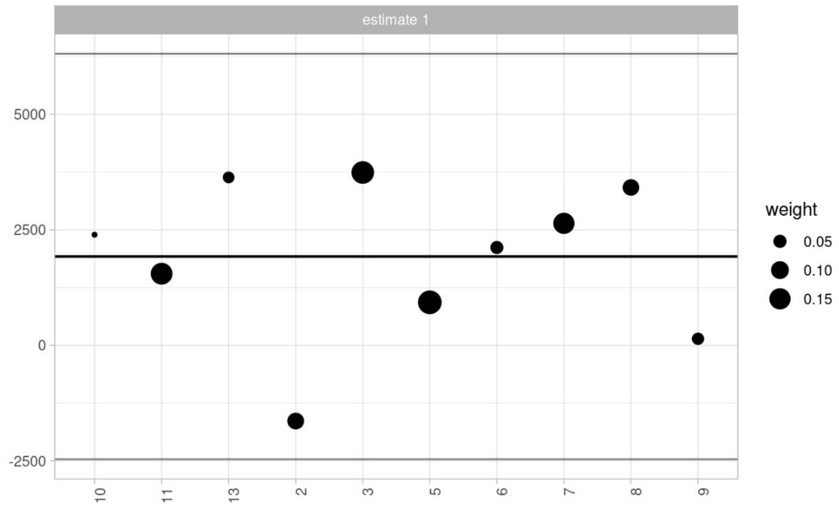

Fig. 21. This table shows the distribution of weights of the donor pool for synthetic DiD performed for apartments.

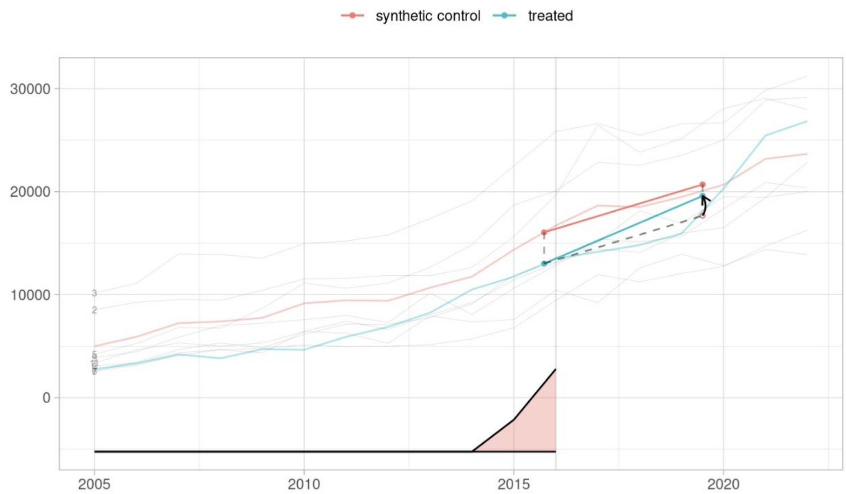

Fig. 22. This spaghetti plot shows the placebo test for the synthetic DiD for apartments.

A.5 Synthetic DiD Figures for Houses

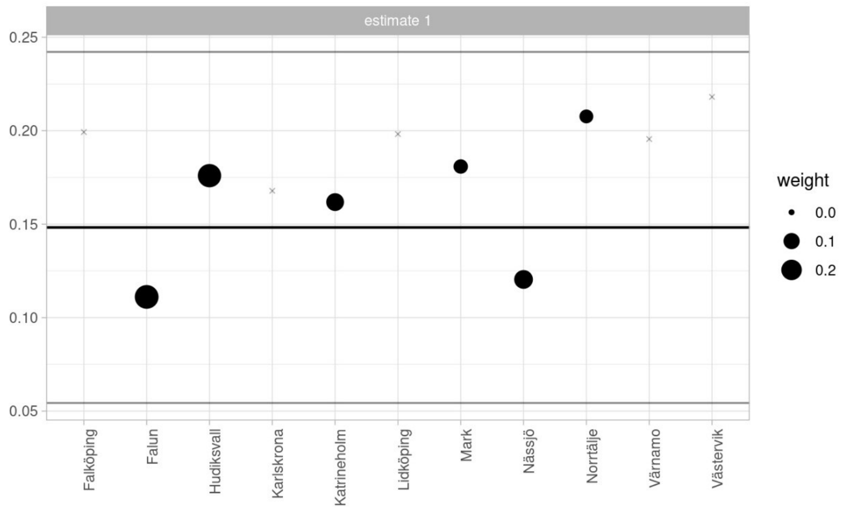

Fig. 23. This table shows the distribution of weights of the donor pool for synthetic DiD for houses.

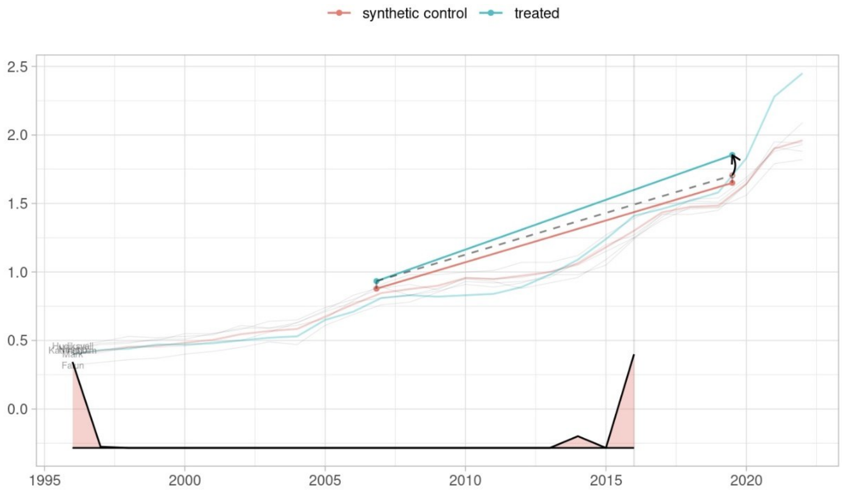

Fig. 24. This spaghetti plot shows the placebo test for the synthetic DiD for houses.

A.6 Donors Map

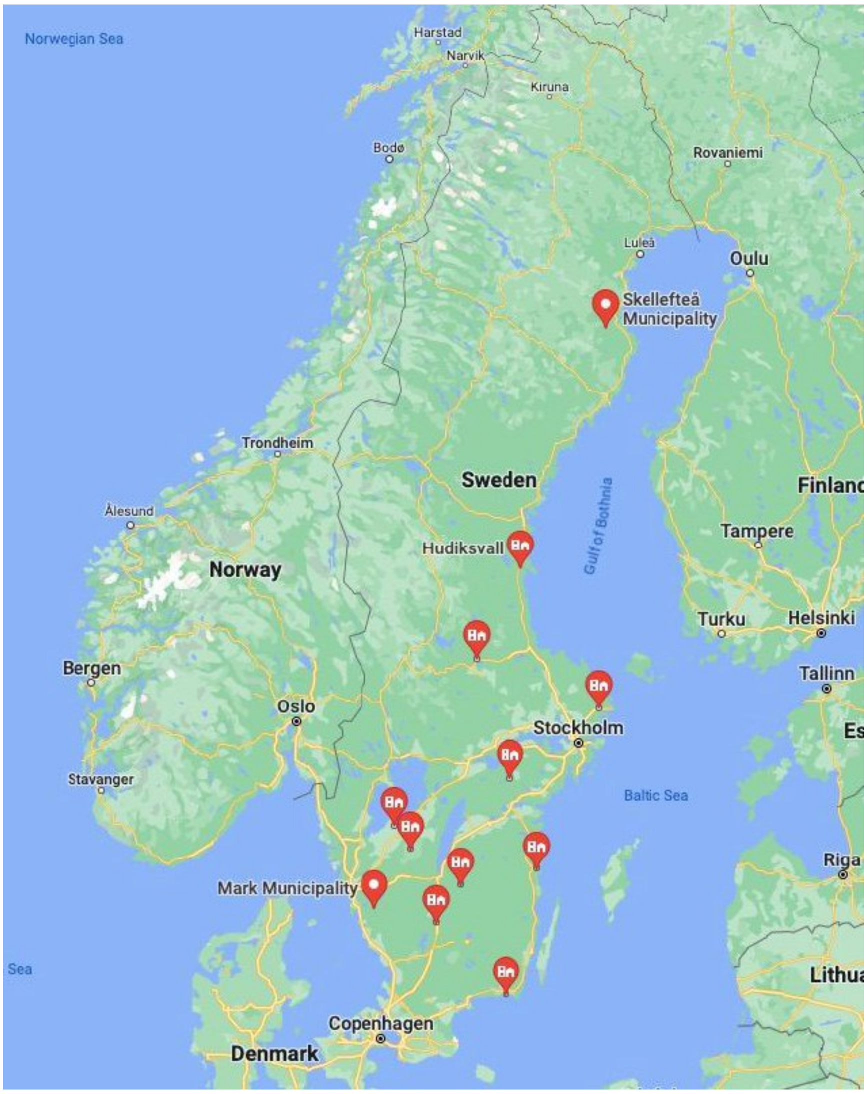

Fig. 25. This is the synthetic control donor map with 13 donor units' locations in Google Maps.