Storm Surges, Informational Shocks, and the Price of Urban Real Estate: An Application to the Case of Hurricane Sandy*

Jason Barr
Department of Economics
Rutgers University-Newark
Email: jmbarr@rutgers.edu

Jeffrey P. Cohen
Center for Real Estate and Urban Economic Studies
School of Business
University of Connecticut
Email: jeffrey.cohen@uconn.edu

Eon Kim
Department of Security and Crime Science
University College of London
Email: eongyung.kim.15@ucl.ac.uk

September 2018

Abstract: A major hurricane's impacts on real estate can be devastating. Recent U.S. hurricanes demonstrate how flooding damage can unexpectedly extend beyond FEMA flood zones. Such flood risk surprises/shocks can provide property owners—including those not flooded—with new future flood risks information. Our innovative approach uses locally weighted regressions to quantify and map separate shock effects for each dry property's price, and demonstrate geographic heterogeneity. We provide an application to non-flooded properties in New York City for Hurricane Sandy. Houses, apartments and commercial properties show the most price volatility within the older, denser urban core, mostly in gentrifying neighborhoods.

Key words: Hurricane Sandy, Locally Weighted Regressions, Flood Risk, New York City, Real Estate Prices, geographic heterogeneity

JEL Classification: R3

*We thank the National Resources Defense Council for providing GIS maps related to Hurricane Sandy. Earlier versions of the paper were presented at the 2017 AREUEA National Conference and the 2017 Eastern Economics Association meetings. We thank Steve Malpezzi, Suleyman Taspinar and session participants for their comments. Any errors belong to the authors.

1. **Introduction**

Major U.S. hurricanes in 2017, including Houston (Harvey) and Florida (Irma), and 2012 in New York City (NYC) (Sandy) are examples of how flooding damage can unexpectedly extend beyond the Federal Emergency Management Agency (FEMA) designated flood zones.[1] Surprises/shocks to flood risks can provide property owners—including those not flooded—with new information about future flood risks, based on how close these dry properties are to the actual flooded areas. We quantify how flood risk shocks impact non-flooded property values. One innovation is that we use information on repeat property sales that sold once before and again after Sandy to estimate a separate shock effect for each dry property. Using locally weighted regressions (LWRs), we investigate heterogeneous effects across NYC, and present results in several maps. We addresses demand and supply effects from the storm and possible sample selection. We find properties showed the most volatility within the older, denser urban core, mostly in those neighborhoods that appear to be gentrifying.

Harvey struck the Houston, Texas area in late-August 2017. Preliminary damage assessments are in the range of $150 billion (McWilliams and Marianna, 2017), with thousands of houses destroyed and many other residential and commercial sustaining major damage. In September 2017, Hurricane Irma hit Florida, with waist-deep flooding in downtown Miami (Sun-Sentinel, 2017), among other areas. Total costs of Irma could reach as high as $300 billion (Wood, 2017). On October 29, 2012, Hurricane Sandy made landfall in NYC; it was arguably the largest and most damaging storm to hit the region. The surge level in lower Manhattan was 13.88 feet, surpassing the old record set in 1960 (CNN, 2013). Loss estimates for NYC were $19 billion, and $33 billion for the entire state.[2]

Studies have focused on estimating the damage costs (ESA, 2013). However no known work has explored the implicit costs of storm surges on the value of NYC real estate for properties *not damaged by the surge*. Understanding how flooding affected dry properties is important because it gives clues to the impacts of higher future storm surge risks. Which neighborhoods react the most and why? We develop new methodologies to investigate real estate price volatility from expectations about future surges, by focusing on real estate price changes for non-flooded properties.

In 1968, Congress created the National Flood Insurance Program (NFIP) to help property owners protect themselves. The NFIP offers flood insurance to homeowners, renters, and business owners if their respective town or city participates in the NFIP. Participating communities agree to adopt and enforce ordinances that meet or exceed FEMA requirements to reduce flood risk (FEMA, 2017b).

---

[1] For Harvey see: https://www.nytimes.com/interactive/2017/09/01/us/houston-damaged-buildings-in-fema-flood-zones.html?mcubz=1&_r=1.
[2] For NYC see: http://www.nyc.gov/html/sirr/downloads/pdf/final_report/Ch_1_SandyImpacts_FINAL_singles.pdf.

FEMA partners with states and communities through the Risk Mapping, Assessment, and Planning (Risk MAP) program to identify flood hazards and assess flood risks. These data are incorporated into Flood Insurance Rate Maps (FIRMs), which support the NFIP and provide the basis for community floodplain management regulations and flood insurance requirements. Most commonly used for insurance purposes are the 100-year floodplain maps, which are regions designated to have a 1% chance of being inundated each year.

Buyers who seek a mortgage are often required to purchase flood insurance if they are within a FEMA-designated floodplain (FEMA, 2017a). These floodplain maps are a publicly-available assessment of the likelihood of a property being flooded. For those outside the floodplain, the distance to the plain can presumably be used to infer the neighborhood's relative flood safety. At 20 feet from a floodplain, a property is potentially at more risk than one 2,000 feet away.

Our goal is to estimate how properties that remained dry may be impacted by storms. If the Hurricane represents an informational shock about the likelihood of future damage then this may be priced into properties, as people reassess the likelihood of future storm shocks and associated potential damage.

Because urban real estate is part of an *urban system*, the effects of one neighborhood's property value changes depend on surrounding neighborhoods. Measuring average effects do not control for or measure the locational interdependencies. For instance, with Sandy, the farthest property away from the surge was only three miles; most parts of the city are relatively low lying. So, vast swaths of the city could be susceptible to future surges.

This paper employs several innovative statistical techniques to assess the expectations of the real estate market. In addition to repeat sales, we estimate changes in the average prices over time for each borough, to control for market-wide fluctuations independent of the storm surge.

Next, we use LWRs, as in Cleveland and Devlin (1988), which has longstanding roots in the geography literature. Because ordinarily least squares (OLS) regression estimates cannot capture heterogeneity in the price response across neighborhoods, we determine heterogeneous coefficient estimates for each property with repeat sales approach using LWR. If price volatility were uniform city-wide, we would not reject the null hypothesis of uniform coefficients across boroughs. In our specific application of NYC, our tests on the sample of repeat sales properties reject this null hypothesis and suggest that while the price of many dry properties across the city were not significantly affected by the storm, others did see price effects due to the storm. LWR enables this exploration of heterogeneous effects for each property.

This paper demonstrates that these effects often can be heterogeneous across geographic locations. Two similar buildings in their age, use, and quality, and the same distance from the storm surge, can be impacted by the storm differently, based on the economic and demographic profile of the neighborhoods. OLS coefficient estimates represent averages for an entire city, and by focusing on averages, one may miss important local variations. When investigating the geographic impacts of storms on cities, it is vital to understand spatial variation; this is crucial not only for measurement reasons but for the policy implications about where to deploy resources before/after storms.

Using LWRs for three different property classes,[3] we find significant variation in price change responses to the storm surge. In Staten Island and the Bronx, we find relatively less reaction to the storm, while in Queens and Brooklyn we find strong price reactions to the surge (relative to the FEMA line). The price effects for Manhattan are weak for one and two family homes (of which there are relatively few) but stronger for apartments buildings and commercial properties.

Later we regress the LWR coefficients on several control variables. Across property types the responses to closer proximity to the surge were much greater closer to the Empire State Building, all else equal. We also find evidence that census tracts with better subway access and higher incomes (for homes and offices) were more responsive to these "shocks." We infer this likely to be due to gentrification and the rising value the residents are placing on proximity to the city center; it is likely that wealthier people moving to the center might be more responsive to these shocks.

The next section provides a literature review on the effects of storms and storm surges on real estate prices. Section 3 provides the background on our methodology that informs the statistical analysis. Section 4 provides information on the data set for the Hurricane Sandy application. Sections 5 and 6 provides evidence on the effects of the storm surge on both flooded non-flooded properties throughout the city, using hedonic regressions and LWRs, respectively. In Section 7, we test for reasons why there were heterogeneous effects across the city, and we also test for spatial stationarity. Finally, Section 8 offers some concluding remarks. The appendices provide information about data sources and LWRs.

2. Literature Review

Hurricanes Harvey in Texas and Irma in Florida demonstrate the FEMA flood zones remain an imperfect measure of flooding likelihoods (Fessenden, et al., 2017). Some studies investigate the impacts of storms or natural disasters on real estate. These include studies of specific hurricanes[4], and the proximity to the coast or flood risks. For instance, Faber (2015) examines how flood risk from Hurricane Sandy impacted various demographic groups differently. He finds that both demographics and economic factors are important considerations for flood risks.[5]

Heitz et al (2009) examine one catchment area in France, and conduct landowner surveys to understand flood risk perceptions. These perceptions vary, depending on whether respondents are inside an "erosion" area or "sedimentation" area. They also find approximately half of landowners trust government flood-risk information.

---

[3] See Appendix A for information about the data set.

[4] Although not directly focused on flood risk, another recent study examines NYC housing prices and Hurricane Sandy. Specifically, Ortega and Taspinar (2016) examine Sandy and the NYC housing market, and address whether housing demand shifted towards less exposed areas.

[5] We consider demographic factors in our analysis presented in Table 8.

Bin and Landry (2013) find unexpected flood risk effects following a major storm disappear after several years. They find a 5%-9% discount following Hurricanes Fran and Floyd in North Carolina. More recent data indicate higher short-term discount rates.

Bin et al. (2011) focus on a similar geographic area in North Carolina to estimate value of lost property from potential flooding. For a 20-to-70 year future period, they forecast between $179-$576 million loss for properties in four North Carolina shore counties, implying hurricanes pose tremendous risk.

An earlier flood risk study is MacDonald et. al (1987), who estimate hedonic house price functions for a flood-prone area (Monroe, Louisiana). They provide examples for a small sample of homes, and find that for these houses a higher flood risk leads to a $2000-$8000 decrease in sales prices.

In another coastal study of flood risks, Atreya and Czajkowski (2016) use a spatial regression model to study price effects of proximity to the coast in Galveston, Texas. They find within ¼ mile from the coast, properties sell for more than farther away properties.

Hammond et al (2015) consider urban flooding impacts by providing a literature review. They argue that understanding flood impacts is crucial for building cities that are flood "resilient" (which is important for mitigating flood risks).

With these studies in mind, we now turn to our approach for measuring changes in dry property flood risk impacts.

3. The Theory of Price Effects

Our aim is to understand how a storm shock can affect those properties that were not damaged by the storm. For many property owners the storm represented new information on the potential damage due to storm surges. *A priori*, however, the effects of a storm on the dry side of the storm surge can be unclear. On the one hand, the surge was a negative shock to the city, and, as a result, it represents the possibility that the city is subject to future surges. Those property owners relatively close to the surge are particularly vulnerable. This can be called the demand effect, where a negative neighborhood shock would reduce the demand for real estate and thus reduce its price.

On the other hand, there are likely to be supply effects. First, properties in the flooded zone were damaged or destroyed, which would reduce the quantity/quality of available properties in a particular neighborhood; this would have the effect of making the remaining structures relatively more valuable. There are also likely to be second order effects of shifting demand from the wet part to dry part. If properties in "gentrifying" neighborhoods experience a loss of real estate due to the surge, but the neighborhood remains in strong demand overall, then the relative demand within a neighborhood might positively shift to the dry properties. Thus, the dry properties may receive two benefits that could raise their prices—the relative scarcity of structures in a neighborhood near the flood will increase their price, and shifting relative demand within the neighborhood could also drive up the prices of dry properties.

Our variable of interest is the difference between the closest distance of a property to the FEMA floodplain relative to the closest distance to the storm surge. That is, we estimate:

$$\Delta lnP_i = \theta shock_i + X_i\zeta + \varepsilon_i$$

where

$$shock_i = surge\ distance_i\ - FEMA\ floodplain\ distance_i$$

for $i = 1, \ldots, N$, non-flooded properties, and where X_i are control variables and ε_i is the error term. ΔlnP_i is the price change given that *the first sale is before the storm and the second sale is after the storm*, i.e., where repeat sales straddle the storm.

Consider two identical houses, A and B, where each is 100 feet from the closest FEMA floodplain boundary. For house A, perhaps the flood approached within 150 feet, for a $shock = 150 - 100 = 50$ (so that any value above 0 is "good news" or a positive shock). For house B, suppose the flood came to within 50 feet of the house, for a $shock = 50 - 100 = -50$; thus a negative shock. In this case, we expect house B to lose value, relative to house A. This would suggest $\theta > 0$, where θ is the effect of a one-foot shock on the housing price change.

To estimate θ, we employ several different empirical strategies. First, we use hedonic regressions to estimate the impacts on the dry and wet sides. But, because of the data set size, we can use repeat sales that allow us to "net out" the static unobservable characteristics that might otherwise be omitted. We use a technique similar to Ries and Somerville (2010), where we develop a price index to estimate price-effects that are independent of the storm. We also include a measure of 2010 census tract building occupancy rate changes to control for supply effects (both because of, and independent of, the storm). We are able to measure changes in the occupancies rates before and after the storm using vacancy data provided by the Department of Housing and Urban Development (HUD), and thus our estimation is able to measure the demand effect—how much of prices changes were due to informational shocks, rather than changes in the real estate stock.

A key issue is not just whether $\theta > 0$, but whether the magnitude of θ varies across the city. Are their spatial differences in how responsive some properties are to the size (and sign) of the shock? If so, this would suggest OLS is not fully capturing spatial variation in the shocks. For this reason we employ LWRs (described below), which gives an estimate, $\hat{\theta}_i$, for each property. Given the complexity of many cities, their heterogeneous demographics, and diverse real estate stock, there is no reason to believe, *a priori*, that the effects of a storm on the dry side is homogenous; rather prices in different neighborhoods are likely to respond differently based on the perceived risk and the characteristics of property owners.

3.1 Methodology

Each period, properties owners have information about the likelihood of flooding affecting their property, *i*. Denote,

$$F_{it} = flooding\ likelihood\ estimate_{it},$$

for each period *t*. If *t* is before October 29, 2012, then

$$F_{it} = distance\ to\ FEMA\ floodplain\ map_{it}$$

If *t* is after the storm (denote this then *t+j*),

$$F_{it+j} = distance\ to\ Sandy\ inundation\ boundary_{it+j}$$

To estimate the informational shock from the storm, we examine properties that sold once before Sandy (in time *t*), and once again after Sandy (in time *t+j*). In order to estimate this model, we first begin with the linear hedonic model:

$$lnP_{nit} = \theta F_{nit} + X\zeta + \varepsilon_{nit}, \qquad (1)$$

where

$$\varepsilon_{nit} = \alpha_{nt} + \nu_{nit}, \qquad \nu_{nit} \sim iid(0, \sigma^2)$$

α_{nt} is a price level (index) in NYC borough *n* at time *t*. F_{nit} is the flooding likelihood measure at time *t*, with parameter θ. *X* (with parameter vector, ζ) is a matrix of observations for physical characteristics of the property (which are assumed to be time-invariant).

Now consider Equation (1), but for period *t+j*, where *t+j* is after the occurrence of Hurricane Sandy:

$$lnP_{nit+j} = \theta F_{nit+j} + X\zeta + \varepsilon_{nit+j}. \qquad (2)$$

Subtracting (1) from (2) yields:

$$ln\left(\frac{P_{nit+j}}{P_{nit}}\right) = \theta\left(F_{nit+j} - F_{nit}\right) + \varepsilon_{nit+j} - \varepsilon_{nit}. \qquad (3)$$

Using the fact that $\varepsilon_{nit} = \alpha_{nt} + \nu_{nit}$, we can rewrite (3) as:

$$ln\left(\frac{P_{nit+j}}{P_{nit}}\right) - \left(\alpha_{nt+j} - \alpha_{nt}\right) = \theta\left(F_{nit+j} - F_{nit}\right) + \nu_{nit+j} - \nu_{nit}. \qquad (3')$$

Our objective is to estimate (3') in order to obtain θ, the effect flood shocks on sale prices. There are a number of possible estimation approaches to generating this estimate of θ, including OLS and LWRs, assuming it is possible to obtain estimates of $\left(\alpha_{nt+j} - \alpha_{nt}\right)$ in a manner that avoids potential endogeneity due to correlation between neighborhood-level shocks and *F* (discussed more below).

Furthermore, the OLS model assumes θ is constant. But, given the heterogeneity of many major cities (such as NYC), both demographically and geographically, there is no reason to believe that θ is constant across the city (and this is confirmed in our Hurricane Sandy results below).

An extension of this model assumes a non-constant relationship between the explanatory variables in (1) and the dependent variable. Specifically, consider the following equation:

$$lnP_{nit} = \theta_i F_{nit} + X\zeta + \varepsilon_{nit}, \qquad (4)$$

When each property has its own coefficient, θ_i, the model is similar to equation (3'):

$$ln\left(\frac{P_{nit+j}}{P_{nit}}\right) - \left(\alpha_{nt+j} - \alpha_{nt}\right) = \theta_i\left(F_{nit+j} - F_{nit}\right) + \nu_{nit+j} - \nu_{nit} \qquad (4')$$

To determine the heterogeneous marginal effects of shocks on prices, we estimate a nonparametric variation of (4') using LWRs. This leads to the following weighted least squares regression equation:

$$w_{ik} \times \left[ln\left(\frac{P_{nit+j}}{P_{nit}}\right) - \left(\alpha_{nt+j} - \alpha_{nt}\right)\right] = w_{ik}\theta_i\left(F_{nit+j} - F_{nit}\right) + w_{ik}\left(\nu_{nit+j} - \nu_{nit}\right) \quad (5)$$

where $w_{ik} = e^{-(d_{ik}/b)^2}$. This method generates a separate parameter estimate, $\hat{\theta}_i$, for each repeat sales observation *i*, *b* is the bandwidth parameter, and d_{ik} is the distance between two properties.

One can explore, in the context of our specific application, various issues such as: does the marginal effect of informational shock vary across the city? Is it higher in say Manhattan than in other boroughs? Does it change for different types of buildings? But to estimate (3') (and (5)), we need an estimate of $\left(\alpha_{nt+j} - \alpha_{nt}\right)$. Note, again, that $\left(\alpha_{nt+j} - \alpha_{nt}\right)$ is the change in repeat sales price index for borough *n* between two periods, *t* and *t+j*.

3.1.1 Estimating Price Indexes

To estimate $\left(\alpha_{nt+j} - \alpha_{nt}\right)$, there are several possible approaches. One is a repeat sales estimator, as in Baily, Muth, and Nourse (1973). A more recent version of the repeat sales estimator is presented in Anenberg and Laufer (forthcoming), or the Fourier price index approach of McMillen (2003) and McMillen and Dombrow (2001). The former estimates the Fourier price index nonparametrically, and the latter uses a parametric Fourier price index approach. A third is a more recently developed approach of a "matching" estimator, as in Deng, et al. (2012). For ease of implementation, we focus on a parametric Fourier price index approach to obtaining the estimates for $\left(\alpha_{nt+j} - \alpha_{nt}\right)$.

One potential concern is that neighborhood-level shocks in estimating α could be correlated with $F_{ni,t}$, thus biasing θ. Here one should use the sample of repeat sales for which *both property sales occur either before or after the flood date*. Then, a different sample—the sample of repeat sales with dates that straddle the flood date—should be used in estimating Equation (5).

Specifically, we construct a parametric Fourier repeat sales price index using properties with repeat sales that were either both before or both after the flood date. Then we use these repeat sales price indexes for each borough of the city to adjust the property-level price ratio. In Appendix A.2 we explain this in more detail.

After estimating the Fourier price indexes and obtaining the fitted values of $\left(\alpha_{nt+j} - \alpha_{nt}\right)$ with the estimates in (7), one substitutes these estimates for $\left(\alpha_{nt+j} - \alpha_{nt}\right)$ into (5) and then regress the independent effects of the changes in sale prices on shock value as follows:

$$ln\left(\frac{P_{nit+j}}{P_{nit}}\right) - \left(\hat{\alpha}_{nt+j} - \hat{\alpha}_{nt}\right) = \theta_i\left(F_{nit+j} - F_{nit}\right) + \nu_{nit+j} - \nu_{nit} \qquad (4')$$

3.1.2 Selection Bias

Since we use repeat sales, it might be the case that there is something unrepresentative about the structures for which there are repeat sales. For this reason, we first estimate an inverse mills ratio (IMR) value (Heckman, 1976, 1979) to include in the repeat sales LWRs. Specifically, using all sales, we first estimate the following probit model, where $I=0$ if a sale is not a repeat, and $I=1$ if a sale is one of a set of repeat sales (i.e. the second or greater sale, if observed)):

$$Prob(I = 1) = \Phi(\gamma_0 + X\beta + \upsilon)$$

where $\Phi(\bullet)$ is the cumulative normal density function; X is the matrix of observations for the characteristics typically used in the hedonic regression; and γ_0 is a constant. The IMR is given by:

$$IMR = \frac{\phi(\gamma_0 + X\beta)}{\Phi(\gamma_0 + X\beta)},$$

where $\phi(\bullet)$ is the standard normal density.

Next, one would return to our sample of repeat sales pairs. If IMR_t is the IMR evaluated at the X value for sale #1, and $IMR_{t\,t+j}$ is evaluated at the X value for sale #2, one would include the difference $(IMR_{t+ji} - IMR_{ti})$, for each observation i, as an explanatory variable in the straddle LWR regression as follows:

$$ln\left(\frac{P_{nit+j}}{P_{nit}}\right) - \left(\hat{\alpha}_{nt+j} - \hat{\alpha}_{nt}\right) = \theta_i\left(F_{nit+j} - F_{nit}\right) + \omega\left(IMR_{t+j,i} - IMR_{ti}\right) + \left(\nu_{nit+j} - \nu_{nit}\right)$$

The probits include a wide variety of property and neighborhood characteristics that are likely to influence the probability of repeat sales (see Appendix A for regression results). Note that the χ^2-statistic for the regressions in our Sandy application (for homes, apartments and commercial properties) all have p-values less than 0.01.[6]

3.1.3. Controlling for Supply

Lastly, one other issue of concern is what we call the supply effect. If dry structures are close to flooded structures, then the reduction in supply may cause our estimates of θ_i to be biased if there is a correlation between the change in F and changes in housing stock. For this reason, we

[6] The OLS regressions for the equations for the repeat sales, as well as the LWR coefficients, show that, on average, $\omega \neq 0$, suggesting we should be concerned about selection bias (see Table 7 and Appendix A).

include the change in occupancy rates of homes in the census tracts of each of the properties (HUD, 2016).

Using LWRs we estimate the following model:

$$ln\left(\frac{P_{nit+j}}{P_{nit}}\right) - \left(\hat{\alpha}_{nt+j} - \hat{\alpha}_{nt}\right) = \theta_i\left(F_{nit+j} - F_{nit}\right) + \omega\left(IMR_{t+j,i} - IMR_{ti}\right) + \beta\left(O_{t+j,n} - O_{tn}\right) + \left(\nu_{nit+j} - \nu_{nit}\right), \qquad (6)$$

where $\left(O_{t+j,n} - O_{tn}\right)$ is the change in occupancy rates in a neighborhood before and after the flood.

4. Hurricane Sandy Application: The Data

Here we provide some basic information about the data; Appendix A gives more details. We began by collecting data on nearly all bona fide open market sales of buildings in NYC between January 2003 and October 2014 (the data set omits sales of condo or coop units). Hurricane Sandy occurred on October 29, 2012, and thus we have about two years of data after the storm to assess the short and medium run effects.

In this application of our technique to Sandy, we investigate three types of properties: one and two family homes, apartment buildings, and commercial properties, in order to compare and contrast the effects of these property classes on real estate prices. This data set comes from the NYC Department of Finance and provides information on the type of property, lot and building square footage, its age, and address. The sales data were then merged with the NYC's Primary Land Use Tax Lot Output (PLUTO) files, which contain additional information about the structures, such as the number of floors, the census tract, and latitude and longitude coordinates. To estimate location-based effects, we also calculated the distance in miles (as the crow flies) of each property to the Empire State Building, which is our measure of the city center (as in Barr and Cohen, 2015).

Next, we utilized GIS shape files related to the storm surge of Hurricane Sandy (see Figures 2-4). These files have been generously provided by the Natural Resources Defense Council (NRDC). The map indicates the location of the storm surge and the location of the FEMA floodplain. Thus the maps show four areas: the area of FEMA floodplain that remained dry, the area in the FEMA floodplain that was hit by the storm surge, the area of the surge that was outside of the FEMA floodplain, and the area that was neither in the floodplain nor the storm surge. Thus, we categorize each property based on it being in one of those four areas.

For each property, we also calculated the closest distance to the floodplain boundary, the closest distance to the shore, and closest distance to the surge boundary. We also have been able to obtain other data sets that are helpful in estimating the effect of the storm on property values, including the elevation of each property relative to mean sea level and the depth of the storm surge beneath each flooded property (see Appendix A for more information).

Finally, we also merged the HUD's quarterly vacancy data from the 4th quarter of 2005 to the last quarter of 2014 (HUD, 2016). This data set provides information on the occupancy rate of structures, yielding estimates for the number of structures in each census tract. Thus the occupancy rate of structures (of all kinds) is our measure of the supply of building space.

A table of the descriptive statistics is in Appendix A. Our data set includes an initial sample of over 371,000 sales. Of those, 13% are from after the Hurricane. About 5,123 properties in our sample experienced flooding from the storm. For these houses, we estimate the mean flood height was 3.24 feet, and with a maximum surge of 13.3 feet. On average, throughout the city, the flood extended about 0.035 miles inland; its maximum extent was 0.9 miles inland. Across the city, the average elevation is 16.3 feet, and the average distance to the closest shoreline is 1.25 miles.

The average sales price for all properties in the data set, unadjusted for inflation, is $240 per square foot, and adjusted to October 2014 prices, it is $273 per square foot (where real prices are used based on the NYC CPI, excluding shelter). The average lot size is 3840 square feet and the average building area is about 6704 square feet. 74% of the sales in the sample are for one or two family houses, 8% are apartment buildings. 4% of the sales are for commercial properties.

5. Assessing the Damage: A Hedonics Approach

Figure 1 shows two indexes of real estate prices throughout NYC from 2003 and 2014—those properties that remain dry and those that were to be or were flooded by Hurricane Sandy on October 29, 2012. The indexes come from two hedonic regressions of the log of the real price of building space per square foot (sales prices divided by the NYC CPI excluding shelter costs) and a series of building and locational controls (further discussed in this section and in Appendix A). The results show, as expected, that the two series moved in tandem until the storm; at that point we observe a sharp reduction the prices in the flooded properties. Subsequently, the flooded areas experienced a rebound, though have remained below the non-flooded properties.[7]

{FIGURE_1_here}

Tables 1-3 present the results of hedonic regressions aimed at assessing how the storm affected real estate prices. Table 1 is for one or two family homes; Table 2 is for apartment buildings, and Table 3 is for commercial properties. All regressions contain a series of building-level controls (square footage of property, year built, square footage of lot, # of structures on the property, and total units, as well as year-quarterly dummies, building type dummies and zip code dummies). The tables present only the variables that, presumably, are storm-related. The standard errors are clustered at the zip code level. Each table provides five specifications. In all tables, Equations (1)-(2) include all the properties in the respective category from January 2003 to October 2014. Equations (3)-(5) include only properties that are either in the flooded area or within a half-mile away, and for sales within two years of the storm (i.e., November 2010 to October 2014).

[7] It is possible that after the storm, certain types of properties in the flooded areas are more likely to appear on the market than others (e.g., those less damaged by the storm). But an Inverse Mills Ratio (IMR) adjustment suggests that sample selection is not a key problem. Details are available upon request.

{TABLES_1_through_3_here}

The explanatory variables are:

1. Distance of property to shoreline (miles), interacted with before and after the storm dummies, respectively;
2. Whether the property was in the FEMA floodplain, interacted with before and after the storm dummies, respectively;
3. The elevation of property (in feet), interacted with before and after the storm dummies, respectively;
4. Dummy variable if property is flooded by storm surge, times a post-storm dummy;
5. Height of storm surge (0 for dry properties), interacted with a post-storm dummy;
6. Distance of dry properties from the storm surge, interacted with post-storm dummy and dry-property dummy; and
7. Quarterly occupancy rate of structures in respective census tracts (2005Q4-2014Q4).

Table 1 shows results for one and two family houses. Based on results from column (1), residential properties, on average, lost 12.7% of their value in the flooded zone. Columns (2)–(5) show a strong negative relationship between the height of the surge and the price after the storm. The results show that, on average, a one-foot increase in the storm surge is associated with 3.1%-3.7% drop in housing prices. The results also suggest higher elevation became more valuable after the storm. In column (5), we do not see evidence that, on average, dry properties close to the storm experienced any price impacts, but we explore this issue in more detail below.

The value of being in a FEMA floodplain district is unclear, given that the signs change across specifications. However, after the storm, all coefficients for the FEMA floodplain dummy are negative (though statistically insignificant), suggesting that being in the FEMA floodplains is an informational disamenity, given that it likely reveals new information about the likelihood of future storm flooding.

Table 2 contains the same regression specifications but for apartment buildings. Here we observe, with the results in column (1), apartment buildings lost about 16.7% of their value, on average, if they were in the flooded area. In columns (2) to (5), we see that a one-foot increase in the surge reduced prices between 6.0%-10.5%, on average. We also see evidence that the value of being in the FEMA floodplain became more negative after the storm. There is also evidence of an elevation premium after the storm—that is, the elevation coefficients increase in value after Sandy, suggesting that apartment buildings on higher ground become relatively more valuable.

Finally, Table 3, shows commercial properties results. On average, inundated commercial properties lost about 9.7% of their value (though it’s statistically insignificant). For flooded properties, each foot of surge height is associated with 8.2%-12.3% loss in value. Being in a FEMA floodplain after the storm yields significant losses for commercial properties, as those properties experience dramatic price drops. Interestingly, these properties do not exhibit an elevation premium.

In summary, across property types, there is strong evidence that the storm surge created significant losses in property values, as would be expected. Those properties with greater flooding lost more value, likely due to the greater damage caused by the storm. For apartment buildings, we see a strong elevation premium after the storm. Finally, across regressions, being in the FEMA floodplain after the storm caused properties to lose value. This suggests that the informational shock about the likelihood of being flooded is larger than the value placed on having flood insurance.

5.1 Dry Property Effects

In this section, using OLS regressions, we investigate how the storm surge might have influenced properties only on the dry side after the storm; that is, properties that flooded. Tables 4 through 6 present regression results.

Regarding the FEMA floodplain coefficients, for one and two family homes, there is not much evidence of an effect. For apartment buildings, the FEMA coefficients are all negative and there is some evidence that, after the storm, the coefficients become less negative, though, again, they are not statistically significant. For commercial properties, the effect becomes more negative after the storm, suggesting that commercial property buyers consider being close the FEMA boundary as bad news for their properties. For homes and apartments (but not commercial buildings), there is an elevation premium.

{Tables_4_thru_6_here}

The coefficient for the distance to the surge boundary is positive (but not significant) across all three dependent variables. This provides weak evidence that those properties further away from the flood zone rose in value, on average, as would be expected if there was an informational shock from the storm.

However, to explore this further, we also interact the borough dummies with the distance from the surge to see if there are heterogeneous effects across the city. Across dependent variables, for Brooklyn and Manhattan there are positive effects from being further away from the surge after the storm. However, there are negative effects for Bronx and Staten Island, which is contrary to what one might expect.

Importantly, in columns (4) and (5), across tables, there is a positive effect related to the distance to surge minus the distance to FEMA zone. This suggests that the informational shock generated changes in prices for these properties, on average. We explore this finding in more detail below to see if there is heterogeneity in the response to shocks across the city.

In summary, the key findings of these regressions are that the distance to the surge boundary, after the storm, for each of the five boroughs, shows positive effects in some cases and negative in others. We would expect positive in all cases—being further away would be better. Finally, we find a positive coefficient estimates for *Sandy Distance-FEMA Distance*, as would be expected. We now turn to the investigating the effects of the storm using LWRs, which allows us to explore in more detail the heterogeneous effects of the storm surge on the dry properties.

6. Repeat Sales and LWRs for Dry Properties: Results for NYC

Here we report the results of the LWRs for the repeat sales. When measuring the effects of Sandy on the dry properties, it is also important to control for other factors that might be correlated with the shock effect. Because the shock is likely to vary across the city (from north to south and east to west), latitude and longitude are useful proxies for this spatial variation. Thus, we included, as spatial controls, the latitude and longitude (in degrees) of each property.

Furthermore, since our data set includes properties with sales either before, during or after the financial crises and Great Recession, it is also important to include time-related controls, so, again, we could better isolate the cause of the shock. To this end, we include two time-related variables: the number of calendar days between the two repeat sales and the year and quarter of the second sale.

Descriptive statistics of the LWR coefficients are given in Table 7. OLS regression results the repeat sales equation for one and two family homes, apartment buildings and commercial structures, respectively, as well as LWR coefficient estimate histograms, are in Appendix B. Table 7 shows that, on average, the LWR coefficients for the surge boundary distance minus the FEMA boundary distance *(Surge – FEMA)* variable are positive, as would be expected. That is, a "positive" shock—when the storm did not come as close as the FEMA line—would mean a relative benefit for those property owners.

The average of the standard deviations of the coefficients are greater than the mean values of the coefficients, suggesting a relatively large degree of variation for the coefficient estimates. The ranges of the coefficients show this as well. Furthermore, a test for the non-stationarity of the coefficients shows that for each of the three dependent variables, we can reject stationarity at greater than the 99% level of confidence.

In sum, the evidence strongly rejects that OLS estimates accurately measure the relationship between the storm shock and price changes; rather the LWR estimates are better able to measure the degree of spatial heterogeneity across the city.

{Table_7_here}

Figures 2 through 4 present maps of the t-statistics of the coefficients (i.e., the LWR coefficient estimates divided by their respective standard errors). There are a few general conclusions to be drawn from these maps. First, there is, again, a substantial degree of variation in the coefficient estimates and their significance levels. For homes, the largest positive t-statistics are mostly found in Brooklyn and Queens. Staten Island, northeastern Queens, and the western Bronx for the most part, have insignificant coefficient estimates. For apartments, the largest t-statistics appear in Brooklyn and Queens, with some positive coefficients along the central spine of the Bronx. For commercial properties, there are large effects in downtown Manhattan and also across Brooklyn.

{Figures_2_through_4_here}

A key finding is that, while the OLS coefficient generates an average coefficient for the shock variable, there are large swaths of the city where prices are unaffected by the surge simultaneous with areas of the city that are affected. In particular, the areas of the city that appear to have the greatest impact are the older residential neighborhoods in Brooklyn. In the next section, we test some hypotheses about what might be driving the variation in the LWR coefficient estimates across the city.

Also, as the t-statistics demonstrate across the figures, while there are negative coefficient estimates, the vast majority of them are statistically insignificant. This again suggests there is no "shock effect" across much of the city and a concentrated shock effect in other parts. Also see Appendix B for histograms of the LWR coefficient estimates.

7. Explaining Heterogeneous Effects

The variation in the LWR coefficients leads to the question of what might account for this variation across the city and across property types. That is, why would some properties exhibit greater volatility than others?

The maps of the t-statistics suggest several possible factors. In particular, in all three property classes, Brooklyn has large pockets of positive and significant coefficients. More broadly, the maps suggest that properties close to the center of the city (here designated as the Empire State Building) have larger values as well.

These areas have undergone a significant degree of gentrification in the last decade or so (Meltzer and Ghorbani, 2017). This might lead one to consider a theory that gentrification is driving the responses to the shocks. Areas with significant new investments will have more to lose (or gain) with future shocks, and, as a result, they will exhibit greater volatility.

This theory would suggest a few testable hypotheses: the closer the property to the center city, the greater the coefficient estimate; properties in denser areas, with better public transportation, would have larger coefficients estimates.

To test this hypotheses, we created a data set where the LWR coefficient estimates are the dependent variables and with several control variables, including 2010 Census data at the tract level (see Appendix C.). The data include the three spatial variables: the distance of the property to the Empire State Building (i.e. distance to city center as the crow flies in miles), the lot's elevation (in feet), and the building's latitude and longitude coordinates (in degrees).

For demographic variables, we include the census tract's median income, and the percent of residents who are white, black and Hispanic, respectively (all other racial/ethnic groups are the omitted category). As measures of neighborhood density, we include the census tract population and the average floor area ratio of each building in the tracts. As a measure of transportation access, we included the number of subway stops within a half mile of each property. Finally, we also included borough dummy variables. The standard errors are clustered by the census tract.

We ran three regressions for each of the three properties. The first equation includes only spatial characteristics and borough dummies. The second regression includes income and population counts. The third equation includes all the variables listed in the previous paragraph.

First, for all property types, the distance to the center of the city is negatively correlated with the LWR coefficient estimates, suggesting that, all else equal, those (higher valued) properties near the center city are more responsive to shocks. For all properties, transportation access is positively related to the shock coefficients. For the one and two family homes and commercial properties, we find that higher income neighborhoods positively correlated with the LWR coefficient estimates. We do not find strong support for our density measures or the racial/ethnic categories.

Taken together, the data support that more centralized locations have higher coefficients, all else equal. However, we leave for future work a more detailed treatment of the causes for the variation in shock effects.

{Table_8_here}

To confirm our estimates are measuring the effect of Sandy on real estate values, and do not emerge from some other source, we perform some LWR statistical robustness tests. Table 9 presents p-values for these tests.

The "straddle" column contains the results of the tests for three data sets with repeat sales pairs that straddle the storm, while the "both after" column is for repeat sales where both sales occur *after the storm*. Since there is no Sandy shock during this time, we expect to find in these cases that we cannot reject the respective null hypothesis of no effect from the *Surge-FEMA* variable.

The first series of tests relates to ordinary least squares where we look at price changes as a function of the "shock variable" and other controls as discussed above. For homes and apartments, there are significant results for the straddle sales, but insignificant effects for the post-Sandy sales. For commercial properties, there is an insignificant effect for both the straddle properties and the two post-Sandy sales (though we do see a rise in the p-value for the "both after" sales).

Next, we test whether LWRs are able to capture the variation in the data better than OLS, using the bandwidth test, which is a test if the LWR regressions perform better than OLS in fitting the data (and essentially tests if at least one right hand side variable should be estimated via LWR). In all cases but one, the answer is "yes." This result suggests that at least one right hand side variable in the specification improves the model fit. However, this leads to the next test of whether the LWRs perform better for our variable of interest—*Sandy distance-FEMA distance*.

In this case, we perform tests of spatial nonstationarity of the *Sandy-FEMA* coefficients. Here the null hypothesis is that there is no variation in the coefficients and hence LWRs would not be necessary for that particular variable. We find that we can reject the null in the straddle cases, but we cannot reject it in the "both after" cases. This finding suggests that there is spatial stationarity in the "both after" case and that the spatial stationary coefficient is statistically insignificant. In short, the spatial stationarity test combined with the OLS test suggests that our straddle data is capturing a true Sandy effect.

{Table_9_here}

8. Conclusion

In this paper we develop a methodology to estimate the effects that a major hurricane has on properties that are not flooded by the storm. Our approach examines how price changes are affected by the distance to the actual flood zone relative to the distance of the FEMA floodplain. Before the storm, the floodplain maps serve to provide information on the likelihood a flood. The FEMA floodplain maps typically differ from actual storm surges. After the storm, the distance to the inundation zone provides new information about future flooding risks from storm surges, which may detrimentally impact property values. Moreover, these shifts in flood risks can be different across locations within a city, because at some locations these “shocks” may be substantial while at other locations they may be negligible.

Our methodology uses LWRs to allow for the possibility that the effects of the surge are heterogonous across a city. Our approach controls for both supply (i.e., building occupancy) and demand effects, as well as sample selection issues with an inverse Mills ratio adjustment. We propose using a smooth local price index, which controls for neighborhood price variation independent of the intendent variable of interest.

We demonstrate our methodology with a dataset of NYC commercial and residential repeat sales properties that sell once before and once after Hurricane Sandy. Neighborhoods in outer parts of NYC experience no changes in prices due to the storm. But we see large and statically significant effects in the inner parts of the city, in particular parts of Brooklyn and Queens near Manhattan, which suggest gentrifying neighborhoods have the largest coefficients. We illustrate these heterogeneous effects with several maps, and these maps demonstrate substantial variation in the effects across NYC.

Our approach has the potential to address the flood risk impacts of other recent, major storms, such as Hurricanes Harvey and Irma. Given the prevalence of several major, devastating hurricanes in the past few years in the U.S., our approach has the potential to aid policy makers in estimating damages from devastating storms, and can provide information on potential benefits that could have been realized if preventative storm surge mitigation had been undertaken. We also anticipate our empirical model may be calibrated to various cities across the U.S. for forecasting the property value impacts of an array of potential future “shock” effects from unexpected hurricanes or floods.

References

Anenberg, E., & Laufer, S. Forthcoming. A more timely house price index. Review of Economics and Statistics.

Atreya, A., & Czajkowski, J. (2016). Graduated Flood Risks and Property Prices in Galveston County. Real Estate Economics.

Bailey, M. J., Muth, R. F., & Nourse, H. O. (1963). A regression method for real estate price index construction. Journal of the American Statistical Association, 58(304), 933-942.

Barr, J., & Cohen, J. P. (2014). The floor area ratio gradient: New York City, 1890–2009. Regional Science and Urban Economics, 48, 110-119.

Bin, O., & Landry, C. E. (2013). Changes in implicit flood risk premiums: Empirical evidence from the housing market. Journal of Environmental Economics and management, 65(3), 361-376.

Bin, O., Poulter, B., Dumas, C. F., & Whitehead, J. C. (2011). Measuring the impact of sea-level rise on coastal real estate: a hedonic property model approach. Journal of Regional Science, 51(4), 751-767.

CNN (2013). http://www.cnn.com/2013/07/13/world/americas/hurricane-sandy-fast-facts/

Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. Journal of the American statistical association, 83(403), 596-610.

Deng, Y., McMillen, D.P., & Sing, T.F. (2012). Private residential price indices in Singapore: a matching approach. Regional Science and Urban Economics. 42(3), 485-494.

ESA (2013). http://www.esa.doc.gov/sites/default/files/sandyfinal101713.pdf . Accessed on 9/16/2017.

Faber, J. W. (2015). Superstorm Sandy and the demographics of flood risk in New York city. *Human ecology*, *43*(3), 363-378.

FEMA (2017a) https://www.fema.gov/media-library-data/1468504201672-3c52280b1b1d936e8d23e26f12816017/Flood_Hazard_Mapping_Updates_Overview_Fact_Sheet.pdf . Accessed on 9/16/2017.

FEMA (2017b). https://www.floodsmart.gov/floodsmart/pages/about/nfip_overview.jsp . Accessed on 9/16/2017.

Fessenden, F., Gebeloff, R. Walsh, M. W. and Griggs, T. (2017). Water damage from Hurricane Harvey extended far beyond flood zones." New York Times, September 1.

Fotheringham, A. S., & Brunsdon, C. (2002). Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Charlton.

Hammond, M. J., Chen, A. S., Djordjević, S., Butler, D., & Mark, O. (2015). Urban flood impact assessment: A state-of-the-art review. *Urban Water Journal*, *12*(1), 14-29.

Heckman, J.J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Annals of Economic and Social Measurement. 5(4): 475–492.

Heckman, J.J. (1979). Sample selection as a specification error. Econometrica. 47 (1): 153–161.

Heitz, C., Spaeter, S., Auzet, A. V., & Glatron, S. (2009). Local stakeholders' perception of muddy flood risk and implications for management approaches: A case study in Alsace (France). *Land Use Policy*, *26*(2), 443-451.

HUD (2016). Aggregated USPS Administrative Data on Address Vacancies. https://www.huduser.gov/portal/usps/home.html Accessed 9/16/2017.

MacDonald, D. N., Murdoch, J. C., & White, H. L. (1987). Uncertain hazards, insurance, and consumer choice: Evidence from housing markets. Land Economics, 63(4), 361-371.

McMillen, D. P., & Redfearn, C. L. (2010). Estimation and hypothesis testing for nonparametric hedonic house price functions. Journal of Regional Science, 50(3), 712-733.

McMillen, D. P., & Dombrow, J. (2001). A flexible Fourier approach to repeat sales price indexes. Real Estate Economics, 29(2), 207.

McMillen, D.P. & McDonald, J.F. (2004). Reaction of house prices to a new rapid transit line: Chicago's midway line, 1983–1999. Real Estate Economics, 32(3), 463-486.

McMillen, D. P. (2003). Neighborhood house price indexes in Chicago: a Fourier repeat sales approach. Journal of Economic Geography, 3(1), 57-73.

McWilliams, G. & Marianna, P. (2017). Texas governor says Harvey damage could reach $180 billion. Reuters. Retrieved September 3, 2017.

Meltzer, R., & Ghorbani, P. (2017). Does gentrification increase employment opportunities in low-Income neighborhoods? Regional Science and Urban Economics.

Ortega, F., & Taspinar, S. (2016). Rising sea levels and sinking property values: the effects of Hurricane Sandy on New York's housing market (No. 10374). Institute for the Study of Labor (IZA).

Ries, J., & Somerville, T. (2010). School quality and residential property values: evidence from Vancouver rezoning. The Review of Economics and Statistics, 92(4), 928-944.

Sun-Sentinel. (2017). Tornado warning issued for Broward and Palm Beach Counties, as thunderstorm spun off by Hurricane Irma". September 10. Retrieved September 10, 2017.

Wood, Z. (2017). Economic cost of Hurricane Irma 'could reach $300bn'. The Guardian. ISSN 0261-3077. September 10. Retrieved September 13, 2017.

Figures and Tables

Figures

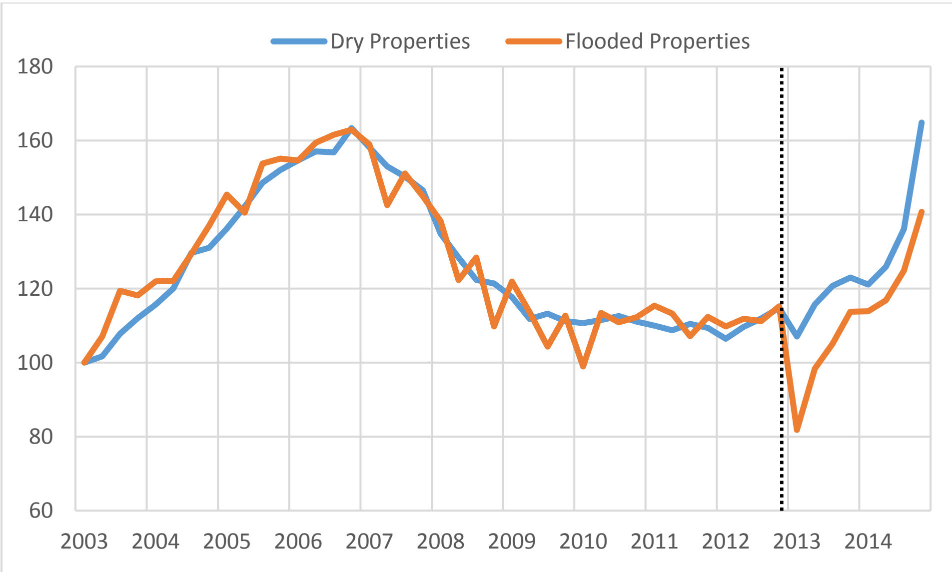

Figure 1: Real Price Index for NYC Real Estate (of all building types), Jan. 2003- Oct. 2014. Vertical Line is date of Hurricane Sandy. See the Appendix A for sources and preparation.

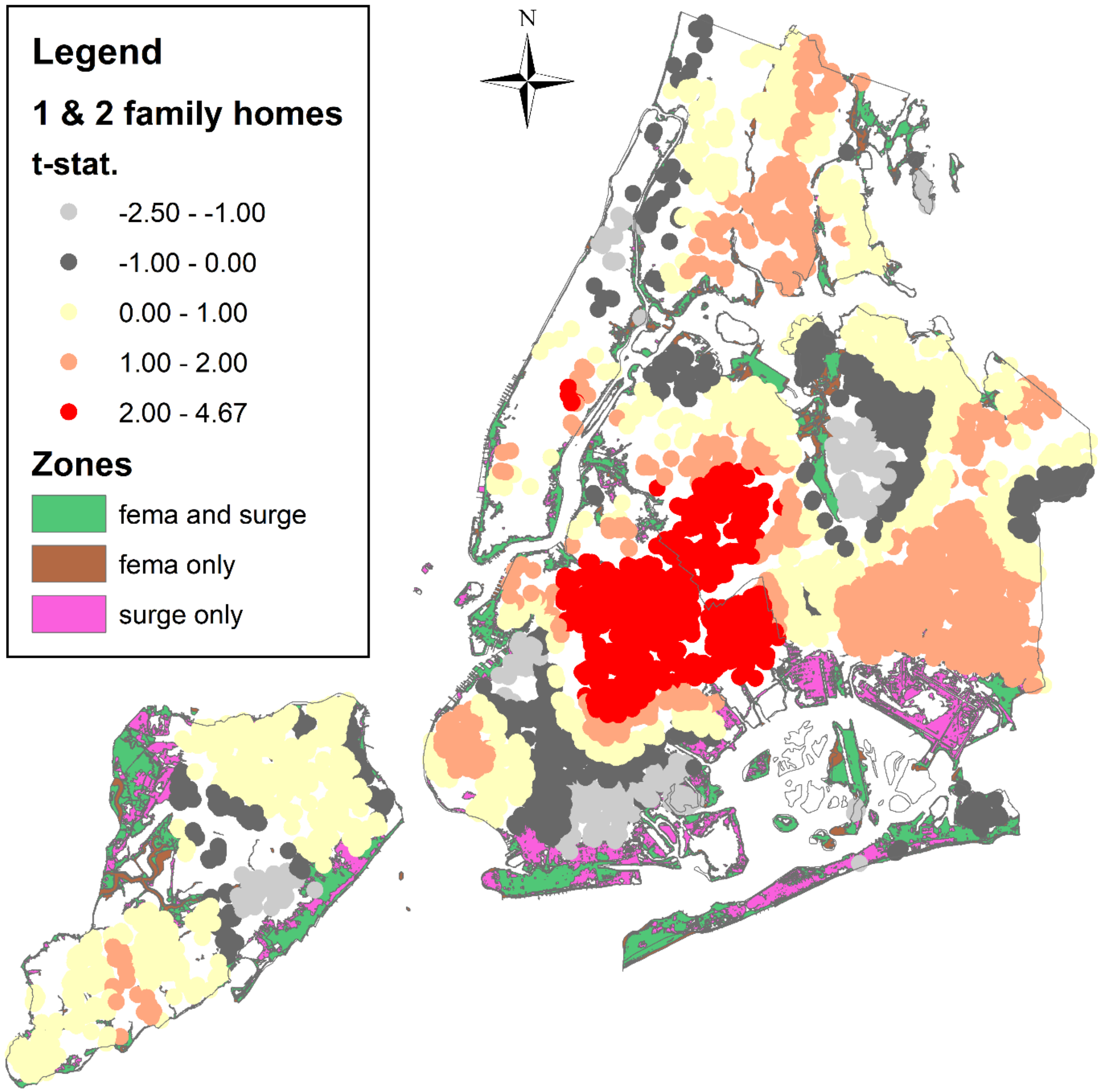

Figure 2: t-statistics for LWR coefficients for one and two family homes. See the Appendix A for data sources.

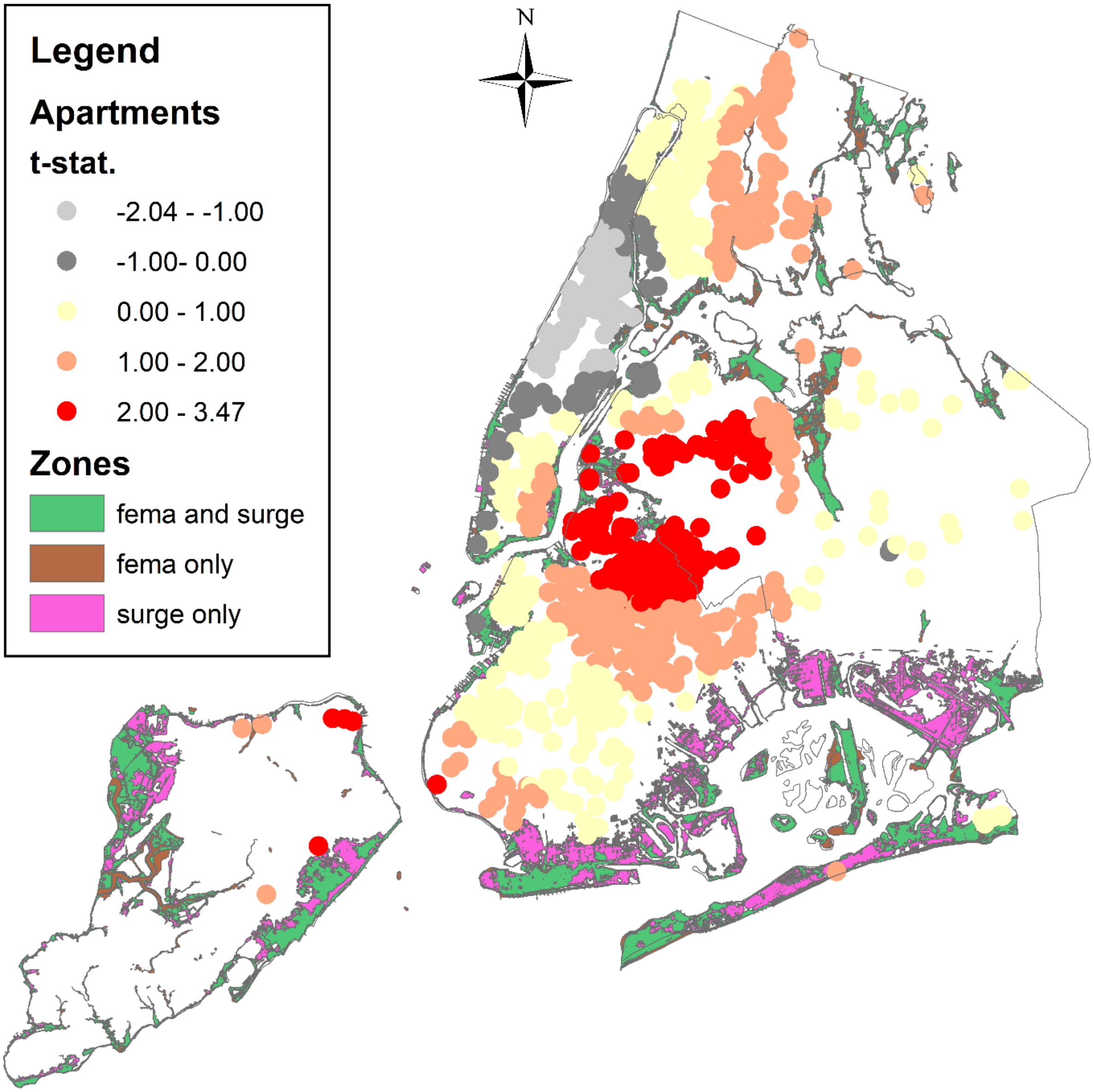

Figure 3: t-statistics for LWR coefficients for apartment buildings. See the Appendix A for data sources.

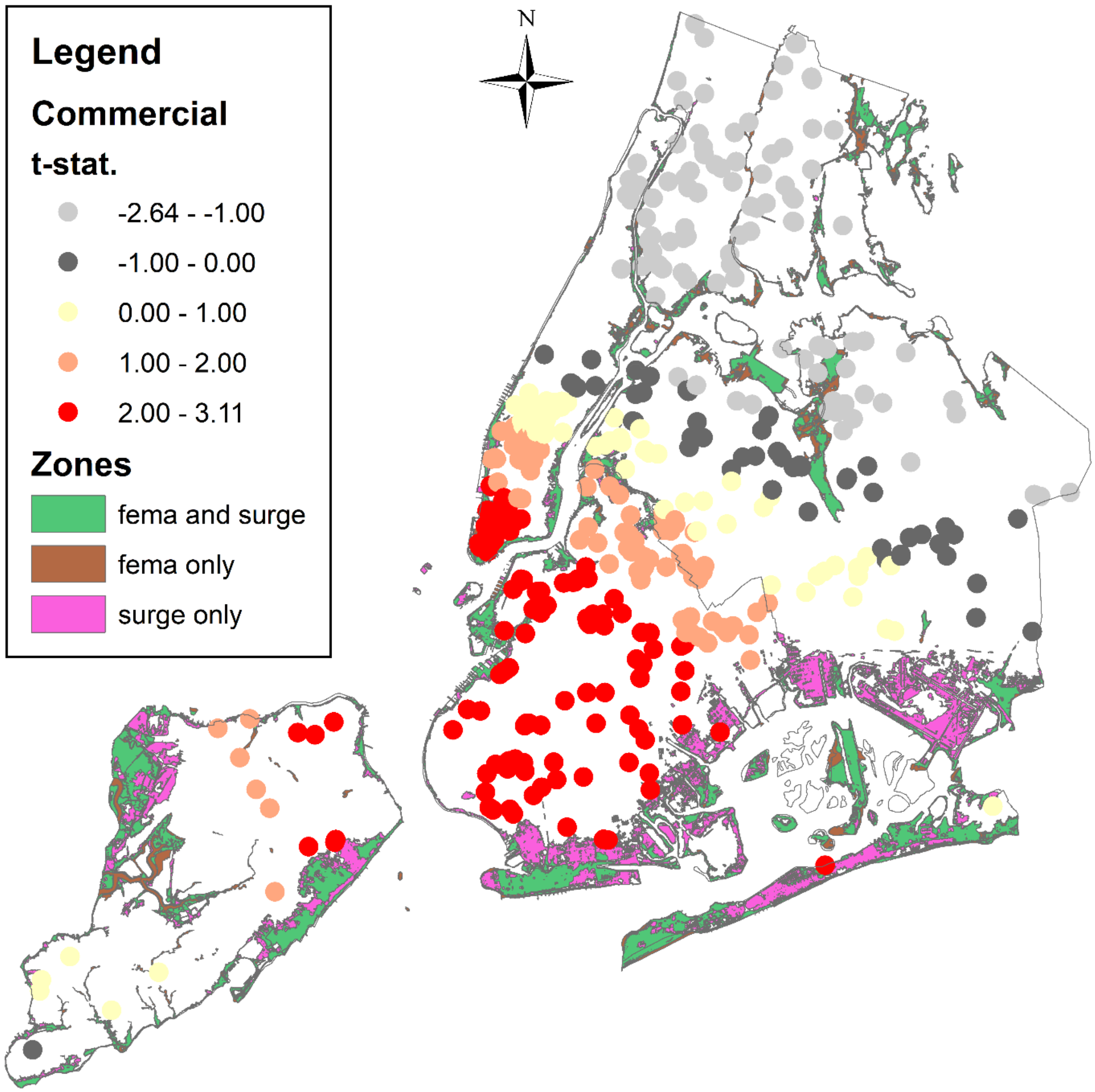

Figure 4: t-statistics for LWR coefficients for commercial properties. See the Appendix A for data sources.

Tables

Table 1: OLS Regressions for one and two family homes. Dependent Var.: Ln(Real Price per Sq. Foot)

	(1)	(2)	(3)	(4)	(5)
Dist. to Shore x Pre-Sandy Dummy	0.028	0.028	0.068	0.069	0.068
	(1.56)	(1.56)	(2.14)*	(2.08)*	(2.02)*
Dist. to Shore x Post-Sandy Dummy	0.014	0.013	0.076	0.079	0.076
	(0.62)	(0.60)	(2.40)*	(2.16)*	(2.06)*
FEMA Zone Dummy x Pre-Sandy Dummy	0.052	0.052	-0.004	-0.004	-0.004
	(1.77)	(1.78)	(0.04)	(0.04)	(0.04)
FEMA Zone Dummy x Post-Sandy Dummy	-0.034	-0.035	-0.012	-0.012	-0.022
	(0.70)	(0.72)	(0.25)	(0.24)	(0.41)
Elevation (feet) x Pre-Sandy Dummy	0.0027	0.0027	0.0022	0.0022	0.0021
	(3.30)**	(3.29)**	(2.56)*	(2.54)*	(2.35)*
Elevation (feet) x Post-Sandy Dummy	0.0038	0.0038	0.0024	0.0026	0.0024
	(3.29)**	(3.27)**	(2.76)**	(2.78)**	(2.63)**
Inundated Dummy x Post-Sandy Dummy	-0.127	-0.007	-0.036	-0.038	-0.035
	(4.40)**	(0.17)	(1.02)	(1.06)	(1.12)
Surge height (feet) x Post-Sandy Dummy		-0.037	-0.032	-0.031	-0.033
		(2.84)**	(2.19)*	(2.19)*	(2.39)*
Dist. to Surge (miles) x Dry Property Dummy x Post-Sandy Dummy				-0.013	-0.013
				(0.37)	(0.37)
Neighborhood Occupancy Rate					0.116
					(1.83)
R^2	0.24	0.24	0.27	0.27	0.27
# obs.	274,263	274,263	50,709	50,709	48,875

Note: Prices are deflated by NYC CPI, excluding shelter. Only storm-related variables shown; full regressions include housing and neighborhood controls (see Appendix A). Standard errors were clustered by zip codes. t-statistics are given below coefficient estimates. **Stat. sig. at 99%. *Stat. sig. at 95%.

Table 2: OLS regressions for apartment buildings. Dependent Var.: Ln(Real Price per Sq. Foot)

	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Dist. to Shore x Pre-Sandy Dummy	0.02	0.02	0.086	0.084	0.085
	(0.72)	(0.72)	(1.03)	(1.01)	(0.99)
Dist. to Shore x Post-Sandy Dummy	-0.05	-0.05	-0.119	-0.178	-0.178
	(1.49)	(1.51)	(1.66)	(1.83)	(1.79)
FEMA Zone Dummy x Pre-Sandy Dummy	-0.055	-0.052	-0.104	-0.107	-0.107
	(0.50)	(0.47)	(0.36)	(0.37)	(0.37)
FEMA Zone Dummy x Post-Sandy Dummy	-0.08	-0.078	-0.183	-0.169	-0.169
	(0.42)	(0.41)	(1.05)	(0.98)	(0.97)
Elevation (feet) x Pre-Sandy Dummy	0.001	0.001	0.000	0.000	0.000
	(0.75)	(0.75)	(0.04)	(0.02)	(0.07)
Elevation (feet) x Post-Sandy Dummy	0.004	0.004	0.004	0.003	0.003
	(2.15)*	(2.14)*	(1.68)	(1.37)	(1.30)
Inundated Dummy x Post-Sandy Dummy	-0.167	0.189	0.082	0.102	0.118
	(1.23)	(1.57)	(0.76)	(0.94)	(1.06)
Surge height (feet) x Post-Sandy Dummy		-0.105	-0.059	-0.06	-0.066
		(2.41)*	(2.00)*	(2.03)*	(2.09)*
Dist. to Surge (miles) x Dry Property Dummy x Post-Sandy Dummy				0.114	0.12
				(1.05)	(1.10)
Neighborhood Occupancy Rate					-0.011
					(0.07)
R^2	0.29	0.29	0.37	0.37	0.37
# obs.	29,277	29,277	7,005	7,005	6,909

Note: Prices are deflated by NYC CPI, excluding shelter. Only storm-related variables shown; full regressions include housing and neighborhood controls (see Appendix A). Standard errors were clustered by zip codes. t-statistics are given below coefficient estimates. **Stat. sig. at 99%. *Stat. sig. at 95%.

Table 3: OLS regressions for commercial properties. Dependent Var.: Ln(Real Price per Sq. Foot)

	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Dist. to Shore x Pre-Sandy Dummy	0.092	0.093	0.012	0.011	0.011
	(2.78)**	(2.80)**	(0.15)	(0.13)	(0.13)
Dist. to Shore x Post-Sandy Dummy	0.04	0.038	0.062	0.054	0.059
	(0.99)	(0.96)	(0.98)	(0.68)	(0.73)
FEMA Zone Dummy x Pre-Sandy Dummy	-0.038	-0.036	0.027	0.027	0.03
	(0.27)	(0.26)	(0.15)	(0.15)	(0.16)
FEMA Zone Dummy x Post-Sandy Dummy	-0.271	-0.274	-0.237	-0.236	-0.238
	(1.31)	(1.32)	(1.19)	(1.19)	(1.23)
Elevation (feet) x Pre-Sandy Dummy	0.001	0.001	-0.003	-0.003	-0.003
	(0.72)	(0.73)	(0.66)	(0.65)	(0.74)
Elevation (feet) x Post-Sandy Dummy	0.001	0.001	-0.003	-0.003	-0.003
	(0.44)	(0.33)	(0.92)	(0.90)	(0.84)
Inundated Dummy x Post-Sandy Dummy	-0.097	0.237	0.148	0.15	0.178
	(1.06)	(1.43)	(0.91)	(0.91)	(1.09)
Surge height (feet) x Post-Sandy Dummy		-0.123	-0.082	-0.082	-0.086
		(2.81)**	(1.73)	(1.73)	(1.86)
Dist. to Surge (miles) x Dry Property Dummy x Post-Sandy Dummy				0.021	0.006
				(0.14)	(0.04)
Neighborhood Occupancy Rate					-0.152
					(0.65)
R^2	0.31	0.31	0.28	0.28	0.28
# obs.	15,923	15,923	4,008	4,008	3,947

Note: Prices are deflated by NYC CPI, excluding shelter. Only storm-related variables shown; full regressions include housing and neighborhood controls (see Appendix A). Standard errors were clustered by zip codes. t-statistics are given below coefficient estimates. **Stat. sig. at 99%. *Stat. sig. at 95%.

Table 4: OLS regressions for homes not flooded. Dependent Var.: Ln(Real Price per Sq. Foot)

	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Dist. To Shore (miles) x Pre-Sandy Dummy	0.022	0.024	0.059	0.013	0.011
	(1.25)	(1.37)	(1.79)	(0.57)	(0.47)
Dist. To Shore (miles) x Post-Sandy Dummy	0.00	-0.016	0.065	0.019	0.016
	(0.01)	(0.59)	(1.77)	(0.69)	(0.58)
Dist. To FEMA Boundary (miles) x Pre-Sandy Dummy	0.039	0.035	-0.017	-0.019	-0.021
	(1.38)	(1.22)	(0.18)	(0.21)	(0.22)
Dist. To FEMA Boundary (miles) x Post-Sandy Dummy	-0.042	-0.023	-0.001	-0.044	-0.056
	(0.89)	(0.48)	(0.02)	(0.94)	(1.09)
Elevation (feet) x Pre-Sandy Dummy	0.003	0.003	0.002	0.003	0.003
	(3.33)**	(3.16)**	(2.51)*	(3.58)**	(3.37)**
Elevation (feet) x Post-Sandy Dummy	0.004	0.006	0.004	0.004	0.003
	(3.23)**	(4.74)**	(3.73)**	(3.17)**	(3.02)**
Dist. to Surge Boundary (miles) x Post-Sandy Dummy	0.019				
	(0.73)				
Dist. to Surg. x Post-Sandy for MN		0.535	0.151	0.148	0.148
		(1.94)	(0.68)	(0.68)	(0.68)
Dist. to Surg. x Post-Sandy for BK		0.071	0.048	0.062	0.061
		(2.81)**	(1.16)	(2.94)**	(2.88)**
Dist. to Surg. x Post-Sandy for BX		-0.192	-0.113	-0.142	-0.142
		(6.39)**	(2.50)*	(4.81)**	(4.97)**
Dist. to Surge x Post-Sandy for QN		0.012	0.001	-0.015	-0.015
		(0.44)	(0.02)	(0.78)	(0.82)
Dist. to Surge x Post-Sandy for SI		-0.072	-0.058	-0.089	-0.083
		(1.80)	(1.26)	(2.49)*	(2.43)*
Sandy Dist. - FEMA Dist. (miles) x Post-Sandy				0.096	0.10
				(2.82)**	(2.88)**
Neighborhood Occupancy Rate					0.147
					(2.79)**
R^2	0.25	0.25	0.28	0.27	0.27
# obs.	253,960	253,960	44,966	73,802	71,856

Note: Prices are deflated by NYC CPI, excluding shelter. Only storm-related variables shown; full regressions include housing and neighborhood controls (see Appendix A). Standard errors were clustered by zip codes. T-statistics are given below coefficient estimates. **Stat. sig. at 99%. *Stat. sig. at 95%.

Table 5: OLS regressions for apartment buildings not flooded. Dependent Var.: Ln(Real Price per Sq. Foot)

	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Dist. To Shore (miles) x Pre-Sandy Dummy	0.015	0.018	0.047	-0.006	-0.006
	(0.55)	(0.65)	(0.57)	(0.11)	(0.11)
Dist. To Shore (miles) x Post-Sandy Dummy	-0.109	-0.215	-0.208	-0.138	-0.141
	(1.80)	(3.32)**	(2.04)*	(1.37)	(1.37)
Dist. To FEMA Boundary (miles) x Pre-Sandy Dummy	-0.063	-0.065	-0.141	-0.167	-0.163
	(0.50)	(0.53)	(0.51)	(0.58)	(0.55)
Dist. To FEMA Boundary (miles) x Post-Sandy Dummy	-0.064	-0.039	-0.119	-0.114	-0.117
	(0.32)	(0.19)	(0.70)	(0.68)	(0.69)
Elevation (feet) x Pre-Sandy Dummy	0.001	0.001	0.000	-0.001	-0.001
	(0.69)	(0.47)	(0.14)	(0.40)	(0.49)
Elevation (feet) x Post-Sandy Dummy	0.004	0.005	0.003	0.002	0.002
	(1.87)	(2.52)*	(1.38)	(1.38)	(1.29)
Dist. to Surge Boundary (miles) x Post-Sandy Dummy	0.079				
	(1.06)				
Dist. to Surg. x Post-Sandy for MN		0.482	0.424	0.41	0.424
		(2.50)*	(2.60)*	(2.17)*	(2.23)*
Dist. to Surg. x Post-Sandy for BK		0.235	0.203	0.22	0.228
		(3.55)**	(1.51)	(2.22)*	(2.26)*
Dist. to Surg. x Post-Sandy for BX		-0.086	-0.002	-0.066	-0.054
		(1.34)	(0.02)	(0.53)	(0.42)
Dist. to Surge x Post-Sandy for QN		0.155	0.007	0.068	0.075
		(1.97)	(0.04)	(0.54)	(0.59)
Dist. to Surge x Post-Sandy for SI		-0.379	-0.184	-0.284	-0.253
		(1.36)	(0.69)	(1.04)	(0.87)
Sandy Dist. - FEMA Dist. (miles) x Post-Sandy				0.082	0.082
				(0.71)	(0.70)
Neighborhood Occupancy Rate					0.115
					(1.02)
R^2	0.29	0.29	0.35	0.31	0.31
# obs.	28,456	28,456	6,721	9,575	9,460

Note: Prices are deflated by NYC CPI, excluding shelter. Only storm-related variables shown; full regressions include housing and neighborhood controls (see Appendix A). Standard errors were clustered by zip codes. t-statistics are given below coefficient estimates. **Stat. sig. at 99%. *Stat. sig. at 95%.

Table 6: OLS regressions for commercial buildings not flooded. Dependent Var.: Ln(Real Price per Sq. Foot)

	(1)	(2)	(3)	(4)	(5)
Dist. To Shore (miles) x Pre-Sandy Dummy	0.087	0.086	-0.006	0.044	0.045
	(2.59)*	(2.50)*	(0.08)	(0.90)	(0.91)
Dist. To Shore (miles) x Post-Sandy Dummy	-0.023	-0.028	0.005	0.013	0.022
	(0.39)	(0.50)	(0.06)	(0.19)	(0.34)
Dist. To FEMA Boundary (miles) x Pre-Sandy Dummy	-0.029	-0.035	0.033	0.067	0.096
	(0.21)	(0.25)	(0.18)	(0.36)	(0.51)
Dist. To FEMA Boundary (miles) x Post-Sandy Dummy	-0.255	-0.221	-0.205	-0.255	-0.257
	(1.17)	(1.03)	(1.00)	(1.19)	(1.28)
Elevation (feet) x Pre-Sandy Dummy	0.001	0.001	-0.002	-0.001	0.000
	(0.83)	(0.78)	(0.53)	(0.42)	(0.15)
Elevation (feet) x Post-Sandy Dummy	-0.001	0.001	-0.002	-0.001	0.000
	(0.46)	(0.29)	(0.50)	(0.32)	(0.02)
Dist. to Surge Boundary (miles) x Post-Sandy Dummy	0.095				
	(1.36)				
Dist. to Surg. x Post-Sandy for MN		0.487	0.20	0.189	0.166
		(1.88)	(0.80)	(0.78)	(0.68)
Dist. to Surg. x Post-Sandy for BK		0.19	0.097	0.124	0.101
		(2.54)*	(0.54)	(1.49)	(1.22)
Dist. to Surg. x Post-Sandy for BX		-0.013	-0.305	-0.138	-0.149
		(0.13)	(1.52)	(1.10)	(1.18)
Dist. to Surge x Post-Sandy for QN		0.053	0.051	0.001	-0.009
		(0.83)	(0.27)	(0.02)	(0.10)
Dist. to Surge x Post-Sandy for SI		-0.135	-0.011	-0.131	-0.184
		(1.29)	(0.06)	(0.83)	(1.18)
Sandy Dist. - FEMA Dist. (miles) x Post-Sandy				0.163	0.174
				(1.29)	(1.37)
Neighborhood Occupancy Rate					-0.164
					(0.93)
R^2	0.32	0.32	0.28	0.28	0.27
# obs.	14,853	14,853	3,630	4,986	4,922

Note: Prices are deflated by NYC CPI, excluding shelter. Only storm-related variables shown; full regressions include housing and neighborhood controls (see Appendix A). Standard errors were clustered by zip codes. t-statistics are given below coefficient estimates. **Stat. sig. at 99%. *Stat. sig. at 95%.

Table 7: Descriptive statistics for LWR coefficients, standard errors, and t-statistics.

Variable	Mean	Std. Dev.	Min.	Max.
One & Two Family Homes (# obs.=5248)				
surge - fema coeff.	0.14	0.29	-4.39	3.35
surge - fema s.e.	0.18	0.22	0.06	3.25
surge - fema t-stat	1.07	1.41	-2.50	4.67
Δoccupancy coeff.	0.83	0.77	-2.77	16.35
Δoccupancy s.e.	0.46	0.43	0.27	12.23
Δoccupancy t-stat	2.22	1.34	-1.03	5.52
Δmills ratio coeff.	-1.89	0.81	-10.59	0.24
Δmills ratio s.e.	0.38	0.19	0.24	3.96
Δmills ratio t-stat.	-5.55	2.78	-10.69	0.57
Δdays coeff.	-0.0005	0.0005	-0.0020	0.0000
Δdays s.e.	0.000	0.000	0.000	0.000
Δdays t-stat.	-12.89	8.67	-30.76	0.17
year_quarter coeff.	-0.61	0.15	-0.96	0.12
year_quarter s.e.	0.06	0.03	0.04	0.46
year_quarter t-stat.	-12.51	5.06	-21.31	0.71
latitude coeff.	0.41	3.19	-29.69	10.27
latitude s.e.	1.81	0.98	0.81	22.30
latitude t-stat.	0.33	1.85	-3.21	5.66
longitude coeff.	-0.56	2.53	-13.22	12.38
longitude s.e.	1.51	1.07	0.74	16.02
longitude t-stat.	-0.24	1.91	-6.82	5.29
Apartment Buildings (# obs.=1553)				
surge - fema coeff.	0.11	0.29	-0.69	5.28
surge - fema s.e.	0.17	0.19	0.10	3.70
surge - fema t-stat	0.88	1.30	-2.04	3.47
Δoccupancy coeff.	0.14	0.27	-1.89	3.09
Δoccupancy s.e.	0.41	0.29	0.29	4.57
Δoccupancy t-stat	0.37	0.49	-0.79	1.64
Δmills ratio coeff.	-0.16	0.57	-3.60	4.13
Δmills ratio s.e.	0.39	0.59	0.27	10.95
Δmills ratio t-stat.	-0.55	1.53	-2.68	2.57
Δdays coeff.	0.00	0.00	0.00	0.00
Δdays s.e.	0.00	0.00	0.00	0.00
Δdays t-stat.	-3.58	1.95	-6.48	2.31
year_quarter coeff.	-0.14	0.10	-0.60	0.23
year_quarter s.e.	0.06	0.09	0.05	1.73
year_quarter t-stat.	-2.53	1.67	-5.02	2.12
latitude coeff.	1.60	1.72	-3.05	22.15
latitude s.e.	1.14	1.21	0.39	17.56

latitude t-stat.	1.67	1.30	-1.08	4.75
longitude coeff.	0.12	2.67	-5.14	11.27
longitude s.e.	1.11	0.78	0.67	15.17
longitude t-stat.	-0.08	2.41	-4.20	4.39
Commercial Properties (# obs.=431)				
surge - fema coeff.	0.07	0.15	-0.60	0.24
surge - fema s.e.	0.17	0.03	0.15	0.40
surge - fema t-stat	0.47	0.77	-1.50	1.22
Δoccupancy coeff.	0.33	0.46	-1.72	1.39
Δoccupancy s.e.	0.47	0.27	0.39	3.39
Δoccupancy t-stat	0.76	0.92	-1.41	2.38
Δmills ratio coeff.	-0.24	0.09	-0.71	-0.03
Δmills ratio s.e.	0.12	0.04	0.10	0.59
Δmills ratio t-stat.	-2.10	0.60	-2.86	-0.22
Δdays coeff.	0.00	0.00	0.00	0.00
Δdays s.e.	0.00	0.00	0.00	0.00
Δdays t-stat.	-3.50	0.34	-4.28	-2.40
year_quarter coeff.	-0.24	0.06	-0.39	0.35
year_quarter s.e.	0.07	0.03	0.06	0.37
year_quarter t-stat.	-3.53	0.72	-4.88	0.95
latitude coeff.	0.08	1.80	-2.28	10.96
latitude s.e.	0.79	0.31	0.67	4.04
latitude t-stat.	-0.12	1.70	-2.78	3.72
longitude coeff.	-2.05	0.98	-8.14	0.36
longitude s.e.	0.66	0.22	0.55	2.63
longitude t-stat.	-3.06	0.85	-4.28	0.46

Table 8: OLS regressions of LWR Coefficients on census tract (CT) level and other controls.

	1 & 2 Family Homes			Apartment Buildings			Commercial Properties		
Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
Dist. To ESB (miles)	-0.03	-0.031	-0.021	-0.011	-0.012	-0.014	-0.011	-0.012	-0.013
	(4.40)**	(4.11)**	(2.75)**	(1.91)	(2.06)*	(2.28)*	(2.20)*	(2.47)*	(2.34)*
Latitude (degrees)	-0.46	-0.49	0.085	-0.18	-0.20	-0.60	-3.14	-3.19	-3.10
	(0.86)	(0.90)	(0.16)	(0.48)	(0.53)	(1.71)	(16.96)**	(18.34)**	(17.07)**
Longitude (degrees)	0.90	0.93	0.50	0.88	0.84	1.29	-0.29	-0.23	-0.23
	(3.77)**	(3.69)**	(1.46)	(2.86)**	(2.67)**	(3.60)**	(1.12)	(0.89)	(0.81)
Elevation (feet)	0.002	0.002	0.002	0.000	0.000	0.000	0.001	0.001	0.001
	(1.91)	(1.88)	(1.62)	(0.05)	(0.15)	(0.11)	(2.09)*	(2.44)*	(2.50)*
ln(CT Median Income)		0.027	0.101		-0.023	-0.015		0.031	0.021
		(0.79)	(3.14)**		(1.01)	(0.55)		(3.30)**	(2.20)*
ln(CT Population)		0.008	-0.003		0.022	0.008		0.005	0.003
		(0.38)	(0.16)		(0.98)	(0.38)		(0.97)	(0.53)
Average CT FAR			0.142			-0.025			-0.009
			(3.15)**			(1.89)			(1.77)
# Subway stops w/in .5 mile			0.022			0.01			0.005
			(2.83)**			(2.92)**			(3.63)**
White (% of CT pop.)			0.001			-0.0005			-0.0002899
			(1.42)			(0.69)			(1.17)
Black (% of CT pop.)			0.003			-0.002			0.0002
			(6.15)**			(3.00)**			(0.90)
Hispanic (% of CT pop.)			0.001			0.001			-0.001
			(1.48)			(0.93)			(3.18)**
Manhattan Dummy	0.20	0.18	-0.28	-2.59	-2.60	-2.58	0.18	0.16	0.14
	(0.74)	(0.67)	(0.79)	(4.31)**	(4.32)**	(4.28)**	(2.26)*	(2.03)*	(1.74)
Bronx Dummy	-0.21	-0.20	-0.30	-2.33	-2.34	-2.36	0.20	0.21	0.20
	(1.75)	(1.65)	(2.43)*	(3.87)**	(3.88)**	(3.92)**	(2.58)*	(2.83)**	(2.67)**
Brooklyn Dummy	-0.18	-0.18	-0.22	-2.22	-2.23	-2.26	0.23	0.24	0.22
	(3.06)**	(3.04)**	(3.19)**	(3.75)**	(3.77)**	(3.81)**	(3.17)**	(3.25)**	(2.94)**
Queens Dummy	-0.26	-0.26	-0.19	-2.16	-2.16	-2.27	0.11	0.10	0.086
	(3.31)**	(3.35)**	(2.13)*	(3.65)**	(3.66)**	(3.84)**	(1.24)	(1.18)	(0.95)
Constant	85.5	89.3	32.6	75.1	72.4	122.4	106.2	112.1	108.8
	(2.62)**	(2.58)**	(0.83)	(2.17)*	(2.09)*	(3.36)**	(4.39)**	(4.72)**	(4.13)**
R^2	0.12	0.12	0.19	0.61	0.61	0.64	0.93	0.93	0.93
# obs.	5,248	5,248	5,248	1,553	1,553	1,553	431	427	427

Note: Omitted borough is Staten Island. Standard errors were clustered by census tracts. t-statistics are given below coefficient estimates. **Stat. sig. at 99%. *Stat. sig. at 95%.

Table 9: p-values for LWR Hypothesis Tests for "Straddle" and "Both After"

Test	Straddle	Both After
Homes		
OLS Regression	**0.011**	0.883
Significance Test for Bandwidth	**0.000**	**0.000**
Significance Tests for Non-Stationarity	**0.000**	0.700
# obs.	5248	1897
Apartments		
OLS Regression	**0.075**	0.183
Significance Test for Bandwidth	**0.000**	**0.000**
Significance Tests for Non-Stationarity	**0.000**	0.375
# obs.	1553	185
Commercial		
OLS Regression	0.280	0.557
Significance Test for Bandwidth	**0.000**	0.250
Significance Tests for Non-Stationarity	**0.000**	0.852
# obs.	431	66

Note: Bold indicates p-values<0.10 (and hence rejection of null hypothesis at greater than 90% level. Also note that Bandwidth and Non-Stationarity tests are based on Monte Carlo simulations. Rep numbers for these MC simulations are as follows Homes-Straddle: 8, Homes-After: 10, Apartments-Straddle & Apartments After: 8, Commercial-Straddle & Commercial After: 20.

Appendices: A, B, and C

Appendix A: Data and Regression Results.

A.1. The Data Set

The full processed data set that we used for our analysis has 326,122 real estate sales between January 2003 and October 2014 throughout the entire city of New York. The data set has nearly every type of building, including one and two family homes, offices, factories, apartment buildings, etc. In short, it has both residential and commercial properties. The source of the real estate transactions is the New York City Department of Finance (DoF) website, http://www1.nyc.gov/site/finance/taxes/property-annualized-sales-update.page. Data include sales price, date of sale, building address, some details about the property, including the square foot of building, the square foot of lot, year constructed, and the type of property. Note that this data is only for building sales and excludes condo or coop unit sales.

Table A.1 lists the different property types and categories that were used in this paper. Building types information can be found in the NYC PLUTO data dictionary, https://www1.nyc.gov/assets/planning/download/pdf/data-maps/open-data/pluto_datadictionary.pdf?r=16v2.

{Table A.1 here—Building Types and Frequencies}

The data downloaded from the DoF includes all transfer of title and does not distinguish open market transactions from the rest. As a result, we were required to process the data and make some assumptions in order to create a data set that seemed to include only open market sales. First, we deleted observations that had no data for lot size, building size, year of construction or were sold for less than $100. Then we generated the price per square foot for each property. In the hedonic regressions (Tables 1-6) we excluded from the regressions those properties that had price per square foot in the bottom one and top one percentiles, respectively, to avoid the influence of outliers.

Each lot in New York City is assigned a unique borough, block and lot (BBL) number. Using the BBL we then merged the sales data with the 2014 Primary Land Use Tax Lot Output (PLUTO) file, which details property characteristics for every BBL in the city, some of which overlap with the salse data and some do not. The PLUTO file includes additional information, such as the property latitude and longitude, and the number of floors of the building. The PLUTO file is available at https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page. For the hedonic regressions, we adjusted the sales price by the NYC CPI without shelter (CUURA101SASL2RS), with the first quarter of 2014 as the base period.

Sandy flooding maps and FEMA insurance flood maps were generously provided by the National Resources Defense Council (NRDC). They provided us with GIS shape files that indicated the locations in the city of the surge flood, and the locations of FEMA floodplain. We used the same files as shown in Figure 1 of their report on Sanday, at https://www.nrdc.org/sites/default/files/hurricane-sandy-coastal-flooding-report.pdf.

Using this information, we then created our Sandy-related variables, which include the distance to the flood zone boundary for all properties, the distance to the shoreline, and the distance to the FEMA floodplain boundary. For flooded properties, we ascertained whether the building was in the FEMA floodplain map that was in effect in 2012. We also used the NRDC shape file to ascertain the distance of each property to the closest shoreline. In addition, we obtained the elevation of each property from the New York City Digital Evolution model, at http://opengeometadata.stanford.edu/metadata/org.opengeoportal/Columbia:Columbia.usgs_nyc1999_1m/fgdc.html. The depth of the surge across the city was ascertained from the file “NYC_Feb14Final1mSurgeDataClipped.zip,” available at https://www.arcgis.com/home/item.html?id=307dd522499d4a44a33d7296a5da5ea0.

Table A.2 gives the descriptive statistics for the data set.

{Table A.2 about here Desc. Stats}

Table A.3 gives regression results for one and two family homes, apartment buildings and commercial properties respectively. For the sake of brevity, we only include one specification for each property type, which is also Equation (5) in Tables 2-4. However, here, we do not restrict the sample period or distance to the shore in Table A.3. Other specifications are available upon request.

Also note that the regressions for the price index (Figure 1) regressed the log of the real price per square foot on the following controls: distance to the Empire State Building, number of floors, number of units, number of buildings on property, ln(Land Area), ln(Building Area), year built, year-quarterly dummies, building class dummies and zip code dummies. Standard errors were clustered at the zip code level. We ran two regression one for only the dry properties and the other for the properties that experienced flooding. In the second equation, we also included an inverse Mill ratio that was calculated from a probit that estimated the probability that a property would be in the surge area. That is the dependent variable was equal to 1 if it was flooded or in the flood zone (prior to flooding), 0 otherwise. Controls included year-quarterly dummies, building elevation, building elevation *x* post-Sandy dummy, the distance to the shoreline, and the distance to shoreline *x* post-Sandy dummy. Results available upon request.

For the index, the base year was 2003Q1. We created the index by taking the exponent of the coefficients for the year-quarterly dummy variables and multiplying them by100, i.e. $index\ value_t = 100 * exp(\hat{\beta}_t)$, where $\hat{\beta}_t$, is the coefficient estimate for year-quarter t. Note $\hat{\beta}_{2002Q1} = 0$ because it was omitted from the regression. Thus 2003Q1=100.

{A.3. Regression Tables for Data Set}

A.2. Repeat Sales and Fourier Price Index

The master data set described above has a significant number of repeat sales. In particular, for each property type (one and two family homes, apartment buildings, and commercial properties), we created new data sets that contained each of these repeat sales. We excluded any property if we observed a change in square footage, construction year, building class type on the assumption that the nature of the property changed over time.

From these remaining properties, we generated sales data sets where for each properties with at least two sales, we included those properties that had at least one sale before the storm and one sale after (i.e., has repeat sales that straddled the storm). These straddle data sets needed to be further processed. First, to simplify the analysis we decided to include only pairs of repeat sales. The included pairs where cases where the first sale occurred closest to before the storm date and the second sale occurred closest to the storm after the date. To mitigate against included property "flips" (i.e., those repeat sales that bought and sold for speculation and may have included renovations), we excluded pairs if the second sale took place within 30 days of the first. Lastly, we excluded those sales where the log of the price changes were either in the top or bottom one percentiles, respectively, within each property category. Table A.4. Presents the descriptive statistics for the difference of the key variables included in the locally weighted regressions.

{Table. A.4. about here: Desc. Stats for Straddle Repeats}

McMillen and Dombrow (2001) obtain the parametric version of the Fourier repeat sales estimator by first estimating the following equation:

$$ln\left(\frac{P_{nt+j}}{P_{n,t}}\right) = \varphi_1(z_{t+j} - z_t) + \varphi_2(z_{t+j}^2 - z_t^2) + \sum_\rho \left[\sigma_\rho\left(sin(\rho z_{t+j}) - sin(\rho z_t)\right) + \delta_\rho\left(cos(\rho z_{t+j}) - cos(\rho z_t)\right)\right] + (\mu_{nt+j} - \mu_{nt}), \quad (6)$$

where ρ is the number of lags, $z_t = 2\pi T_t / max(T)$, and T_t represents the numerical day in the sample at time t.[8] After using least squares regressions to estimate the parameters ϕ_1 , ϕ_2, σ_ρ, and δ_ρ, one then calculates the fitted values of the following equation at various time points to obtain the price index:

$$(\hat{\alpha}_{nt+j} - \hat{\alpha}_{nt}) = \hat{\varphi}_1(z_{t+j} - z_t) + \hat{\varphi}_2(z_{t+j}^2 - z_t^2) + \sum_\rho \left[\hat{\sigma}_\rho \left(sin(\rho z_{t+j}) - sin(\rho z_t)\right) + \hat{\delta}_\rho \left(cos(\rho z_{t+j}) - cos(\rho z_t)\right)\right] \qquad (7)$$

The lag length (ρ) is determined through minimization of the Schwarz information criterion (SIC).[9]

A.3. Probits for Inverse Mills Ratio for Repeat Sales

As described in Section 3, we first ran probit regressions (for the dry properties only) to estimate the probability of repeat sales occurring. The dependent variable takes on the value of 1 if a second or third sales takes place for a particular property in the data set, 0 otherwise. Control variables included the census tract occupancy rate, the distance to the Empire State Building, the number of floors, ln(Land Area), ln(Building Area), year built, distance to shore *x* pre-Sandy dummy, distance to shore *x* post-Sandy dummy, FEMA dummy *x* pre-Sandy dummy, FEMA dummy *x* post-Sandy dummy, elevation *x* pre-Sandy dummy, elevation *x* post-Sandy dummy, distance to the surge *x* outside surge dummy *x* post-Sandy dummy, year-quarterly dummies and building class dummies. Table A.5 gives the results.

{Table A.5 about here: Probits for Repeat Sales}

Appendix B: Locally Weighted Regressions

B.1: LWR Methodology

"Locally weighted regression" (LWR) is the name of the general procedure where d_{ik} can be any distance measure (not just geographic. Geographically weighted regression (GWR) denotes that d_{ik} is a function of geographic distances, which here is the Cartesian distance between two latitude-longitude points. We implement LWR using the "spgwr" package in R and the "gwr" package in Stata.

In this paper, we use locally weighted regressions, which is a version of weighted least squares, as discussed in Cleveland and Devlin (1988). Implementation of the model gives an estimated parameter for each target observation (i.e., building):

$$\hat{\beta}_i = (\sum w_{ik} X_k Y_k)\left(\sum w_{ik} X_k X'_{jk}\right)^{-1},$$

[8] As McMillen and Dombrow (2001) note, this essentially lines up the dates in the sample, in our case starting at January 1, 2003 as *t=1*, January 2, 2003 as *t=2*, etc., and rescales the time variable on the interval between *0* and *2π*.
[9] x_i should be adapted accordingly for the situation where $\rho>1$. In our application, $\rho=1$ minimizes the SIC.

where X_k is a vector of control variables including the constant for each observation except i; Y_k is the dependent variable for all observations except i; w_{ik} is the weight that building k is given for building i; and the summations given by $\sum$ are taken over all buildings, k, and w_{ii} =0.

We use a Gaussian (standard normal) weighting function (kernel) given by

$$w_{ik} = \boldsymbol{K}\left(\frac{d_{ik}}{b}\right) = \boldsymbol{e}^{-\left(\frac{\boldsymbol{d_{ik}}}{\boldsymbol{b}}\right)^{\boldsymbol{2}}} ,$$

where d_{ik} is the Euclidian distance between building i and j (as measured in degrees latitude and longitude).

McMillen and Redfearn (2010) notes that the choice of the kernel has little effect on the results since most kernel choices have rapid decay with distance. $b>0$ is the bandwidth parameter. The bandwidth parameter determines the "variance" of the weights. A larger b means that, *ceteris paribus*, observations further away will have larger weight values, compared to a smaller value of b.

For the LWRs, the bandwidth value was selected using the standard cross-validation (C-V) method. The C-V algorithm runs a LWR for each observation for a given bandwidth value. Then a statistic is generated that is the mean squared residual of the LWR, where the residual is the difference between an observed value of the dependent variable and the predicted value, after omitting the ith observation from the model. The bandwidth that minimizes this statistic is used. See McMillen and McDonald (2004) for more information. As an example, if two properties are one mile apart, then d_{ik} is about 0.185. If, say, $b=0.02$, then the weight is about *0.424*. Two properties that are 0.1 miles apart means that $d_{ik}=0.00146$, so then $w_{ik}=0.995$.

For each estimated coefficient, a standard error is also produced, given by equation 2.21 in Fotheringham et al. (2002). t-statistics are generated by taking the coefficient estimates divided by the standard errors.

To tests hypotheses about the coefficients we use the tests from the 'gwr' package in Stata. See: https://www.staff.ncl.ac.uk/m.s.pearce/stbgwr.htm Two tests are performed. The first is the significance test of the bandwidth, which tests if the locally weighted regression model is a significantly better model than the OLS regression model. Second is the significance tests for non-stationarity, which tests if the LWR coefficients for a particular independent variables are the same or not. If the null hypothesis is rejected, it suggests that LWR is better able to fit the data than OLS. Note that if the non-stationarity test does not reject the null hypothesis and an OLS regression also does not reject the null hypothesis for a particular right hand side variable, it suggests that variable has no explanatory power with respect to the dependent variable.

Appendix B.2. Additional Results for Straddle Data Set

Table B.1 gives OLS results for the repeat sales data sets—the same ones used for the LWRs. The dependent variables is $\Delta lnp_i - \Delta\alpha_i$; that is, the change in sale price minus the change in the price index. The independent variables are Sandy-FEMA, change in census tract occupancy rate, change in the inverse Mills ratio, latitude, longitude, number of days between sales, and the year-quarter of the second sale.

{Table B.1 about here: OLR on straddle data set}

Table B.2. gives the bandwidths used for the LWRs, based on the cross-validation method.

{Table B.2 about here: Bandwidths}

Figure B.1 gives the histograms of the LWR coefficients for each property type.

{Figure B.1 about here: Histograms of LWRs coefficients}

Appendix C: Census Tract Level Data Sources

For each property included in the repeat sales LWRs that straddle the storm, we also have the respective 2010 census tract for that property. We thus merged the LWR coefficient estimates with census tract level data. Table C.1 gives descriptive statistics for this data set. Race and ethnicity is from the 2010 Census File DEC_10_SPF_P11_with_ann. Median income is from ACS_10_5YR_S1903_with_ann. Average Built FAR: NYC PLUTO FILE, 2016. Subway stops: https://www.baruch.cuny.edu/confluence/display/geoportal/NYC+Mass+Transit+Spatial+Layers. (Note also include Staten Island rail stops).

{Table C.1 about here—Desc. Stats for variables for LWR regressions}

Appendix Tables and Figures

Table A.1: Building Types for each Class. Source: NYC DOF Sales File.

Class #	Building Class	Nobs.
	One and Two Family Homes	
1	ONE FAMILY HOME	146,098
2	TWO FAMILY HOME	128,381
	Apartments	
7	RENTALS - WALKUP APARTMENT	26,127
8	RENTALS - ELEVEVATOR APARTMENTS	3,265
	Commercial Buildings	
21	OFFICE BUILDINGS	2,295
22	STORE BUILDINGS	5,806
23	LOFT BUILDINGS	561
27	FACTORIES	1,985
29	COMMERCIAL GARAGES	2,691
30	WAREHOUSES	1,942
32	HOSPITAL AND HEALTH FACILITIES	285
33	EDUCATIONAL FACILITIES	196
34	THEATRES	48
35	INDOOR PUBLIC AND CULTURAL FACILI	185
38	ASYLUMS AND HOMES	71
39	TRANSPORTATION FACILITIES	8

Table A.2: Descriptive Statistics for Data Set

Variable	Mean	Std. Dev.	Min.	Max.	# Obs.
Building & Neighborhood Variables					
PPSF	240.02	138.47	0.42	1006.51	371,992
Real Price Per Square Foot (2014Q4)	273.28	154.58	0.42	1297.92	371,992
Dist. to the Empire State Bldg (miles)	9.18	3.68	0.003	21.98	371,992
# Floors	2.81	3.65	1	114	371,322
# Units	3.6	31.32	0	8800	371,992
# Buildings	1	3.40	0	1929	371,322
Lot Area (sq. ft.)	3,840	351667	0	1.23E+07	371,322
Building Area (sq. ft.)	6,704	38532	0	8942176	371,322
Year Built	1,942	32.55	1800.00	2014	371,992
Occupancy Rate by Census Tract (%)	95.6	6.19	18.9	100	236,051
1 & 2 Family Homes Dummy	0.738				371,992

Commercial Dummy	0.043				371,992
Apartments Dummy	0.079				371,992
Manhattan Dummy	0.048				371,992
Bronx Dummy	0.108				371,992
Brooklyn Dummy	0.313				371,992
Queens Dummy	0.404				371,992
Staten Island Dunny	0.127				371,992
Sandy-Related Variables					
Elevation (feet)	16.3	11.98	-0.47	113.4	371,992
Dist. To Shoreline (miles)	1.25	0.90	0.00	4.13	371,992
Surge Height if Flooded	3.25	2.26	0.00	13.33	5,119
Inundated Dummy x Post Sandy Dummy	0.008				371,992
Post Sandy Dummy	0.132				371,992
FEMA floodplain Dummy x Post Sandy Dummy	0.0004				371,992
Distance to FEMA Floodplain (after Sandy)	0.821	0.653	-0.072	2.99	79,245
Distance to Surge boundary (after Sandy)	0.743	0.625	-0.903	3.11	79,245
Sandy Disantce - FEMA distance (miles; dry properties after Sandy)	-0.071	0.309	-1.11	1.76	74,122

Table A.3: Full OLS regressions. Dep. Var.: ln(Reap Price per Square Foot)

	(1)	(2)	(3)	(4)	(5)	(6)
	1&2 Family Homes	**Apart-ments**	**Commercial**	**1&2 Family Homes**	**Apart-ments**	**Commercial**
	All	**All**	**All**	**Dry only**	**Dry only**	**Dry only**
Occupancy Rate of Census Tract (%)	0.21	0.015	-0.223	0.228	0.066	-0.282
	(4.20)**	(0.11)	(1.30)	(4.55)**	(0.51)	(1.71)
Dist. Empire State Bldg. (miles)	0.022	-0.048	-0.031	0.028	-0.048	-0.032
	(1.48)	(2.04)*	(1.15)	(1.68)	(2.11)*	(1.12)
# of Floors	0.016	0.007	0.001	0.017	0.009	0.001
	(1.98)*	(0.70)	(0.23)	(2.03)*	(0.86)	(0.21)
Total Units	-0.02	0.000	-0.001	-0.019	0.000	-0.001
	(4.66)**	(2.59)*	(2.19)*	(4.62)**	(2.23)*	(2.12)*
Number of Buildings	0.014	-0.064	-0.055	0.014	-0.006	-0.061
	(2.92)**	(2.09)*	(2.91)**	(3.61)**	(0.40)	(3.01)**
ln(Land Area) (sq. ft.)	0.248	0.225	0.283	0.239	0.184	0.281
	(20.33)**	(5.44)**	(10.31)**	(20.19)**	(4.53)**	(9.79)**
ln(Building Area) (sq. ft.)	-0.683	-0.399	-0.465	-0.694	-0.373	-0.461
	(51.75)**	(10.57)**	(18.04)**	(51.79)**	(9.87)**	(17.52)**
Year Built	0.002	0.001	0.000	0.002	0.001	0.000
	(12.15)**	(1.72)	(0.70)	(11.96)**	(1.00)	(0.25)

Dist. To Shore (miles) x Pre-Sandy Dummy	0.021	0.129	0.08	0.014	-0.01	0.07
	(0.60)	(2.34)*	(2.08)*	(0.78)	(0.28)	(1.81)
Dist. To Shore (miles) x Post-Sandy Dummy	-0.001	-0.007	0.003	-0.008	-0.065	0.02
	(0.03)	(0.11)	(0.05)	(0.28)	(0.77)	(0.34)
FEMA Zone Dummy x Pre-Sandy Dummy	0.07	-0.04	-0.009	0.046	-0.13	0.035
	(1.43)	(0.19)	(0.05)	(1.01)	(0.60)	(0.20)
FEMA Zone Dummy x Post-Sandy Dummy	-0.028	-0.037	-0.284	-0.07	-0.146	-0.231
	(0.51)	(0.19)	(1.44)	(1.40)	(0.73)	(1.16)
Elevation (feet) x Pre-Sandy Dummy	0.004	0.163	0.003	0.002	-0.001	0.003
	(0.12)	(2.51)*	(1.61)	(2.57)*	(0.49)	(1.51)
Elevation (feet) x Post-Sandy Dummy	0.006	0.171	0.001	0.005	0.004	0.001
	(0.16)	(2.59)*	(0.33)	(4.03)**	(1.94)	(0.48)
Dist. to Surge Boundary (miles) x Post-Sandy Dummy	0.015	0.116	0.092			
	(0.59)	(1.49)	(1.41)			
Inundated Dummy x Post-Sandy Dummy	-0.002	0.284	0.202			
	(0.06)	(2.37)*	(1.31)			
Surge height (feet) x Post-Sandy Dummy	-0.034	-0.1	-0.115			
	(2.46)*	(2.70)**	(2.75)**			
Inverse Mills Ratio	-0.003	-0.186	0.128			
	(0.06)	(2.54)*	(1.08)			
Sandy Dist. - FEMA Dist. (miles) x Post-Sandy				0.15	0.274	0.206
				(3.19)**	(2.17)*	(1.88)
Dist. to Surg. x Post-Sandy for MN				0.272	0.266	0.365
				(1.20)	(1.21)	(1.41)
Dist. to Surg. x Post-Sandy for BK				0.051	0.106	0.144
				(2.08)*	(1.25)	(1.93)
Dist. to Surg. x Post-Sandy for BX				-0.265	-0.253	-0.158
				(6.58)**	(1.96)	(1.38)
Dist. to Surge x Post-Sandy for QN				-0.012	-0.03	0.018
				(0.46)	(0.29)	(0.24)
Dist. to Surge x Post-Sandy for SI				-0.139	-0.557	-0.249
				(2.87)**	(2.01)*	(1.87)
Constant	3.951	5.427	7.216	4.251	6.362	7.639
	(8.25)**	(5.63)**	(7.97)**	(8.83)**	(6.42)**	(8.06)**
R^2	0.23	0.3	0.27	0.23	0.3	0.27
N	169,128	18,470	10,428	159,235	18,770	9,766

t-stats. below estimates (clustered by zip code). Note year-quarterly dummies and building class dummies included. F-stats for dummy groups are all statistically significant. **Statistically significant at 99% level; *Statistically significant at 95% level.

Table A.4: Descriptive Statistics for Straddle Repeats Data Set

Variable	Mean	Std. Dev.	Min.	Max.	Nobs.
One & Two Family Homes					
Δlnprice	0.148	0.571	-1.797	2.262	5,248
Δα	0.855	0.525	-0.093	2.306	5,248
Δlnprice - Δα	-0.707	0.808	-3.798	2.001	5,248
ΔOccupancy Rate	-0.018	0.074	-0.803	0.466	5,248
ΔInv. Mills Ratio	-0.070	0.112	-0.589	0.275	5,248
Surge - FEMA	-0.102	0.353	-1.108	1.704	5,248
Apartments					
Δlnprice	0.329	0.655	-2.911	3.258	1,553
Δα	0.473	0.343	-0.202	1.976	1,553
Δlnprice - Δα	-0.144	0.694	-4.098	2.748	1,553
ΔOccupancy Rate	-0.026	0.086	-0.760	0.086	1,553
ΔInv. Mills Ratio	-0.250	0.162	-0.728	0.199	1,553
Surge - FEMA	-0.064	0.318	-1.106	1.634	1,553
Commercial					
Δlnprice	0.294	0.704	-2.579	2.940	431
Δα	0.323	0.338	-0.762	1.351	431
Δlnprice - Δα	-0.029	0.769	-2.649	2.928	431
ΔOccupancy Rate	-0.038	0.094	-0.659	0.185	431
ΔInv. Mills Ratio	-0.269	0.371	-5.353	0.058	431
Surge - FEMA	-0.087	0.297	-1.107	1.478	431

Table A.5: Probits for Repeats Sales. Dep. Var.=1 if sale is a repeat, 0 otherwise.

	1 & 2 Family Homes	Apartments	Commercial
Occupancy Rate of Census Tract (%)	-0.64	-0.602	-0.233
	(2.92)**	(2.47)*	(1.24)
Dist. Empire State Bldg. (miles)	0.018	0.027	0.009
	(2.18)*	(2.19)*	(1.47)
# of Floors	0.052	0.002	-0.003
	(2.75)**	(0.19)	(0.51)
ln(Land Area) (sq. ft.)	-0.361	-0.189	-0.167
	(10.29)**	(3.86)**	(5.97)**
ln(Building Area) (sq. ft.)	-0.14	0.106	0.126
	(2.96)**	(2.13)*	(5.00)**
Year Built	-0.002	0	0.003
	(2.93)**	(0.05)	(4.22)**

Dist. To Shore (miles) x Post-Sandy Dummy	0.125	-0.075	-0.096
	(2.89)**	(1.18)	(1.36)
Dist. To Shore (miles) x Pre-Sandy Dummy	0.118	0.083	0.001
	(3.94)**	(2.25)*	(0.04)
FEMA Zone Dummy x Post-Sandy Dummy	0.004	-0.154	-0.102
	(0.03)	(0.46)	(0.34)
FEMA Zone Dummy x Pre-Sandy Dummy	-0.146	-0.468	-0.279
	(1.54)	(2.01)*	(1.34)
Elevation (feet) x Pre-Sandy Dummy	-0.006	0.001	-0.001
	(2.56)*	(0.34)	(0.48)
Elevation (feet) x Post-Sandy Dummy	-0.003	0.005	-0.001
	(1.50)	(1.49)	(0.35)
Dist. to Surge Boundary (miles) x Post-Sandy Dummy	-0.062	0.124	0.151
	(1.07)	(1.39)	(1.49)
Constant	7.152	0.366	-5.749
	(5.63)**	(0.24)	(4.16)**
# obs.	158,336	18,578	9,578
Pseudo R2	0.044	0.031	0.049

z-stats. below estimates (clustered by zip code). Note year-quarterly dummies and building class dummies included. F-stats for dummy groups are all statistically significant. **Statistically significant at 99% level; *Statistically significant at 95% level.

Table B.1: OLS Regressions for Straddle Data Sets. Dep. Var: ∆lnprice - ∆α

	1 & 2 Family Homes	Apartments	Commercial
Surge - FEMA	0.327	0.136	-0.01
	(7.18)**	(1.85)	(0.08)
∆Occupancy Rate	0.708	0.113	-0.007
	(5.89)**	(0.59)	(0.01)
∆Inv. Mills Ratio	-1.94	-0.069	-0.196
	(13.55)**	(0.30)	(1.72)
# days between sales	-0.001	-0.0002	-0.0002
	(15.49)**	(4.46)**	(4.78)**
Year-Quarter	-0.636	-0.137	-0.271
	(28.33)**	(4.38)**	(4.29)**
Latitude (degrees)	-0.378	1.93	-0.556
	(1.47)	(5.68)**	(1.04)
Longitude (degrees)	-0.683	-0.482	-1.39
	(3.88)**	(1.01)	(2.57)*
Constant	1,244.8	162.6	466.4
	(23.92)**	(2.17)*	(3.06)**
R^2	0.56	0.15	0.14
# obs.	5,248	1,553	431

t-stats. below estimates (clustered by zip code). **Statistically significant at 99% level; *Statistically significant at 95% level.

Table B.2 Table of bandwidths.

	Bandwidth	
	Straddle	Both After
Homes	0.0186	0.0423
Apartments	0.0369	0.4019
Commercial	0.0996	0.1415

Note: The “straddle” column is for the sales that straddled Hurricane Sandy. The "Both After" are for the repeat sales that both occurred after the storm. The bandwidths were selected using the CV-method described above.

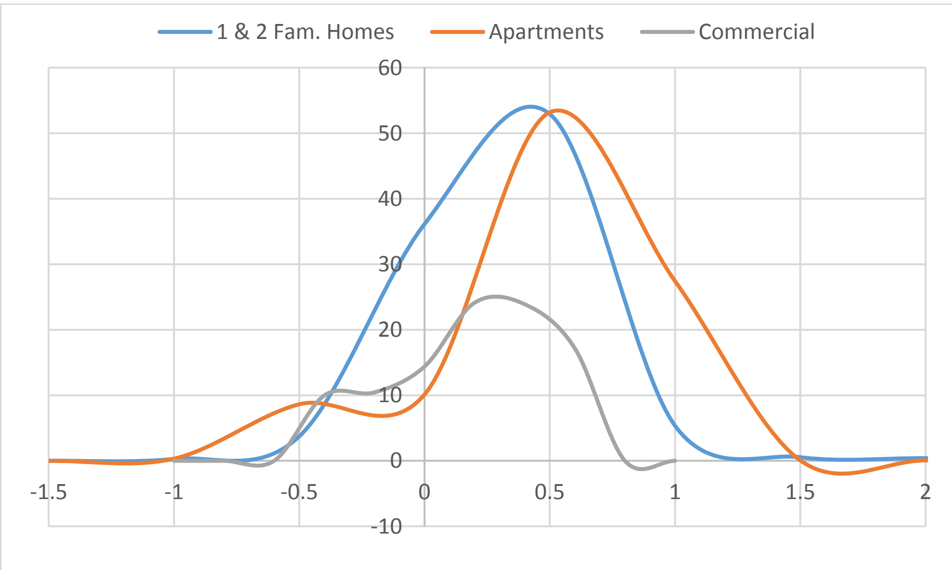

Figure B.2: Relative Frequency Histograms (%) of LWR Coefficients.

Table C.1 Descriptive Statistics for Variables used in LWR Coefficient Regressions

Variable	Mean	St. Dev.	Min.	Max.	# obs.
CT Population	3770.8	46.8	0	26588	2168
CT Hispanic (%)	26.5	22.3	0	96.3	2137
CT White (%)	33.5	30.9	0	100	2137
CT Black (%)	24.5	29.7	0	100	2137
CT Median Household Income	53447	27386	0	250000	2165
CT Avg. Built FAR	1.44	1.38	0.29	14.8	1686
# Subway Stops w/in half mile	1.79	2.27	0	19	7232

Note: CT=Census Tract.