ESTIMATING THE ANNOUNCEMENT EFFECT OF A METRO EXPANSION ON HOUSING PRICES: EVIDENCE FROM VIENNA

By

Felix Maierhofer

Submitted to
Central European University
Department of Undergraduate Studies

In partial fulfillment of the requirements for the Bachelor of Arts in Philosophy, Politics and Economics (PPE)

Supervisor: Prof. Robert Lieli

Vienna, Austria

2025

Author's Declaration

I, the undersigned, **Felix Maierhofer**, candidate for the Bachelor's degree in Philosophy, Politics and Economics (PPE) declare herewith that the present thesis is exclusively my own work, based on my research and only such external information as properly credited in notes and bibliography. I declare that no unidentified and illegitimate use was made of the work of others, and no part of the thesis infringes on any person's or institution's copyright. I also declare that no part of the thesis has been submitted in this form to any other institution of higher education for an academic degree.

Vienna, May 1, 2025

Felix Maierhofer

Copyright Notice

Abstract

Urban policies, for instance changes to a city's public transport infrastructure, usually change the attractiveness of locations within a city, impacting housing prices. In this thesis, I attempt to estimate the causal effect of the U2 x U5 metro expansion in Vienna, announced in 2014, on housing prices in locations in spatial proximity to future stations. Since construction has not yet been completed, I focus on the effect of the announcement, which should alter people's expectations regarding the future value of a location and thus already impact prices. To estimate the effect, I use a binary Difference-in-Differences (DiD) design and a continuous quasi-DiD design with various sensitivity checks. I also use a Random Forest Model to attempt to isolate the causal effect through a different route. The dataset used contains data on apartment sales surrounding prospective stations and spanning the years 2012-2014, which I collected using a private firm's software. My findings suggest that the announcement decreased housing prices close to new stations by around 9-11%. Although the estimated magnitude of the effect changes depending on the specification, the negative sign of the effect is robust to different specifications. This research indicates that the supposed positive effects of public transport construction projects on nearby residents may be outweighed by negative side-effects after construction is finished or nuisance during construction, which must be taken into consideration by policymakers.

Acknowledgements

CEU eTD Collection

Contents

CEU eTD Collection

List of Figures

List of Tables

Chapter 1

Introduction

Any urban intervention will likely render locations within the city more or less attractive. As a result, housing prices should move up or down, because they reflect both the value of the location's plot of land, which is affected by the intervention, and of the structure built on it, which is unaffected. This theoretical prediction has been investigated by a large empirical literature that has examined different kinds of urban interventions, for instance parks (Wu et al. 2017) or garbage incinerators (Kiel and McClain 1995). Among them, changes to cities' public transport infrastructure, for instance extensions of the metro system, have received much attention, partly because of their crucial importance for transportation in many modern cities. However, there are potential theoretical reasons for both rising and falling housing prices in response to a metro expansion. On the one hand, the metro might improve the connectivity of locations close to new stations, pushing prices up. On the other hand, metro stops are often hotspots for homelessness and small-scale crimes. In addition, better connectivity might render surrounding neighborhoods more busy, which may be seen as negative by some and consequently decrease prices.

To determine which theoretical mechanisms prevails, previous studies have examined the effects of metro extension projects on housing prices in several cities around the world, including Beijing (Li, L. Chen, and Zhao 2019), Hangzhou (Wen et al. 2018), Wuhan (R. Tan et al. 2019), Shanghai (Zhou et al. 2021), Shenzhen (He et al. 2024), Singapore (Diao, Leonard, and Sing 2017), Santiago (Agostini and Palmucci 2008; Lopez-Morales et al. 2023), Bogotá (Cárdenas, Gallego, and Urrutia 2023), New York City (Guan, M. J. Tan, and Peiser 2021), Los Angeles (Keeler and Stephens 2023), Warsaw (Trojanek and Gluszak 2018), Istanbul (Demircan and Oguztimur 2023), Sydney (Y. Chen et al. 2019), Madrid (Dorantes, Paez, and Vassallo 2011) and Vienna (Wieser 2011; Dengg 2018; Hamberger 2023).

This thesis adds to the existing literature by studying the case of Vienna's most recent metro expansion project, which consists in building a new U5 line and extending the existing U2 line. Since construction has not been completed, it is not yet possible to estimate the causal effect after new stations became operational. However, one would expect prices to already adapt between the announcement and completion of a project, because new information from the an-

nouncement alters the expected future utility from owning an apartment. This prediction has been confirmed by some of the aforementioned empirical studies that have noted announcement effects, including Agostini and Palmucci (2008), Dengg (2018), Y. Chen et al. (2019), and Cárdenas, Gallego, and Urrutia (2023). Thus, the research question I will exclusively focus on concerns the effect of the announcement of the U2 x U5 project, made on June 27, 2014 (wien.at 2014), on housing prices. The announcement included the location of new U2 and U5 stations, the planned construction timeline, as well as changes to the tram and bus networks over the following decade. Construction of the U2 x U5 project would start in 2018 and would be carried out in two stages, with the first finishing in 2023 and the second at an indeterminate later time.

The U5 had previously been missing among Vienna's existing U1, U2, U3, U4 and U6 lines and will connect the city center with the western part of Vienna. The U2 line, which used to connect the east of Vienna with the city center, will be extended to the south. The project comprises a total of 15 new stations, among them 8 exclusively for the U5 (the stations Hernals, Elterleinplatz, Michelbeuern AKH, Anne-Karlsson-Park, Frankhplatz, Rathaus, Volkstheater, Museumsquartier and Karlsplatz) and 6 exclusively for the U2 (the stations Wienerberg, Gußriegelstraße, Matzleinsdorfer Platz, Reinprechtsdorfer Straße, Pilgramgasse and Neubaugasse). The two lines will cross at the station Rathaus.

It is important to note that already in the preceding years and decades, rumors had been surrounding the possible construction of Vienna's missing U5 metro line. In the months leading up to the announcement, plans were at times leaked to the press. In particular, on October 4, 2013, an article claimed to have received insider information on the location of future stations (Natmessnig 2013). A similar article was published on January 29, 2014 (Stuhlpfarrer 2014). Because these leaks might already have influenced people's beliefs and thus the demand for apartments before the official announcement date, prices might have reacted earlier, possibly rendering the effect after the official announcement date negligible. For this reason, it is also necessary to examine the possible causal effects of the two leak dates.

Another important point to consider is that changes to a city's metro system are often accompanied by changes in other means of public transport. It might thus be the case that the benefits of proximity to a new metro station are consumed by the ensuing lack of other means of public transport previously present in a neighborhood. Although the announcement studied also included changes to the bus and tram infrastructure, these changes were not in the vicinity of new stations wien.at (2014).

The identification strategy I will use is a Difference-in-Differences (DiD) design. This method has frequently been used in this field, including in the aforementioned studies by Agostini and Palmucci (2008), Trojanek and Gluszak (2018), Cárdenas, Gallego, and Urrutia (2023), and He et al. (2024). Compared to hedonic pricing models familiar from research on urban valuation, this method allows for a more convenient estimation of causal effects because it does not require data on all potential confounding factors for estimates to be unbiased. By contrast, controls are

2

only necessary insofar as they increase the credibility of the key identifying assumption, the parallel-trends assumption (PTA). In addition to performing various sensitivity checks, I also estimate the causal effect using a random forest machine learning model, which does assume that data on all price-relevant factors is available.

Among previous studies of Vienna, Wieser (2011) examined the effect of the metro expansion over the years 1987-2004, which included the construction or extension of the lines U1, U2, U3, and U6, on land prices by estimating a hedonic land price model using non-parametric kernel regressions. For apartments, they find that positive price effects are identifiable as far as 2000 meters from stations. Dengg (2018) estimated the effect of the previous extension of the line U2 on the prices of vacant land in Vienna using a DiD design and data for the years 2000-2017. In contrast to Wieser (2011), they find a positive and significant effect only within 370 meters from stations. Within 100 meters, the estimated causal effect is about 300%. However, the study does not include a test of the parallel trends assumption on pre-trends, and it is thus unclear whether the DiD-estimate of the causal effect is indeed unbiased.

The study closest to mine was conducted by Hamberger (2023), which also estimated the effect of the U5 construction on housing prices using a DiD approach. However, there are several key differences between the two studies. First, the two papers seek to answer related, but nonetheless different research questions. Hamberger (2023) was concerned with estimating the impact of the lengthy construction period on prices; this paper is solely interested in the immediate announcement effect of the construction project. One main theoretical reason behind this shift of focus is that a longer timeframe increases the incidence of other events that might obscure the causal effect of the announcement itself. Furthermore, while it is true that nuisance during the lengthy construction period makes locations close to it temporarily less attractive, one would also expect prices to be forward-looking and to already reflect the expected value of the location after construction is finished. This expectation is based on a classic no-arbitrage argument since - if prices did not already reflect expected future values - real estate investors could make riskless after-tax profits by buying an apartment at low prices during the construction period and selling it at higher prices after construction is finished. However, since transaction costs are substantial in real estate transactions, this no-arbitrage argument might not fully hold. This implies that new information might not be fully priced in immediately after the announcement but gradually as the opening date gets closer. On the other hand, an important detail to consider is that construction did not start right after the announcement, but around 4 years later in 2018. Thus, a price decrease after the announcement in anticipation of nuisance during construction seems less likely, since apartments may be bought to be sold again within the next four years before the nuisance of construction takes effect. For these reasons, the focus on the immediate announcement effect seems justified.

The second important difference is that similar to Wieser (2011) and Dengg (2018), Hamberger (2023) uses data for the years 2008-2019 from a publicly-available dataset of real estate transactions maintained by the Viennese city administration, which only includes houses and apart-

3

ment buildings, but not individual apartments. By contrast, this study uses a dataset I compiled from private firm data and that includes only apartment sales for the years 2012-2014. Because apartments are disproportionately located in the districts in spatial proximity to future metro stations, this implies a larger sample size and more statistical power, despite spanning a shorter timeframe.

Third, although both studies use a DiD design, the definitions of the treatment and control group are different, which is another potential explanation for the violation of the parallel trends assumption in Hamberger (2023) as explained below, but not in this study. While Hamberger (2023) considered observations as treated if they were within a 1000 meter radius around a new metro station, and classified them as control if within a 1000 meter radius around "counterfactual" stations that were included in previous plans but not in the final ones, my treatment/control distinction relies on concentric circles around a metro station: if an observation is located within the smaller circle, it is a treated observation; if it is located between the smaller and larger circle, it is a control observation. Since treated and control observations are thus located in spatial proximity to one another, it is more likely that they are alike in important respects and that their prices will follow parallel trends.

Fourth, Hamberger (2023) focused only on metro stations of the new line U5, while I am also considering new U2 metro stations, which were planned and are currently under construction as part of the same metro extension project.

Hamberger (2023)'s estimates yield a negative and significant coefficient but it must be noted that the key underlying assumption of the DiD approach, the parallel trends assumption, did not pass a formal test in the pre-treatment period. Therefore, this estimated coefficient is likely biased by confounding factors that affect the treated and the control group differentially over time.

In light of the differences mentioned and of Hamberger (2023)'s observation that the parallel trends assumption seems not to hold, the present study offers a relevant and novel contribution to the existing literature along a few dimensions. First, because it studies the announcement effect of the project rather than an effect upon completion, it can potentially offer insights into the informational efficiency of the Viennese housing market. Second, the study is especially interesting since research on metro expansions in European cities is still relatively scarce although many European cities rely heavily on public transport infrastructure. This thesis thus offers partial insights into how such projects are valued by residents, especially in contexts where public transport infrastructure is already well developed. Of course, potentially rising prices cannot be taken to reflect all the value-added of such a project, since other residents - as well as non-residents like tourists - will benefit from better connectivity in the city, despite living further away from new stations. To some extent, these additional benefits, which make the city as a whole more attractive, should drive up prices in the entire city, but this effect is outside of the scope of investigation of this study. Third, due to the significant data collection effort, this study benefits from more fine-grained apartment sale data to estimate the causal effect more reliably.

I find a negative effect of the metro expansion announcement on housing prices close to stations, which is in accordance with Hamberger (2023)'s results, despite the shortcomings of her study.

The remainder of the thesis is organized as follows. In Chapter 2, I will mention further relevant theoretical considerations and explain the method and specifications used to estimate the causal effects of the announcement on housing prices. Chapter 3 will describe the data collection process, highlight key descriptive statistics, and formally test the parallel-trends assumption. Chapter 4 will present and discuss the results of the DiD analysis, while Chapter 5 will conclude.

Chapter 2

Methodology

There are a few mechanisms through which the metro expansion project could have affected the value of locations and thus housing prices. First, there may be a direct positive effect in anticipation of the metro connecting a location better to other parts of the city, which reduces the future time commuting to work and leisure activities. But better connectivity might also attract small businesses like shops or restaurants in anticipation of more customers, which can itself contribute to the attractiveness of a location and have positive second-order effects on housing prices. However, there are also possible mechanisms through which the metro construction project could have rendered locations less attractive, pushing prices down upon announcement. For instance, metro stops are often hotspots for drug abuse, small-scale crimes, and homelessness. Furthermore, better connectivity might render a location more busy, which may be perceived as negative by prospective residents. Last, during construction, locations close to future stations may be less attractive due to construction noise, impaired views, traffic disruptions or obstructed footpaths. Since the construction of the metro would be spanning several years, nuisance during construction might significantly decrease buyers' willingness to pay for apartments, even if the benefits upon completion would outweigh the costs (see Hamberger (2023)).

To empirically estimate the causal effect of the metro construction announcement on housing prices, it is insufficient to conduct a naive comparison between average prices of apartments close to stations before and after the announcement date. This is because the causal effect of the announcement might be obscured by aggregate temporal price trends in the housing market or price trends specific to neighborhoods close to stations. In other words, even in the absence of the announcement, prices might have changed close to stations, simply reflecting changes to the attractiveness of these locations or of the city as a whole. This also introduces a potential reverse-causality problem, because it might have been the case that the location of prospective stations was selected precisely because of certain price trends in candidate neighborhoods.

Therefore, to estimate the causal effect I use a classic Difference-in-Differences (DiD) design with binary treatment definitions, as well as a quasi-DiD design with a continuous treatment definition. The binary specifications require a division of observations into two groups along two dimensions - a temporal before-after dimension and a treatment-control dimension. There

6

are a few relevant considerations regarding the DiD setup that merit attention. First, what should ultimately matter for the price that potential buyers of an apartment are willing to pay for its location is not its distance to the metro rails, but its distance to the nearest metro station. Second, in the presence of urban infrastructure like buildings, streets, etc., it is the walking distance to the nearest station that should matter, not the straight-line distance. Third, since metro stations in Vienna have at least two entrances, it would seem ideal to calculate the walking distance from each apartment to the nearest station's closest entrance.

To accommodate these concerns, in the binary DiD design, observations will be classified as part of the treatment group if they are within a certain walking distance from the nearest metro station. Control observations, on the other hand, are those outside the given walking distance. Despite multiple entrances to metro stations, I resorted to one address for the location of metro stations. However, this imprecision is unlikely to significantly bias the estimates of the announcement effect, since information on the exact location of entrances was not contained in the announcement.

In the first specifications presented in Table 3, treatment observations are those with a walking distance of at most 600 meters to the nearest stations, while control observations are further away. This walking distance roughly corresponds to an average walking duration of 7 minutes, which seems a reasonable cutoff within a city with a well-connected public transport system. Because this cutoff is nonetheless arbitrary, I perform sensitivity checks with treatment-control distinctions based on 400 and 800 meter walking distances (see Table A1). It also seems likely that there is no clear walking distance cutoff, but that the effect phases out the further away from a station an apartment is located. I accommodate this concern through a continuous quasi-DiD design where the treatment variable of interest is simply the walking distance from an apartment to the nearest metro station, which will give a better idea of when the treatment effect has faded out.

In addition, since it is possible that nuisance during the construction period, for instance caused by construction noise, can lead to a negative immediate effect of the announcement, I also experiment with a donut-like treatment definition. According to this definition, observations are classified as treated if they are between a walking distance of 200 and 600 meters away from the nearest station, and as control if they are further away. Observations within a walking distance of 200 meters of the nearest station, where one would expect the largest disutility during construction, are dropped.

Another important issue is the definition of periods after and before the project's announcement. The project was announced by the Viennese city administration on June 27, 2014 (wien.at 2014). However, there might have been cases where the buyer and seller had agreed to a price prior to the announcement date, but where the purchase agreement was only signed after that date. Such potential issues might lead to a wrong classification of apartments as belonging to the AFTER group, although they were unaffected by the announcement and should thus be part of the BEFORE group.

7

To circumvent such issues, in the main specifications of the DiD model, apartments are classified as belonging to the BEFORE group if the purchase agreement was signed before June 1, and as belonging to the AFTER group if it was signed after September 1. This leaves a time-frame of around a month (May 31st - June 27) between the BEFORE group cutoff date and the announcement date, and a timeframe of around 2 months (June 27 - September 1) between the announcement date and the AFTER group cutoff date. Thus, it should be ensured that there is no mischaracterization of apartments as belonging to the BEFORE or the AFTER group. As a sensitivity test, I also extended the period between the announcement from September 1 to October 1 (see Table A2).

Another important issue is the possibility that the location of future stations was anticipated before the official announcement was made. This could lead to price changes already before the announcement and possibly render the effect after the announcement date negligible. In this context, this is a real possibility since on October 4, 2013 (Natmessnig 2013), and January 29, 2014 (Stuhlpfarrer 2014), newspaper articles claimed to have insights into the locations of future stations.

In retrospect, these articles identified only some of the stations correctly. What matters more for the evolution of housing prices, however, is not the truth or falsity of such leaks, but the beliefs of potential buyers and sellers of apartments regarding their truth and falsity. If they were sufficiently confident in the truth of these leaks, one would expect the leaks to already have an effect on housing prices. Depending on the size of this "leak" effect, it is possible that the effect at the official announcement date might become negligible, because prices already reacted to the news when they were leaked. In order to examine this potential issue, I will adapt the definition of the AFTER variable in the model with the treatment variable based on a walking distance of 600 meters. The AFTER variable used to check the first leak date takes the value 1 if the apartment was sold after December 1, 2013, roughly 2 months after the article was published, and the value 0 if it was sold before the publishing date on October 4, 2013. For the second leak date, it takes the value 1 if the apartment was sold after April 1, 2014, again roughly 2 months after the publishing of the article, and the value 0 if it was sold before January 29, 2014, the publishing date (see Table A2).

Another potential issue is that insiders might have bought apartments at locations close to stations prior to the official announcement in anticipation of later price increases. Although this hypothesis is hard to test, it seems unlikely that this phenomenon could have been of sufficient magnitude to significantly bias the model estimates.

To assess the statistical significance of the coefficient estimates, I will implement the DiD design in several regressions of the form:

$$Y_i = \beta_0 + \beta_1 \cdot AFTER_i + \beta_2 \cdot TREATED_i + \delta \cdot AFTER_i \times TREATED_i + X_i + \epsilon_i \quad (1)$$

where Y_i is the outcome variable of interest, price per m², and $AFTER_i$ is a dummy variable that is 1 if observation i is part of the AFTER group and 0 if it is part of the control group. $TREATED_i$ is the treatment variable in the given specification, which is either a dummy variable that is 1 if observation i is part of to the treated group and 0 if it is part of the control group, or a quantitative variable capturing the walking distance from the nearest station. X_i is a vector of controls that consists of 14 station dummy variables and 5 non-station control dummies. Among station dummies, the dummy for the station closest to a given apartment is assigned the value 1 for that observation, while all others are assigned the value 0. Non-station control dummies take the value 1 if the observation was taken not from the vicinity of a station, but from around 1 of 5 other locations in Vienna (see Chapter 3, Section 1). Note that apartment size is already effectively controlled for by choosing price per m² as the outcome variable. δ is the DiD estimate of the Average Treatment Effect on the Treated (ATT) and thus the coefficient of interest. If the parallel-trends assumption (PTA) holds, i.e. the mean price per m² in the treatment and control groups would have evolved in a parallel fashion in the AFTER period in a counterfactual scenario without the metro announcement, then any difference in the AFTER period between the treatment and control group that is above and beyond the initial ("baseline") difference must be due to the metro announcement. In other words, if the PTA holds, then δ captures the causal effect of the announcement on housing prices. Potential violations of the PTA could be caused by systematic differences between treatment and control group apartments that lead to a larger change in demand for one than the other over time, differentially affecting treatment and control group prices. For instance, it could be the case that apartments closer to stations are systematically older, and that older apartments experience a larger increase in demand over time.

The plausibility of the PTA in the given context is discussed in detail in Chapter 3, Section 2, where I present plots of the temporal evolution of mean treatment and control group prices and evidence from a formal test of parallel pre-trends.

It is also possible that there is treatment effect heterogeneity on the station level, with the announcement effect being stronger close to stations with certain characteristics than for other stations. First, the stations Karlsplatz, Museumsquartier and Volkstheater were U2 stations already prior to the U2 x U5 project, and will simply be turned into U5 stations. Thus, one would expect a price change only insofar as residents value the U5 connection to the west of Vienna more than the existing U2 connection to the east. As a sensitivity check, I check the results of the main specifications change if observations closest to the stations Karlsplatz, Museumsquartier and Volkstheater are dropped (see Table A3).

Second, it seems plausible that the magnitude of the effect will depend on the prior connectivity of an area. Other things equal and assuming diminishing marginal returns to connectivity, the effect might be smaller for stations built in better connected areas, which will likely be the center. It thus is possible that more central stations systematically exhibit smaller treatment effects than outer stations. On the other hand, since construction for the more central

9

stations (Karlsplatz, Museumsquartier, Volkstheater, Rathaus, Frankhplatz, Neubaugasse, Pil-gramgasse, Reinprechtsdorfer Straße and Matzleinsdorfer Platz) was to be finished earlier than for outer stations (Anne-Karlsson-Park, Michelbeuern AKH, Elterleinplatz, Hernals, Gußriegel-straße, Wienerberg), apartments close to inner stations will be affected by construction nuisance for a shorter period of time, which might imply a larger announcement effect. Since the two theoretical mechanisms pull in different directions, it is unclear whether to expect a larger or smaller announcement effect for apartment close to central stations compared to outer stations. Due to time and space constraints, I will not further pursue this question in this thesis.

To test the robustness of the DiD estimates, I also train a random forest machine learning model on control group data from before and after the announcement, and treatment group data before the announcement. This allows the model to study connections between a range of variables and the price per m² of an apartment, but in the absence of any announcement effect. The variables used include the walking distance from an apartment to the nearest station, the apartment's sale date and station dummies as well as dummies denoting control locations not directly associated with any station, which account for neighborhood characteristics. I then used the model to predict counterfactual prices per m² in the treatment group in the period after the announcement. The average difference between actual prices per m² and the counterfactual prices per m² predicted by the random forest model then gives another estimate of the ATT, which should capture the causal effect of the announcement (see Table A4).

Chapter 3

Data

Previous studies in Vienna (Wieser 2011; Dengg 2018; Hamberger 2023) used a publicly available dataset of housing prices maintained by the Viennese city administration. Unfortunately, this dataset has since been taken offline due to data protection reasons and could not be used for this research. Thus, I built a new dataset from data provided by the company ImmoUnited.

3.1   Data Collection and Cleaning

The software I had access to contained data on all Austrian real estate transactions, taken from the official purchase agreements, and allows searching for transactions in a given timeframe within circles of different radii around a target location. The timeframe was set to 2012-2014 to get data both after the announcement date on June 27, 2014, and a longer before-period. The latter is needed for a test of parallel pre-trends. The target locations were all 15 new stations of the metro lines U2 (Wienerberg, Gußriegelstraße, Matzleinsdorfer Platz, Reinprechtsdorfer Straße, Pilgramgasse and Neubaugasse) and U5 (Hernals, Elterleinplatz, Michelbeuern AKH, Anne-Karlsson-Park, Frankhplatz, Volkstheater, Museumsquartier and Karlsplatz), as well as the station Rathaus for both lines U2 and U5. The station addresses taken were Wienerbergstraße 15, Gußriegelstraße 32, Matzleinsdorfer Platz 1, Bacherplatz 4, Pilgramgasse 17, Mariahilfer Straße 71, Heigerleinstraße 84, Elterleinplatz 1, Michelbeuernsteg 1, Spitalgasse 33, Frankhplatz 4, Landesgerichtsstraße 4, Burggasse 2, Mariahilfer Straße 1, Karlsplatz 14 and Landesgerichtsstraße 4, respectively. Since the radii available were 100-400, 500, 1000 and 1500 meters, I set the radius to 500 meters and randomly picked treatment group observations from the list of observations within the circle both in the before and after periods. In the same manner, I picked a total of 125 control observations from five random locations in Vienna that seem too far from new stations to be affected, but are still in close neighborhoods so that it is more likely that prices would follow parallel trends. These locations are Keinergasse 14, Flachgasse 7, Ruckergasse 50, Karolinengasse 20 and Thimiggasse 22. Additional control observations were then picked manually on a map within a 1000 meters circle, but outside of the smaller 500

11

meter circle, again both in the before and after periods. The fact that these control observations are close to treatment observations should again render it more likely that aggregate price trends are parallel over time. For the spatial distribution of observations see Figure 1.
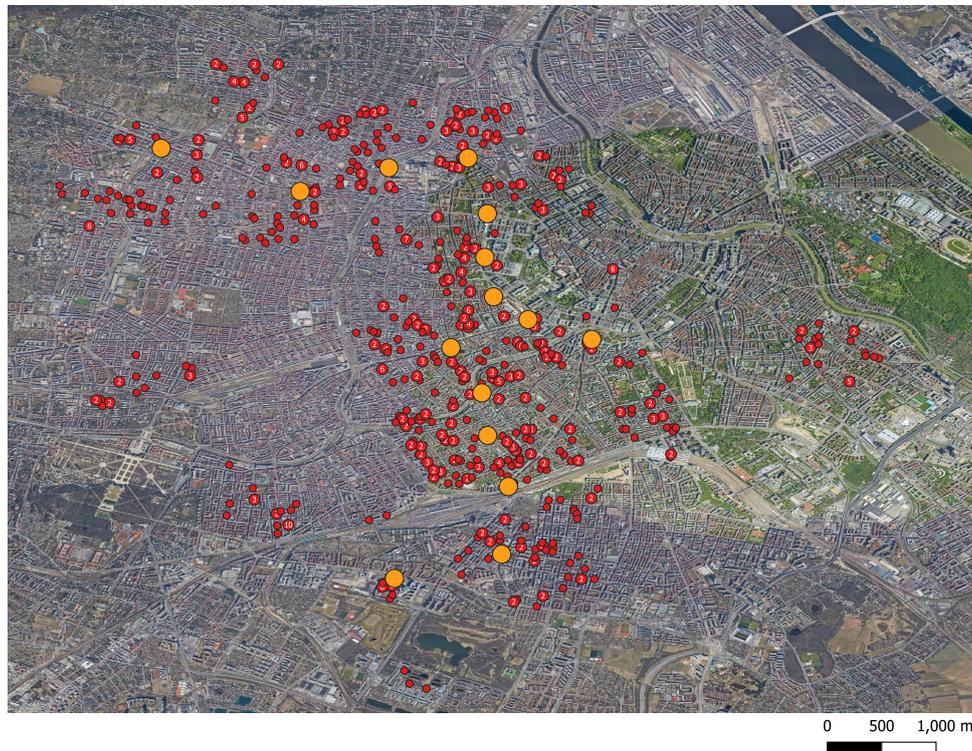

Figure 1: Map of Vienna generated using QGIS. Orange dots indicate the location of new metro stations while red dots indicate the location of apartments in the sample. Numbers indicate the numbers of apartments if they are in very close spatial proximity to one another.

In contrast to the dataset used by the previous Viennese studies conducted by Wieser (2011), Dengg (2018), and Hamberger (2023), which only includes transaction of private houses and apartment buildings, this dataset solely comprises apartments. Variables in the data set include the sale price of the apartment (€), area (m²), the exact address, the date on which the purchase agreement was signed, and a range of other apartment characteristics. Unfortunately, too many missing values rendered the apartment characteristics unfit for use. Regarding area, the dataset includes two variables: usable floor area and usable floor area according to the Nutzwertgutachten (NWGA). Because the NWGA is an Austrian regulation to provide an objective basis for official valuations, I decided to use usable floor area according to the NWGA wherever available, and usable floor area where it was unavailable. Based on the new area variable, I calculated each apartment's price per m², which will be the outcome variable. Since the dataset included outliers with unreasonably high or low prices per m² that seemed to have been data errors, I dropped observations with prices per m² below 400€ and above 10,000€. The final distribution of prices per m² is shown in Figure 2.
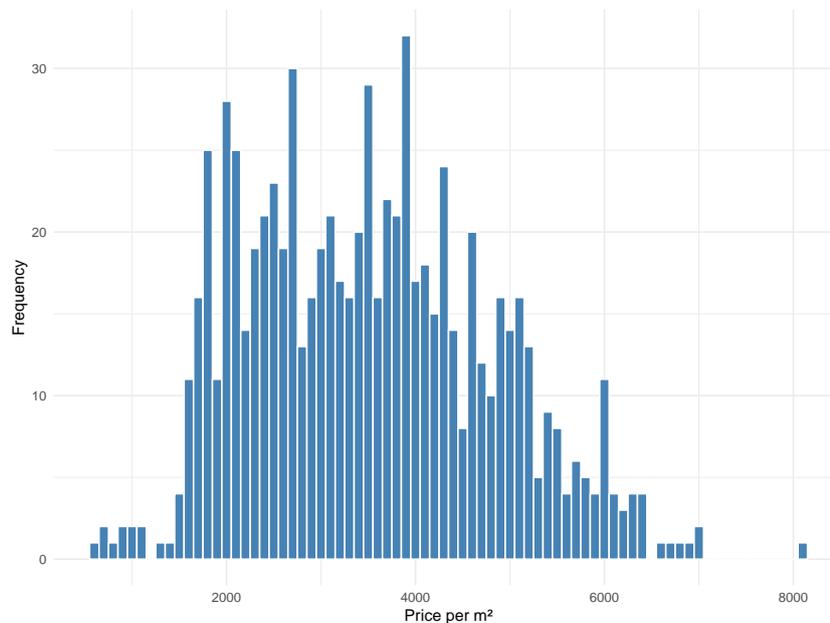
Figure 2: Histogram showing the distribution of apartment prices per m² in the dataset after dropping outliers

Next, I used the Google Maps Geocoding API to geocode apartment and station addresses into latitude and longitude coordinates. This then allowed me to calculate the haversine (straight-line) distance from each apartment to all stations, and single out the station closest to each observation. I then generated station dummies with the value 1 for the station nearest to an observation, and the value 0 for all other stations. Next, I used the Google Maps Distance Matrix API to calculate the walking distance from each apartment to its nearest station as measured by the straight-line distance. This enabled me to define treatment dummies based on a certain walking distance cutoff (e.g. 600 meters) for the binary treatment specifications. For the continuous treatment specification, I used the walking distance variable directly.

3.2 Descriptive Statistics

The final dataset includes 771 observations, of which 489 are from the BEFORE period, i.e. from January 1, 2012 to June 1, 2014, and 241 are from the AFTER period, i.e. from September 1, 2014 to December 31, 2014. 41 observations are in the window in between and are dropped to avoid a possible misidentification of BEFORE and AFTER observations. To examine pre-treatment differences between the treated and the control group, Table 1 presents descriptive statistics for the outcome variable price per m² as well as for a central station dummy, an effect station dummy, and a random location dummy. Among treated observations based on the 600 meter walking distance specification (Panel A), the mean price per m² is around 3630€, which is roughly 370€ higher than the control group mean price per m².

Among treated observations based on the 200-600 donut specification, the difference between

treatment and control group mean prices per m² is less pronounced at around 290€. Standard deviations range from around 1251€ to around 1280€ for the treated and control groups according to the two treatment specifications. The means of the central station dummy and the effect station dummy represent the proportion of observations in the treated and control group that are taken from near central stations and effect stations, respectively. The difference in these means, largest for effect stations in Panel A with around 12%, suggests that treatment and control group are similarly distributed in space. The random location dummy has a mean of 0 in the treated group because only control observations were taken from one of five random locations, which constitute approximately 30% of all pre-treatment control group observations.

In spite of the similar spatial distribution, it seems likely that the difference in the mean price per m² between the treated and the control group reflect differences in location - especially given that a substantial portion of control observations, but no treated observations, were taken from random locations.

Table 1: Descriptive Statistics for the Pre-Treatment Period

Panel A: 600 radius	Treated			Control			Diff. Mean
	N	Mean	SD	N	Mean	SD	
Price per m^2	227	3628.46	1278.90	262	3260.57	1271.25	367.89
Central station dummy	227	0.569	0.496	262	0.538	0.499	0.031
Effect station dummy	227	0.771	0.421	262	0.885	0.319	−0.115
Random location dummy	227	0.000	0.000	262	0.302	0.460	−0.302
Panel B: 200-600 donut	N	Mean	SD	N	Mean	SD	
Price per m^2	198	3545.61	1251.62	250	3252.22	1281.44	293.39
Central station dummy	198	0.556	0.499	250	0.548	0.499	0.008
Effect station dummy	198	0.819	0.387	250	0.884	0.321	−0.066
Random location dummy	198	0.000	0.000	250	0.308	0.463	−0.308

Note: Panel A shows descriptive statistics for the treatment and control group in the pre-treatment period based on the 600 meter walking distance specification, while Panel B does the same based on the 200-600 meter donut definition. Diff. Mean is the difference in means between the treatment and control group. The central station dummy takes the value 1 for observations with the stations Karlsplatz, Museumsquartier, Volkstheater, Rathaus, Frankhplatz, Neubaugasse, Pilgramgasse, Reinprechtsdorfer Straße and Matzleinsdorfer Platz being closest, and 0 otherwise. The effect station dummy takes the value 0 for observations with the closest station being Karlsplatz, Museumsquartier, and Volkstheater, and 1 otherwise. The random location dummy takes the value 1 for control observations not taken from the vicinity of stations, and 0 for those taken from the vicinity of stations.

It is important to note that a DiD design can tolerate a baseline difference in mean prices per m² between the treated and the control group. Rather, it relies on the assumption that this difference is time-invariant, which will be investigated in the next section.

3.3   Tests of the Parallel Trends Assumption

I will now turn to the discussion of the plausibility of the parallel-trends assumption, according to which, in the absence of the announcement, prices in the treatment and the control group would have followed parallel trends. This assumption is of critical importance, because only if

it holds can the DiD estimate be interpreted as the causal effect of the announcement on metro prices. Based on a visual inspection, it is plausible that prices in the treatment and the control group followed parallel trends in the timeframe before the announcement date, i.e. from early 2012 to June 2014 (see Figure 3 for the walking distance 600 specification and Figure 4 for the 200-600 donut specification).
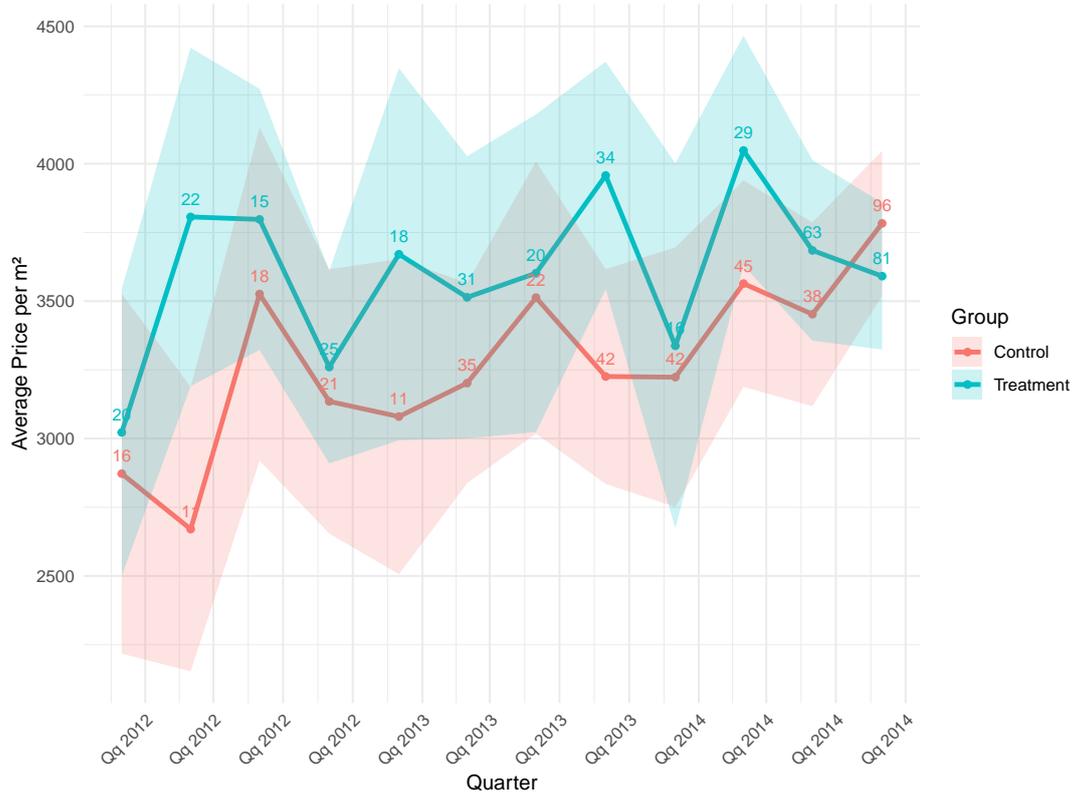

Figure 3: Evolution of average prices in the treatment and control groups based on the 600 meter walking distance specification. Shaded areas are 95% Confidence Intervals. Numbers indicate the number of observations in treatment or control group for a given quarter.
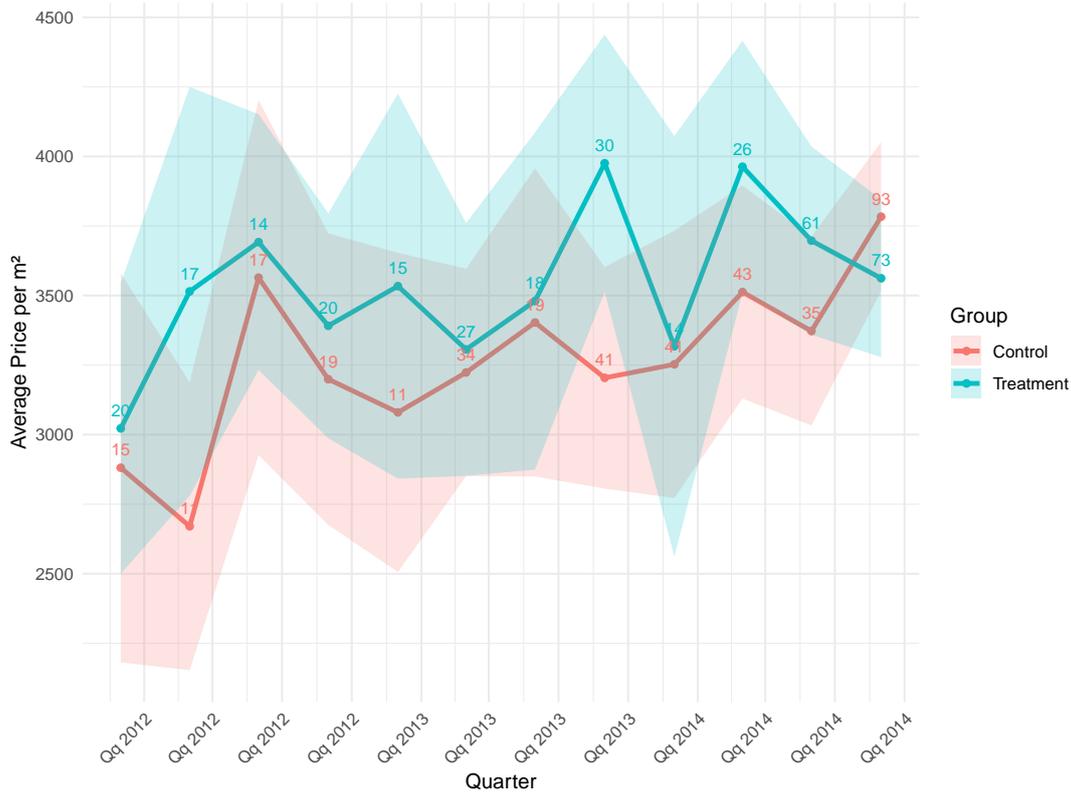
15

Figure 4: Evolution of average prices in the treatment and control groups based on the 200-600 donut specification. Shaded areas are 95% Confidence Intervals. Numbers indicate the number of observations in treatment or control group for a given quarter.

To test the PTA formally, I estimate regressions of the form:

$$Y_i = \beta_0 + TREATED_i \times \sum_{t=1}^{T} \beta_t \cdot Q_{it} + X_i + \varepsilon_i \qquad (1)$$

where Y_i is price per m², X_i is a vector of controls including both station dummies for all stations except for 1, and dummies denoting that an observation was picked from a 500 meter circle around one of the five random control locations, and ε_i is an error term. $TREATED_i$ is the treatment variable interacted with a total of 11 quarterly dummies Q_{it} (from 2012Q1 to 2014Q1) that take the value 1 if the observation was sold in that quarter and 0 otherwise.

If any of the coefficients β_1 to β_2 is statistically significant, then this sheds doubt on the validity of the PTA. This is because this would imply that there was a statistically significant difference between the prices per m² of treatment and control observations in that quarter, but not in others. Thus, it cannot have been the case that treatment and control observations evolved in parallel prior to the announcement. In the two specifications based on the 600 meter walking distance dummy and the 200-600 meter donut there are two and one quarter(s), respectively, with a statistically significant coefficient, which undermines the plausibility of parallel pre-trends in that quarter (see Table 2).

16

To further test the PTA, I estimated a joint F-test of the null hypothesis of all of the coefficients being 0. This allows me to test whether it the possibility that the treatment and control group were on parallel trends in the whole pre-treatment period can be rejected. If this null hypothesis cannot be rejected at reasonable significance levels, it is plausible that treatment and control group prices per m² evolved on parallel trends before the announcement. The p-values of the F statistic are 0.352 and 0,471, respectively, which implies that the null hypothesis of parallel trends in the whole pre-treatment period cannot be rejected (see Table 2). Therefore, it is plausible that the treatment and the control group were on parallel trends in the period before the announcement.

While testing for parallel pre-trends is common in the DiD literature (for a summary see Roth (2022)), Roth (2022) notes two limitations of such tests. First, they often fail to detect trends in the pre-treatment data. While this analysis may suffer from this issue, correcting for it was not possible within the scope of this thesis. The second issue identified is that bias can be introduced by performed the analysis only conditional on the passing of a pre-trends test, which is not an issue in this context.

It is also important to note that the PTA is a stronger assumption than merely an assumption about parallel pre-trends, because it is an assumption of parallelism of the actual control group after-period trend and the counterfactual - and thus unobservable - treatment group after-period trend. Thus, the PTA cannot be established beyond doubt in any DiD design, no matter the plausibility of parallel trends before the announcement. Yet it seems more likely that if the assumption of parallel trends of prices per m² held before the treatment, it would have continued to do so in the after-period, even in the absence of the announcement. The estimated random forest model essentially tries to circumvent these issues surrounding the PTA by predicting the counterfactual treatment group trend directly.

Table 2: Test of the Parallel Trends Assumption

	Dependent variable: price per m²	
	TREATED specification:	
	600 radius	200-600 donut
	(1)	(2)
TREATED × Q1_2012	150.083	83.004
	(378.957)	(391.540)
TREATED × Q2_2012	728.334*	546.894
	(350.717)	(380.988)
TREATED × Q3_2012	−92.510	−265.461
	(391.902)	(402.627)
TREATED × Q4_2012	−79.893	112.074
	(303.602)	(323.230)
TREATED × Q1_2013	498.309	299.331
	(421.248)	(431.899)
TREATED × Q2_2013	131.592	−171.127
	(327.864)	(305.580)
TREATED × Q3_2013	−36.952	−146.256
	(350.950)	(376.797)
TREATED × Q4_2013	561.319*	594.950**
	(277.015)	(286.449)
TREATED × Q1_2014	−100.233	−118.749
	(384.314)	(409.562)
Observations	771	714
R^2	0.281	0.293
Adjusted R^2	0.244	0.253
p-value of F Statistic	0.352	0.471

Note: All standard errors are robust. Controls are station dummies that equal 1 for the station closest to the apartment, and 0 for all other stations, as well as dummies that take the value 1 if an apartment was taken from one of the five control locations, i.e. not from around a station, and the value 0 otherwise. The differences in the number of observations are explained by different definitions of the treated group. *p<0.1; **p<0.05; ***p<0.01

Chapter 4

Results

The estimation results from the DiD regressions indicate a negative causal effect of the U2 x U5 metro expansion announcement on housing prices. In the first specification based on a walking distance of 600 meters from the nearest station, the DiD estimate of the ATT δ is around -503. The coefficient is significant at the 5% level, and almost at the 1% level with a p-value of around 0.013. The inclusion of controls, which consist of station dummies and dummies indicating whether observations were taken from one of the five control locations not associated with a station, decreases the estimate's magnitude somewhat to around -444. Again, the coefficient is almost significant at the 1% level with a coefficient of around 0.013. The latter estimate is to be trusted more, since the parallel trends assumption is more plausible when holding these variables, which reflect neighborhood characteristics, constant by controlling for them. Thus, according to this estimate, the metro announcement decreased prices per m² of apartments within a walking distance of 600 meters of a new station by around 444€, on average. Given that the mean sale price per m² in the treatment group after the announcement is roughly 3560€, this is a economically significant negative effect of around 11.1% percent.

In the second specification, where the treatment variable takes the value 1 if an observation is between 200 and 600 meters away from a station and 0 if it is further away, the estimates of δ are less negative than before. Without including controls, prices seem to have decreased by roughly 428€ and, after the inclusion of controls, by roughly 373€. At a mean sale price per m² in the treatment group after the announcement of roughly 3540€, this corresponds to a negative causal effect of around 9.5%. Both of these estimates are significant at the 5% level. The fact that the magnitude of the effect seems to be smaller compared to the first specification with controls suggests that apartments within 200 meters of walking distance from the nearest station are a strong driver of the negative announcement effect. However, excluding these apartments still yields a large negative estimate. This sheds doubt on a possible explanation that credits solely localized noise during the construction period for the negative announcement effect. Rather, there must be other mechanisms that also render apartments further than 200 meters away from the nearest station substantially less attractive.

In the third specification, where the treatment variable is simply the walking distance from the

nearest station in meters, the coefficient of interest is 0.313. Note that the difference in sign from the previous two specifications is expected: while previously, an increase in the treatment dummy from 0 to 1 indicated observations closer to a station, an increase in the continuous walking distance variable means getting further away from a station. Without controls, the marginal causal effect of an additional meter away from a new station is estimated to be around 0.31. The coefficient is significant at the 1% level. Thus, in accordance with estimates from the first two specifications, the metro announcement had a negative effect on real estate prices close by. Including controls decreases the marginal effect to approximately 0.18, which is only significant at the 10% level. This translates into an average increase of prices per m² of around 18 euros for every 100 meters further away from the nearest station, or of 108 euros for 600 meters further away due to the announcement (see Table 3).

Table 3: DiD Regression Results

	Dependent variable: price per m²					
	TREATED specification:					
	600 radius	600 radius	200-600 donut	200-600 donut	continuous	continuous
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	3,278.939***	3,510.367***	3,271.522***	3,489.623***	3,509.432***	3,918.011***
	(78.983)	(227.156)	(80.751)	(234.961)	(91.746)	(237.117)
AFTER	451.450***	490.647***	446.004***	503.117***	−70.471	134.657
	(138.678)	(118.674)	(141.387)	(117.751)	(156.348)	(137.739)
TREATED	331.244***	397.028***	252.255**	376.982***	−0.087	−0.378***
	(115.870)	(110.674)	(120.424)	(116.880)	(0.068)	(0.119)
AFTER × TREATED	−503.424**	−444.246**	−428.771**	−373.680**	0.313***	0.182*
	(203.196)	(178.631)	(211.388)	(184.221)	(0.116)	(0.106)
Controls	No	Yes	No	Yes	No	Yes
Number of Observations	730	730	675	675	714	714
R^2	0.019	0.273	0.016	0.285	0.013	0.275
Adjusted R^2	0.015	0.250	0.012	0.261	0.009	0.252

Note: All standard errors are robust. Controls, where included, are station dummies that equal 1 for the station closest to the apartment, and 0 for all other stations, as well as dummies that take the value 1 if an apartment was taken from one of the five control locations, i.e. not from around a station, and the value 0 otherwise. The differences in the number of observations are explained by different definitions of the treated group.

*p<0.1; **p<0.05; ***p<0.01

In addition to the main specifications presented in this table, I consider various alternative specifications that alter different aspects of the regression model. For instance, since one would likely expect the effect of distance to be nonlinear, I decided to include a walking-distance-squared term, the coefficient on which is insignificant. I also adapted the walking distance cutoff between treatment and control observations to 400 and 800 meters to test whether the negative result of the 600 meter specification is sensitive to changes to the treatment-control distinction. Although the magnitude of the effect changes depending on the walking distance cutoff, qualitatively the results stay the same (see Table A1). As explained in Chapter 2, I also vary the definition of the AFTER variable to allow for a more lagged effect of the announce-

ment, and to test if the information leaked in two newspaper articles a few months prior to the announcement already had an effect. Again, although the magnitude of the effect changes when allowing for a lagged effect, the negative sign persists. As for the leak dates, although the estimated coefficients of δ are negative, they are all insignificant (see Table A2), which suggests that the leaks before the official announcement did not influence housing prices.

Furthermore, I estimate the effect for a subsample of stations, excluding the stations Karlsplatz, Museumsquartier and Volkstheater that merely changed from being U2 stations to being U5 stations and thus presumably affected prices the least. The effect is less negative and less significant, but qualitatively yields the same result (see Table A3).

The random forest model suggests a minor negative effect, but the results are highly insignificant (see Table A4). It is important to note, however, that random forest models rely on the assumption that data on all possibly relevant controls is observed. In the present context, this assumption does not hold because data on many apartment characteristics that might influence prices was scarce. Therefore, the DiD estimates of δ, which do not rely on the assumption that all confounders are observed, but only on the PTA, are more credible.

Chapter 5

Conclusion

In this thesis, I attempted to estimate the causal effect of the announcement of a Viennese metro expansion involving the lines U2 and U5 on housing prices using different DiD variations. According to my estimates, the announcement seems not to have increased prices close to new stations, but to have decreased them. Furthermore, the effect is large at around 9.5-11.1%, on average, depending on the chosen specification, and statistically significant at close to the 1% level. Although the magnitude of the estimates varies, the negative effect is robust to changes in the treatment group and AFTER definitions and consistent with the results of a continuous quasi-DiD regression. The random forest model to predict counterfactual treatment group observations in the period after the announcement returns slightly negative point estimates for the causal effect, which are, however, insignificant. Moreover, there does not seem to have been an anticipation effect despite leaks to newspapers in the months preceding the announcement.

The negative effect on housing prices suggests that locations close to new metro stations became less attractive to prospective residents after the metro construction was announced, despite the expectation of better future connectivity. This might be due prospective residents' fears that unpleasant side-effects once construction is completed will outweigh the benefits, for instance a higher incidence of homelessness or small-scale criminal activity like drug dealing around metro stations and in the neighborhood. Alternatively, it may be that a busier neighborhood is viewed as negative by residents.

It is also possible that the anticipation of nuisance during construction - as opposed to expected negative side-effects after construction is finished - for instance due to noise or traffic disruptions, explains the negative effect. There are a few caveats to this explanation. First, it seems unlikely that noise during the construction period is the only driver of the negative effect, because the negative effect is robust to the donut specification where apartments within a walking distance of 200 meters were dropped. Since noise nuisance is probably not substantial more than 200 meters away from a station, nuisance has to encompass other factors like traffic disruptions to be plausible as an explanation of the negative effect. Although from today's perspective, traffic disruptions during construction seem not to be major enough to justify such a decrease, prospective residents might have expected more substantial disruptions, which might

22

have led prices to fall.

Second, as noted before, based on a after-tax no-arbitrage argument, prices should be forward-looking and already at the time of the announcement reflect the positive amenity after construction is finished. In spite of the presence of transaction costs that nuance this argument, it is questionable whether transaction costs indeed match the magnitude of the negative effect.

The findings of this thesis are important because they imply that apparently innocuous urban policies can render locations close to them less attractive, which raises crucial questions about the design and implementation of large-scale multi-year infrastructure projects in particular. Once the construction of the U2 x U5 metro expansion has been completed, further research is needed to see whether locations close to stations became more attractive once the metro is operational. If this is the case, this could suggest that potential buyers discount their future utility heavily, which is why nuisance during construction for them outweighs the benefits after construction has been finished. Another possible avenue for further research is a more thorough investigation of the heterogeneity of treatment effects by stations, possibly due to different construction time horizons or difference in the baseline connectivity level of the respective neighborhoods in which the stations are located.

Bibliography

Agostini, Claudio A. and Gastón A. Palmucci (2008). "The Anticipated Capitalisation Effect of a New Metro Line on Housing Prices". In: *Fiscal Studies* 29.2, pp. 233–256. DOI: 10.1111/j.1475-5890.2008.00074.x.

Cárdenas, Jeisson, Juan Miguel Gallego, and María Alejandra Urrutia (2023). "Announcement of the First Metro Line and Its Impact on Housing Prices in Bogotá". In: *Case Studies on Transport Policy* 11, p. 100941. DOI: 10.1016/j.cstp.2022.100941.

Chen, Yuer et al. (2019). "The Impact on Neighbourhood Residential Property Valuations of a Newly Proposed Public Transport Project: The Sydney Northwest Metro Case Study". In: *Transportation Research Interdisciplinary Perspectives* 3, p. 100070. DOI: 10.1016/j.trip.2019.100070.

Demircan, Kamil and Senay Oguztimur (2023). "The Effect of Metro Stations on Housing Prices, Istanbul Case". In: *MEGARON / Yildiz Technical University, Faculty of Architecture E-Journal*, pp. 453–465. DOI: 10.14744/megaron.2023.84555.

Dengg, Stephanie (2018). "How a Metro Development Project Affects Land Values - Evidence from Vienna". Master's thesis. University of Groningen.

Diao, Mi, Delon Leonard, and Tien Foo Sing (2017). "Spatial-Difference-in-Differences Models for Impact of New Mass Rapid Transit Line on Private Housing Values". In: *Regional Science and Urban Economics* 67, pp. 64–77. DOI: 10.1016/j.regsciurbeco.2017.08.006.

Dorantes, Lucia Mejia, Antonio Paez, and Jose Manuel Vassallo (2011). "Analysis of House Prices to Assess Economic Impacts of New Public Transport Infrastructure: Madrid Metro Line 12". In: *Transportation Research Record* 2245.1, pp. 131–139. DOI: 10.3141/2245-16.

Guan, ChengHe, Mark Junjie Tan, and Richard Peiser (2021). "Spatiotemporal Effects of Proximity to Metro Extension on Housing Price Dynamics in Manhattan, New York City". In: *Journal of Transport and Land Use* 14.1, pp. 1295–1315.

Hamberger, Caroline (2023). "Mind the Gap: Effects of the U5 Subway Line on Real Estate Values in Vienna". Master's thesis. Central European University.

He, Dongsheng et al. (2024). "New Metro and Housing Price and Rent Premiums: A Natural Experiment in China". In: *Urban Studies* 61.7, pp. 1371–1392. DOI: 10.1177/00420980231208560.

Keeler, Zachary T. and Heather M. Stephens (2023). "The Capitalization of Metro Rail Access in Urban Housing Markets". In: *Real Estate Economics* 51.3, pp. 686–720. DOI: `10.1111/1540-6229.12411`.

Kiel, Katherine A. and Katherine T. McClain (1995). "The Effect of an Incinerator Siting on Housing Appreciation Rates". In: *Journal of Urban Economics* 37.3, pp. 311–323. DOI: `10.1006/juec.1995.1016`.

Li, Shengxiao, Luoye Chen, and Pengjun Zhao (2019). "The Impact of Metro Services on Housing Prices: A Case Study from Beijing". In: *Transportation* 46.4, pp. 1291–1317. DOI: `10.1007/s11116-017-9834-7`.

Lopez-Morales, Ernesto et al. (2023). "Land and Housing Price Increases Due to Metro Effect: An Empirical Analysis of Santiago, Chile, 2008–2019". In: *Land Use Policy* 132, p. 106793. DOI: `10.1016/j.landusepol.2023.106793`.

Natmessnig, Elias (2013). *Wien drückt bei der U-Bahnlinie U5 aufs Tempo*. Published online on October 4, 2013. Accessed: March 20, 2025. URL: `https://kurier.at/chronik/wien/wien-drueckt-bei-der-u-bahnlinie-u5-aufs-tempo/29.493.935`.

Roth, Jonathan (2022). "Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends". In: *American Economic Review: Insights* 4.3, pp. 305–322. DOI: `10.1257/aeri.20210236`.

Stuhlpfarrer, Martin (2014). *Chronologie der U5: Wiens neue Linie in Startposition*. First published in print on January 29, 2014. Accessed: March 20, 2025. URL: `https://www.diepresse.com/1554952/chronologie-der-u5-wiens-neue-linie-in-startposition`.

Tan, Ronghui et al. (2019). "The Effect of New Metro Stations on Local Land Use and Housing Prices: The Case of Wuhan, China". In: *Journal of Transport Geography* 79, p. 102488. DOI: `10.1016/j.jtrangeo.2019.102488`.

Trojanek, Radoslaw and Michal Gluszak (2018). "Spatial and Time Effect of Subway on Property Prices". In: *Journal of Housing and the Built Environment* 33.2, pp. 359–384. DOI: `10.1007/s10901-017-9569-y`.

Wen, Haizhen et al. (2018). "Subway Opening, Traffic Accessibility, and Housing Prices: A Quantile Hedonic Analysis in Hangzhou, China". In: *Sustainability* 10.7, p. 2254. DOI: `10.3390/su10072254`.

wien.at (2014). *City of Vienna*. Accessed: November 21, 2024. URL: `https://web.archive.org/web/20140629152752/http://www.wien.gv.at/verkehr-stadtentwicklung/fahrplan/oeffi-paket.html`.

Wieser, Robert (2011). "Zahlungsbereitschaften für städtische Verkehrsinnovationen Eine hedonische Bodenpreisanalyse der U-Bahn-Effekte in Wien". In: *Der öffentliche Sektor* 37.1. DOI: `10.34749/oes.2011.628`.

Wu, Chao et al. (2017). "Spatial Effects of Accessibility to Parks on Housing Prices in Shenzhen, China". In: *Habitat International* 63, pp. 45–54. DOI: 10.1016/j.habitatint.2017.03.010.

Zhou, Zhengyi et al. (2021). "The Effect of a Subway on House Prices: Evidence from Shanghai". In: *Real Estate Economics* 49.S1, pp. 199–234. DOI: 10.1111/1540-6229.12275.

Appendix

Table A1: DiD Regression Results: Variation of the TREATED variable

	Dependent variable: price per m²					
	TREATED specification:					
	400 radius	400 radius	800 radius	800 radius	continuous dist.²	continuous dist.²
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	3327.831***	3559.719***	3309.667***	3532.631***	3835.300***	4160.220***
	(64.376)	(218.236)	(87.211)	(228.211)	(143.000)	(266.500)
AFTER	251.404**	306.009***	570.764***	571.733***	−370.610	−103.800
	(110.575)	(98.404)	(161.073)	(144.053)	(251.382)	(232.800)
TREATED	468.717***	342.750**	211.756*	271.384**	−0.780***	−0.901***
	(143.598)	(139.129)	(116.868)	(120.148)	(0.241)	(0.280)
TREATED^2					0.000215***	0.00022**
					(0.000067)	(0.00011)
AFTER × TREATED	−90.043	17.580	−587.126***	−468.772**	0.095**	0.68*
	(271.419)	(232.552)	(207.112)	(182.780)	(0.042)	(0.38)
AFTER × TREATED^2					−0.000198*	−0.000157
					(0.000114)	(0.000104)
Controls	No	Yes	No	Yes	No	Yes
Number of Observations	730	730	730	730	714	714
R^2	0.025	0.243	0.017	0.240	0.025	0.280
Adjusted R^2	0.021	0.225	0.013	0.221	0.018	0.255

Note: All standard errors are robust. Controls, where included, are station dummies that equal 1 for the station closest to the apartment, and 0 for all other stations, as well as dummies that take the value 1 if an apartment was taken from one of the five control locations, i.e. not from around a station, and the value 0 otherwise. The differences in the number of observations are explained by different definitions of the treated group. *p<0.1; **p<0.05; ***p<0.01

Table A2: DiD Regression Results: Variation of the AFTER variable

	Dependent variable: price per m²					
	AFTER specification:					
	AFTER October 1, 2014	AFTER October 1, 2014	AFTER 1st leak date	AFTER 1st leak date	AFTER 2nd leak date	AFTER 2nd leak date
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	3278.939***	3262.580***	3240.930***	3398.537***	3263.244***	3425.135***
	(79.104)	(240.900)	(100.640)	(237.243)	(91.153)	(243.482)
AFTER	503.875***	525.710***	294.760*	311.631***	392.665***	399.185***
	(155.960)	(133.750)	(131.800)	(115.614)	(131.441)	(118.580)
TREATED	331.244**	337.240***	229.490	348.758***	306.882*	387.114***
	(115.900)	(110.250)	(143.750)	(131.177)	(129.968)	(122.753)
AFTER × TREATED	−523.240*	−413.080**	−103.500	−67.901	−261.183	−198.536
	(223.433)	(197.130)	(189.270)	(165.379)	(187.011)	(165.467)
Controls	No	Yes	No	Yes	No	Yes
Number of Observations	730	730	730	730	714	714
R^2	0.022	0.279	0.013	0.252	0.018	0.256
Adjusted R^2	0.017	0.254	0.008	0.229	0.014	0.232

Note: All standard errors are robust. TREATED is based on a walking distance of 600 meters. Columns (1) and (2) check a more lagged effect; columns (3) and (4) check the first leak date on October 4, 2013; columns (5) and (6) check the second leak date on January 29, 2014. Controls, where included, are station dummies that equal 1 for the station closest to the apartment, and 0 for all other stations, as well as dummies that take the value 1 if an apartment was taken from one of the five control locations, i.e. not from around a station, and the value 0 otherwise. The differences in the number of observations are explained by different definitions of the treated group. *p<0.1; **p<0.05; ***p<0.01

Table A3: DiD Regression Results: Variation with only Effect Stations included

	Dependent variable: price per m²					
	TREATED specification:					
	600 radius	600 radius	200-600 donut	200-600 donut	continuous	continuous
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	3191.912***	3465.398***	3191.912***	3476.520***	3302.800***	3779.900***
	(86.413)	(229.710)	(86.425)	(233.121)	(95.869)	(234.500)
AFTER	418.142***	535.181***	418.142***	539.798***	184.330	325.170*
	(158.447)	(127.973)	(158.468)	(127.691)	(161.160)	(140.110)
TREATED	239.738*	356.935***	205.214	327.052***	−0.005	−0.250**
	(124.152)	(116.648)	(126.668)	(119.113)	(0.075)	(0.123)
AFTER × TREATED	−272.061	−331.531*	−254.852	−293.607	0.1408	0.7049
	(219.864)	(186.889)	(226.210)	(192.405)	(0.1283)	(1.1604)
Controls	No	Yes	No	Yes	No	Yes
Number of Observations	720	720	697	697	720	720
R^2	0.0182	0.3051	0.0174	0.307	0.0139	0.298
Adjusted R^2	0.0132	0.2835	0.0123	0.2845	0.0089	0.2762

Note: Effect Stations are those that did not merely switch from being U2 stations to being U5 stations, which excludes the stations Karlsplatz, Museumsquartier and Volkstheater. All standard errors are robust. Controls, where included, are station dummies that equal 1 for the station closest to the apartment, and 0 for all other stations, as well as dummies that take the value 1 if an apartment was taken from one of the five control locations, i.e. not from around a station, and the value 0 otherwise. The differences in the number of observations are explained by different definitions of the treated group. *p<0.1; **p<0.05; ***p<0.01

Table A4: Random Forest Model ATT Estimation Results

| | TREATED specification: | |
	600m	200–600m Donut
ATT	−5.75	−24.12
	(116.08)	(127.82)
N (Train)	597	568
N (Live)	117	107
R^2 (Train)	0.483	0.497
RMSE (Train)	928.41	907.32
MAE (Train)	737.02	715.92
Mean Predicted	3562.27	3564.40
Mean Actual	3558.21	3541.01

Note: ATT is the Average Treatment Effect on the Treated, which was calculated as the average difference between observed and counterfactual prices per m² after the announcement for observations in the treatment group. Counterfactual prices were predicted using random forest models trained on pre and post-treatment control observations as well as pre-treatment treated observations. Covariates were the walking distance to the nearest station, the transaction date, and the usual station and random control location dummy variables. RMSE is the Root Mean Squared Error, MAE is the Mean Absolute Error. Standard errors are based on 1,000 bootstrap replications.