Market Response to Flood Risk: An Empirical Study of Housing Values Using Boundary Discontinuities

Wang, Haoying

New Mexico Tech

15 April 2017

Market Response to Flood Risk: An Empirical Study of Housing Values Using Boundary Discontinuities

Haoying Wang

(Haoying.Wang@nmt.edu)

Department of Business and Technology Management

New Mexico Tech

Abstract:

This paper presents one of the first studies on flood risk evaluation in the US Northeast - a region where we are likely going to see increasing precipitation variability and associated risk of flood in the coming decades. In the paper, a spatial difference-in-differences framework based on floodplain boundary discontinuities is proposed to control for unobserved heterogeneities. Using parcel level data from Juniata County and Perry County in Pennsylvania, the paper finds that on average there is a 5-6 percent housing value reduction due to exposure to 1 percent annual chance of flooding within the FEMA (Federal Emergency Management Agency) 100-year flood zone. For Juniata County, it shows that on average there is a \$3.28/sqft (in 2015 USD) discount for a full-time SFR (single family residential) property located within the flood zone. For an average housing unit of 1430 sqft living space in the sample, the estimate translates to a \$4690 housing value reduction. For Perry County, the corresponding estimates are \$4.00/sqft (in 2015 USD) and \$6320 for an average housing unit of 1580 sqft. The paper also shows that with similar specifications, a standard hedonic price model underestimates the flood risk impact on housing value by a substantial amount as a result of failing to control for unobserved heterogeneities.

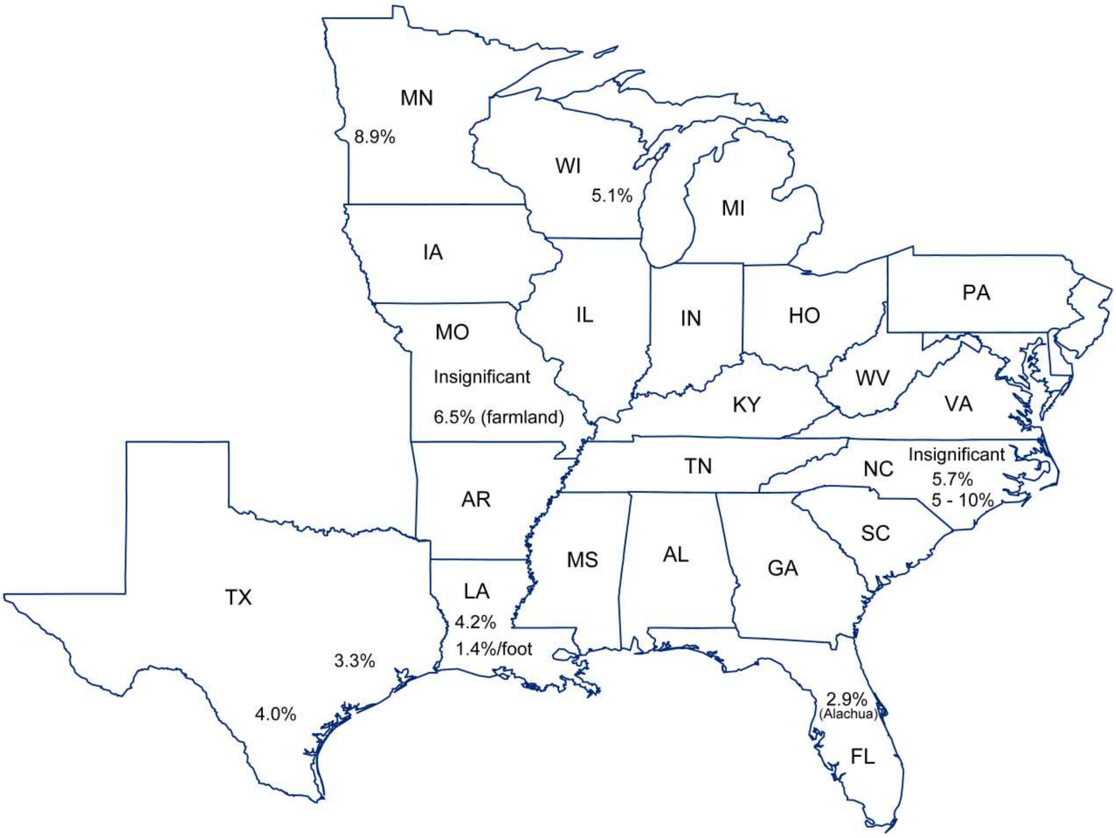
Keywords:

Flood Risk, Natural Hazards, Housing Value, Housing Market, Boundary Discontinuity, Susquehanna River Basin

1 Introduction

The ongoing and expected climate change likely alters the precipitation pattern and runoff in many regions around the world (Labat et al., 2004; Trenberth, 2011). In the US Northeast, for example, the changing climate has been found correlated with less winter precipitation falling as snow and more as rain, as well as earlier spring snowmelt resulting in earlier peak river flows (Frumhoff et al., 2007). One of the consequences proceeds from the growing precipitation variability is the increasing risk of flood, which poses serious threat to economic development and human lives (McMichael et al., 2006; Jongman et al., 2012; Kousky, 2014). According to Jongman et al. (2012), the total global exposure to river and coastal flood risks is 46 trillion USD in 2010, which could rise to a projected 158 trillion USD by 2050. In many states throughout the US, flooding is the lead cause of death among all types of natural disasters. Both mitigation and adaptation strategies are developed towards such potential risk, and one common local strategy is floodplain management and land use planning. To ensure that policy tools play an effective role in managing flood risk, evaluating market response to flood risk becomes a key component in policymaking process for many local governments. Such valuation is of fundamental importance because policy tools often need to provide right incentives for desired behavioral changes as a way to enhance community resilience.

In the literature, hedonic valuation is commonly used to estimate and derive the flood risk premium. By evaluating the relationship between flood zone status or inundation depth and its resulting capitalization in property values, household willingness to pay for avoiding flood risk can be derived (Niskanen and Hanke, 1977). Results from existing studies vary spatially given the nature of their study regions. The results can be simply grouped into two categories: coastal flood risk and inland (river) flood risk. Figure 1 shows estimates from some of the representative studies (by no means an exhaustive list) across eastern, northern, and southern US. These valuations are all based on residential housing market sales without intervention from major flood events (e.g. hurricanes). In other words, these estimates reflect a normal valuation of flood zone status. Flood risk valuations after a major flood event tend to be substantially higher due to the unexpected catastrophic effect from the exogenous shock and rising flood risk perceptions (Hallstrom and Smith, 2005; Carbone et al., 2006; Morgan, 2007; McKenzie and Levendis, 2010; Bin and Landry, 2013). As Bin and Landry (2013) find, however, this large effect is diminishing over time as affected communities recover from the natural disasters.



Data source: compiled from Struyk (1971), Skantz and Strickland (1987), Speyrer and Ragas (1991), Bartosova et al. (2000), Harrison et al. (2001), Shultz and Fridgen (2001), Bin and Polasky (2004), Bin and Kruse (2006), Kousky (2010), McKenzie and Levendis (2010), Zhang et al. (2010), Bin and Landry (2013).

Figure 1: Flood risk impacts on housing values found in the literature

Another worth mentioning trend from the cross-section of estimates in Figure 1 is that, the flood risk valuations are relatively smaller in southern coastal areas (TX, LA, and FL) comparing to eastern central and northern states (NC, MO, MN, and WI). There are two potential explanations for the observation. First, households and housing markets along the Gulf coastal region may have developed higher tolerance to flood risk over years, and more likely had protections and preparedness in place to prevent losses from regular flood events (e.g. Lindell and Prater, 2003). In this regard, institutional resilience may have also been forged over years, which is found to be important in shielding the population from natural-disaster losses (Kahn, 2005). Another explanation is that, along the coastal line, amenity values are often confounded spatially with flood risk. Such a correlated spatial trend in amenities could cause identification issues in empirical estimation and lead to underestimation of flood risk premium (Bin and Kruse, 2006; Carbone et al., 2006). Bin and Kruse (2006) point out that, for example, the coastal flood risk tends to play small to insignificant role in property valuation due to the significant premium from capitalization of the proximity to coastal water and wave action.

Efforts have been devoted in the literature to cope with the identification issue of flood risk impacts caused by unobserved heterogeneity. In many cases these spatial amenity effects cannot be well represented by distance to the shore, for instance due to topography, which complicates the empirical investigation. One solution is to use GIS-based view measures to disentangle the variation of amenities from the distance-based measure of risk exposure (Bin et al., 2008; Hamilton and Morgan, 2010). This approach is subject to the availability of three-dimensional topographic image (e.g. Light Detection and Ranging (LIDAR) Data). Another solution to this problem is to utilize the spatial and temporary variations in market responses to natural events (Carbone et al., 2006). Difference-in-Differences framework is often employed to implement the empirical estimation, where much of the time-invariant spatial heterogeneity (in the sense of relatively short time windows) can be differenced out (Hallstrom and Smith, 2005; Bin and Landry, 2013). When a proxy variable can be interpreted as measuring two locational effects, amenity effect and flood risk in this case, an exogenous treatment or source of information is necessary to distinguish these two effects (Hallstrom and Smith, 2005). In coastal region flood risk studies, disastrous events like typhoons and hurricanes can be used as the exogenous treatment or source of information. Therefore, a temporal difference-in-differences approach works.

For non-coastal regions especially areas near inland water bodies (rivers and lakes), however, natural events like typhoons and hurricanes are less frequently to be observed. In some of these regions, the increase of flood risk happens gradually. The water level of Devils Lake in North Dakota (US), for example, has risen about 10 meters in last two decades (Zheng et al., 2014). Since the 10 meter water level rise did not happen in a very short time window, and two decades of time gives a span enough for many time-invariant locational factors to change, a temporal difference-in-differences framework no longer works. In cases where temporal variation is not sufficient to eliminate unobserved heterogeneities, spatial variation can still be utilized. Based on the similar idea, spatial discontinuities have been used in other literature to deal with biases due to unobserved heterogeneities. Many of the school quality valuation studies, for example, have used boundary discontinuities along school district boundaries to difference out unobserved heterogeneities (e.g. Black, 1999; Gibbons et al., 2013).

In this paper, the researcher applies the rationale of boundary discontinuity to flood risk valuation for a non-coastal area along inland water body. The idea is to use floodplain boundary as cut-off to difference the data among close-neighboring properties to eliminate any location-specific unobserved heterogeneities which affect housing values. Presumably, along the sides of floodplain boundary, the only major price difference between two properties after controlling for land size and structural differences is due to the disparity in flood zone status. In some cases, the floodplain boundary could coincide with major roads or municipal boundaries. Nevertheless, this is less likely a concern due to the fact that floodplain is designed based mainly on topographical factors like soil saturation capacity and land elevation. Note that this empirical strategy only works for identifying the impact of flood risk if there is a sharp discontinuity in its price effects between close-locating properties along the two sides of floodplain boundary. The condition holds because the flood zone status implies different costs to homeowners regarding aspects like flood insurance requirement, local housing regulation, and flood preparedness.

Households living in high risk flood area are mandated by federal law to have flood insurance if the property and buildings have mortgages from federally regulated or insured lenders. The policy does protect both households and lenders from major financial losses in the case of extreme events. The mandate, however, also has significant consequences on insurance affordability and property values over the time (Kousky and Kunreuther, 2014). Especially in recent years, the flood insurance premium has been increasing, often as a response to recent extreme flood events, which leads to an even larger gap in property values. On the top of increasing premium, additional costs related to flood insurance may include getting flood elevation certificate and supplemental coverage for basement contents. Infrastructure is another cost that makes the difference. As part of planning process, projects like storm drains, pumps, runoff ponds, and levees are designed to make many flood-prone residential areas more flood resistant, though not necessarily flood proof. In general, it is unrealistic to expect people and businesses to move out of the flood zone. The additional costs of living in the flood zone, however, have to be accounted into the budget of housing consumption. Its impact on property values is often an empirical question, varying from region to region.

Using parcel level data from Juniata County and Perry County in Pennsylvania, this paper finds that there is a 5-6 percent housing value reduction due to potential exposure to 1 percent annual chance of flooding in the study region. Specifically, for Juniata County, the proposed empirical approach shows that on average there is a \$35.31 discount per square meter of living space (or \$3.28/sqft, in 2015 USD) for a full-time SFR property located within the flood zone. For an average housing unit of 133 m^2 (1430 sqft) living space in the sample, the estimate translates to a \$4690 housing value reduction. For Perry County, the corresponding estimates are \$40.06 per square meter (or \$4.00/sqft, in 2015 USD) and \$6320 for an average housing unit of 147 m^2 (1580 sqft). The paper also shows that with similar specifications, a standard hedonic price model underestimates the flood risk impact on housing value by a substantial amount. This confirms with the literature that correlated flood risk and amenities could lead to underestimation of flood risk premium (Bin and Kruse, 2006; Carbone et al., 2006). It also suggests that the proposed difference-in-differences framework using boundary discontinuities of floodplain can effectively correct potential omitted variables bias due to unobserved heterogeneities.

The paper is organized as follows. Section 2 introduces both an economic model of flood risk valuation and the empirical strategy in details. Section 3 describes study area and data. Section 4 discusses estimation results. Section 5 concludes the paper.

2 Method

2.1 Economic Model

This section introduces an economic model of housing valuation, which includes *N* households. A housing unit is valued as a bundle of three types of attributes: physical attributes (structures), locational attributes, and components observable to the homeowner and local residents but not to the researcher. Flood risk is one of the locational attributes expected to reduce property value. First, following Gayer et al. (2000), the researcher establishes a household risk perception function which depends on two types of risk assessment: objective and subjective. Household objective risk assessment (p) is derived based on public information and knowledge, and it is observable to the researcher. Household subjective risk assessment (q) is formed based on local information and private knowledge, which is unobservable to the researcher but could be well-observed in the neighborhood (e.g. through social interactions). The researcher also introduces two information parameters, $\delta_0 > 0$ and $\lambda_0 > 0$, associated with objective assessment and subjective assessment, respectively. The information parameter measures information content associated with the particular risk assessment.

The household risk perception function is defined as:

$$\pi = \pi(p, q) = \frac{\delta_0 p + \lambda_0 q}{\delta_0 + \lambda_0} = \delta p + \lambda q \tag{1}$$

where $\delta = \delta_0/(\delta_0 + \lambda_0)$, and $\lambda = \lambda_0/(\delta_0 + \lambda_0)$. In this paper, p represents risks that can be measured and well quantified by the 100-year floodplain map, which is public information. q represents risks that are not measured by the floodplain map. The assessment of such risks totally depends on local information and private knowledge. This type of risks may include: landslide, groundwater damage, and other environmental hazards like brownfields. Note that these non-flood related risks may present threats to both households in the floodplain and those outside the floodplain. Failing to account for subjective risk assessment can lead to biased estimates in hedonic valuation of flood risk, in a way similar to the influence of unobserved spatial heterogeneity.

Given household level risk perception, a household maximizes its expected utility over two states of the world: U_1- utility in the risk state; U_2- utility in the no risk state. Before moving forward, similar to Gayer et al. (2000), three assumptions are necessary to establish the household decision making problem: (1) For any given level of income, households prefer being safe, i.e. $U_2 > U_1$; (2) Within each state of the world, households are risk-neutral or risk-averse; (3) Marginal utility of income is higher when there is no risk. Household utility in each state of the world is defined as:

$$U = U(X, Z, S) \tag{2}$$

where X denotes consumption of a composite good with price standardized to 1. Z represents a set of housing characteristics, and S a bundle of locational attributes (amenities and disamenities) which include flood zone status. Housing price (relative to the composite good) is a function of locational attributes, risk perception, and housing characteristics:

$$h = h(\pi, Z, S) \tag{3}$$

Given household income level Y, the household maximizes expected utility by solving the following optimization problem:

$$
\begin{aligned}
\max \quad V &= \pi(p,q)U_1(X,Z,S) + (1-\pi(p,q))U_2(X,Z,S) \\
s.t. \quad Y &= X + h(\pi(p,q),Z,S)
\end{aligned}
\tag{4}
$$

By construction from (1), it gives $\partial\pi/\partial p > 0$ and $\partial\pi/\partial q > 0$. Therefore, the researcher can determine the sign of the marginal effects of risk assessments on housing price:

$$
\begin{aligned}
\frac{\partial h}{\partial q} &= \frac{(U_1 - U_2)\dfrac{\partial\pi}{\partial q}}{\pi\dfrac{\partial U_1}{\partial X} + (1-\pi)\dfrac{\partial U_2}{\partial X}} < 0 \\
\frac{\partial h}{\partial p} &= \frac{(U_1 - U_2)\dfrac{\partial\pi}{\partial p}}{\pi\dfrac{\partial U_1}{\partial X} + (1-\pi)\dfrac{\partial U_2}{\partial X}} < 0
\end{aligned}
\tag{5}
$$

Combining two equations from (5), the researcher can derive the marginal effect of household risk perception on housing price:

$$
\frac{\partial h}{\partial \pi} = \frac{(U_1 - U_2)}{\pi\dfrac{\partial U_1}{\partial X} + (1-\pi)\dfrac{\partial U_2}{\partial X}} < 0 \tag{6}
$$

The result in equation (6) indicates that household risk perception has a negative effect on housing price, which presents an empirically testable hypothesis. A few things to note about this result. First, the magnitude of the effect depends on several factors: current levels of utility (or quality of life) in both the risk state of the world and no risk state of the world, marginal utilities of non-housing consumption (simplified to composite good X) in two states of the world, and current level of household risk perception. Second and conceptually, the marginal price derived here gives an expression of the marginal willingness-to-pay for an incremental reduction of household flood risk perception. This is a useful framework, with which one can then compute the welfare effect of a marginal change in measureable objective flood risk from the price gradient (Gayer et al., 2000). Since the objective risk assessment and the subjective risk assessment are additive and separable according to (1), this gives the hedonic valuation of flood risk.

2.2 Empirical Strategy

The goal of this paper is to quantify the effect of flood risk on housing values based on the hedonic framework described above. The theoretical result established in (6) does not give any direct hint on handling unobserved heterogeneities in an empirical implementation. The empirical design relies on boundary discontinuity to eliminate unobserved heterogeneities. Let h_i be the price of an observed home sale on housing unit i, and housing characteristics denoted by matrix Z_i, a hedonic housing price model can be given as:

$$\ln(h_i) = \alpha + Z_i\beta + f_i\theta + g_i + \varepsilon_i \tag{7}$$

where α is the intercept term, f_i is a [0,1] dummy variable indicating flood zone status (flood zone = 1). Similar as in Gibbons et al. (2013), g_i represents unobserved influences on housing prices that are correlated and distributed continuously across neighboring locations. ε_i is an idiosyncratic error to the household i, which is known to the household but unobservable to the researcher.

The key task of estimating equation (7) is to identify coefficient θ. The empirical strategy is to take the difference of two hedonic equations from two close-neighboring locations, say i and j, which gives:

$$\ln(h_i) - \ln(h_j) = (Z_i - Z_j)\beta + (f_i - f_j)\theta + g_i - g_j + \varepsilon_i - \varepsilon_j \tag{8}$$

Without loss of generality, let i denotes the property in the flood zone, and j the property outside the flood zone. Therefore, $f_i - f_j = 1$ and θ reduces to the intercept term in the new equation. Consistent estimation of the implicit hedonic price θ requires the unobserved heterogeneity term $g_i - g_j$ to be effectively random and uncorrelated with the difference in flood risks, $f_i - f_j$. This condition naturally holds in this paper, because $f_i - f_j$ is constant by design. To effectively eliminate unobserved heterogeneities the researcher makes the following assumption: $Var(g_i - g_j) \rightarrow 0$ as the Euclidian distance between two locations i and j approaches zero (Gibbons et al., 2013). In practice, the condition likely holds given that many of the factors and amenities influencing housing value are distributed continuously in the space unless their discontinuity points coincide with the floodplain boundary. As being discussed before, such coincidence is rare. The researcher will further explore these empirical concerns when discussing the results. Let $\omega_i = g_i - g_j + \varepsilon_i - \varepsilon_j$ be the new idiosyncratic error and be independently and identically distributed, the new estimation equation becomes:

$$\Delta \ln(h_i) = \theta + \Delta Z_i \beta + \omega_i \tag{9}$$

Note that here for simplicity the researcher suppressed the subscript notations so that i represents a pair of close-neighboring (as geographically close as possible) properties with one in the flood zone and another one outside. Conveniently, in the new estimation equation intercept term θ becomes the key coefficient of interest. The new empirical framework effectively helps to reduce the impacts from unobserved heterogeneities and potential correlations between socio-economic factors and flood risk in estimation. Fixed effects can still be included in model (9) to control for certain institutional or jurisdictional effects, the researcher will further explore this option in results section.

Another aspect that one would argue is that, the empirical strategy and reliability of estimates of θ hinge on an implicit assumption that households do not sort into or outside the 100-year floodplain in a systematic way. In general, residents sort across the space based on school quality, racial identify, labor markets, and other socio-economic factors. In the study region of this paper, this is less likely a concern. Residents in Juniata County are served only by two school districts: Juniata County School District and Greenwood School District. Among these two, Juniata County School District covers a majority of residential area in the county. According to estimates from the US Census, 97.5% of the population in Juniata County is white in 2015. Therefore, racial segregation is unlikely to be a factor of sorting.

Perry County shares a similar demographics, with 97.1 % of its population being white in 2015 according to the US Census. Perry County has one major school district (West Perry School District), two small school districts (Newport School District and

Susquenita School District), and one school district (Greenwood School District) shared with Juniata County. It is also necessary to point out that the area of flood zone is considerably smaller than any school districts. In general, the researcher argues that in the study region flood zone status is unlikely to be a key determinant of overall neighborhood quality. Therefore, sorting is not a concern for identification when separating willingness-to-pay for flood risk reduction from willingness-to-pay for neighborhood quality. If, however, households do sort across flood zone boundary based on other socio-economic factors such as race and school quality, the consequence is that it becomes difficult to identify the willingness-to-pay for avoiding flood risk in a clean way. In this case, neighborhood or household demographic variables can be included to control for potential bias, which requires detail household demographic information.

3 Study Area and Data

This study assembles a unique housing transaction data set between 1980 and 2015 from the assessor offices of two neighboring Pennsylvania counties: Juniata and Perry, both in close proximity of the state capitol Harrisburg and located in the Susquehanna River Basin (SRB). Juniata County contains the Tuscarora Creek watershed and part of the Juniata River watershed, mostly in the Juniata sub-basin of the SRB. Perry County contains the Buffalo Creek watershed, the Sherman Creek watershed, part of the Juniata River watershed, and part of the main Susquehanna River watershed, mostly in the Lower Susquehanna sub-basin of the SRB. Figure A1 in the appendix shows the relative location of the study area in the basin.

Based on the rural-urban continuum codes from the Economic Research Service of USDA and the Office of Management and Budget (OMB), Juniata County is categorized as a rural non-metro county. With a relatively low population density, the residential landscape is fairly fragmented across the county. Located in a traditionally agricultural region, farmland and forest make up a large portion of the county's land cover. Perry County is categorized as a rural metro county but also with low population density, the land cover of Perry County is similar to that of

Juniata County. Flood events are observed in the region frequently. For instance, Juniata County had 6,374 acres flooded during Tropical Storm Agnes in 1972, and 57 families were seriously affected by the disaster (Juniata County Planning Commission, 2009). For Perry County, 29 of its 30 municipalities are flood prone.

Table A1 and A2 in the appendix report a list of major flood events occurred in both counties within recent decades, mainly the most dangerous kind of floods – flash flood. When excessive amount of rainfall water fills up dry creeks or river beds (often smaller tributaries) connected to currently flowing creeks and rivers, it leads to flash floods and can cause rapid rises of water in a short time. Flash floods often occur with no forewarning. Juniata County and Perry County make it an ideal region to study flood risk impact on rural housing markets. The housing markets in both counties belong to the greater Harrisburg metropolitan area housing market. As shown in Figure A2, the real unit housing price (converted to constant value using Housing Price Index (HPI) of Harrisburg metropolitan area) on average increases over the study period in both counties.

Following the rationale of identification through boundary discontinuity, the researcher matched 597 pairs of SFR properties (1194 observations) which had at least one sale between 1980 and 2015 with both sale price and sale year recorded in Juniata County. For Perry County, similarly, 912 pairs of SFR properties are matched among properties which had at least one sale between 1980 and 2015. The matching of properties is based on the nearest distance. The algorithm is described in details in Appendix B with a sample computer code. Figure 2 shows the location of all the matched SFR properties located in the 100-year floodplain (area inundated by 1 percent annual chance of flooding) created by FEMA. The shaded area within the county boundary is the flood prone area (including all land use types) drawn based on the 100-year floodplain.

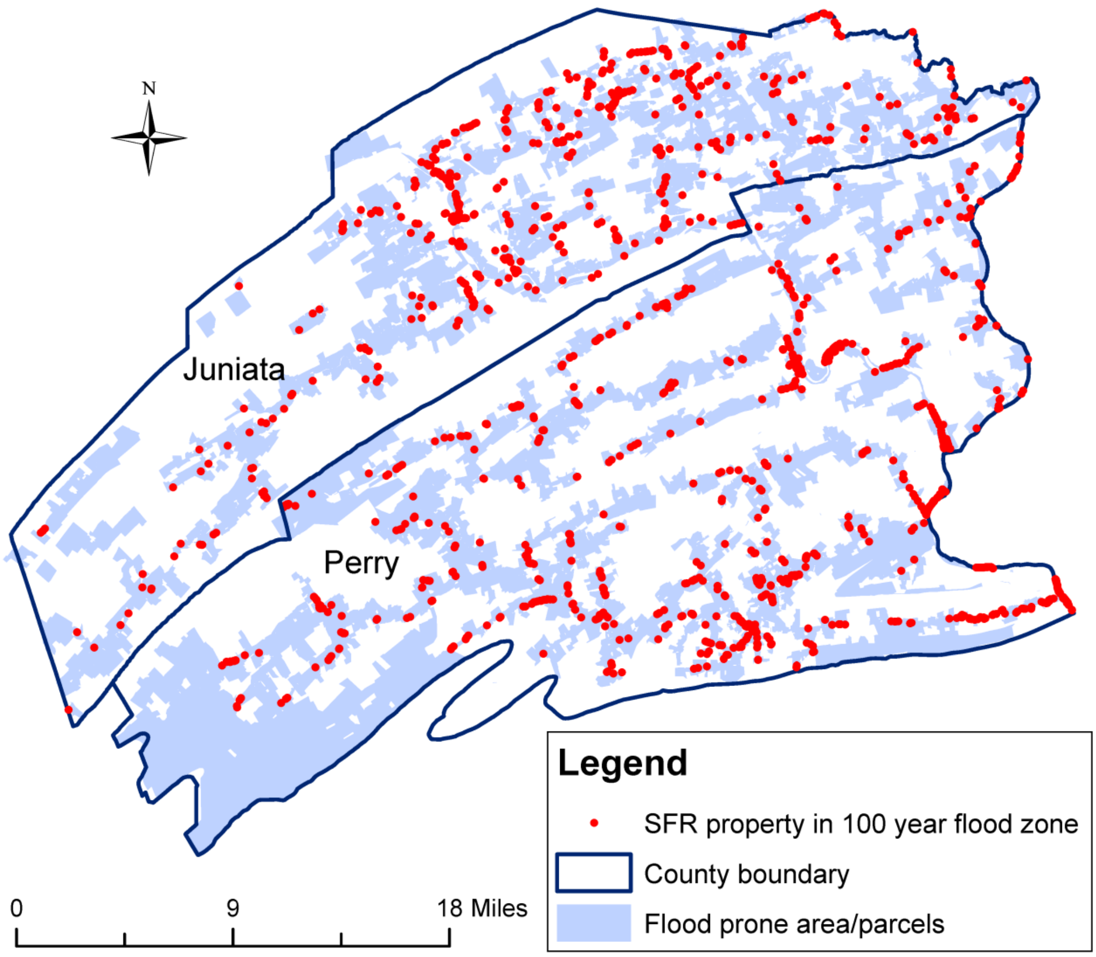


Figure 2: Flood prone area and location of observations

Note that this study only includes full-time SFR properties, since the valuation of other types of residential properties (e.g. condos, multi-families, apartments, vacation homes, and mobile homes) may be structurally different. Among all the SFR property transactions, the researcher also excluded all of the love and affection sales (usually with a transaction price of $1). In the estimation sample, the researcher dropped all the sales with transaction price less than $500. Due to data limitation, the researcher only has access to the most recent sale record if multiple sales were recorded on a property in the past. All the sale prices are converted to constant values in 2015 USD using the HPI of Harrisburg metropolitan area (both counties are in commuting distance to the city of Harrisburg) published by the Federal Housing Finance Agency. Table 1 summarizes all of the variables used to estimate the proposed empirical model in (9). Note that the assessed land value and the assessed building value do not add up to the housing price. This is normal because the determination of property assessment value follows a system (e.g. negotiation between homeowners and assessor's office) quite different from the market mechanism. However, the assessment value in general is highly correlated with structural attributes (e.g. number of bedrooms and bathrooms) of the property. The basic model is estimated using ordinary least squares (OLS) with heteroskedasticity robust standard errors. The main control variables in the model are housing characteristics which are supposed to capture any structural differences between paired properties, as well as the distances to county seat and the city center of Harrisburg.

Table 1: Summary statistics of variables

County	Juniata	Perry	Juniata	Perry
Variable	In 100 year flood zone		Outside 100 year flood zone	
	mean (s. d.)	mean (s. d.)	mean (s. d.)	mean (s. d.)
Housing price (in $1000)	86.92(54.33)	100.33(65.19)	88.71(54.67)	109.56(65.09)
Unit housing price (in $/sqft)	64.95(38.44)	67.74(39.65)	68.83(41.13)	73.18(40.74)
Lot size (in acres)	2.95(4.25)	1.84(2.13)	1.81(2.90)	1.62(1.92)
Assessed Land value (in $1000)	2.42(1.81)	40.08(18.84)	2.20(1.01)	40.23(17.61)
Assessed Building value (in $1000)	12.82(8.58)	88.95(56.16)	14.10(8.31)	98.35(54.82)
Living space (in 1000 sqft)	1.44(0.63)	1.57(0.65)	1.41(0.62)	1.58(0.58)
Distance to Harrisburg (in miles)	32.79(3.49)	18.99(7.47)	32.80(3.48)	18.99(7.45)
Distance to county seat* (in miles)	7.57(5.98)	9.24(4.13)	7.58(5.98)	9.20(4.12)
Number of observations	597	912	597	912

* The seat of Juniata County is in Mifflintown, Pennsylvania. The seat of Perry County is in New Bloomfield, Pennsylvania.

From the summary statistics, one can see that properties in the 100-year flood zone on average is cheaper than their counterparts outside the 100-year flood zone, in terms of both overall price and unit price. One should also note that housing price in Juniata County is relatively lower than housing price in Perry County, which likely reflects the fact that Perry County locates closer to the city of Harrisburg. This suggests that distance to the central business district should be controlled in the empirical model. To control for structural differences between matched properties as much as possible, the researcher uses housing price per sqft living space as dependent variable. Figure 3 plots the distribution of unit housing prices (note that actual dependent variable in the empirical model takes logarithm transformation) inside and outside the flood zone, obtained through nonparametric kernel density estimation. As shown in the graph, the distribution of unit housing prices outside the 100-year flood zone (red/dash line) shifts to the right of its counterpart, which is consistent with the summary statistics.

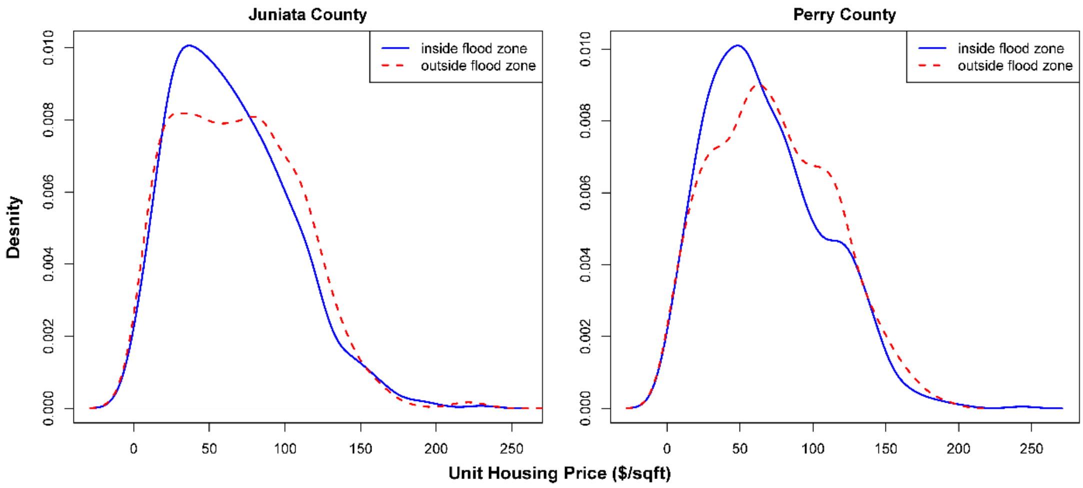


Figure 3: Distribution of unit housing price

In the assessment data, the structure of each housing unit is described by one of more than 100 structure codes, which makes them difficult to quantify. Instead, the researcher used the assessed building value as a within-county proxy for the richness of housing structure (e.g. number of bedrooms and number of bathrooms). The empirical model also controls for property lot size, assessed land value, housing living space, and distances to county seat and the nearest metropolitan city - Harrisburg. Due to a large portion of missing values, the number bedrooms and bathrooms, though available in Perry County dataset, cannot be included in the empirical model. As shown in Figure A3, however, the size of living space is highly correlated with the number of bedrooms and bathrooms. Therefore, living space can help to control for major structural differences of properties in the empirical model. Note that, as shown in Table 1, the level of assessment values are significantly different in two counties. This implies that two counties have essentially different assessment processes, which suggests that the regression analysis for two counties has to be implemented separately. However, the estimation results on flood risk premium should still be comparable. Because, arguably, two counties belong to one large housing market and locate within a similar environment.

4 Results and Discussion

Table 2 reports main estimation results from the difference-in-differences model. The OLS columns report standard linear regression results with heteroskedasticity robust standard errors. Feasible GLS (generalized least squares) estimation is used to gain potential efficiency improvement over OLS estimates, which is reported under the GLS columns. The OLS estimates and GLS estimates are very close to each other in terms of magnitude. Several consistent observations can be made across the estimation results for two neighboring counties. First, being located in the 100-year flood zone has a significant negative impact on unit housing price. The magnitude of such an impact will be discussed later in the section after addressing other potential estimation bias in the estimation. Second, assessed land values and building structure values have a significant positive relationship with unit housing price. This is in tune with the fact that assessment values tend to consider most of the structural attributes and amenities associated with the property, not just the sizes of the property and the structure. Third, as one would expect, the negative significant estimates on living space indicate a diminishing effect of living space on the whole property value given that the dependent variable is measured by per sqft living space.

Property size (lot size) has a positive impact in Juniata County, but a small to none negative impact in Perry County. Such a variation may be explained by the very different average plot size in two counties, as indicated in the summary statistics in Table 1. Distance to the central business district (Harrisburg) has a strong negative impact in Perry County as the urban economic theory suggests, but is found insignificant in Juniata County. This may be explained by the relative location of two counties. Perry County is right next to the city of Harrisburg in the west. Juniata

County is much further away from the city. It is natural to expect the effect of monocentricity becomes less pronounced as we move away from the central business district. The distance to county seat has a slight positive impact in both counties. The county seats in both cases are very small town or borough. Therefore, its impact on the housing market may not be explained by the theory of monocentricity. One possible interpretation is that, the variable captures more of the relatively high level of natural amenities in the area further away from county seat, which is positively correlated with the distance to county seat. Another possible interpretation is that, in most of these Northeast counties, county seat is usually not the only major town in the county. In these cases, a polycentric structure describes the pattern of land development better than the monocentric structure. For example, in Perry County, Duncannon and Newport are also major towns with even better access to major highways.

Table 2: Estimation results of difference-in-differences models

Log(price/living space)	Model			
	Juniata County		Perry County	
	OLS	GLS	OLS	GLS
Intercept (flood risk impact)	**-0.0727* (0.0447)**	**-0.0703*** (0.0028)**	**-0.0644* (0.0369)**	**-0.0654*** (0.0009)**
Lot size	0.0251** (0.0104)	0.0242*** (0.0007)	-0.0106 (0.0281)	-0.0094*** (0.0009)
Assessed Land value	0.0852*** (0.0326)	0.0896*** (0.0025)	0.0083*** (0.0034)	0.0082*** (0.0001)
Assessed Building value	0.0066 (0.0054)	0.0068*** (0.0004)	0.0027*** (0.0008)	0.0027*** (0.0001)
Living space	-0.5138*** (0.0652)	-0.5208*** (0.0062)	-0.4936*** (0.0560)	-0.4941*** (0.0033)
Distance to Harrisburg	-0.0371 (0.1748)	0.0316 (0.0197)	-0.3833** (0.1645)	-0.3792*** (0.0076)
Distance to county seat	0.1526 (0.2488)	0.1172*** (0.0145)	0.1294 (0.1479)	0.1271*** (0.0073)
R^2	0.1445	-	0.0929	-
Sample size	597	597	912	912

Note: throughout the paper, asterisks (*,**,***) indicate statistical significance at 10%, 5%, and 1% level, respectively, unless otherwise noted.

One of the major concerns on using boundary discontinuities to identity policy parameter is the potential selection effect around the boundary (Black, 1999; Gibbons et al., 2013). In this study, such a selection effect could happen if the flood zone boundary coincides with or is very close to municipal boundary or school district boundary. The associated estimation bias comes from the fact that there may be omitted variables that vary at the municipality or school district level, in particular property tax rates and public goods provision. In the study region of this paper, the municipal boundary and the school district boundary overlaps each other. In general, one school district covers several municipalities. Therefore, we only consider the impact from municipal boundaries.

To capture the change of municipalities between two matched properties, a single dummy variable indicating if a matched pair of properties come from different municipalities (yes=1, otherwise 0) is included. This different municipality dummy variable is similar to including fixed effects for all municipalities in the standard hedonic price model from equation (7) without an intercept term. If, for all pairs, two properties in each pair happen to locate in the same municipality, then the difference-in-differences model effectively cancels out all of the municipality fixed effects. In this case, the dummy variable (always equals to 0) is dropped out of the regression due to no variation.

In the estimation sample, there are some matched properties which do not belong to the same municipality. Therefore, the differencing procedure cannot cancel out all of the fixed effects. The different municipality dummy variable plays a role of absorbing the municipality fixed effects that are not netted out by the differencing procedure. Still and practically, in the proposed difference-in-differences model, a new set of municipality fixed effects can be included. However, their interpretation is not straightforward. In Gibbons et al. (2013), boundary fixed effects are accounted for. In this paper, the irregular shape of floodplain boundaries makes them difficult to quantify, and their implications are relatively homogeneous across the space. The preferred GLS estimation results with the same municipality dummy variable are reported in Table 3.

Table 3: GLS estimation results with control for municipal fixed effects

Log(price/living space)	Model	
	Juniata County	Perry County
Intercept (flood risk impact)	**-0.0622*** (0.0029)**	**-0.0720*** (0.0017)**
Lot size	0.0232*** (0.0008)	-0.0113*** (0.0013)
Assessed Land value	0.0868*** (0.0025)	0.0084*** (0.0002)
Assessed Building value	0.0067*** (0.0004)	0.0027*** (0.0001)
Living space	-0.5154*** (0.0040)	-0.4950*** (0.0025)
Distance to Harrisburg	0.0249 (0.0193)	-0.3703*** (0.0118)
Distance to county seat	0.1281*** (0.0162)	0.1315*** (0.0108)
Different municipalities (yes=1)	-0.0605*** (0.0051)	0.1780*** (0.0068)
Sample size	597	912

The municipal boundary selection effect can bias estimates in either directions. As one can see from Table 3, the different municipality dummy has a significant negative estimate of -0.0605 for Juniata County, while it has a significant positive estimation of 0.1780 for Perry County. As a result of these significant municipal boundary selection effects, our key flood risk impact estimates (intercepts) are corrected correspondingly. For Juniata County, the estimate changes from -0.0703 to -0.0622 (a change of 13%). For Perry County, the change is 10% (from -0.0654 to -0.0720). With the corrected estimates, we can now derive the economic implications of the estimated flood risk impact on unit housing price.

From the results in Table 3 one can see that, as indicated by the intercept estimates, flood risk has a significant negative effect on unit housing price which the theoretical model suggests. Given the mean of log(price/living space) being 3.9639 in the sample for Juniata County and a flood risk impact estimate of -0.0622, a back-of-the-envelope calculation gives a corresponding flood risk premium of \$3.28/sqft:

$$0.5\left|\exp(3.9639+0.0622)-\exp(3.9639)\right| + 0.5\left|\exp(3.9639-0.0622)-\exp(3.9639)\right| = 3.28\ (\$/\mathit{sqft}) \quad (10)$$

Considering that an average SFR housing unit has about 1430 sqft (133 m^2) living space in the estimation sample of Juniata County, the estimated flood risk premium translates to a $4690 (or 5.34% given the average housing price around $87,800) price difference in 2015 housing market.

Similarly, given the mean of log(price/living space) being 4.0168 in the sample for Perry County and a flood risk impact estimate of -0.0720, a back-of-the-envelope calculation gives a corresponding flood risk premium of $4.00/sqft:

$$0.5\left|\exp(4.0168+0.0720)-\exp(4.0168)\right| + 0.5\left|\exp(4.0168-0.0720)-\exp(4.0168)\right| = 4.00\ (\$/\mathit{sqft}) \quad (11)$$

Again, considering that an average SFR housing unit has about 1580 sqft (147 m^2) living space in the estimation sample of Perry County, the estimated flood risk premium translates to a $6320 (or 6.02% given the average housing price around $105,000) price difference in 2015 housing market.

These estimates are consistent with other findings in the literature (see Figure 1). The estimates are also the first set of empirical flood risk evaluation for the US Northeast - a region likely with increasing precipitation variability and flood risk in the coming decades (Collins, 2009). The significance of the estimated flood risk premium suggests that households in the study region do value potential flood risk in a consequential way, which gets reflected through price variations in the housing market. In some regions across the US, the Mississippi River valley for example, the price differential along floodplain boundaries can also be due to artificial structures (e.g. berms and dykes) providing enhanced protection against flood risk. In that case, the price differential reflects both a willingness-to-pay for the flood prevention infrastructures and a flood risk premium. In the study region of this paper, however, there are no such flood prevention infrastructures, given its relatively low frequency of flood events comparing to coastal areas. Therefore, the estimated price differential reveals mainly a premium for potential flood risk.

Given the methodological advantage of the proposed difference-in-differences framework in controlling for unobserved heterogeneities, it is still interesting to compare results from a standard hedonic price model to the proposed model. Table 4 reports the results from the standard hedonic price model (with municipality fixed effects) with all matched SFR properties pooled together. The estimates are comparable to those of Table 3. Note that the flood zone status dummy variable in the standard hedonic model corresponds to the intercept in the difference-in-differences model. The flood zone status in the standard hedonic model is statistically insignificant at 10% level in OLS estimation. By comparing the GLS estimation results between Table 3 and Table 4, it is clear that the standard hedonic price model underestimates the flood risk impact for both counties.

Table 4: Estimation results of hedonic price model with municipal fixed effects

Log(price/living space)	Model			
	Juniata County		Perry County	
	OLS	GLS	OLS	GLS
Flood zone (yes=1)	**-0.0481 (0.0431)**	**-0.0538*** (0.0052)**	**-0.0494 (0.0360)**	**-0.0516*** (0.0034)**
Lot size	0.0179** (0.0072)	0.0173*** (0.0009)	-0.0256 (0.0182)	-0.0270*** (0.0011)
Assessed Land value	0.0147*** (0.0038)	0.0148*** (0.0004)	0.0030*** (0.0005)	0.0030*** (0.0001)
Assessed Building value	0.0748*** (0.0265)	0.0790*** (0.0033)	0.0100*** (0.0022)	0.0101*** (0.0002)
Living space	-0.5740*** (0.0524)	-0.5745*** (0.0039)	-0.5055*** (0.0420)	-0.5010*** (0.0043)
Distance to Harrisburg	0.0211 (0.0205)	0.0179*** (0.0020)	0.0010 (0.0139)	0.0026 (0.0017)
Distance to county seat	-0.0033 (0.0135)	-0.0044*** (0.0016)	-0.0401*** (0.0140)	-0.0382*** (0.0022)
R^2	0.2094	-	0.1647	-
Fixed Effects	Municipality			
Sample size	1194	1194	1824	1824

This suggests that the proposed difference-in-differences framework using boundary discontinuities of floodplain can potentially eliminate the bias due to omitted variables in the hedonic price model. If the researcher does the same back-of-the-envelope calculation, for Juniata County, an estimate of -0.0538 in the standard hedonic price model gives a flood risk premium estimate of $2.83, which is 13.7% less than the estimate of $3.28. Similarly and for Perry County, an estimate of -0.0516 in the standard hedonic price model gives a flood risk premium estimate of $2.87, which is 28.3% less than the estimate of $4.00. The differences imply that the flood zone status is not orthogonal to other unobserved spatial heterogeneities, such as amenity value of access to the river bank, and the proposed empirical strategy helps to correct potential biases in the estimates to a large extent. Note that the overall model fit (measured by R^2 in OLS cases) between two models here are not directly comparable, given that their dependent variables are not defined in the same way. Also note that, when estimating the model with GLS, the total sum of squares cannot be broken down in the same way as in OLS, making the R^2 statistic less useful as a diagnostic tool for model fit. In fact, an R^2 statistic computed from GLS needs not be bounded between zero and one.

An important question regarding the implications of the results in this paper is how generalizable the results are to other rural regions and the entire Northeast region. Studies have shown that flood risk awareness has been rising in rural America (e.g. Knocke and Kolivras, 2007). Existing quantitative studies have rarely concerned about the flood risk impacts in rural regions. Among a few existing studies, Bin and Polasky (2005) find insignificant impacts of flood risk on housing values in a rural region (Carteret County, NC). Using data from rural Connecticut (part of Hartford County), Paterson and Boyle (2002) find that close proximity to rivers and streams has a significant negative impact on housing values. In general, the empirical findings in this paper is in line with the existing studies. However, few of the existing studies explicitly control for unobserved heterogeneities. The associated omitted variables bias could lead to either underestimation or overestimation of the key coefficient on flood risk impact. Such a bias can be substantial, especially considering the fact that certain omitted amenity measures may be correlated with the flood risk exposure measure. In this study, the researcher strives to control and minimize the impact of unobserved heterogeneities on the estimate of flood risk impact. The paper provides the first quantitative evaluation on the flood risk impact associated with the SRB region. It is also one of the first studies focusing on flood risk impact in the US Northeast region.

5 Concluding Remarks

The literature on flood risk capitalization has been focusing on coastal housing market and effects of major natural disasters along the coastal area. Much less attention has been attached to flood risk in non-coastal areas where river flood risk also poses serious threat to economic development and human lives. This paper presents one of the first studies on flood risk evaluation in the US Northeast - a region where we are likely going to see increasing precipitation variability and associated flood risk in the coming decades, which also contributes to the literature with one of the first studies of flood risk impacts on rural housing market.

In this paper, a spatial difference-in-differences framework based on floodplain boundary discontinuities is proposed to control for unobserved heterogeneities. Using parcel level data from Juniata County and Perry County in Pennsylvania, the paper finds that there is a 5-6 percent housing value reduction due to exposure to 1 percent annual chance of flooding in the study region. Specifically, for Juniata County, the proposed empirical approach shows that on average there is a \$35.31 discount per square meter of living space (or \$3.28/sqft, in 2015 USD) for a full-time SFR property located within the flood zone. For an average housing unit of 133 m^2 (1430 sqft) living space in the sample, the estimate translates to a \$4690 housing value reduction. For Perry County, the corresponding estimates are \$40.06 per square meter (or \$4.00/sqft, in 2015 USD) and \$6320 for an average housing unit of 147 m^2 (1580 sqft).

The paper also finds that with similar specifications, a standard hedonic price model underestimates the flood risk impact on housing value by a substantial amount. This further confirms with the literature that correlated flood risk and amenity could lead to underestimation of flood risk premium (Bin and Kruse, 2006; Carbone et al., 2006). The bias is a type of omitted variables bias likely caused by the unobserved heterogeneities that drive housing value. Since such a bias is not negligible, its implications for policy (e.g. planning policy and flood insurance policy) can be important. The empirical strategy proposed in this paper provides an effective alternative to correct such potential biases in flood risk evaluation.

The current empirical framework does not explicitly control for any subjective component of the overall household risk perception due to data limitation. It is arguable that for two close enough households (as they are matched by the nearest distance) their subjective risk perceptions should not differ from each other substantially given the strong likelihood of social interactions. The question then becomes to what extent neighborhood spillovers and social interactions can affect household risk perception regarding natural hazards. Data collection (e.g. survey data) on household subjective flood risk assessment can be valuable in pursuing this aspect of the research question, which points to a fruitful direction for future research. Among few existing studies with survey data, Petrolia et al. (2013) find a significant positive relationship between risk aversion and decision to purchase flood insurance.

Another interesting aspect to improve the current research is to further explore the role of flood insurance. The decisions to purchase flood insurance (sometime it is required by mortgage lenders pursuant to federal law) and choose insurance deductible and coverage can reveal important information regarding household subjective risk assessment. Kousky and Walls (2014) argue that spatial targeting can substantially increase the net benefit from floodplain management. A similar idea of spatial targeting has been proposed by Baade et al. (2007). A deep understanding of household level risk perception is certainly an important part of spatial targeting, on which the framework proposed in this paper can shed light. In terms of building and enhancing local resilience to flood risk, many opportunities can be identified ranging from strengthening social institutions (Baade et al., 2007) to smart growth development patterns (Brody et al., 2013).

References

Baade RA, Baumann R and Matheson V (2007) Estimating the economic impact of natural and social disasters, with an application to Hurricane Katrina. *Urban Studies* 44(11): 2061-2076.

Bartosova A, Clark D, Novotny V, et al. (2000) Using GIS to evaluate the effects of flood risk on residential property values. Proceedings of The EPA Conference 2000: 1-35.

Bin O and Polasky S (2004) Effects of flood hazards on property values: Evidence before and after Hurricane Floyd. Land Economics 80(4): 490-500.

Bin O and Polasky S (2005) Evidence on the amenity value of wetlands in a rural setting. Journal of Agricultural and Applied Economics 37(3): 589-602.

Bin O and Kruse JB (2006) Real estate market response to coastal flood hazards. Natural Hazards Review 7(4): 137-144.

Bin O, Crawford TW, Kruse JB and Landry CE (2008) Viewscapes and flood hazard: Coastal housing market response to amenities and risk. Land Economics 84(3): 434-448.

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

Black SE (1999) Do better schools matter? Parental valuation of elementary education. Quarterly Journal of Economics 114(2): 577-599.

Brody S, Kim H and Gunn J (2013) Examining the impacts of development patterns on flooding on the Gulf of Mexico coast. Urban Studies 50(4): 789-806.

Carbone JC, Hallstrom DG and Smith VK (2006) Can natural experiments measure behavioral responses to environmental risks? Environmental and Resource Economics 33(3): 273-297.

Collins, MJ (2009) Evidence for changing flood risk in New England since the late 20th Century. Journal of the American Water Resources Association 45(2): 279–290.

Frumhoff PC, McCarthy JJ, Melillo JM, Moser SC and Wuebbles DJ (2007) Confronting climate change in the US Northeast: A report of the northeast climate impacts assessment. Union of Concerned Scientists: Cambridge, Massachusetts.

Gayer T, Hamilton JT and Viscusi WK (2000) Private values of risk tradeoffs at superfund sites: housing market evidence on learning about risk. Review of Economics and Statistics 82(3): 439-451.

Gibbons S, Machin S and Silva O (2013) Valuing school quality using boundary discontinuities. Journal of Urban Economics 75: 15-28.

Hallstrom DG and Smith VK (2005) Market responses to hurricanes. Journal of Environmental Economics and Management 50(3): 541-561.

Hamilton SE and Morgan A (2010) Integrating lidar, GIS and hedonic price modeling to measure amenity values in urban beach residential property markets. Computers, Environment and Urban Systems 34(2): 133-141.

Harrison D, Smersh GT and Schwartz AL (2001) Environmental determinants of housing prices: The impact of flood zone status. Journal of Real Estate Research 21(1-2): 3-20.

Jongman B, Ward PJ and Aerts JC (2012) Global exposure to river and coastal flooding: Long term trends and changes. Global Environmental Change 22(4): 823-835.

Juniata County Planning Commission (2009) Juniata County Comprehensive Plan, http://www.co.juniata.pa.us/departments/planning/comprehensive-plan/, accessed on 03/12/2016.

Kahn ME (2005) The death toll from natural disasters: The role of income, geography, and institutions. Review of economics and statistics 87(2): 271-284.

Knocke ET and Kolivras KN (2007) Flash flood awareness in southwest Virginia. Risk analysis 27(1): 155-169.

Kousky C (2010) Learning from extreme events: Risk perceptions after the flood. Land Economics 86(3): 395-422.

Kousky C (2014) Informing climate adaptation: A review of the economic costs of natural disasters. Energy Economics 46: 576-592.

Kousky, C and Kunreuther H (2014) Addressing affordability in the national flood insurance program. Journal of Extreme Events 1(01): 1-28.

Kousky C and Walls M (2014) Floodplain conservation as a flood mitigation strategy: Examining costs and benefits. Ecological Economics 104: 119-128.

Labat D, Goddéris Y, Probst JL, et al. (2004) Evidence for global runoff increase related to climate warming. Advances in Water Resources 27(6): 631-642.

Lindell MK and Prater CS (2003) Assessing community impacts of natural disasters. Natural Hazards Review 4(4): 176-185.

McKenzie R and Levendis J (2010) Flood hazards and urban housing markets: The effects of Katrina on New Orleans. Journal of Real Estate Finance and Economics 40(1): 62-76.

McMichael AJ, Woodruff RE and Hales S (2006) Climate change and human health: Present and future risks. The Lancet 367(9513): 859-869.

Morgan A (2007) The impact of Hurricane Ivan on expected flood losses, perceived flood risk, and property values. Journal of Housing Research 16(1): 47-60.

Niskanen WA and Hanke SH (1977) Land prices substantially underestimate the value of environmental quality. Review of Economics and Statistics 59(3): 375-377.

Paterson RW and Boyle KJ (2002) Out of sight, out of mind? Using GIS to incorporate visibility in hedonic property value models. Land Economics 78(3): 417-425.

Petrolia DR, Landry CE and Coble KH (2013) Risk preferences, risk perceptions, and flood insurance. Land Economics 89(2): 227-245.

Shultz SD and Fridgen PM (2001) Floodplains and housing values: Implications for flood mitigation projects. Journal of the American Water Resources Association 37: 595-603.

Skantz T and Strickland T (1987) House prices and a flood event: An empirical investigation of market efficiency. Journal of Real Estate Research 2(2): 75-83.

Speyrer JF and Ragas WR (1991) Housing prices and flood risk: An examination using spline regression. Journal of Real Estate Finance and Economics 4(4): 395-407.

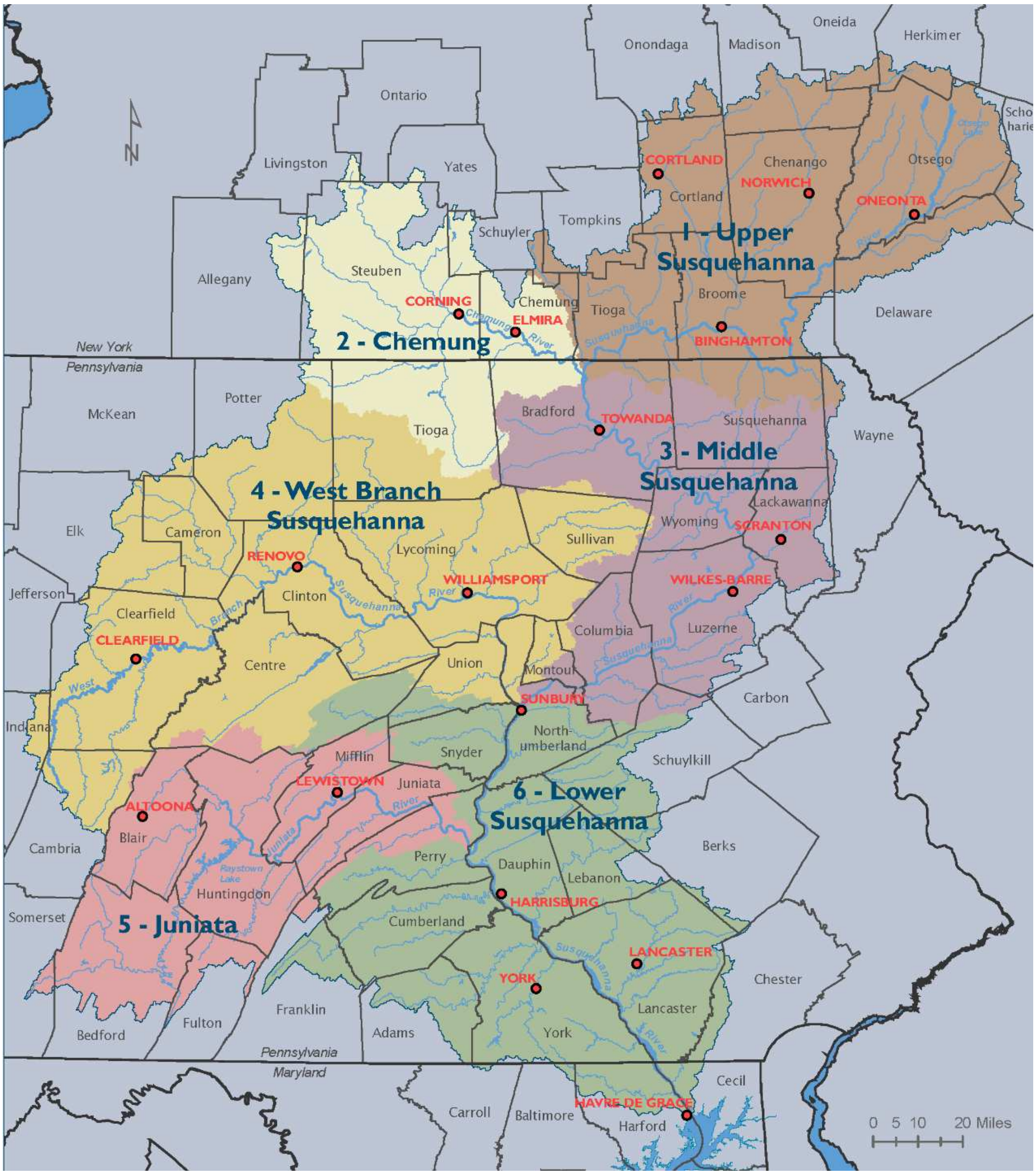
Struyk RJ (1971) Flood risk and agricultural land values: A test. Water Resources Research 7(4): 789-797.

Trenberth KE (2011) Changes in precipitation with climate change. Climate Research 47(1): 123-138.

Zhang Y, Hwang SN and Lindell MK (2010) Hazard proximity or risk perception? Evaluating effects of natural and technological hazards on housing values. Environment and Behavior 42(5): 597-624.

Zheng H, Barta D and Zhang X (2014) Lesson learned from adaptation response to Devils Lake flooding in North Dakota, USA. Regional Environmental Change 14(1): 185-194.

Appendix A: Supplementary Figures and Tables



Data Source: Susquehanna River Basin Commission

Figure A1: Susquehanna River Basin and county boundary in the watershed

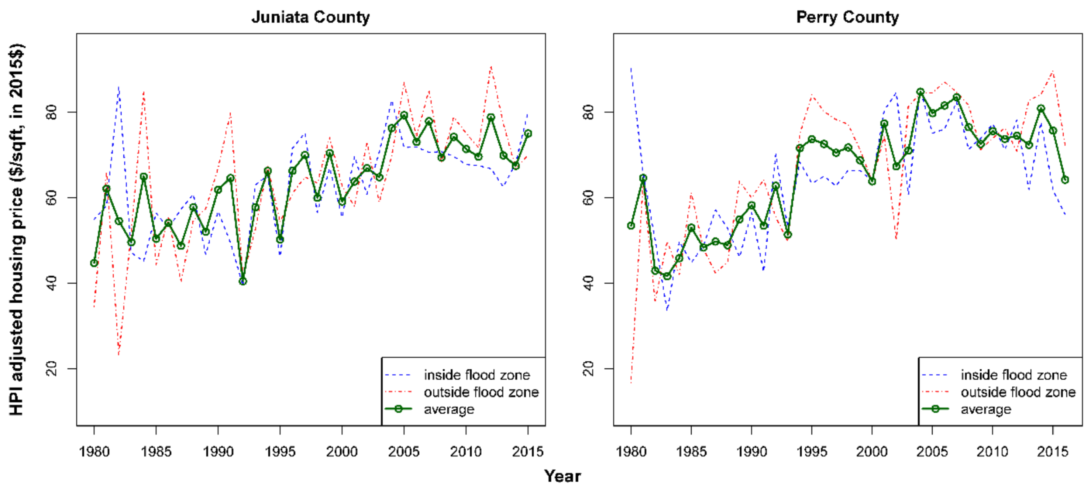
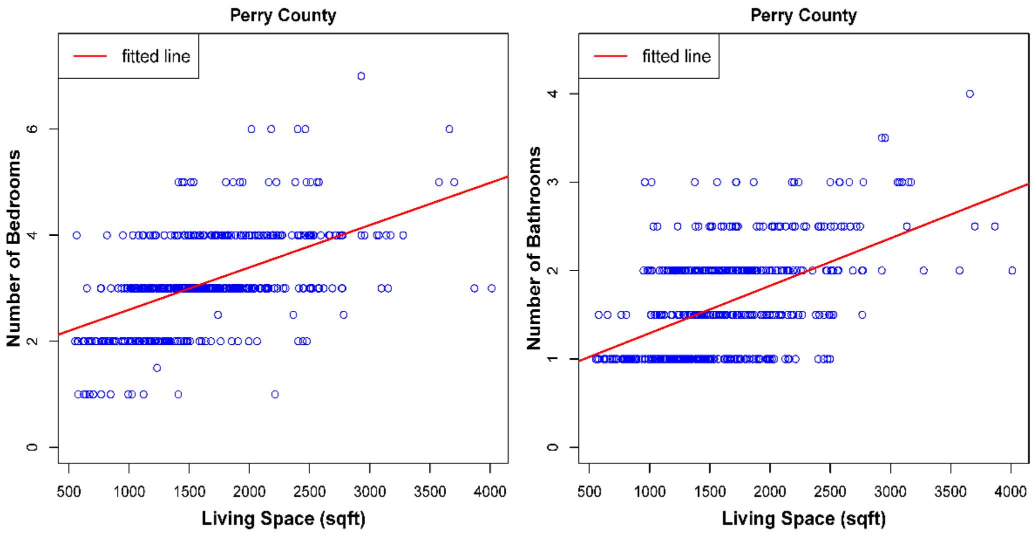


Figure A2: HPI adjusted unit housing prices in Juniata County and Perry County



Note: The graph is drawn based on a limited set of observations (561 out of 1824 matched properties)

Figure A3: Relationship between the size of living space and number of bedrooms/bathrooms

Table A1: Juniata County Flooding Event History

Flood Date	Flood Type
10/21/1995	Flash Flood
01/19/1996	Flood
01/19/1996	Flash Flood
09/06/1996	Flash Flood
09/13/1996	Flash Flood
12/13/1996	Flash Flood
09/11/1997	Flash Flood
01/08/1998	Flash Flood
04/19/1998	Flash Flood
01/23/1999	Flash Flood
08/20/1999	Flash Flood
06/20/2001	Flash Flood
09/17/2004	Flood
03/28/2005	Flood
06/27/2006	Flash Flood

Data source: National Climatic Data Center and Juniata County Multi-Jurisdictional Hazard Mitigation Plan

Table A2: Perry County Flooding Event History

Flood Date	Flood Type	Flood Date	Flood Type
11/28/1993	Flood/Flash Flood	12/11/2003	Flood
01/20/1995	Flood	07/12/2004	Flash Flood
10/21/1995	Flood/Flash Flood	09/17/2004	Flood
01/19/1996	Flood/Flash Flood	09/18/2004	Flood
09/06/1996	Flash Flood	09/28/2004	Flood
09/13/1996	Flash Flood	03/28/2005	Flood
12/13/1996	Flash Flood	03/29/2005	Flood
09/11/1997	Flash Flood	04/02/2005	Flood
11/08/1997	Flash Flood	11/30/2005	Flood
01/08/1998	Flash Flood	06/27/2006	Flash Flood
03/21/1998	Flash Flood	03/05/2008	Flood
09/06/1999	Flash Flood	05/28/2009	Flash Flood
09/16/1999	Flash Flood	05/29/2009	Flood
09/01/2000	Flash Flood	03/10/2011	Flood
08/14/2001	Flash Flood	04/16/2011	Flood
08/09/2003	Flash Flood	09/07/2011	Flood

Data source: National Climatic Data Center and Perry County Multi-Jurisdictional Hazard Mitigation Plan

Appendix B: Matching Algorithm

Matching between properties within the flood zone and outside the flood zone is implemented using the following procedure:

1. Select all the SFR properties which have market transactions during the study period, and categorize them into two subsamples: flood free and flood prone;

2. For each SFR property within the flood prone sample, compute its distance to all other properties in the flood free sample within a given radius (e.g. 10 miles, to reduce computational burden). This process creates a large matrix with three columns: flood prone property ID (INPUT_FID), flood free property ID (NEAR_FID), and distance.

3. Rank all of the rows in the matrix by the computed distance, extract and save the pair of IDs corresponding to the row with the smallest distance. Next, delete all the rows with either the saved flood prone property ID or the saved flood free property ID. This makes sure that the paired IDs are always unique, and each ID is only used once in each subsample.

4. Repeat step 3 with the rest of rows, until all the rows have been extracted or deleted.

Sample code for matching SFR properties in R:

```
data <- read.csv("data.csv",header=TRUE) # three columns: INPUT_FID, NEAR_FID, DISTANCE
matched <- matrix(NA,length(unique(data$INPUT_FID)),3)
i <- 1
reduced_data <- data
while(i <= length(unique(data$INPUT_FID))){
  tempdata <- reduced_data[reduced_data$DISTANCE==min(reduced_data$DISTANCE),]
  tempdata <- tempdata[1,] # keep the first row in case of duplicates
  matched[i,] <- as.numeric(tempdata)
  reduced_data <- reduced_data[reduced_data$INPUT_FID!=tempdata$INPUT_FID &
reduced_data$NEAR_FID!=tempdata$NEAR_FID,]
  i <- i+1
  print(i)
}
newdata <- as.data.frame(matched)
names(newdata) <- names(data)
write.csv(newdata,file="output.csv", row.names=F)
```