Julian Plough
Term Project
GIS in Water Resources
12/7/2018

Purpose:

Recent models following the hedonic methodology have demonstrated that flood risk is negatively capitalized into property transaction prices. The goal of this project is to integrate GIS hydrology tools with STATA software in order to recover improved spatial estimates of flooding extent and corresponding housing market shifts versus more spatially naïve models. Using Harris County as a test site, changes in repeat sale home prices from the period 2000-2014 are regressed on flood indicators at various levels of spatial aggregation, and this difference in differences model allows for a recovery of the estimated impact of flooding proximity on home price.

Background:

In order to recover monetary estimates of the value of a phenomenon such as flooding (or of avoiding it) where no explicit market exists, economists turn to non-market evaluation solutions. There are many methods that have been developed in various applied economic and econometric fields, but a dominant one in the analysis of spatially localized environmental (dis)amenities is the hedonic methodology.

Rosen (1974) developed the idea that a home sale can be viewed economically as a sale of a bundle of home qualities, such as square footage, home age, neighborhood crime, etc., and that each characteristic of a home has a corresponding implicit marginal price. Using housing market equilibrium arguments, these marginal prices reflect rational economic valuations of their corresponding non-market phenomena. Hedonic analysis is a quasi-experimental method that exploits natural temporal variation in a housing characteristic of interest, while the remaining characteristics of the home are held, theoretically, constant. This variation must also occur in the period between home sales, allowing the marginal implicit price of the characteristic of interest to be recovered as a determinant of the change in sale price over the period.

In this project, a fixed effects panel data model allows for the control of time invariant components of home price, while flooding rate and extent vary. This is a similar set up to most major papers in the literature (Bin & Landry 2013, Beltran et al 2018), where county level massive events (hurricanes) happen between home sales and the corresponding price changes are regressed on flood zone proximity. These estimates consider large events, and consider parcels as treated/flooded if they are in the 100 year flood zone and in an affected county. This creates issues of establishing proper controls given that massive events usually do not miss neighboring counties, and thus properties outside the 100 year flood zone (or 500) are typically considered the control.

This project seeks to investigate the changes in estimates resulting from a more granular spatial approach than the one mentioned above. This is done using two new approaches to get a more spatially refined estimate of which parcels flooded, as compared to the "industry standard" approach.

Data:

- NOAA Storm Event Database https://www.ncdc.noaa.gov/stormevents/
  - These are fairly large storm events as collected and recorded by NOAA. These are sourced from the NWS, local municipalities and other sources giving them some variability as to the extent to which they are capturing smaller events, however they capture all hurricane events.
  - The events taken from this list are floods ("normal", flash, coastal) tropical storms (hurricanes, typhoons) and heavy rain.
  - These events are at the county level
- USGS Stream Monitors https://waterdata.usgs.gov/nwis
  - These are daily gage height readings from the various stations set up on stream segments.
- Core Logic Systems Housing Database https://www.corelogic.com/
  - The housing transactions over the period were accessed via this private database, as licensed by Dr. Douglas Wrenn, Penn State University.
- Harris County Appraiser's Office http://hcad.org/
  - Spatial parcel data
- USGS National Hydrography Dataset https://viewer.nationalmap.gov/basic/#productSearch
  - Digital Elevation Model
  - Flowline Dataset

Methods:

Econometric Methods and Variable Definition:

First Model:

- This is the "naïve" model, where county level flood indicators (NOAA Storm Event Database) are tallied over 1, 3 and 5 year lags prior to sale. Changes in prices for parcels with repeat sales over the period of study are regressed on the number of events that occurred within some lag prior to sale as below.

  This is also interacted with a 100-year floodplain indicator to determine flooded versus control.

  $$\ln(price)_{it} = \alpha + \beta(F_{CT} * floodplain_i) + \delta_i + \theta_{ct} + \varepsilon_{it}$$

  F_{it}: The number of NOAA defined flooding events for county C (currently there is only one county under study). within some lag T $\in\{1, 3, 5\}$ years prior to sale.

  $floodplain_i$: This is a parcel level indicator that is 1 if the parcel is in 100-year flood zone, 0 otherwise

  δ_i: A (time invariant) fixed effect indicator at the parcel level

  θ_{ct}: A time varying, county by year of sale indicator

Second Model:

- This is an updated model, where USGS flood gages are spatially joined to catchment basins corresponding to the stream reach they appear in. This allows for a spatial linkage to housing parcels that share a catchment basin. A USGS defined flooding events for a station is then determined via an assumed normal distribution, where those gage readings in the upper right tail are tallied as floods. This rule can change, but it is currently set at +2.58 standard deviations above the mean to capture about the top .5% of events per gage.

  For now, the analysis is restricted to those "gage level" events that happened concurrently with "NOAA level" events, where the intuition is to be able to look at a decomposition of flooding effects by definition of localness. There is certainly intuition to look at events that were missed by NOAA as well, but that is beyond the scope for this semester.

  Again, flood zone indicators determine treatment, but flood zones are now subsected by the flood indicator variables within each catchment basin. This can be thought of as a medium grain approach.

$$\ln(price)_{it} = \alpha + \beta(F_{CT} * floodplain_i) + \gamma(F_{CT} * floodplain_i * cat_{bT}) + \delta_i + \theta_{ct} + \varepsilon_{it}$$

cat_{bt}: This is an indicator for the number of USGS gage defined flooding events within t (1, 3 or 5) years prior to home sale. It is defined at the catchment basin level.

Third Model:

- The final model has the finest granularity, where gage heights are spatially interpolated for the stream reaches in Harris County that do not have a USGS station, via those that do. Here, however, the floodplain indicator is replaced by a HAND (height above nearest drainage) indicator that is calculated at the parcel level.

$$\ln(price)_{it} = \alpha + \beta(F_{it} * floodplain_i) + \gamma(F_{bt} * HAND_i) + \delta_i + \theta_{ct} + \varepsilon_{it}$$

GIS Methods and Spatial Analysis

USGS Stream Gages:

- A point file of Harris County USGS stream monitoring stations with gage height readings over the period was assembled via latitude and longitude data. This was then joined with corresponding gage height data.
- These were spatially joined with catchment basins, and then exported to STATA.
- Here, the missing data for unrecorded dates was interpolated for each gage over the 14 year period of study, plus 5 additional years of lag (1995-2014).

Parcel Data:

- The parcel shapefile (~1.5 million parcels) was joined (non spatially) in GIS with the list of parcels with multiple sales over the study period (2000-2014). This reduced it to around 250,000 parcels.
- These parcels were then spatially joined with catchment basins (or reaches), the 100-year flood zone, and HAND estimates.

HAND

- Using a DEM downloaded from USGS, translated into the NAD 1983 State Plane Texas South Central coordinate system (as advised by Harris County), a HAND map was created.
- This DEM was extracted by mask, plus a buffer distance, via a shapefile of Harris County and filled.
- The NHD flowline dataset was also clipped for the analysis, where the dangling vertices were used to weight the flow accumulation and direction raster files that were created via d8 and dinfinity methods, as per exercise 5.
    - The stream link and watershed tools created the stream reach/catchment basin definitions used as well.
- The resolution of this was around 30m cell size, which seemed fine given that most parcels are larger than that so there is no real loss of precision here.
- The final HAND file was converted to a point file, and the minimum value was spatially joined to the parcel shapefile.

Stream Height Interpolation (Progress Report)

Coming into this class, I was imagining making a bunch of flooding extent maps, coming out of the parcel shapefile and changing over time. After trying my hand at the method in this application, I have now realized that the HAND maps are static over time, and it is the changing stream heights that I need to worry about.

This can be done via spatial interpolation that takes the 46 USGS gages and their corresponding 46 reaches, and interpolates heights at the missing 4,000 reaches (an unideal ratio to be sure). While interpolating a raster surface or something to that extent is not too difficult, I need a reading at each reach for every day over the 19 years that cover the flooding data. That means I need to automate this process via an external software to generate the data that I need.

I have experimented with STATA's interpolation toolkit given that I feel much more comfortable running for loops etc., as I will need to spend time figuring out how to automate processes in GIS. While I have code that *works*, I have come across some issues with the USGS data. Even after ensuring that all things are in the same units (meters), there are a few gages that read consistently much higher than others, and have mean heights that are just shy of the maximum HAND values generated in their respective catchment basins. This is perplexing and makes me think that these might be readings above the datum (so total elevation), not water elevation above the riverbed. With this issue, as well as my relative inexperience interpolating such a massive amount of missing data, I do not have results from the 3rd model specification at this juncture. I will include what I have been able to accomplish, but this is certainly the future direction of this project.

Maps:

Harris County Hand vs. Calculated Flow Lines

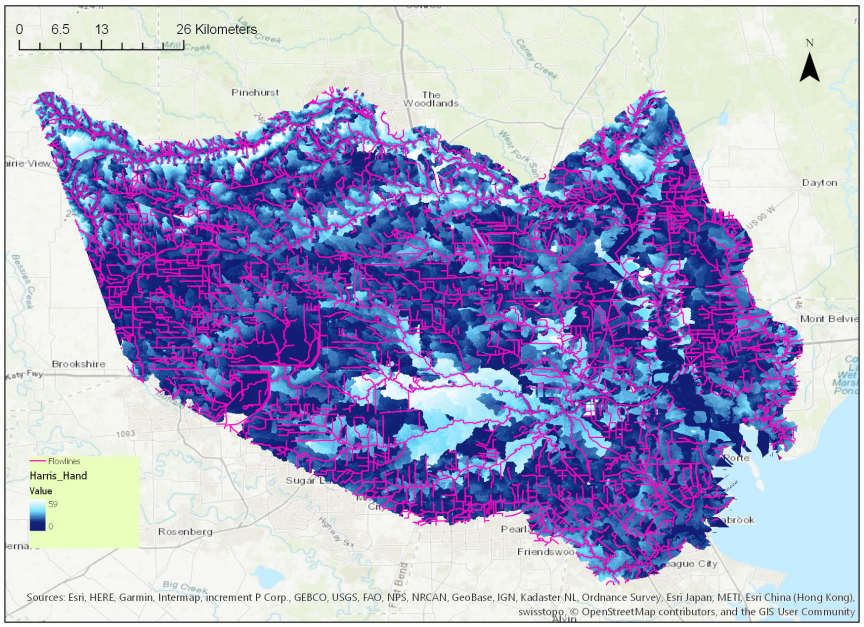

Figure 1

Harris County Flowlines, Catchment Basins and USGS Monitors

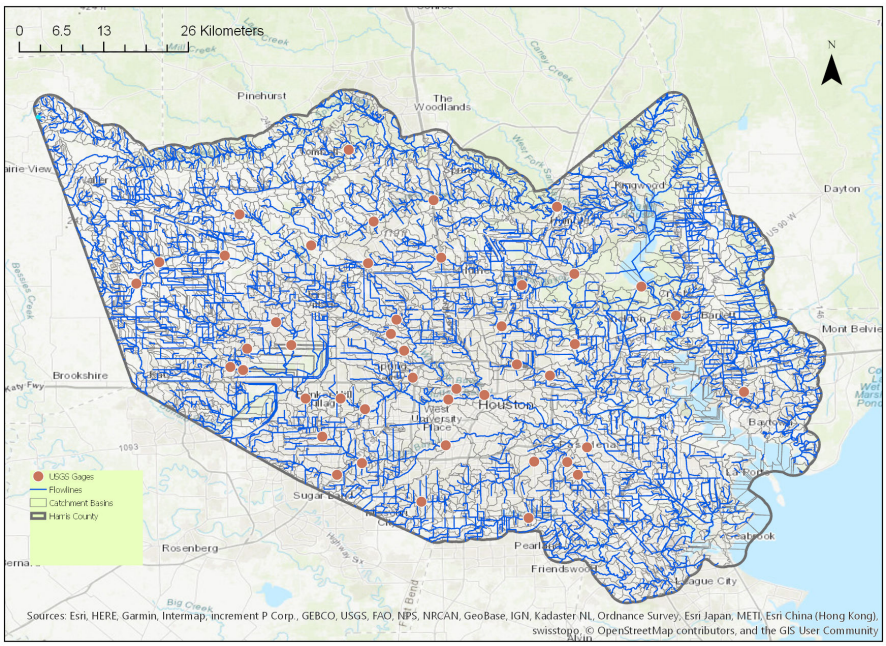

Figure 2

Harris County Repeat Sales Parcels 2000-2014 vs 100 Year Flood Zone

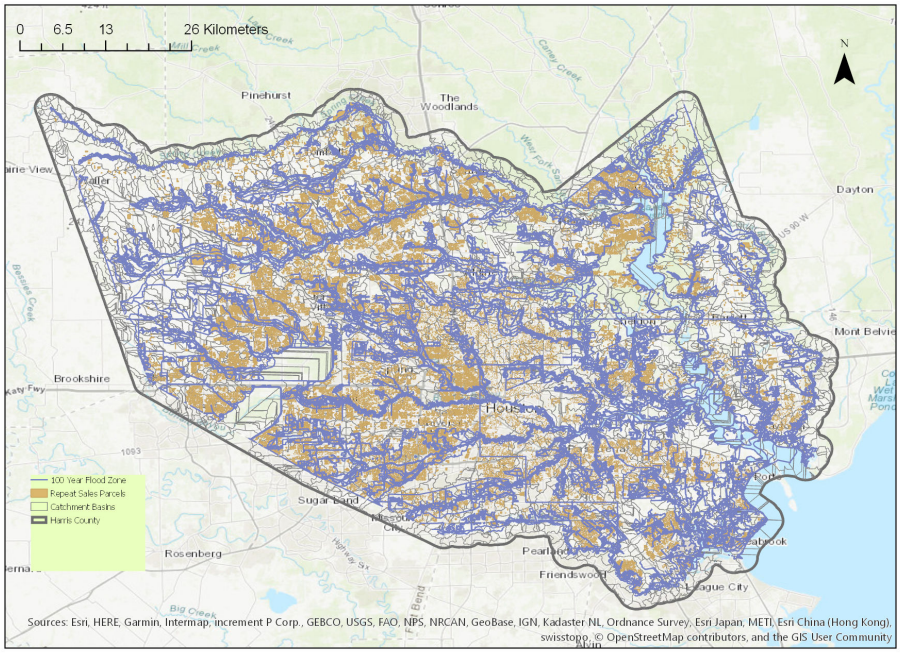

Figure 3

Harris County Repeat Sales Parcels 2000-2014 vs HAND

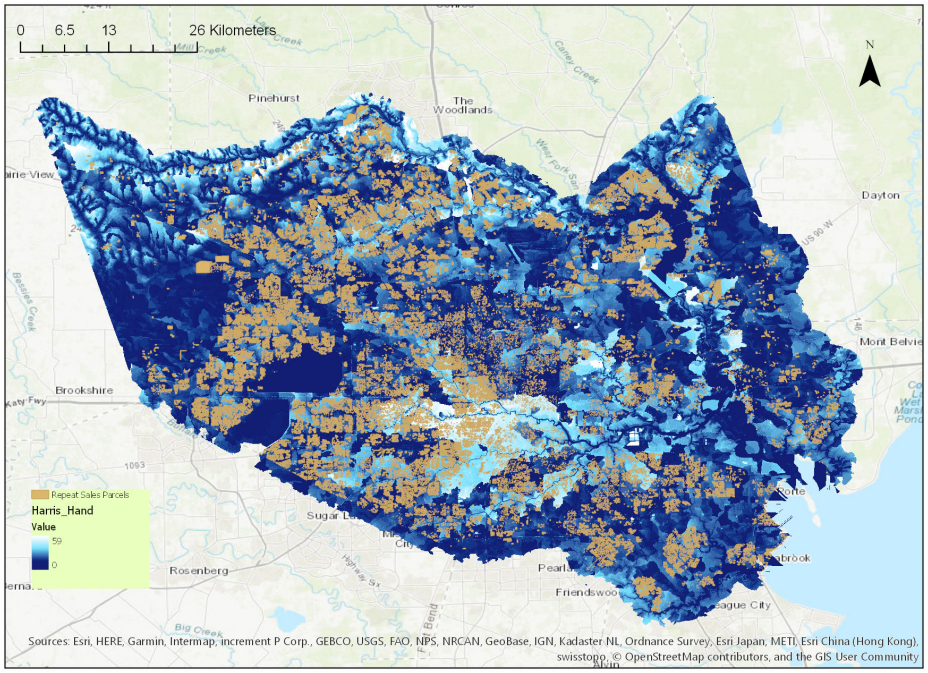

Figure 4

Results:

	(1) lnprice	(2) lnprice	(3) lnprice	(4) lnprice
Total Events*Floodplain (1 Year Prior to Sale)	0.00814***	0.0111***	0.00956***	0.0119***
	(0.00155)	(0.00170)	(0.00160)	(0.00163)
Total Events*Floodplain (2-3 Years Prior to Sale)		0.00186		-.00107
		(0.00127)		(0.00126)
Total Events*Floodplain (3-5 Years Prior to Sale)		0.00699***		.00559***
		(0.00116)		(0.00115)
Station Events*Floodplain (1 Year Prior to Sale)			-0.00753*	-0.00641*
			(0.00295)	(0.00302)
Station Events*Floodplain (2-3 Years Prior to Sale)				0.00102
				(0.00221)
Station Events*Floodplain (3-5 Years Prior to Sale)				-0.000345
				(0.00205)
N	479795	479795	479795	479795
adj. R-sq	0.062	0.062	0.062	0.062

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

For these results, columns 1 and 2 correspond to model 1, and columns 2 and 3 correspond to model 2. In column 1, there is a significant and positive association between additional flooding events and home price, but this is misleading. There is no inter county variation (at this stage) so this is really more of an indicator for flood zone presence and lacks a full interpretation, but given the log linear specification, this can (read: shouldn't) be interpreted as an additional NOAA defined flooding event within the year prior to sale increases home sales prices by .81% for homes within the 100-year flood zone. This is likely picking up the premium for water proximity. Without inter county variation for water-proximate properties allowing some to flood and others to stay dry, this model cannot properly capture the marginal effect of flooding as untangled from water frontage premiums. A similar assessment can be made for column 2 which adds lags for events in 2-3, and 4-5 year bins prior to sale.

Column 3 uses 1 year lags for events prior to sale, but incorporates catchment basin level flooding indicators for homes. This allows for inter basin variation in flooding to control for the water-proximity premium as described above, and can be more confidently interpreted as a marginal effect of an additional flooding event. The coefficient was found to be significant at the 95% level, it is in the negative direction and is on the same order of magnitude as the county level event coefficient. In column 4, additional lags for events in 2-3 and 4-5 year bins have been added. The estimate for USGS events doesn't attenuate much, and remains statistically significant. However, events for the 2-3 year bin and 4-5 year bin are not statistically different from zero. Column 4 is the preferred specification theoretically, and the coefficient on USGS events can be interpreted as follows; an additional USGS defined flooding event in the catchment basin of a 100 year flood zone home within a year prior to sale reduces sale price by .641%.

Discussion:

Columns 1 and 2 of the results table can be viewed as fairly similar to the kind of results in the extant literature, save for the lack of inter county variation. This limits the interpretation of the coefficient, but does not limit the interpretation that this level of indicator is likely missing something. Adding in a measure of localized flooding proximity at the catchment basin level in columns 3 and 4, it is apparent that negative and significant effects are coming out of finer grained estimates of flooding proximity. This could be because the inter-catchment variation allows for some control of the water proximity premium, but is likely an effect that would persist even if other counties were included (I have confirmed this while looking at Tampa Bay previously, as presented in class).

The effects of flooding are considered, at least econometrically, in vague terms. It is unclear if the interest is in the effect of property damage, nuisance flooding that shuts down roads/schools etc, larger level stigmas, or something else. It is also unclear where government programs like NFIP may be insulating property owners or municipalities in general. However, the results here suggest that finer grained spatial analysis can begin to unpack this behavioral picture that appears confusing from the outside-that people continue to live in flood risk that they should be aware of.

From these results, it is clear that there are additional localized effects of flooding that extant valuation efforts fail to pick up. This suggests that continuing work towards a full definition of model 3 could continue to refine estimates. Once estimates can be calculated at a parcel level, the picture should become clearer of how exactly home owners are behaving in regards to perceived flood risk. After that point, deviations from rational expectations can be compared to NFIP rates, informational shocks/deficits and other possible explanatory variables to continue to unravel this initially unintuitive set of results.

Future work will ensure that flood gage levels are being interpreted properly (I am not concerned about the estimates from model 2 since flood is indicated by standardized deviance and not absolute) in order to pursue interpolation at a large scale over the 4,000+ gages and 7,000+ days. Results suggest fruitful returns to integrating HAND estimates into the analysis, but a thorough investigation of interpolation methodologies will be necessary a priori.

References :

Beltran, Allan, et al. "Is Flood Risk Capitalised Into Property Values?" Ecological Economics, vol. 146, 2018, pp. 668–685., doi:10.1016/j.ecolecon.2017.12.015.

Bin, Okmyung, and Craig E. Landry. "Changes in Implicit Flood Risk Premiums: Empirical Evidence from the Housing Market." Journal of Environmental Economics and Management, vol. 65, no. 3, 2013, pp. 361–376., doi:10.1016/j.jeem.2012.12.002.

Rosen, S. (1974) Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. Journal of Political Economy, 82, 34-55.
https://doi.org/10.1086/260169