Take the Q Train: Value Capture of Public Infrastructure Projects[*]

Arpit Gupta[†] Stijn Van Nieuwerburgh[‡] Constantine Kontokosta[§]

July 22, 2020

Abstract

Transit infrastructure is a critical asset for economic activity yet costly to build in dense urban environments. We measure the benefit of the Second Avenue Subway extension in New York City by analyzing local real estate prices which capitalize the benefits of transit spillovers. We find 10% price increases, creating $7 billion in new property value. Using cell phone ping data, we document substantial reductions in commuting time especially among subway users, offering a plausible mechanism for the price gains. The increase in prices reflects both higher rents and lower risk. Infrastructure improvements lower the riskiness of real estate investments. Only 30% of the private value created by the subway is captured through higher property tax revenue, and is insufficient to cover the cost of the subway. Targeted property tax increases may help governments capture more of the value created, and serve as a useful funding tool.

Keywords: Infrastructure Finance, Real Estate, House Prices, Public Finance, Urban Transit, Value Capture, Commuting
JEL code: G10, G50, R32, R41, R42, R51

---

[*] We gratefully acknowledge financial support from the Lincoln Institute and the NYU Stern Infrastructure Initiative, as well as comments from conference participants at the CREFR Spring Conference, the PREA Institute Conference in NYC, the AREUEA National Conference 2020, the Stanford Hoover Institute infrastructure conference, and the 2019 Urban Economics Association conference. We thank Yuan Lai, Michael Leahy, and Joshua Coven for superb research assistance.

[†] Department of Finance, Stern School of Business, New York University, 44 W. 4th St, New York, NY 10027, `arpit.gupta@stern.nyu.edu`.

[‡] **Corresponding author**: Department of Finance, Graduate School of Business, Columbia University, Uris Hall 809, 3022 Broadway, New York, NY 10027; NBER; CEPR; `svnieuwe@gsb.columbia.edu`.

[§] Marron Institute for Urban Management, New York University, 60 5th Ave, 2nd Floor, New York, NY 10011, `ck2218@nyu.edu`.

1 Introduction

Adequate infrastructure is an essential input into economic growth. In developing countries, the demands from economic development and a rising population drive investment needs in transportation, power, water, telecom, and real estate assets. In the developed world, aging infrastructure needs maintenance and upgrades as urban density grows and technologies change. In 2015, the world spent $2.5 trillion on economic infrastructure investment, and $9.5 trillion on a broader asset definition that includes real estate, oil and gas, mining, and social infrastructure. Yet, to keep pace with GDP growth, the world will need to an invest an average of $3.7 trillion on economic infrastructure each year between 2017 and 2035, or 4.1% of GDP. This is nearly $70 trillion in aggregate projected future spending, $20 trillion of which is in the United States (McKinsey, 2017). These massive spending needs take place against a reality of fiscal constraints in the public sectors in the U.S., Europe, and many other countries. Private capital will need to play a materially larger role in infrastructure finance (Walter, 2016). Indeed, assets under management in private equity funds specialized in infrastructure investments have already grown tenfold over the past decade to nearly $500 billion (Andonov, Kräussl, and Rauh, 2019).

The largest category of economic infrastructure is transportation. The costs of building public transportation improvements are high. Subway investments, which offer the prospect of carrying the most passengers in the densest urban locations have particularly large costs. Nowhere is this more true than in the largest cities in the United States. Recent extensions of the 7 and the Q subway lines in New York City cost about $2.5 billion per mile. These high costs make it essential to measure the benefits of additional subway construction in order to assess whether the expansion is worthwhile. This paper focuses on measuring the benefits from the 2nd Avenue Subway (Q-train) expansion, the most substantial investment in public subway infrastructure in the United States over the past several decades.

Prior literature has identified several potential benefits of subway construction projects. New transportation lines can improve access to workplaces and amenities due to shorter commuting times (Kahn and Baum-Snow, 2000, 2005; Severen, 2018), which in turn can boost labor force participation (Black, Kolesnikova, and Taylor, 2014), reduce traffic congestion on roads and other public transportation, and reduce pollution (Anderson, 2014). Other associated benefits of new transit linkages include improved urban amenities such as increased retail presence (Kahn, 2007), noise and crime reductions around stations (Bowes and Ihlanfeldt, 2001), and less drunk driving (Jackson and Owens, 2011). These diffuse benefits are difficult to measure directly, complicating a straightforward cost-benefit calculation. The public return to infrastructure investment, defined narrowly as the user fees net of operational expenditures and more broadly to include the incremental property, sales, and labor income tax revenues, captures only a part of the total benefit of infrastructure investment because it ignores the positive externalities that the infrastructure generates for the private sector. Failure to appropriately account for private-sector benefits may result in important infrastructure investments remaining unfunded.

Our analysis provides an alternate measure to assess the benefits of infrastructure improvements. First, we provide novel estimates on the commuting time benefits of subway construction. We then take advantage of the fact that transportation infrastructure and real estate assets are complements; as a result, real estate values in the vicinity of public transportation hubs capitalize the present value of all future benefits that accrue to households and business from transportation gains. To perform this calculation, we measure how residential and commercial real estate asset values change after the extension of public transportation.

We define geographical areas that are "treated" by the subway extension. We compare the changes in real estate values in the treated areas to the changes in real estate prices in the "control" areas in a difference-in-difference setup. Our baseline treatment definition selects all properties in a rectangular area between 59th and 100th streets and between First and Third Avenues (the "2nd Avenue corridor"). The control area is the rest of the Upper East Side (UES) of Manhattan, the remaining properties between 59th and 100th streets between Fifth Avenue and the East River. We consider three alternative treatment definitions. The second treatment is defined as the area within a 0.3 mile walking distance from one of the three new stations that were added as part of the subway expansion. The third treatment definition considers buildings whose distances to the nearest station on any subway line are reduced after the opening of the new subway stations. The fourth treatment looks at the intersection of the first three treatment definitions.

The Q-line extension opened on January 1st 2017. We start the Post period four years earlier, to capture the fact that there was little residual uncertainty over eventual subway completion as early as 2013. Since real estate prices are forward-looking, they should anticipate the benefits from the future subway extension. In a second specification, we break up the pre-2013 period into the pre-2006 and the 2007–2013 periods. This specification allows for six additional years of potential anticipation effects. It also captures potential disamenities (noise, pollution, business disruption) from heavy construction, which was concentrated in 2007–2013.

Our data combines deeds and property tax records from NYC's Department of Finance with unit and building characteristics scraped from StreetEasy, an online real estate listing platform. Our final sample covers about 50,000 arms length transactions of condo and coop units on the UES. From the same data source, we also collect rental listing information on about 100,000 rental units. We augment this sample with high-frequency geolocation information from mobile phones, which allows us to track exact commute lengths at the individual level, before and after the subway opening, as well as the mode of commuting.

We find compelling evidence that the 2nd Avenue Subway expansion led to strong changes in commuting patterns. Using our benchmark difference-in-difference specification, we find that residents in areas served by the new subway expansion experience a decline in commute lengths of 3–5 minutes (7.5%). These gains increase to 14 minutes among subway commuters. We find evidence that new migrants into the area, who are likely to be marginal price setters in the real estate market, are disproportionately likely to take the Q-train.

We then link the subway expansion to a sizable increase in real estate values. Our benchmark difference-in-difference specification estimates a 10.2% increase in real estate values when comparing the prices ten years before 2013 to the prices six years after. Prices on the 2nd Avenue corridor increase 14.8% relative to 2003–2006, with more than half of this gain (10.2%) manifesting during the construction period 2007–2013. The three alternative treatment definitions result in similar point estimates: 8.4%, 9.5%, and 9.9% when comparing the post-period to the entire pre-period and 13.2%, 145.7%, 13.7% when comparing the post-period to the pre-2006 period.

We also estimate specifications which control more finely for building amenity effects through the use of building fixed effects or unit-specific characteristics through a repeat-sales approach. Though smaller, the 2–6% price increases we find in these specifications still suggest substantial value creation in the area around subway construction.

We find evidence that larger and newer housing units experience a larger value gain. We conjecture that one channel through which the subway created increases in real estate values was the stimulus of real estate development. Building permit data confirms a positive housing supply response that is (at least directionally) consistent with this channel.

Next, we ask whether the price increase reflects a cash flow or discount rate effect (Campbell and Shiller, 1988). We collect asking rents, and unit and building characteristics for rental apartment buildings from StreetEasy. Our sample contains about 100,000 rental units. Using the same difference-in-difference model as for sales, we also show that rents increase in response to subway construction. Applying a naive constant discount rate produces implied price increases that are quantitatively consistent with the observed price increase. We go further and ask whether infrastructure investment affects the expected return, the discount rate, of real estate investments.

We find that about half of the price increase is due to changes in rents, and half is due to changes in the price-rent ratio, a proxy for the (inverse of the) discount rate. This finding is consistent with results in Campbell, Davis, Gallin, and Martin (2006) that discount rate variation drives U.S. residential real estate markets, and similar results in Plazzi, Torous, and Valkanov (2010) for commercial real estate. Our contribution is to emphasize that infrastructure improvements lower the riskiness of real estate investments. While intuitive, this is a novel point in the literature.

Finally, we estimate the aggregate real estate value created by the subway extension, and how much of that value flows back to government coffers in the form of higher property taxes. This analysis proceeds in several steps. The first task is to value the stocks of owner-occupied residential, renter-occupied residential, and commercial real estate in the treatment area prior to the subway (as of 2012). To that end, we combine our main data set on residential units that are sold or rented in our sample period and on the total number of units in the building with a data set on property tax assessments, and with our dynamic DiD estimation. Our approach estimates a \$32 billion aggregate valuation for owner-occupied residential, \$26 billion for renter-occupied residential, and \$12 billion for commercial real estate properties on the 2nd Avenue corridor in 2012. Second, we apply our baseline 10.2% price increase estimate to the \$70 billion in aggregate property value, resulting in a \$7.1 billion windfall to private real estate owners. Third, we analyze how much of this value creation flows back to the government. To the extent that the property tax system is able to recoup some of these expenses, this provides a natural mechanism for local governments to finance infrastructure investments. However, there are good reasons to think that the local government captures only part of the value created. Detailed analysis of property tax data shows that NYC recuperates 30.6% of the increase in market values in present value terms. This amounts to \$2.1 billion in extra property tax revenue. As a result, though the subway generated more value than the \$4.5 billion cost of construction, this value largely accrues to private landowners, rather than the city government. The city's own cost-benefit shortfall is \$2.4 billion.

This analysis motivates the possibility for additional value capture taxes which may help recoup an additional component of the investment cost, and thereby make possible additional public infrastructure investments. Cities like Tokyo and Hong Kong have successfully employed such value capture in the past. Our findings are policy-relevant and timely given ongoing debates in New York City on the future extension of the 2nd Avenue Subway line, the repair of the L line, and the East Side access project. They also have ramifications for the broader debate on how to finance an upgrade to U.S. infrastructure assets and how to provide new infrastructure in developing countries whose governments have limited borrowing and taxation capacities. Given that infrastructure projects entail enormous expenditures of public resources, it is essential to have a full accounting of the total benefits resulting from these infrastructure expansions, which our work helps to provide.

Literature Review Our paper relates to a large literature investigating the effectiveness of infrastructure investments. Previous research has found a wide range of estimates for the return on infrastructure investment, depending on the assumptions made on the efficiency of an expansion of the public capital stock, the strength of the crowd-out effect on private investment, and the timing vis-à-vis the business cycle (Cadot, Röller, and Stephan, 2006; Andonov, Kräussl, and Rauh, 2019; Castells and Solé-Ollé, 2005; Finkenzeller, Dechant, and Schäfers, 2010).[1] The uncertainty over these estimates suggests that the approach of inferring the returns to infrastructure investment from real estate return is a useful complement to the traditional approach.

Our paper also belongs to an active literature that studies the land or house price capitalization of urban rail.[2] Price premium estimates for real estate surrounding transit hubs typically range from 3% to 10%, with some outliers at the upper end of 40-45%. Kahn (2007) finds that new pubic transit has the biggest impact on real estate prices when the new transit connects an area to a vibrant downtown, which is the case for the New York City 2nd Avenue subway expansion. A few studies have identified negative relationships between distance to transit stations and prices (Bowes and Ihlanfeldt, 2001; Pan and Zhang, 2008), reflective of disamenities of transit stations (e.g. crowding, noise, and crime). Our paper is the first to study the recent subway expansion in New York City. The New York City subway system is one of the oldest and most widely used public transit systems, and the one with the most stations. As argued, this expansion was the most expensive per-mile expansion in U.S. transportation history. The urban density and pre-existing transportation network make for an important and interesting context.

We contribute further by investigating the interplay between the ownership market

---

[1] Bom and Ligthart (2014) conduct a meta-regression analysis, summarizing decades of research measuring the effects of public infrastructure spending on economic output. The authors find that on average for the United States, a 1% increase in the public capital stock (about $134 billion in 2015) would increase private-sector economic output by 0.03% in the short term ($12.7 billion) and by 0.12% in the long term ($18.7 billion). The Congressional Budget Office estimates the effect of the same-size expansion at 0.06% ($9.2 billion).

[2] See Dewees (1976) for Toronto, McDonald and Osuji (1995); McMillen and McDonald (2004); Diao, Leonard, and Sing (2017) for Chicago, Cervero and Duncan (2002) for San Jose, Lin and Hwang (2004) for Taipei, Hess and Almeida (2007) for Buffalo, sixteen cities among which Atlanta, Boston, Chicago, Portland, and Washington DC by Kahn and Baum-Snow (2005), Zheng and Kahn (2013) for Beijing, Fesselmeyer and Liu (2018) for Singapore, and Zhou, Chen, Han, and Zhang (2020) for Shanghai.

(condos and coops) and the rental market. Our results indicate that infrastructure improvements affect both the cash flows and their riskiness, allowing us to connect to the asset pricing literature on the role of cash flows and discount rates in the stock market (Campbell and Shiller, 1988; Koijen and van Nieuwerburgh, 2011; Cochrane, 2011) and in real estate markets (Campbell, Davis, Gallin, and Martin, 2009; Plazzi, Torous, and Valkanov, 2010; Van Nieuwerburgh, 2019). We find that infrastructure investments increase cash flows and lower the risk of real estate investments.

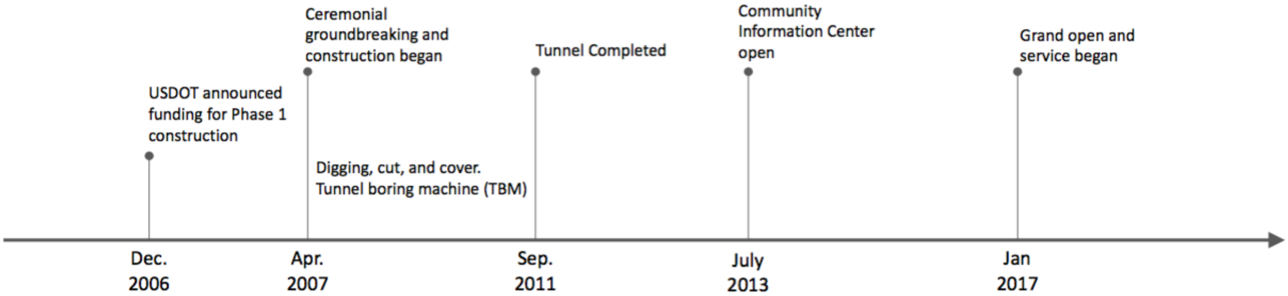
A new literature, including Athey, Ferguson, Gentzkow, and Schmidt (2019), Chen, Haggag, Pope, and Rohla (2019), and Chen and Rohla (2018), has begun to use rich geolocation data from smart phone pings to track individual trajectories. Our paper is the first to use this data to study commuting lengths. This data are uniquely well-suited to this task because ping data allow us to capture actual commuting lengths, rather than the estimated commuting lengths from surveys used in prior research (as in Couture, Duranton, and Turner (2018)). Doing so allows us to quantify the transportation gains resulting from the subway extension, which we then tie to complementary real estate valuation gains.

The rest of the paper is organized as follows. Section 2 provides institutional background. Section 3 contains the empirical specification. Section 4 discusses the data. Section 5 analyzes our commuting results. The main real estate valuation results are in Section 6. Section 7 contains the analysis on rents and price-rent ratios. Section 8 computes the aggregate value creation from the subway extension and how much of it flows back to the government. Section 9 concludes. The Appendix contains some auxiliary empirical results.

2 Institutional Background

Elevated rail lines were formerly running on 2nd and 3rd Avenues in New York as a part of citywide system of "el" trains operated by privately managed and jointly funded companies. This network was gradually replaced with underground subways starting in 1904. A 2nd Avenue Subway, in particular, was a major component of a subway expansion proposed in 1920 by the Independent Subway System (IND), a publicly owned and operated managed entity. Ultimately, the IND was combined with two other private companies and placed under government control. The elevated 2nd Avenue line was torn down in 1942 in anticipation of a new underground 2nd Avenue Subway. However, construction plans hit numerous difficulties across several decades, including the Great Depression, World War II, and the NYC funding crisis of the 1970s, and remained a "pipe

Figure 1: Time Line of Construction



dream."

The Metropolitan Transportation Authority started a new study exploring various options for the 2nd Avenue subway in 1997 and approved an environmental impact statement in 2004. New York voters passed a crucial transportation bond issue to fund the expansion in November 2005. The Department of Transportation authorized funding for construction in 2006. Construction work on the line started in 2007. Construction of the subway tunnel was completed in 2011. By 2013, it was clear that the end of construction was on the horizon and a Community Information Center opened up on the UES. The grand opening of the subway was on January 1, 2017. Figure 1 shows the time line.

Figure 2 highlights the subway line in the context of the local area. The Q-line runs for 8.5 miles, including the 1.8 mile stretch of the completed 2nd Avenue Subway extension between 59th Street and 96th Street. The construction included three new subway stations on 2nd Avenue at 72nd Street, 86th Street, and 96th Street, as well as a subway tunnel connection to the existing Q-line stop on 59th Street and Lexington Avenue. The total cost of the 2 mile expansion project was $4.5 billion, making the expansion the most expensive subway construction project per mile in history.[3]

---

[3] An interesting question, outside the scope of this article, is why construction was so expensive. An investigation by the New York Times explores several possibilities. See https://www.nytimes.com/2017/12/28/nyregion/new-york-subway-construction-costs.html.

Figure 2: Subway Map on the Upper East Side of Manhattan

3 Empirical Specification

3.1 Baseline Definition of Treatment Areas

The key empirical challenge is that the value of real estate depends on a myriad of factors beyond the opening of a new subway line. Other changes in the local economic environment may confound the effects from the transit improvements on real estate values. To address this challenge, we conduct a differences-in-differences analysis, comparing valuations on the 2nd Avenue corridor before and after the subway extension, relative to outcomes in a control group.

In our baseline specification, we define the treatment group to be all the land parcels between 59th street and 100th street and between First Avenue and Third Avenue, taking the midpoint of the avenues as the demarcation line. This is what we call the 2nd Avenue corridor. Our control group consists of three corridors that make up the rest of the UES. The Lexington Avenue corridor is the collection of parcels between 59th street and 100th street and between Third and Park Avenues. The Madison Avenue corridor is the collection of parcels between 59th street and 100th street and between Park and Fifth Avenues. Its western border is Central Park. Finally, the York Avenue corridor is the collection of parcels between 59th street and 100th street to the east of the midpoint of First Avenue. Its eastern border is the East River.

This choice of baseline treatment and control group is driven by a trade-off between minimizing the treatment effect on the control group and maximizing the similarity in terms of common drivers of real estate valuations. By differencing out trends in real estate values in the control group, we remove common drivers of real estate prices that affect the entire area (UES) and isolate the effects of the subway extension. The Lexington Avenue corridor is geographically the closest to the 2nd Avenue and may be affected the strongest by the neighborhood trends that affect real estate valuation on 2nd Avenue other than the subway extension. However, the Lexington Avenue control group may also be directly affected by the subway extension. Residents in the Lexington corridor benefit from the new subway line, either because it directly shortens their commutes or because it alleviates congestion on Manhattan's busiest line, the 4-5-6, which runs under Lexington Avenue and parallel to the Q-line. The resulting improvement in transportation from the 2nd Avenue subway extension may affect real estate values in the Lexington Avenue corridor. Removing those effects tends to bias downward our estimate of the value created by the subway extension. A countervailing effect that tends to bias our treatment effects estimation upward is that the subway expansion may have made 2nd Avenue more competitive in terms of attracting residential, retail, and other commercial tenants away from Lexington Avenue.

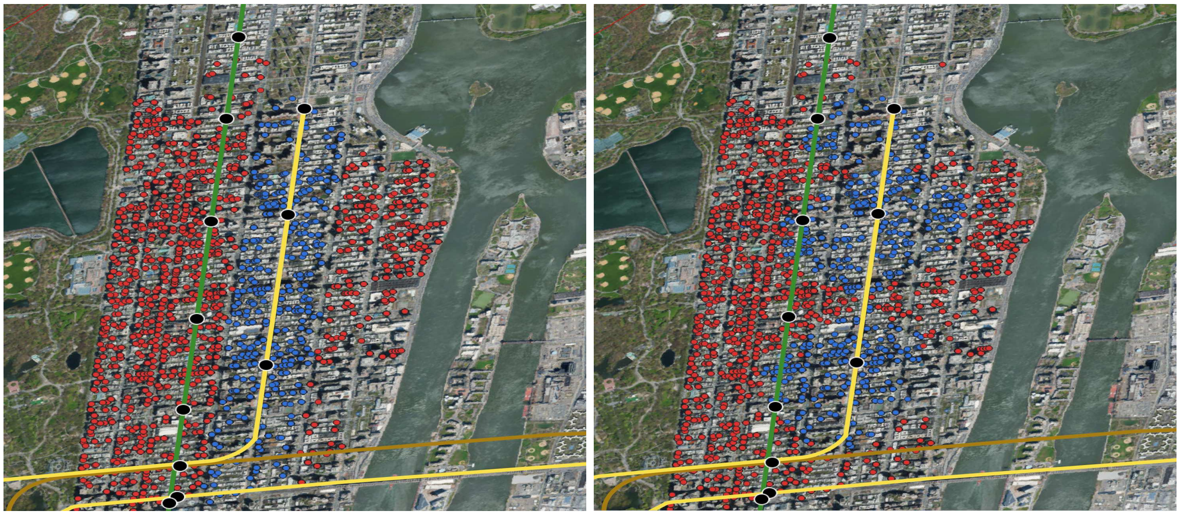
Residents living in the York Avenue corridor also potentially benefit from the Q-line extension. Indeed, for most of them, the new 2nd Avenue subway stations are the closest ones. We consider York Avenue corridor residents to be in the control group in our baseline specification because they are fairly far from the new subway stations, but study alternative treatment definitions below in section 6.4 where properties in the York Avenue corridor are part of the treatment group.

Panel A of Figure 3 indicates the buildings where we have at least one apartment transaction in our sample. Apartments in treated buildings are colored in blue while buildings in the control sample are in red. The large black dots indicate subway stations on the UES, including the three new stops on the 2nd Avenue subway.

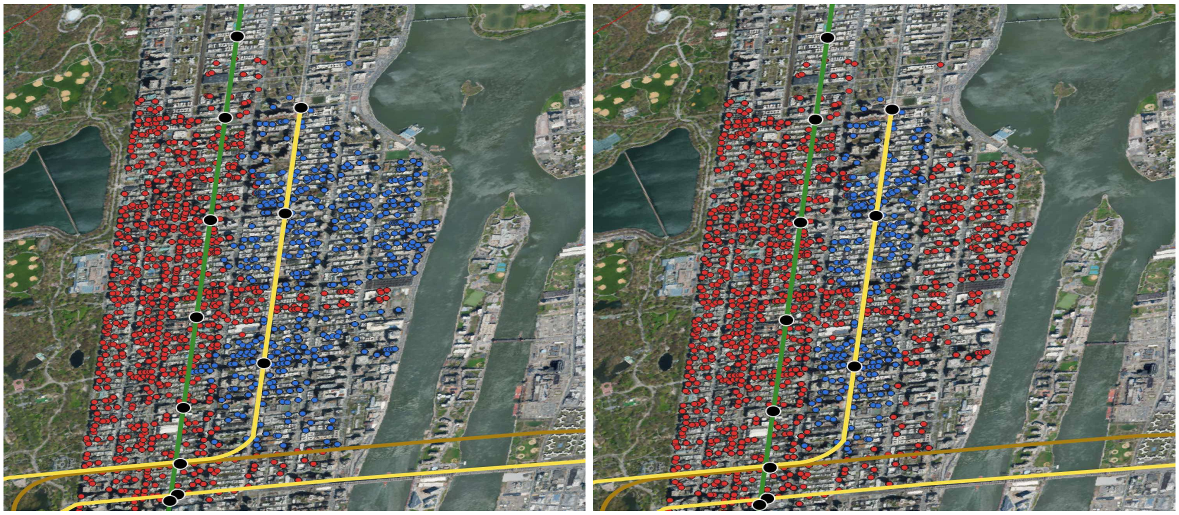
A second research design question is where to draw the demarcation line between the pre- and post-treatment periods. The subway went into operation on January 1st, 2017. While there was considerable uncertainty about the exact opening date until the last minute, eventual project completion was long anticipated. Construction started in April 2007. Tunnel excavation began in May 2010 and blasting concluded in March 2013. In 2011, the original 2013 completion date was pushed back to December 2016. Forward-looking developers and property owners willing to tolerate the inconvenience of the construction project could capture some of the potential future benefits by acting prior to the subway opening. These anticipatory effects should be reflected in real estate prices, which reflect the expected discounted value of future rents. In our benchmark analysis,

Figure 3: Treatment Based on Distance to New Stations

Panel A: Second Ave Corridor (Treat1) *Panel B: Within 0.3 Miles (Treat2)*



Panel C: Change in Distance to Nearest Station (Treat3) *Panel D: Combination Treatment (Treat4)*



Notes: Panel A shows treatment definition 1 which corresponds to properties that are on the 2nd Avenue Corridor defined as between 1st and 3rd Avenues. Panel B shows treatment 2 which consists of properties that are within 0.3 miles in walking distance of one of the new Second Avenue stops. Panel C shows treatment 3 which captures properties with a reduction in distance to the nearest subway station. Panel D shows treatment 4 which is the intersection of the first three treatments.

we strike a middle ground and take January 1st 2013 as the demarcation line between the before and after. This allows for four years of anticipation effects prior to the inauguration of the new subway line. A subway community information center was opened in 2013, signaling that project completion was no longer in doubt. This choice also provides a large enough sample in the before and after period.

3.2 Empirical Specification

Our core empirical specifications are difference-in-difference specifications defined across two dependent variables: commute times and real estate transaction prices. Our commuting regressions can be expressed as:

$$y_{it} \quad = \quad \alpha + \gamma_1 \cdot \text{Treatment}_i + \delta_1 \cdot \text{Post}_{it} + \beta_1 \cdot \text{Treatment}_i \times \text{Post}_t + \boldsymbol{X}'_{it} \cdot \theta + \varepsilon_{it}, \qquad (1)$$

in which y_{it} represents commute time for a person i in seconds. The omitted pre-period in this analysis refers to the period June 2016–Dec 2016; and the "Post" period refers to the time after subway construction, from Jan 2017–Aug 2017. The resulting β_1 coefficient captures the impact of subway construction on commuting times.

We also estimate a triple-interaction specification with an indicator for subway usage:

$$\begin{aligned} y_{it} \quad &= \quad \alpha + \gamma_1 \cdot \text{Treatment}_i + \delta_1 \cdot \text{Post}_t + \beta_1 \cdot \text{Treatment}_i \times \text{Post}_t + \boldsymbol{X}'_{it} \cdot \theta \qquad (2) \\ &+ \quad \delta_2 \cdot \text{Subway}_{it} + \beta_2 \cdot \text{Subway}_{it} \times \text{Post}_t + \delta_3 \cdot \text{Subway}_{it} \times \text{Post}_t \times \text{Treatment}_i + \varepsilon_{it}. \end{aligned}$$

In this specification, a key coefficient is δ_3, which captures the differential effect of being in the treatment area, in the post period, for subway users.

Our main specification for the real estate pricing regressions is:

$$\ln(p_{it}) \quad = \quad \alpha + \gamma_1 \cdot \text{Treatment}_i + \delta_1 \cdot \text{Post}_{it} + \beta_1 \cdot \text{Treatment}_i \times \text{Post}_t + \boldsymbol{X}'_{it} \cdot \theta + \varepsilon_{it}, \quad (3)$$

where p_{it} reflects the transaction sale price of a unit i in period t. We consider a much longer time span in our real estate analysis, with the pre-period making up January 2003–December 2012; and our post-period January 2013–March 2019. The key parameter of interest is β_1, which corresponds to the treatment effect corresponding to our various treatment definitions (for instance, properties along the 2nd Avenue corridor), in the period.

To investigate the presence of additional anticipation effects, we also consider an empirical specification using our real estate outcomes which splits the "Pre" period into two subperiods: January 2003–December 2006 and January 2007–December 2012. We call the latter period the Construction Period because it coincides with the period of heavy tunnel blasting. In those specifications, real estate prices in the Construction and Post periods are estimated relative to the omitted 2003–06 period. This specification is:

$$\ln(p_{it}) \quad = \quad \alpha + \gamma_1 \cdot \text{Treatment}_i + \delta_1 \cdot \text{Post}_t + \beta_1 \cdot \text{Treatment}_i \times \text{Post}_t + \boldsymbol{X}'_{it} \cdot \theta$$

$$+ \quad \delta_2 \cdot \text{Construction Period}_t + \beta_2 \cdot \text{Treatment}_i \times \text{Construction Period}_t + \varepsilon_{it}. \quad (4)$$

The additional parameter of interest is β_2, which corresponds to the relative price increase in the construction period (2007–12) relative to the earlier period (2003–06). The coefficient is the net effect of early anticipatory price effects and disamenity effects resulting from the construction.

Our difference-in-difference analysis accounts for the level differences in prices between treatment and control areas. However, if there are changes over time in the average characteristics of transacted properties which differ between treatment and control group, then that could affect the estimate of the subway extension. Therefore, our main specifications will control for building and housing unit characteristics $\boldsymbol{X}_{it}$. We also consider a specification that adds building fixed effects.

We focus on whether we observe convergence in prices. If the value gap for the 2nd Avenue corridor is driven by scarce access to public transportation options, we expect price convergence after subway construction.

4 Data

4.1 Location Data

We obtain unique data from smart phone Global Positioning System (GPS) pings. Our data provider aggregates GPS signals from approximately 50 million smart phone users across the United States. GPS data were combined across applications for a given user to produce pings corresponding to time stamp-location pairs. Ping data include both background pings (location data provided while the application is running in the background) and foreground pings (activated while users are actively using the application). Ping data provides nearly continuous-time location information (every 1-3 seconds) throughout the day. Our sample period covers June 2016–October 2017, an ideal time frame since the subway opened on January 1st, 2017, right in the middle of this time frame.

To identify commuting lengths, we use the panel dimension of our mobile phone data. We use a Microsoft open-source data set to define the physical footprint of buildings.[4] We isolate possible home locations by first selecting all nighttime pings by a building's users (from midnight to 7am). We require that users have a minimum presence in the buildings of three night-time pings on five different days. Then, to identify homes for these users, we require that these users ping at possible home locations at least twice on two different

---

[4] See `https://github.com/microsoft/USBuildingFootprints`.

nights. We then pick the home location as the building in which individuals ping most often over the sample. Similarly, we define possible work locations as the building in which individuals ping most often between the hours of 10am–1pm and 2pm–7pm. We select the building with the most frequent day-time ping activity as the work location. This classification produces a list of home and work locations, from which we select those with home locations on the UES.

We define morning commute length as the time difference between the last ping observed in the home location, and the first ping observed in the work location. Evening commutes are similarly defined as the difference in time between the last ping observed at work and the first ping observed at home. We require that commutes be at least 0.4 km in distance, and so exclude individuals who work at home or have minimal commutes. Commutes are expressed in seconds. The final sample contains 27,549 commutes.

We define a subway commuter as an individual who pings close to a subway stop on either the Q line or the 4-5-6, and one other station in NYC during the commute time window.

We define a recent mover as a user whose home location is in the UES after January 1, 2017 and elsewhere before.

4.2 Sales Data

We build a new dataset of all residential transactions on New York City's UES from January 2003 until March 2019. The two primary data sources are the New York City deeds records and StreetEasy.

The deeds records have information on the the sale price, sale date, address, as well as a tax ID (the BBL code). From StreetEasy we collect information on all past residential real estate sales on the UES via web scraping. We add properties between 96th Street and 100th Street, which StreetEasy considers to be part of East Harlem. We also eliminate properties that are above 100th Street along Fifth Avenue, which StreetEasy considers to be part of the UES.

StreetEasy has apartment unit and building characteristics, which are absent in the deeds records. We obtain the following building characteristics: exact street address, latitude and longitude, year of construction of the building, and building amenities. The amenity vector contains: doorman, bike room, gym, elevator, laundry room, concierge, live-in super, pool, storage room, roof deck, children's playroom, parking. Based on the exact location, we use Google Map's API to compute walking distance to Central Park and walking distance to Grand Central Terminal, a major employment center.

The unit characteristics we have are apartment unit name (e.g. 17A), the number of bedrooms, number of bathrooms, an indicator variable for condo, an indicator variable for coop, an indicator variable for studio, the square footage of the unit, and of course the transaction date and the transaction price. We infer the floor of the unit based on the apartment unit name.

A text field in the StreetEasy data describes the transaction in more detail. Based on the text field, we eliminate transactions that are commercial space, storage units, maid's rooms, parking spots, or garages. We also eliminate units that have zero bathrooms and zero bedrooms but are not studios. Importantly, we remove all "sales" which are neither reported as "sold" nor as "recorded closing." Cross-checking against the deed records database reveals that these "sales" are not actual sales but merely removed listings.

We express all transaction prices in real terms by scaling by the Consumer Price Index based in December 2017. We then eliminate all transactions with a real price below $100,000 and above $10 million. Transactions below $100,000 in 2017 dollars are unlikely to be arms-length transactions for actual apartment units in Manhattan. Transactions above $10 million are unlikely to be affected by the 2nd Avenue subway and distort sample averages. The final sample contains 49,673 transactions.

Table 1 provides summary statistics from our data. The top panel reports properties on the 2nd Avenue Corridor, which are treated according to our baseline treatment area definition. The bottom panel reports properties in the baseline control group (Madison Ave, Lexington Ave, and York Ave corridors). We have 19,941 sales in the treatment group and 29,732 in the control group, so that 40.1% of transactions are treated observations. The average property on 2nd Aenue costs $1.09million, is about 1039 square feet large, costs $1062 per sqft, has 1.5 bedrooms bathrooms, and is in a building that is 46 years old at the time of transaction. The treatment group has 40% condos and 60% coops. Buildings in the control group cost substantially more. The typical sale price is $1.84million or $1244per sqft, are 200 sqft larger, have 1.9 bedrooms and 1.8 bathrooms, and are older (59 years). There is a smaller fraction of studios (5% vs. 9%), while the condo-coop breakdown tilts more towards coops at 30%–70%.

4.3 Rental Data

We also collect data from StreetEasy on all rental buildings in the UES. For each apartment unit in the rental data (with a rental listing at some point between 2001 and 2019) we obtain the same unit and building characteristics as for the sales transactions sample: exact location (in treatment area or not, distance from Central Park, distance from Grand

Table 1: Summary Statistics

Panel A: Treatment Group

	N	Mean	St.Dev	p1	p25	p50	p75	p99
saleprice	19941	1090000	1020000	189000	509000	761000	1280000	5520000
sqft	13355	1039.486	670.708	392	670	850	1250	3158
ppsf	13330	1062.336	442.979	332.336	779.935	979	1277.826	2444.582
bedrooms	19918	1.501	0.968	0	1	1	2	4
bathrooms	19384	1.495	0.825	1	1	1	2	5
condo	19941	0.375	0.484	0	0	0	1	1
coop	19941	0.625	0.484	0	0	1	1	1
studio	19941	0.092	0.289	0	0	0	0	1
building age	19941	45.791	24.388	1	28	44	57	105
NewConstr	19941	0.059	0.235	0	0	0	0	1
closest pre	19941	0.324	0.114	0.057	0.245	0.313	0.395	0.551
closest post	19941	0.183	0.084	0.007	0.111	0.186	0.247	0.364
dist change	19941	0.14	0.128	0	0.011	0.112	0.249	0.429
treat2	19941	0.803	0.398	0	1	1	1	1
treat3	19941	0.79	0.408	0	1	1	1	1
treat4	19941	0.728	0.445	0	0	1	1	1

Panel B: Control Group

	N	Mean	St.Dev	p1	p25	p50	p75	p99
saleprice	29732	1840000	1790000	203000	646000	1180000	2330000	8730000
sqft	15527	1271.255	862.084	379	725	1050	1569	4034
ppsf	15449	1243.767	610.658	335.328	838.746	1101.92	1472.258	3381.886
bedrooms	29678	1.882	1.063	0	1	2	2.192	5
bathrooms	28875	1.83	1.03	1	1	1.5	2.5	5
condo	29732	0.304	0.46	0	0	0	1	1
coop	29732	0.696	0.46	0	0	1	1	1
studio	29732	0.053	0.223	0	0	0	0	1
building age	29732	59.009	27.97	1	42	56	83	109
NewConstr	29732	0.041	0.198	0	0	0	0	1
closest pre	29732	0.343	0.221	0.022	0.162	0.283	0.503	0.851
closest post	29732	0.265	0.14	0.022	0.158	0.247	0.357	0.603
dist change	29732	0.078	0.127	0	0	0	0.13	0.429
treat2	29732	0.219	0.414	0	0	0	0	1
treat3	29732	0.341	0.474	0	0	0	1	1
treat4	29732	0	0	0	0	0	0	0

Central), number of bedrooms, number of bathrooms, studio flag, floor, the same building amenities as listed above, year built, and total number of units in the building.

5 Commuting Length Results

We begin with an analysis of how the extension of the 2nd Avenue Subway affected commute lengths. Table 2 shows the results from the difference-in-difference estimation of equation (1) with commute length (expressed in seconds) as the dependent variable. The Post period refers to January–October 2017, the period after subway opening. We use four treatment definitions corresponding to the benchmark Second Avenue corridor treatment, defined above, and three alternative definitions of treatment, defined in more detail in Section 6.4.

Panel A of this table shows the effect of subway extension on commute times for all affected residents, regardless of their choice of commuting method. Our baseline specification, in column 1, shows a reduction in typical commute lengths of 193 seconds (over 3 minutes) for smart phone users who live in the treated corridor. This is a 7.4% reduction relative to a pre-treatment mean commuting time of 43.6 minutes in the treatment group. We find comparable treatment effects between 160 and 251 seconds when looking at alternate treatment definitions in the remaining columns. The effects are estimated precisely. Before the Q-line extension, residents in the Second Ave corridor commute 359 seconds (6 minutes) longer than other residents in the UES. The new subway line closes the average commuting gap by more than half, effectuating substantial convergence.

Panel B of Table 2 breaks out the effect by commuting mode, as in equation (2). We are particularly interested in the triple interaction of "Subway × Post × Treatment." Our results show that subway users experience a substantial reduction of 850 seconds (14 minutes) in commute lengths in the treated areas in the aftermath of the Q-line opening. We define subway commuters by their commute choices in the post period. As a result, our measure includes reductions in commute lengths from individuals who shift to subway commutes from another mode of transportation in the pre-period, as well as from those who were already commuting by subway. The large improvement in commute times, concentrated among actual subway commuters in the treatment area, points to the large impact of the Second Ave subway extension on residents' access to work.

We further note that the interaction of "Subway × Post" is also significantly negative (at the 5% or 10% level depending on the treatment definition) and estimated to be around 180-210 seconds. This shows that the Q-line extension reduced commuting times also for subway users in the control area, either because they too started using the Q train to commute to work or because the Q train alleviated congestion on the 4-5-6 line.

Next, we analyze the choice of commuting by splitting residents into recent movers to the UES and everyone else. Figure 4 shows that recent movers are substantially more likely to use the Q-line as their primary commuting choice. The difference is 16.5% points, and statistically different from zero (*t*-stat of 2.29). Since recent movers are more likely to be the marginal buyers and renters, the large gains in commuting suggest one important channel through which the Second Avenue subway extension may have increased prices and rents in real estate markets. We investigate this in the following section.

Table 2: Effect of Subway Construction on Commute Times

Panel A: Treatment Corridor

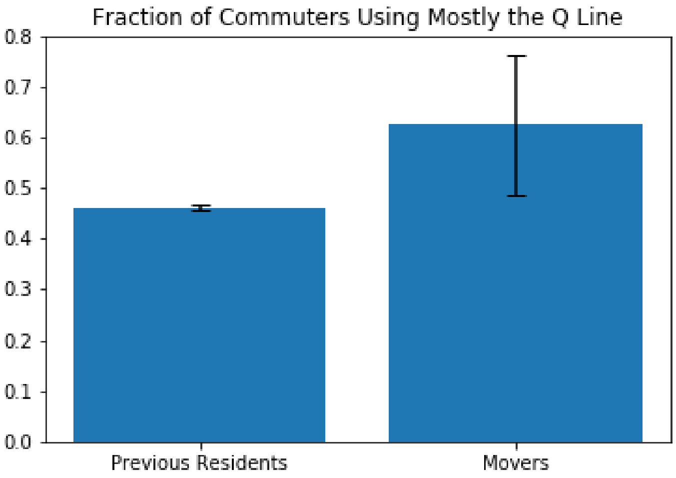
	Commute Time (sec)			
VARIABLES	On 2nd Ave	Walking Distance	Closer Subway	Intersection
Post	-3	10	-2	8
	(35)	(36)	(37)	(33)
Treatment	359***	356***	383***	448***
	(48)	(48)	(47)	(50)
Post x Treatment	-193***	-199***	-160***	-251***
	(55)	(54)	(54)	(57)
Observations	27549	27549	27549	27549
R-squared	0.004	0.004	0.006	0.005
Treatment Def.	1	2	3	4

Panel B: Interacted with Subway Use

	Commute Time (sec)			
VARIABLES	On 2nd Ave	Walking Distance	Closer Subway	Intersection
Post	144	149*	138	175**
	(91)	(86)	(91)	(86)
Treatment	-324*	153	99	-13
	(189)	(241)	(182)	(248)
Subway	-324***	-262***	-277***	-263***
	(88)	(85)	(90)	(83)
Post x Treatment	592***	631**	446**	563**
	(200)	(254)	(195)	(260)
Subway x Treatment	749***	248	330*	505**
	(195)	(246)	(189)	(254)
Subway x Post	-182*	-191**	-181*	-211**
	(99)	(94)	(100)	(93)
Subway x Post x Treatment	-850***	-854***	-653***	-864***
	(208)	(260)	(203)	(267)
Observations	27549	27549	27549	27549
R-squared	0.013	0.016	0.016	0.015
Treatment Def.	1	2	3	4

Notes: Post is an indicator variable for the period from January 1st 2017–October 2017. Treatment is an indicator variable for units exposed to the subway extension that varies by column. Treatment definition 1 corresponds to properties that are on the 2nd Avenue Corridor defined as between 1st and 3rd Avenues. The second treatment definition consists of properties that are within 0.3 miles in walking distance of one of the new 2nd Avenue stops. The third treatment definition captures individuals with a reduction in distance to the nearest subway station. The fourth treatment definition is a composite requiring that all three treatments hold. Panel A runs a difference-in-difference specification, following equation 1, across these four treatment definitions before and after subway extension on commute times. Commutes are defined as the time difference between pings observed at home and work locations, as described in the text. Panel B shows a triple interaction with the effects broken out by whether users use the subway. Subway usage is defined as whether individuals (in the post-period) ping close to either the 4-5-6, Q-line, and one other station in NYC during the commute time window. Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Figure 4: Subway Impact on Commuting



6 Real Estate Capitalization Results

6.1 Corridors: Baseline Treatment and Control

The previous section established the strong impact of the Q-line construction on commuting patterns. Individuals in treated areas saw substantial declines in commuting, driven by subway commuters. New residents were disproportionately likely to use the Q-line train. Given the complementarity between transportation improvements and real estate, we investigate the hypothesis that these transportation improvements led to valuation gains in real state markets.

Table 3 presents our main treatment estimates to measure the real estate capitalization effect of subway construction. The Post variable in this specification captures the price impact after January 2013, and so accounts for any general time-series increase in price; relative to the entire pre-period of January 2003–December 2012. Our specifications suggest that these time trends are generally important. In column 1, for instance, the coefficient on Post is 0.0903 on log price. This suggests that the post-period is associated with a price premium of $\exp(0.0903) - 1 = 9.45\%$. This variable accounts for the general increase in valuation of UES apartments. The Treat coefficient captures the value differential associated with being "On 2nd Avenue" in general. This effect is quite negative. Properties in the 2nd Avenue corridor generally transact for 37.5% less than properties in the control group, i.e, in the rest of the UES, before considering controls.

The key coefficient of interest is that on the interaction effect "Post × On 2nd Ave." This coefficient measures the differential price impact of being on the 2nd Avenue corridor

Table 3: Main Difference-in-Difference Results - Baseline Treatment Definition

VARIABLES	(1) Log Price	(2) Log Price	(3) Log Price	(4) Log Price	(5) Log Price
Post x On 2nd Ave	0.138***	0.0970***	0.0432***	0.138***	0.0597***
	(0.0154)	(0.00957)	(0.00866)	(0.0112)	(0.0103)
Constr. Period x On 2nd Ave				0.0845***	0.0317***
				(0.0115)	(0.0104)
Post	0.0903***	0.123***	0.111***	0.177***	0.159***
	(0.00982)	(0.00610)	(0.00550)	(0.00717)	(0.00652)
On 2nd Ave	-0.469***	-0.203***		-0.246***	
	(0.00927)	(0.00612)		(0.00849)	
Constr. Period				0.101***	0.0882***
				(0.00721)	(0.00652)
Observations	49,673	49,673	49,673	49,673	49,673
R-squared	0.068	0.643	0.739	0.648	0.741
Controls	NO	YES	YES	YES	YES
Building FE	NO	NO	YES	NO	YES

Notes: Post is an indicator variable for the period after January 1st 2013. Constr. Period is an indicator variable for the construction period between January 1st 2007 and December 31, 2012. On 2nd Ave is an indicator variable for a unit located in the Second Avenue Corridor as defined in the main text. Controls include: an indicator variable for a condo transaction; an indicator variable for a studio; number of bedrooms; number of bathrooms; the floor of the building; the year of construction; distance to Central Park; distance to Grand Central Terminal; indicator variables for building amenities (doorman, bike room, gym, elevator, laundry room, concierge, live-in super, pool, storage room, roof deck, children's playroom, parking); as well as indicators if the control variables are missing. Standard errors in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

after 2013, the time period when subway completion was either imminent or achieved. This period captures at least some of the anticipatory effects of subway completion on real estate values, namely those between January 1st 2013 and subway opening on January 1st of 2017. It also contains the subsequent price effects in 2017, 2018, and the first quarter of 2019. The coefficient on the interaction term in column 1 suggests that the 2nd Avenue Subway resulted in a statistically significant and economically large price rise of 14.8% for properties transacting on the avenue. This number suggests that the construction of the subway was associated with a substantial value creation. We observe convergence in prices: Subway construction closes over 1/3 of the gap in valuations between the 2nd Ave corridor and the rest of the UES.

Our main specification is reported in column 2. It adds a number of important controls to account for the differences in unit and building characteristics documented above. Controls include: an indicator variable for a condo transaction, an indicator variable for a studio, categorical variables for the number of bedrooms (1BR, 2BR, 3BR, 4+BR), the number of bathrooms, the floor of the building, the year of construction, the distance to Central Park (an important recreational amenity), the distance to Grand Central Terminal (an important central business district), and indicators for the various building amenities described above; as well as indicators if the control variables are missing. These control variables boost the R^2 value from 6.8% in column 1 to to 64.3% in column 2. The lower coefficient (in absolute value) of "On 2nd Ave" indicates that about half of the unconditional difference in valuations between the treatment and control group is accounted for by different average characteristics. However, the estimate of Post × On 2nd Ave remains large and precisely estimated at $\exp(.097) - 1 = 10.2\%$. It indicates even faster convergence of property prices than in column 1: nearly 1/2 of the price difference between 2nd Ave properties and properties in the rest of the UES is eliminated around the time of subway completion.

One possibility is that there are additional property characteristics beyond those included in column 2, and unobserved to us, that matter for real estate values. We capture constant latent differences across neighborhoods and buildings by including building fixed effects in column 3 of Table 3.[5] This specification compares transactions in the same building before and after the subway. Our sample is dominated by transactions in large buildings; 92% of observations are in buildings that contain at least five transactions in the Pre period and at least five transactions in the Post period. Thus, we have enough power to identify most building fixed effects accurately. Adding building fixed effects in column 3 increases the R^2 to 73.9%. In this specification, property values are 4.4% higher on Second Avenue in the Post relative to the Pre period and relative to the control group. The estimate is significant at the 1% level and remains economically large.

6.2 Additional Anticipation Effects

We consider the possibility of additional anticipation effects as far back as 2007, the year when the decade-long subway construction endeavor began. We include an indicator variable "Constr. Period" which takes the value of 1 for transactions between January 2007 and December 2012, allowing for six more years of potential anticipation effects. This being also the period of heaviest construction, it is plausible that this period experienced a reduction in property values due to disamenities (noise, pollution, closure of retail) related to construction activity. The interaction effect of "Constr Period × On 2nd Ave" estimates the net effect of additional anticipation and disamenities on prices in the 2nd Ave corridor, relative to the omitted category of 2003-06. The coefficient on "Constr. Period" shows the general price trend on the entire UES during this period, relative to

---

[5] The coefficient on the treatment variable is not separately identified from the building fixed effects so we drop it in the specifications with building fixed effects.

the omitted category of 2003–2006. Under this specification, the "Post × On 2nd Ave" coefficient measures the price change between the period 2013–2019 and the earlier period 2003–2006 (rather than relative to 2003–2012 in columns 1 and 2).

Column 4 of Table 3 shows that the construction period was associated with a substantial increase in real estate values in general on the UES. Prices were 10.6% higher in real terms in 2007–12 relative to 2003–06, after controlling for property characteristics. Properties on the 2nd Ave corridor appreciated by 8.8% more than properties in the control group over this period. The point estimate is statistically significant and demonstrates the presence of additional anticipation effects, strong enough to outweigh the disamenity effects from construction.

In the Post period, properties on 2nd Ave are 14.8% more valuable than in 2003–06, relative to the control group. In sum, subway construction triggered an initial appreciation of 8.8% in 2007–12 and a further appreciation of 6% (14.8% − 8.8%) in 2013–2019.

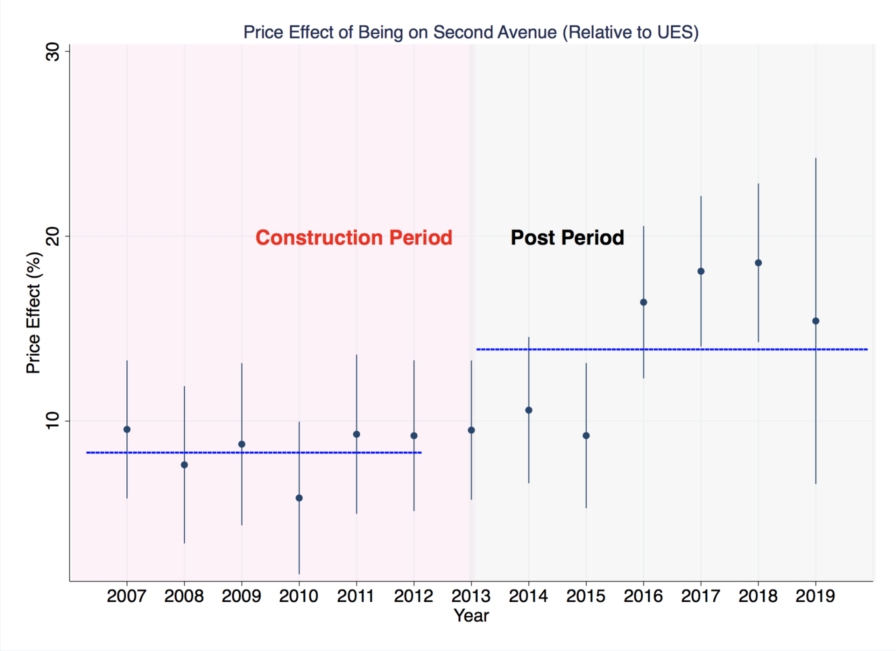
Figure 5 illustrates this result graphically. It plots the coefficient estimates from a dynamic difference-in-differences specification, in which each calendar year is allowed to have its own treatment effect. We see positive price coefficients that are stable around 10% in the construction period of 2007–2012. The price effects grow stronger after 2013, and are especially large in 2016–2018, a period centered around subway opening. This helps alleviate the concern that other trends are driving the effect. The graph also illustrates that our results are not sensitive to various choices of demarcation between Pre and Post periods between 2007 and 2015. By the end of the sample in 2019.Q1, the treatment effect ceases to grow, suggestion that the market has largely priced in the full impact of subway construction.

In column 5 of Table 3, we add building fixed effects to the specification of column 4. The early anticipation effect during the construction period is smaller at 3.3% but remains statistically precisely estimated. Property values in the Post period are 6.2% higher than in the 2003–06 period on 2nd Avenue compared to the control group. This is an economically and statistically significant difference.

6.3 Unpacking the Control Group

In Table 4, we revisit our main specification but unpack the control group into its constituent corridors. The omitted corridor is the Madison Ave corridor (spanning from Fifth Ave to Park Ave), so that all changes are measured relative to that Madison Ave corridor. Since this corridor is the farthest removed from the 2nd Ave subway and since it contains very wealthy residents who are less likely to use public transportation, this

Figure 5: Dynamic Treatment Effects - Baseline Treatment



is a natural choice for the omitted category. Column 1 shows that property prices in the pre period (after controlling for building and unit characteristics) were the lowest on 2nd Ave, closely followed by York Ave, then Lex Ave, and highest on Madison Ave (omitted category).

We continue to see our main treatment effect: prices appreciate by 13% more in the 2nd Ave corridor in the Post period relative to the Madison Ave corridor. In contrast, we see no change for the Lexington Ave corridor. The null effect on Lexington Ave suggests that the potential benefits from reducing congestion on the 4-5-6 line did not affect relative property prices on Lexington Avenue, or were entirely offset by by reductions in relative prices due to increased competition in the neighboring real estate submarket of 2nd Ave. The York Avenue corridor sees a substantial 7.0% price change. The estimate is about half as large as the treatment effect for the 2nd Ave corridor and is significant at the 1% level. This evidence suggests that York Ave may have been at least partially affected (treated) by the subway extension. We study this possibility in more detail below.

Column 2 adds building fixed effects. It finds a 6.3% price gain on Second Ave, that is precisely measured, and a 3.4% price gain on York Ave which is significant at the 5% level.

The last two columns consider the specification with the construction period broken out. Column 3 shows a strong 18.5% capital gain on 2nd Ave, relative to Madison Ave

Table 4: Unpacking the Control Group

VARIABLES	(1) Log Price	(2) Log Price	(3) Log Price	(4) Log Price
Post x On 2nd Ave	0.122***	0.0610***	0.170***	0.0726***
	(0.0138)	(0.0127)	(0.0161)	(0.0149)
Post x On Lexington Ave	0.0103	0.0157	0.0126	0.00270
	(0.0153)	(0.0140)	(0.0179)	(0.0164)
Post x On York Ave	0.0677***	0.0326**	0.0877***	0.0318*
	(0.0155)	(0.0140)	(0.0181)	(0.0165)
Constr. Period x On 2nd Ave			0.0969***	0.0210
			(0.0162)	(0.0147)
Constr. Period x On Lexington Ave			0.00391	-0.0254
			(0.0180)	(0.0163)
Constr. Period x On York Ave			0.0386**	-0.00406
			(0.0181)	(0.0164)
Post	0.0981***	0.0931***	0.144***	0.146***
	(0.0117)	(0.0107)	(0.0137)	(0.0125)
On 2nd Ave	-0.498***		-0.545***	
	(0.0133)		(0.0156)	
On Lexington Ave	-0.236***		-0.237***	
	(0.0106)		(0.0141)	
On York Ave	-0.443***		-0.460***	
	(0.0189)		(0.0209)	
Constr. Period			0.0859***	0.0989***
			(0.0136)	(0.0123)
Observations	49,673	49,673	49,673	49,673
R-squared	0.649	0.739	0.653	0.741
Controls	YES	YES	YES	YES
Building FE	NO	YES	NO	YES

Notes: "Post" is an indicator variable for the period after January 1st 2013. "Constr. Period" is an indicator variable for the construction period between January 1st 2007 and December 31 2012. Control variables are the same as in Table 2. Standard errors are in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

and relative to the pre-construction era of 2003-006. The gain of 10.2% in the construction period underscores early anticipation effects. Lexington Ave shows no change in either period, relative to Madison. Property prices on York Ave appreciate in the 2007–12 period relative to Madison Ave (3.9%). York Ave prices continued to catch up relative to Madison Ave in the Post period, for a combined effect of 9.2%. Finally, in column 4, we add building fixed effects. While the post-period real estate capital gain remains strong at 7.5% and precisely estimated, the construction-era effect disappears.

6.4 Alternative Treatment Definitions

6.4.1 Distance to New Stations

One drawback of our baseline definition of treatment is that we assume that all properties along the 2nd Avenue Corridor are equally treated by new subway construction. This may not be the case if areas far from the subway stops, along 2nd Ave, do not find much of a benefit from using the new subway. To analyze this possibility, we consider a second treatment definition which includes all properties within 0.3 miles of one of the three new 2nd Avenue subway stops.[6] If the properties which benefit the most from the subway construction, they should expect the greatest property price appreciation. But, disamenities from construction may also have been greatest closest to the subway stops.

Table 1 refers to this alternative treatment definition as "treat2". It shows that 80.3% of the transactions on the 2nd Avenue corridor and 21.9% of the transactions in the Madison, Lexington, and York Ave corridors fall within 0.3 miles of one of the new subway stations. In other words, this treatment is strongly but not perfectly correlated with our baseline treatment. Figure 3, Panel B shows the treated and control buildings. The 0.3-mile distance requirement traces diamond-shaped areas around the three new subway stations.

Table 5: Difference-in-Difference Estimates: Alternative Treatment Definitions

VARIABLES	(1) Log Price	(2) Log Price	(3) Log Price	(4) Log Price	(5) Log Price	(6) Log Price
Post x Treat	0.0711***	0.0398***	0.0862***	0.0297***	0.0819***	0.0372***
	(0.00947)	(0.00851)	(0.00945)	(0.00849)	(0.0103)	(0.00937)
Post	0.129***	0.110***	0.115***	0.113***	0.137***	0.117***
	(0.00641)	(0.00578)	(0.00683)	(0.00614)	(0.00564)	(0.00506)
Treat	-0.137***		-0.151***		-0.165***	
	(0.00592)		(0.00769)		(0.00656)	
Observations	49,673	49,673	49,673	49,673	49,673	49,673
R-squared	0.639	0.739	0.638	0.739	0.640	0.739
Controls	YES	YES	YES	YES	YES	YES
Building FE	NO	YES	NO	YES	NO	YES
Treatment Def.	2	2	3	3	4	4

Notes: "Post" is an indicator variable for the period after January 1st 2013. "Treat" is an indicator variable which take son the value of 1 if a transaction is in the treatment area. The table considers three alternative treatment definitions, as indicated in the last row. Columns 1 and 2 use treatment definition 2 which takes the value of 1 for a transaction located within 0.3 miles of one of the three new subway stations on the Second Avenue subway and 0 otherwise. Columns 3 and 4 use the change in distance definition (treatment 3) which is 1 for a transaction located in an area that experienced a change in distance to the closest station after the Second Avenue subway and 0 otherwise. Columns 5 and 6 use the all of the above definition (treatment 4) which is 1 for a transaction located in treatment areas 1, 2, and 3 and 0 otherwise. Controls are the same as in Table 2. Standard errors in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Columns 1 and 2 of Table 5 revisit our main difference-in-differences estimation for this alternative treatment definition and for our preferred specifications with building and unit controls, and with building fixed effects. We find a strongly positive and statistically

---

[6] Distance is defined by walking distance as calculated by Google Maps. For each of our buildings, we feed in the street address into the Google Maps API and obtain the distance to each subway station entrance (multiple per station) on the UES, to Central Park, and to Grand Central Terminal.

significant increase in value due to the subway for those properties that are within 0.3 miles of one of the three new Q-line stations. The headline increase is 7.4%, while the increase with building fixed effects is 4.1%. The corresponding numbers for the baseline treatment were 10.2% and 4.4%. This comparison suggests that properties in the 2nd Avenue corridor that are not within 0.3 miles from a new station benefitted slightly more from the subway than properties in the Lexington Ave or York Ave corridors that are within 0.3 miles of a new 2nd Ave subway station.

Further investigation, reported in Appendix Table 10, breaks down the treatment group into transactions that are between 0 and 0.10 miles, between 0.10 and 0.20 miles, and between 0.20 and 0.30 miles from a new Q-line station. The overall 7.4% price gain results from a large and precisely estimated gains of 9.1% in properties between 0.2 and 0.3 miles away from the station and 6.8% in properties between 0.1 and 0.2 miles away. The gain closer by is 3.4% and only significant at the 10% level. The analysis also shows a small price decline closest to the station during the construction period. This is exactly where we expect the disamenities from construction to show up. In contrast, prices in the 0.2–0.3 mile ring appreciate 10.7% during the construction period and an additional 4.3% (for a total effect of 15%) in the Post period.

6.4.2 Closest Subway Station Becomes Closer

We explore a second alternative treatment definition which places greater weight on peripheral properties which experienced possibly large gains in transit access. For every apartment in our sample, we compute the distance to the nearest subway station on any line serving the UES, both before and after the addition of the three stations on the Second Avenue subway line (8 stations in total). Distance is calculated as walking distance based on Google Maps taking into account that each station has multiple entrances.

Table 1 reports that for the average unit in the 2nd Ave corridor, the closest station was 0.32 miles away before the Q-line extension and 0.18 miles after, for an average distance reduction of 0.14 miles (225 meters). For the residents of the other three corridors in Panel B, the average reduction was smaller at 0.08 miles (129 meters). The latter is the combination of a zero reduction for all residents of the Madison corridor and most residents of the Lexington corridor, on the one hand, and a large reduction for the residents on the York Ave corridor, on the other hand. We define an apartment as treated if there is a strictly positive distancef reduction to the nearest subway station on the UES. Table 1 refers to this alternative treatment definition as "treat3". It shows that 79% of the transactions in the 2nd Avenue corridor and 34.1% of the transactions in the Madison, Lexington, and York

Ave corridors are in a building which experiences a reduction in distance to the nearest station. Again, this treatment is strongly but not perfectly correlated with our baseline treatment. Figure 3, Panel C shows the treated and control group buildings according to this second alternative treatment definition. The largest change with the baseline and this alternative treatment is that nearly all properties east of Second Avenue are now treated.

Columns 3 and 4 of table 5 shows the difference-in-difference estimates. For our main specifications, we find a similar effect from the subway extension: 9% without and 3% with building fixed effects.

Further investigation, reported in Appendix Table 11, breaks down the treatment group into units that experienced a reduction in distance (i) between 0 and 0.10 miles, and (ii) greater than 0.10 miles. The latter group consists mostly of units east of 2nd Ave. The 9% overall price effect is the average of 19% estimated gains in the former group, and 5.1% in the latter group. Both are significant at the 1% level. While one might think that properties experiencing a larger reduction in distance are "more intensively" treated, the data suggest that the gains are largest for those who experience a modest reduction in distance. For some far east residents, it is possible that the 2nd Ave subway remains too far away to be useful. Alternate transportation options may dominate even after the new subway becomes available. Also, properties close to the East River are 10.5% more expensive in the Pre period, suggesting a wealthier clientele that may have lower utilization of public transportation in the first place. Nevertheless, even the 5.1% price gain is substantial and helps to put in context the York Ave results presented above.

6.4.3 All of the Above

A final alternative treatment definition combines the first three treatments. We consider a unit treated if it is treated under all three previous definitions. This treatment isolates properties on the 2nd Ave corridor, close to a new subway station, for which one of the new stations is the closest subway option (i.e., there is a distance reduction). Table 1 reports that 72.8% of units on the 2nd Ave corridor satisfy this requirement ("treat4") and none of the units on the other corridors, by construction. About 29.2% of the overall sample receives this combination treatment. Figure 3, Panel D shows the treatment and control groups according to the combination treatment definition.

Columns 5 and 6 of Table 5 show the difference-in-difference estimates. For our main specifications, we find a 8.5% and 3.8% subway effect, both of which are precisely estimated. In conclusion, the analysis in this section confirms large and robust estimated effects from the Q-line subway extension.

6.5 Heterogeneous Treatment and Supply Response

Though our results suggest substantial effects of the Q-line construction on prices on average, we also consider the possibility that the subway extension had different effects on newer buildings. We define newer buildings to be those constructed after January 2003. The categorical variable "NewConstr" isolates transactions in these buildings. Table 1 shows that 5.9% of units transacted in the treatment group are in newer buildings compared to 4.1% in the control group. Table 6 estimates the triple interaction effect "Post x Treat x NewConstr." We find a 10.1% larger appreciation for units in newer buildings in the treatment area after subway construction than for older buildings. The appreciation is 9.6% for older buildings and 19.7% for newer buildings. The additional 10.1% is precisely estimated despite the relatively small share of transactions in buildings built after 2003. The remaining columns of Table 6 show an even larger treatment effect for recently constructed units when using the alternative treatment definitions. The treatment effect for units built before 2003 remains statistically and economically large in all specifications, however. In sum, one channel through which the 2nd Ave subway has resulted in convergence in real estate values between the 2nd Ave subway corridor and the rest of the UES is by promoting the development of new residential units. Units in newer buildings trade at a substantial premium to existing units, as can be seen in the $\exp(.403) - 1$ =49.6% estimate on "NewConstr." While the new-building premium fell substantially in the Post period in the control group (-13.7%), it fell much less in the treatment group (-13.7% + 10.1% = -3.6%). A larger prevalence of new units on the 2nd Ave corridor then contributes to the convergence.

Motivated by this result, we investigate further whether the 2nd Ave subway extension triggered a housing supply response. We measure new construction from permit data. We obtain data on the number of building permits issued, the number of building permits for new construction jobs, the total estimated cost of the permitted projects, and the number of new residential buildings that receive a permit. The difference-in-difference analysis is in Appendix Table 12. We find economically meaningful housing supply responses; all four variables increase between 3.8% and 36.6%. However, none of the effects are measured precisely, since variables that count the number of permits are only available at the level of the treatment and control area (34 annual observations). The fourth variable, the estimated cost of the permitted construction, is available at the unit level but is a noisy estimate of new apartment construction. It includes all renovation jobs—which dwarf new construction permits in number—and the variable is just an initial estimate of the cost of the work.

Table 6: Heterogeneous Treatment for New vs. Old Buildings

VARIABLES	(1) Log Price	(2) Log Price	(3) Log Price	(4) Log Price
Post x Treat	0.0914***	0.0694***	0.0704***	0.0729***
	(0.00977)	(0.00969)	(0.00967)	(0.0106)
Post x Treat x NewConstr	0.0962***	0.128***	0.295***	0.162***
	(0.0342)	(0.0343)	(0.0358)	(0.0347)
Post x NewConstr	-0.128***	-0.154***	-0.282***	-0.154***
	(0.0290)	(0.0286)	(0.0316)	(0.0266)
Post	0.115***	0.120***	0.113***	0.130***
	(0.00621)	(0.00654)	(0.00694)	(0.00574)
Treat	-0.213***	-0.149***	-0.168***	-0.176***
	(0.00613)	(0.00593)	(0.00771)	(0.00657)
Newconstr	0.403***	0.391***	0.390***	0.397***
	(0.0143)	(0.0144)	(0.0144)	(0.0144)
Observations	49,673	49,673	49,673	49,673
R-squared	0.641	0.637	0.636	0.638
Controls	YES	YES	YES	YES
Building FE	NO	NO	NO	NO
Treatment Def.	1	2	3	4

Notes: "Vintage2" is an indicator variable which is 1 for units in buildings constructed in 2003 or later and zero otherwise. All other variables are as in Table 3. Each column uses an alternative definition of the treatment area, as highlighted in the last row of the table. The alternative treatment definitions 2, 3, and 4 are the same as in Table 4. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

We explore a second source of heterogeneity in the treatment effect. Appendix Table 13 reports treatment effects by apartment size, measured as the number of bedrooms. The omitted category is units with zero bedrooms (studios). We find significant treatment effects across all apartment sizes but the percentage gains are monotonically increasing in the size of the apartment: around 7.6% for one-bedroom units rising to 17.1% for 2BR, 25.0% for 3BR, and 47.1% for units that have four or more bedrooms. The estimates are similar for all four treatment definitions.

The results in this section suggest that gains were unequally distributed and were strongest for larger and newer housing units.

6.6 Repeat Sales

In Table 7, we perform a repeat-sales analysis. This commonly used approach in real estate valuation compares the prices of properties with the previous price paid for the same property. It has the virtue of holding (most) unit characteristics constant. It has the well-known limitation that we are only able to analyze properties that do, in fact, repeatedly transact in this period. We have 16,883 repeat sales, representing only 34% of the total number of transactions, confirming a large reduction in sample size. Column 1 repeats the earlier analysis (main specification with controls) on the subset of apartments that transacts at least twice.[7] The repeat-sales sample features a smaller estimate of the baseline treatment effect: a 3% value creation estimate from the subway extension compared to a 10.2% effect for the full sample. In other words, this one-third subsample with repeat transactions displays a baseline treatment effect that is one-third as large as the full-sample estimate.

With that new baseline estimate in mind, column 2 adds the log residual sale price of the previous transaction of the same unit, i.e., from the first leg of the repeat sale. This residual sale price is the unexplained component from a regression of the log sale price on Post, Treat, Post × Treat, and controls. This procedure removes the subway effect from the transaction price paid in the first leg of the transaction. The residual contains all other unmeasured unit and building characteristics that impact valuation. The lagged residual price enters strongly significantly with a coefficient around 0.6 and boosts the regression R^2 from 74.2% to 85.3%. The last six columns repeat the same two specifications for the three alternative treatment definitions. In all cases, we continue to find significant treatment effects with point estimates on Post × Treat between 1.6% and 2.8%. In sum, controlling for additional unit characteristics via repeat sales shaves off about 1% point from the baseline gain estimate.

---

[7] When determining whether a transaction in our 2003–2019 dataset is a repeat sale, we look for transactions in StreetEasy before January 2003 to avoid selection on properties that transact twice within the 2003–2019 time frame. Despite limited data coverage prior to 2003, this results in several hundred additional repeat sales included in the analysis. Also, if a property is the subject of two (or more) repeat sales, both (all) repeat-sales transactions for which the second leg of the trade pair is in our sample period 2003–2019 enter the repeat sales sample.

Table 7: Repeat Sales Subsample

VARIABLES	(1) Log P	(2) Log P	(3) Log P	(4) Log P	(5) Log P	(6) Log P	(7) Log P	(8) Log P
Post x Treat	0.0299**	0.0191**	0.0432***	0.0276***	0.0544***	0.0211**	0.0265**	0.0164*
	(0.0119)	(0.00900)	(0.0119)	(0.00894)	(0.0120)	(0.00897)	(0.0128)	(0.00961)
Post	0.109***	0.0537***	0.100***	0.0478***	0.0913***	0.0485***	0.113***	0.0555***
	(0.00794)	(0.00600)	(0.00840)	(0.00633)	(0.00903)	(0.00676)	(0.00729)	(0.00549)
Treat	-0.158***	-0.181***	-0.131***	-0.144***	-0.112***	-0.126***	-0.126***	-0.149***
	(0.00881)	(0.00664)	(0.00865)	(0.00650)	(0.0105)	(0.00785)	(0.00951)	(0.00713)
Lagged Log P Resid		0.589***		0.592***		0.597***		0.593***
		(0.00520)		(0.00519)		(0.00517)		(0.00518)
Observations	16,883	16,883	16,883	16,883	16,883	16,883	16,883	16,883
R-squared	0.742	0.853	0.739	0.853	0.736	0.852	0.738	0.853
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Building FE	NO	NO	NO	NO	NO	NO	NO	NO
Treatment Def.	1	1	2	2	3	3	4	4

Notes: This is the subsample of sales transactions for which we observe a prior transaction in the data. The lagged log price residual is the residual from a first-stage regression of the log price in the first transaction of the repeat-sales pair on Post, Treat, Post × Treat, and controls. Standard errors in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

7 Rental and Valuation Analysis

The real estate value creation effects from the subway extension, found in the prior analysis, not only manifest themselves in price gains on owner-occupied units but also in rent increases in rental buildings. We use the universe of rental listings to repeat the difference-in-differences analysis on log asking rents. We include the same, long list of property and unit characteristics to control for observable differences in order to isolate the subway effect. One caveat to this analysis is that the data set contains asking rents not contract rents. To the extent that this creates measurement error, it would attenuate the coefficient of interest. We only include one rental observation per unit-year to avoid double-counting repeated listings of the same unit. The final sample contains 99,034 rental unit-year observations.

Columns 1 and 2 of Table 8 show the treatment estimates for the rental sample. In column 1, we repeat the main specification from the sales analysis, with the Post period starting in 2013 and controls included. We find that rents are 1.8% higher on the 2nd Ave corridor in the Post period. This rental increase closes nearly 1/3 of the 6.19% gap in rent levels between the 2nd Ave corridor and the rest of the UES. The effect is economically large and precisely estimated.

Table 8: Rentals: Difference-in-Difference Results - Baseline Treatment Definition

	Property-level		Block-level indices					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Log R	Log R	Log R	Log P	Log P/R	Log R	Log P	Log P/R
Post x Treat	0.0177***	0.00685***	0.0203	0.0465**	0.0262	0.0274	0.0696***	0.0421
	(0.00255)	(0.00241)	(0.0149)	(0.0212)	(0.0219)	(0.0185)	(0.0264)	(0.0271)
Post	0.0322***	0.00849***	0.0172*	0.0841***	0.0670***	-0.00293	0.0812***	0.0842***
	(0.00186)	(0.00171)	(0.00919)	(0.0131)	(0.0135)	(0.0115)	(0.0164)	(0.0168)
Treat	-0.0601***	-0.0498***	-0.111***	-0.195***	-0.0838***	-0.106***	-0.186***	-0.0799***
	(0.00221)	(0.00147)	(0.0110)	(0.0156)	(0.0162)	(0.00852)	(0.0122)	(0.0125)
Observations	99,034	99,034	1,853	1,853	1,853	1,853	1,853	1,853
R-squared	0.808	0.806	0.404	0.422	0.105	0.400	0.414	0.108
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Building FE	NO	NO	NO	NO	NO	NO	NO	NO
Post Year	2013	2017	2013	2013	2013	2017	2017	2017
Treatment Def	1	1	1	1	1	1	1	1

Notes: The dependent variable in columns 1 and 2 is log asking rents at the unit level. The dependent variable in columns 3 to 8 are log rents, log prices, and log price/rent ratios at the tax block level. Log rents (log prices) at the tax block level are obtained as the fixed effects in a first-stage regression of log rents (log prices) of individual apartment units on tax block x year fixed effects and a vector of unit and building controls, except for distance to Central Park and to Grand Central. 'Controls' indicate different control variables in each specification. In columns 1 and 2, the controls refer to the same unit characteristics used for our main regressions with sales data. In columns 3-8, controls refer to the tax block-level distance from Central Park and Grand Central Station.

In column 2, we redefine the Post period as the period after January 1, 2017. This date marks the opening of the subway. In this specification, we find annual rents that are 0.69% higher in our main treatment area. The effect is precisely estimated. Comparing columns 1 and 2, we find some anticipation effects in the rental market as well. This is consistent with tenants that expect to stay for multiple years and move in anticipation of the subway opening. It is also consistent with a rebound in local area amenities (e.g., street-level retail) after 2013, which were temporarily depressed during the heavy construction phase from 2007 to 2012.

A naïve but common valuation approach in commercial real estate is to divide the rent (net of operating expenses) by a constant cap rate (rent-price ratio) to arrive at the property value. The average multi-family cap rate for the UES over our sample period is around 5%. Assuming operating expenses are proportional to rents and that the rental increase due to the subway is a permanent level shift, the 0.69% increase in rents translates into a 13.8% increase in prices $(0.0069/0.05)$. This is in the same ballpark as our estimated treatment effect on sales transaction prices.

The asset pricing (Campbell and Shiller, 1988; Cochrane, 2011) and real estate literatures (Campbell, Davis, Gallin, and Martin, 2009; Plazzi, Torous, and Valkanov, 2010; Van Nieuwerburgh, 2019) strongly suggest that discount rates may not be constant, however.

Having shown positive responses of both prices and rents to the 2nd Ave subway construction, an interesting question is what happens to valuation ratios, which are a good proxy for the (inverse of the) discount rate.

When forming price-rent ratios, it is important to compare similar units that are for sale and for rent. This is feasible in a dense urban neighborhood like the Upper East Side where both owner- and renter-occupied units are prevalent, often of similar type and quality, on nearly every block. To construct the log price-rent ratio in a given tax block and year for a comparable property, we first estimate separate regressions for log prices and log rents on a full set of tax block × year fixed effects and full set of control variables, using our sales transactions and rental listing data sets, respectively.[8] We then subtract the block-year fixed effect, estimated from the log price regression, from the corresponding block-year fixed effect, estimated from the log rent regression, to form the log price-rent ratio for each block-year. We sort tax blocks into Treatment and Control areas based on their location, using our main treatment definition 1.

We then regress the log rent, the log price, and the log price-rent at the tax block level on Post, Treat, Post×Treat, and controls for distance to Central Park and to Grand Central Terminal. In columns 3-5 the Post period is post-2013, while in column 6-8, the Post period is post-2017. Since the analysis is at the block level, there are fewer observations (1,853) and consequently less power. The earlier regressions of log price and log rent at the unit level already established significance, so that we can focus on economic magnitudes for this exercise. In both columns 5 and 8, we find that at least half of the treatment effect on log price is accounted for by an increase in the log price-rent ratio. Thus, the Second Ave subway expansion reduced the discount rate on residential real estate by about 2-3% points. In sum, the subway not only increased future cash flows but also—in at least equal part—lowered risk premia on real estate. The finding that infrastructure investment lowers risk in real estate markets is novel to the literature, and points to an interesting complementarity.

8 Value Capture

In this section, we take our baseline estimates for the value created by the subway based on the observed transactions and use them to compute the aggregate value creation for the stock of residential real estate on the UES. We then use property tax data to compute how much of this value creation flows back to the city in the form of higher taxes. We

---

[8] We omit the controls distance to Central Park and distance to Grand Central Terminal in these first-stage regressions since these controls are not separately identified from the block-year fixed effects.

find that while there is an overall gain, there is a significant public shortfall compared to the cost of the subway extension.

8.1 Baseline Valuation of the Stock of Real Estate

We start by valuing the stock of real estate in the treatment area in the period before subway construction. We choose 2012 as a base year, the last year of our "Pre" period. This stock consists of owner-occupied residential real estate, renter-occupied residential real estate, and commercial real estate.

8.1.1 Owner-occupied Residential Buildings

Imputing the value of owner-occupied residential real estate occurs in three steps.

Step 1: Transacted Units For each apartment in the baseline treatment area (2nd Ave corridor) for which we observe at least one sale, we use the dynamic difference-in-differences specification with controls to impute an annual valuation for the year 2012. The imputation uses the actual apartment unit and building characteristics alongside the estimated coefficients. Since the regression specification includes a condo indicator variable, valuation differences between coop and condo units are taken into account.

Step 2: Other Units in Coop Buildings with Transactions Even though we observe more than 16 years of transactions in a liquid market, many coop units never transact in our sample. Based on building-level data, we know how many units there are in each coop building and therefore what fraction f of units we are missing. We obtain the valuation of the entire building by multiplying the cumulative value of the units for which we have trades by $1 + f$. The underlying assumption is that the average characteristics of the missing coop units are the same as those of the transacted units.

Step 3: All Other Units Based on a master list of all tax identifiers (Borough-Block-Lot or BBL codes) from the New York City Department of Finance, we obtain a list of all condo units and coop buildings in the Second Avenue corridor and their 2012 "estimated market value" (EMV). After comparing this master list against our transactions data, we obtain the BBLs for which we see no transactions. Each condo unit has its own BBL whereas all units in a coop building share the same BBL. For each condo unit and coop building valued in steps 1 and 2, we calculate an EMV multiple. The EMV multiple is the ratio of our 2012 valuation to the 2012 EMV in the tax roll data. We then average the EMVs separately for condos and coops and for each tax block. There are 83 tax blocks in our Second Avenue treatment area. The 2012 value of a missing condo unit is its 2012 EMV from the city records times the average EMV multiple for condos in that tax block. The value of a missing coop building is the 2012 EMV for that coop building times the EMV multiple for coop buildings in that tax block.

8.1.2 Renter-occupied Buildings

Next, there is a large stock of rental buildings to consider. After all, the home ownership rate on the UES is only 41%. For each unit in our rental building sample, we obtain a 2012 value by combining its own unit and building characteristics and the dynamic difference-in-difference coefficients, estimated from the condo and coop transactions. We set the condo flag equal to 0.5, assuming that rentals are valued at the average of coops and condos. To obtain the total value of the building, we scale up the cumulative value of the transacted units by $1 + f$, where f is the fraction of missing units in the building.

For every rental building thus valued, we compute the EMV multiple as the ratio of our 2012 valuation to the city's 2012 EMV. We average the EMV ratios for rental buildings by tax block. We value the rental buildings (BBLs) for which we have no StreetEasy rental data by multiplying their EMV from the tax roll data by the EMV multiple for rental buildings in that tax block. Our valuation approach is consistent with New York City's Department of Finance approach which values all owner-occupied buildings as if they were rental buildings.

8.1.3 Commercial Properties

The final property type is commercial, non-residential real estate: retail, office, and industrial properties. Since the 2nd Ave corridor is largely a residential neighborhood, the dominant type of commercial real estate is street-level urban retail (shops and restaurants), followed by parking garages. Since we observe very few transactions data for these properties and lack sufficient building characteristics for the transactions we do observe, we exclusively use the EMV approach. Specifically, we calculate the average EMV ratio in each tax block pooling all condo, coop, and rental BBLs. We then value the commercial BBLs as the product of their 2012 EMV by the average EMV multiple in the corresponding tax block.

As shown in the first column of Table 9, we estimate the total 2012 market value of real estate in our treatment area at $70 billion across the three categories of real estate.

8.2 Tax Pass-through

To assess the amount of property taxes that typically passes through to the city government in response to property appreciation, we make use of tax assessment records for New York City. For owners of condos and coops, the city assess property taxes on a portion of the property's market value, the so-called assessed value. This assessed value is calculated using several steps.

First, the property's EMV is calculated as follows. The city imputes the annual Net Operating Income (NOI) per sqft based on comparable rental buildings, typically the average of three buildings that are geographically close to the building in question, of similar size and similar vintage. This annual NOI is then divided by a cap rate, the ratio of NOI to price, to produce the EMV. The city's records indicate that the cap rate was set uniformly at 12.42% in January 2018. The true market cap rate at that time was around 4%, so that the EMV is about three times smaller than the actual market value. Next, the property assessed value is set at 45% of the EMV, and owners pay a 12.9% tax rate on the assessed value, minus exemptions. Absent exemptions, the tax rate is 5.8% of EMV. Changes in property taxes over time are gradually phased in over a five year period. While we do not observe exemptions, we have tax paid in 2015 for all properties. This data suggests a non-trivial role for exemptions, and indicates that actual tax paid is 4.8% of EMV.

To understand how the subway construction affects tax revenue, we start with a simple example for the typical condo building in the 2nd Ave corridor. Suppose a building has 90 units, and a total of 140,000 sqft. Suppose the true market value is $175 million, or $1,250 per sqft. This is the observed average square foot price in the treatment area in Table 1. Given a NOI of $50 per foot, this valuation corresponds to a 4.0% cap rate. The EMV is based on a 12.42% cap rate and so is $37.65 million, or $269 per sqft. The assessed value is 45% of EMV or $16.94 million. This becomes $14 million after the 17.5% condo abatement, a common form of exemption. Annual tax paid is $1.8 million for the building or $20,000 per unit, which is 4.8% of EMV and 1.0% of true market value.

Suppose now that the 2nd Ave subway increases the value of this building by 10.2%, the (exponentiated) point estimate in column 2 of Table 3, or $17.85 million. The EMV increases by $3.84 million, and the assessed value by $1.73m. Taxes paid will increase annually by $183,798 in year 5 and beyond, and gradually be phased in before that. Assuming a government discount rate of 3%, corresponding to NYC's municipal bond yield, the subway results in $5.46 million in extra tax revenue in present value terms. The estimate of value capture, or how much of the price increase accrues to the city government is $5.46m / $17.85m = 30.6%. This pass-through estimate is not far from the nation-wide average long-run elasticity of property tax revenue to house prices, estimated at 0.4 by Lutz (2008).

We adopt this 30.6% pass-through estimate to calculate the additional present value tax revenue increase NYC may expect due to the Second Avenue extension. The first row of Table 9 shows the estimated log change in market value across our main specifications from Table 3, repeated for convenience. The second row exponentiates these numbers to obtain percentage changes. Rows 3-5 apply these percentage gains to the estimated 2012 market value of real estate in our treatment area, per the calculations detailed above. The assumption is that the value gain from subway construction was uniform across property types. We estimate that the subway construction led to a total value increase of $7.1 billion in our benchmark specification. For different specifications in columns 3-5, estimates range between $3.1 and $10.4 billion. However, the city is able to capture only 30.6% of this value in the form of higher taxes. This table displays our estimates of the amount captured by the city government in present value terms from increased property taxation under the row "Property Tax Receipts." The baseline specification predicts a $2.2 billion increase; the other specifications produce estimates ranging from $0.95 to $3.81 billion. We contrast these numbers with the construction cost of $4.5 billion, and show in the last row the shortfall in revenue. The baseline estimate is a $2.3 billion public shortfall. Even though the value generated from subway construction was substantial enough to exceed the (very large) subway construction cost, the gains largely accrued to private owners of condo and co-op units and landlords managing rental and commercial real estate properties. The city suffered a substantial shortfall, especially under the more conservative value gain estimates.

8.3 Value Capture Through Micro-targeted Property Taxes

Our paper demonstrates that it is technically feasible to determine how much each housing unit benefited from the new transit infrastructure, taking into account its exact location, and its unit and building characteristics. In theory, local government could levy a unit-specific property tax surcharge equal to the value created. Such micro-targeted property tax surcharges would not only be based on objectively measurable value increases and property characteristics, and hence be fair, they could also become an important financing tool to fund future infrastructure needs.

Strikingly, nearly all of our estimates of the value gain from the Second Ave Subway construction itself exceed the cost of construction. Our estimates suggest that while the cost of construction of the subway is quite high; so is the value generation, at least in

Table 9: Estimates of Value Creation

Value Add Under:	Value in 2012 (in bn $)	(2) Standard Controls	(3) Building FE	(4) Constr. Period	(5) Constr. Period + Building FE
Treatment Effect:		0.097*** (0.01)	0.043*** (0.00866)	0.138*** (0.01123)	0.060*** (0.01026)
Percentage Change:		10.2	4.4	14.8	6.2
Owner-Occupied Residential ($b)	32	3.24	1.41	4.72	1.97
Renter-Occupied Buildings ($b)	26	2.67	1.16	3.88	1.62
Commercial Non-residential ($b)	12	1.23	.53	1.78	.75
Total ($b):	70	7.1	3.1	10.38	4.3
Property Tax Receipts ($b):		2.17	.9500000000000001	3.18	1.32
Net Gain to Govt ($b):		(2.33)	(3.55)	(1.32)	(3.18)

densely populated areas such as the Upper East Side.

Two caveats are in order. First, it may be politically difficult to levy micro-targeted property taxes. Second, it is an empirical question how large the elasticity of tax revenue is to increases in property taxes. Haughwout, Inman, Craig, and Luce (2004) provide evidence that property prices fall in response to higher property tax rates. They find that New York City was close to the peak of its tax revenue hill in the late 1990s. The extent to which these estimates are still relevant thirty years later is an open question.

9 Conclusion

Mass public transit is a critical infrastructure asset in dense urban environments, but construction costs have risen to enormous amounts. To justify further expansion, transit must demonstrate significant returns either directly or through the capitalization of externalities in real estate prices. Exploiting one of the most expensive extensions in one of the oldest and largest subway systems in the world, the Second Ave subway extension in New York City, we find evidence of such capitalization using a difference-in-difference framework. Our data set allows us to control finely for building and unit characteristics. Our estimates suggest price appreciation of 5–10% across specifications. Much of the value gain occurs in anticipation of the subway opening.

Using new mobile phone location data, we document substantial improvements in commuting lengths, which are concentrated among individuals who live near the new subway stations and take the subway. Q-line subway usage is also higher among new residents, suggesting that the composition of residents was also affected by the subway construction. Such immigrants are likely marginal buyers and renters in the area. The commuting results provide one plausible channel for the price effects.

We also find significant increases in rents that are consistent with the capitalization effects. Increases in price-rent ratios in the treatment area reflect not only higher rents but also lower risk premia on real estate brought about by the infrastructure investment.

Valuing the total stock of treated real estate at $70 billion pre-treatment, our baseline estimate suggests a $7.1 billion gain from the 2nd Ave subway extension to private landlords. We estimate that the city will only recoup about 30% of the gain, or about $2.1 billion, in the form of future property taxes. The former number well exceeds the $4.5 billion cost of the project, while the latter number falls significantly short. This suggests that additional taxation, in the form of targeted property tax increases, might be useful to fill the gap. More broadly, value capture could prove a useful instrument in the financing tool box to help fund the large future infrastructure needs.

References

Anderson, M., 2014, "Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion," *American Economic Review*, 104(9), 2763–796.

Andonov, A., R. Kräussl, and J. Rauh, 2019, "The Subsidy to Infrastructure as an Asset Class," working paper.

Athey, S., B. Ferguson, M. Gentzkow, and T. Schmidt, 2019, "Experienced Segregation," Working paper.

Black, D. A., N. Kolesnikova, and L. J. Taylor, 2014, "Why Do So Few Women Work in New York (And So Many in Minneapolis)? Labor Supply of Married Women across U.S. Cities," *Journal of Urban Economics*, 79, 59–71.

Bom, P. R., and J. E. Ligthart, 2014, "What Have We Learned from Three Decades of Research on the Productivity of Public Capital?," *Journal of economic surveys*, 28(5), 889–916.

Bowes, D. R., and K. R. Ihlanfeldt, 2001, "Identifying the Impacts of Rail Transit Stations on Residential Property Values," *Journal of Urban Economics*, 50(1), 1–25.

Cadot, O., L.-H. Röller, and A. Stephan, 2006, "Contribution to productivity or pork barrel? The two faces of infrastructure investment," *Journal of public Economics*, 90(6-7), 1133–1153.

Campbell, J. Y., and R. J. Shiller, 1988, "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1, 195–227.

Campbell, S., M. Davis, J. Gallin, and R. Martin, 2006, "What Moves Housing Markets: A Trend and Variance Decomposition of the Rent-Price Ratio," Working Paper, Federal Reserve Board of Governors.

Campbell, S. D., M. A. Davis, J. Gallin, and R. F. Martin, 2009, "What Moves Housing Markets: A Variance Decomposition of the Rent-Price Ratio," *Journal of Urban Economics*, 66(2).

Castells, A., and A. Solé-Ollé, 2005, "The regional allocation of infrastructure investment: The role of equity, efficiency and political factors," *European Economic Review*, 49(5), 1165–1205.

Cervero, R., and M. Duncan, 2002, "Transit's Value-Added Effects: Light and Commuter Rail Services and Commercial Land Values," *Transportation research record*, 1805(1), 8–15.

Chen, M. K., K. Haggag, D. G. Pope, and R. Rohla, 2019, "Racial Disparities in Voting Wait Times: Evidence from Smartphone Data," Working Paper 26487, National Bureau of Economic Research.

Chen, M. K., and R. Rohla, 2018, "The effect of partisanship and political advertising on close family ties," *Science*, 360(6392), 1020–1024.

Cochrane, J., 2011, "Discount Rates," *Journal of Finance*, 64(4), 1047–1108.

Couture, V., G. Duranton, and M. A. Turner, 2018, "Speed," *Review of Economics and Statistics*, 100(4), 725–739.

Dewees, D. N., 1976, "The Effect of a Subway on Residential Property Values in Toronto," *Journal of Urban Economics*, 3(4), 357–369.

Diao, M., D. Leonard, and T. F. Sing, 2017, "Spatial-difference-in-differences Models for Impact of New Mass Rapid Transit Line on Private Housing Values," *Regional Science and Urban Economics*, 67, 64–77.

Fesselmeyer, E., and H. Liu, 2018, "How much do users value a network expansion? Evidence from the public transit system in Singapore," *Regional Science and Urban Economics*, 71, 46–61.

Finkenzeller, K., T. Dechant, and W. Schäfers, 2010, "Infrastructure: a new dimension of real estate? An asset allocation analysis," *Journal of Property Investment & Finance*.

Haughwout, A. F., R. Inman, S. Craig, and T. Luce, 2004, "Local Revenue Hills: Evidence and Lessons from Four U.S. Cities," *Review of Economics and Statistics*, 86, 570–585.

Hess, D. B., and T. M. Almeida, 2007, "Impact of Proximity to Light Rail Rapid Transit on Station-area Property Values in Buffalo, New York," *Urban Studies*, 44, 1041–1068.

Jackson, C. K., and E. G. Owens, 2011, "One for the Road: Public Transportation, Alcohol Consumption, and Intoxicated Driving," *Journal of Public Economics*, 95(1), 106–121.

Kahn, M., 2007, "Gentrification Trends in New Transit Oriented Communities: Evidence from Fourteen Cities," *Real Estate Economics*, 35(2), 155–182.

Kahn, M. E., and N. Baum-Snow, 2000, "The Effects of New Public Projects to Expand Urban Rail Transit," *Journal of Public Economics*, 77(2), 241–63.

——— , 2005, "Effects of Urban Rail Transit Expansions: Evidence from Sixteen Cities, 1970-2000," *Brookings-Wharton papers on urban affairs*, pp. 147–206.

Koijen, R., and S. van Nieuwerburgh, 2011, "Predictability of Stock Returns and Cash Flows," *Annual Review of Financial Economics*, 3, 467–491.

Lin, J.-J., and C.-H. Hwang, 2004, "Analysis of Property Prices before and after the Opening of the Taipei Subway System," *Annals of Regional Science*, 38(4), 687–704.

Lutz, B. F., 2008, "The Connection Between House Price Appreciation and Property Tax Revenues," *National Tax Journal*, 61(3), 555–572.

McDonald, J. F., and C. I. Osuji, 1995, "The Effect of Anticipated Transportation Improvement on Residential Land Values," *Regional Science and Urban Economics*, 25(3), 261–278.

McKinsey, 2017, "Bridging Infrastructure Gaps: Has the World Made Progress?," working paper, McKinsey Global Institute.

McMillen, D. P., and J. McDonald, 2004, "Reaction of House Prices to a New Rapid Transit Line: Chicago's Midway Line, 1983–1999," *Real Estate Economics*, 32(3), 463–486.

Pan, H., and M. Zhang, 2008, "Rail Transit Impacts on Land Use: Evidence from Shanghai, China," *Transportation Research Record*, 2048(2048), 16–25.

Plazzi, A., W. Torous, and R. Valkanov, 2010, "Expected Returns and the Expected Growth in Rents of Commercial Real Estate," *Review of Financial Studies*, 23, 3469–3519.

Severen, C., 2018, "Commuting, Labor, and Housing Market Effects of Mass Transportation: Welfare and Identification," *Working Paper (Federal Reserve Bank of Philadelphia)*.

Van Nieuwerburgh, S., 2019, "Why Are REITs Currently So Expensive?," *Real Estate Economics*, 47(1), 18–65.

Walter, I., 2016, "The Infrastructure Finance Challenge," working paper, New York University stern School of Business.

Zheng, S., and M. Kahn, 2013, "Does Government Investment in Local Public Goods Spur Gentrification? Evidence from Beijing," *Real Estate Economics*, 41(1), 1–28.

Zhou, Z., H. Chen, L. Han, and A. Zhang, 2020, "The Effects of a Subway on House Price: Evidence from Shanghai," *Real Estate Economics*, forthcoming.

A Appendix

Table 10: Within Distance Broken Down

VARIABLES	(1) Log Price	(2) Log Price	(3) Log Price	(4) Log Price
Post x Within 0 - .1 mi	0.0333*	0.0231	0.0293	0.0260
	(0.0179)	(0.0158)	(0.0210)	(0.0187)
Post x Within .1 -.2 mi	0.0655***	0.0317**	0.0828***	0.0508***
	(0.0141)	(0.0127)	(0.0167)	(0.0150)
Post x Within .2 -.3 mi	0.0871***	0.0507***	0.140***	0.0715***
	(0.0117)	(0.0105)	(0.0137)	(0.0125)
Constr. Period x Within 0 - .1 mi			-0.00767	0.00644
			(0.0215)	(0.0191)
Constr. Period x Within .1 -.2 mi			0.0334*	0.0358**
			(0.0170)	(0.0154)
Constr. Period x Within .2 -.3 mi			0.101***	0.0346***
			(0.0139)	(0.0126)
Post	0.129***	0.110***	0.186***	0.157***
	(0.00641)	(0.00578)	(0.00751)	(0.00683)
Constr. Period			0.109***	0.0875***
			(0.00762)	(0.00683)
Within 0 - .1 mi	-0.148***		-0.144***	
	(0.0111)		(0.0156)	
Within .1 -.2 mi	-0.126***		-0.147***	
	(0.00883)		(0.0127)	
Within .2 -.3 mi	-0.140***		-0.195***	
	(0.00723)		(0.0102)	
Observations	49,673	49,673	49,673	49,673
R-squared	0.639	0.739	0.644	0.741
Controls	YES	YES	YES	YES
Building FE	NO	YES	NO	YES

Table 11: Change in Distance Broken Down

VARIABLES	(1) Log Price	(2) Log Price	(3) Log Price	(4) Log Price
Post x Chg. dist 0-0.10mi	0.174***	0.0586***	0.261***	0.103***
	(0.0140)	(0.0129)	(0.0161)	(0.0151)
Post x Chg. dist > 0.10mi	0.0496***	0.0190**	0.0650***	0.0307***
	(0.0103)	(0.00920)	(0.0122)	(0.0109)
Constr. Period x Chg. dist 0-0.10mi			0.189***	0.0849***
			(0.0165)	(0.0151)
Constr. Period x Chg. dist > 0.10mi			0.0276**	0.0205*
			(0.0124)	(0.0111)
Post	0.115***	0.113***	0.165***	0.157***
	(0.00682)	(0.00614)	(0.00801)	(0.00726)
Constr. Period			0.0949***	0.0807***
			(0.00816)	(0.00729)
Chg. dist 0-0.10mi	-0.223***		-0.317***	
	(0.00943)		(0.0126)	
Chg. dist > 0.10mi	-0.101***		-0.118***	
	(0.00860)		(0.0108)	
Observations	49,673	49,673	49,673	49,673
R-squared	0.639	0.739	0.644	0.741
Controls	YES	YES	YES	YES
Building FE	NO	YES	NO	YES

Table 12: Construction Activity

VARIABLES	(1) Log Estim. Cost	(2) Log Num. Permits	(3) Log Num. New Constr. Jobs	(4) Log Num. New Res. Build.
Post x On Second Avenue	0.0648	0.0369	0.0390	0.312
	(0.0598)	(0.185)	(0.209)	(0.446)
Post	0.288***	0.107	-0.0497	-0.0762
	(0.0254)	(0.131)	(0.148)	(0.311)
On Second Avenue	-0.447***	-0.964***	-0.826***	-0.314
	(0.0389)	(0.119)	(0.134)	(0.284)
Observations	19,175	34	34	32
R-squared	0.019	0.786	0.674	0.047

Notes: "Log Estim. Cost" is the natural log of the estimated initial cost of a particular permitted construction job. "Log Num. Permits" is the natural log of the number of building permits issued in a treatment-year. "Log Num. New Constr. Jobs" is the natural log of the number of new construction jobs with permits issued for it in a treatment-year (each job may receive multiple permits). "Log Num. New Res. Build." is the natural log of the number of new residential buildings that receive a permit in a treatment-year. Standard errors in parentheses. In the last three columns there is one annual observation for the treatment group and one for the control group. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 13: Heterogenous Treatment Effect by Number of Bedrooms

VARIABLES	(1) Log Price	(2) Log Price	(3) Log Price	(4) Log Price
Post x Treat	-0.0153	-0.0284	0.00558	-0.0137
	(0.0234)	(0.0234)	(0.0214)	(0.0272)
Post x Treat x 1BR	0.0737***	0.0703***	0.0421*	0.0571*
	(0.0249)	(0.0247)	(0.0224)	(0.0293)
Post x Treat x 2BR	0.158***	0.119***	0.110***	0.131***
	(0.0264)	(0.0259)	(0.0237)	(0.0314)
Post x Treat x 3BR	0.223***	0.180***	0.155***	0.215***
	(0.0331)	(0.0310)	(0.0290)	(0.0400)
Post x Treat x 4BR+	0.386***	0.348***	0.310***	0.408***
	(0.0461)	(0.0434)	(0.0393)	(0.0540)
Treat x 1BR	-0.00984	-0.0629***	-0.131***	-0.0480**
	(0.0206)	(0.0207)	(0.0211)	(0.0220)
Treat x 2BR	-0.0971***	-0.123***	-0.253***	-0.149***
	(0.0215)	(0.0213)	(0.0217)	(0.0232)
Treat x 3BR	-0.227***	-0.232***	-0.336***	-0.300***
	(0.0257)	(0.0245)	(0.0246)	(0.0284)
Treat x 4BR+	-0.238***	-0.272***	-0.328***	-0.350***
	(0.0345)	(0.0323)	(0.0317)	(0.0373)
Post	0.124***	0.129***	0.117***	0.138***
	(0.00609)	(0.00641)	(0.00681)	(0.00562)
Treat	-0.140***	-0.0325*	0.0320	-0.0597***
	(0.0192)	(0.0193)	(0.0200)	(0.0204)
1BR	0.390***	0.423***	0.483***	0.415***
	(0.0131)	(0.0133)	(0.0157)	(0.0116)
2BR	0.823***	0.861***	0.955***	0.852***
	(0.0137)	(0.0140)	(0.0162)	(0.0122)
3BR	1.154***	1.198***	1.284***	1.178***
	(0.0157)	(0.0161)	(0.0181)	(0.0143)
4BR+	1.209***	1.255***	1.325***	1.250***
	(0.0190)	(0.0195)	(0.0214)	(0.0178)
Observations	49,673	49,673	49,673	49,673
R-squared	0.645	0.640	0.640	0.642
Controls	YES	YES	YES	YES
Building FE	NO	NO	NO	NO
Treatment Def.	1	2	3	4