The Effects of PFAS Contamination on the Michigan Housing Market

Eric Fegan

ejfegan@umich.edu

Senior Thesis

Economics 495

Professor and Advisor Martha Bailey

16 December 2019

Abstract

This paper seeks to quantify the effects of PFAS contamination, and public knowledge about it, on the Michigan housing market. Given the rapid absorption of news in today's media environment, and that the contamination has already occurred, this presents a unique opportunity for insight into the effect of new information on consumers. The study finds that, while the effect on prices is statistically insignificant, there is an effect on home turnover; it concludes that the trend of home turnover appears to decrease in contaminated zip codes after PFAS information is released to the public, relative to trend changes in non-contaminated zip codes.

I. Introduction

In the past two years, new media reports have arisen regarding contamination of Michigan water supplies with per- and polyfluoroalkyl substances, or PFAS. Often found near manufacturing sites and military bases, these chemicals have unknown effects, but concerns about the potential harms to health from PFAS are widespread.[1] The Michigan Department of Environment, Great Lakes, and Energy (EGLE) has conducted testing and identified at least 71 sites statewide that are above the Environmental Protection Agency (EPA) limit for acceptable levels of PFAS in groundwater of 70 parts per trillion (ppt),[2] and other

---

[1] While a definitive causal relation may be unclear at this time, "[e]xposure to PFAS has been linked in human studies to some cancers, thyroid disorders, elevated cholesterol and other diseases." Ellison, Garret. 2018. "Michigan atop national PFAS site list." *MLive*, April 18. https://www.mlive.com/news/2018/04/ewg_2018_pfas_map_report.html (accessed 5 November 2019).

[2] This number is current as of 5 November 2019; new sites continue to be added. Michigan PFAS Action Response Team. 2019. "Michigan_PFAS_Sites_-_71_-_Table_669885_7." Michigan Department of Environment, Great Lakes, and Energy. https://www.michigan.gov/pfasresponse/0,9038,7-365-86511_95645---,00.html (accessed 5 November 2019).

groups have found some other sites not listed by EGLE.[3] Test results, combined with additional reports that even the 70ppt threshold may be too high,[4] have given rise to concerns that there are dangerous levels of chemicals in much of the drinking water of Michigan residents.

While there have been numerous studies into the effect on property values of pollution, few instances exist of pollution that was not discovered until long after the potential positive impacts of whatever created the pollution had been internalized by residents. Notable exceptions include the Flint, Michigan, water crisis, though this did not present the opportunity for controls that a statewide analysis of dozens of contamination sites provides.[5] Further, this study is apparently the first statewide analysis of the effects of PFAS specifically.

The situation in Michigan motivates the question: what is the information effect of PFAS contamination on property values, home buying, and home selling decisions?

This study will attempt to quantify the effect, if any, of the discovery of these PFAS sites on surrounding property values, as well as other housing market outcomes. The PFAS sites are specific to the street address, but the analysis will focus on the zip code level.[6] I am particularly interested on what mobility exists, if any, for residents fearful of PFAS contamination in their drinking water.

---

[3] Current as of 5 November 2019. PFAS Project. 2019. "Public SSEHRI PFAS Contamination Site Tracker – last update 9.25.2019." Northeastern University Social Science Environmental Health Research Institute. https://pfasproject.com/pfas-contamination-site-tracker/ (accessed 5 November 2019).

[4] The Michigan PFAS Action Response Team "recommends setting drinking water limits as low as six parts per trillion." Thorp, Ben. 2019. "Science advisory group recommends much stricter PFAS standards for Michigan." *Michigan Radio*, June 27. https://www.michiganradio.org/post/science-advisory-group-recommends-much-stricter-pfas-standards-michigan (accessed 5 November 2019).

[5] Notable exceptions also include the discovery of CO_2 as a greenhouse gas, but that is a mostly global pollutant, and well beyond the scope of this paper.

[6] Zip codes are generally small in areas of high population density, and large in more rural areas. This may mean that any response will be stronger in denser zip codes.

Ultimately, I would like to quantify the revealed value of potential PFAS contamination; given that the PFAS pollution has occurred over time, and related positive effects have presumably been internalized, this will provide insight into the pure information effect on consumers. If the analysis reveals that, as compared to consumers in the control group, consumers fail to sell homes in zip codes near contamination sites, or that housing prices fell in the months following the first reports of contamination, this could indicate that knowledge of PFAS contamination has a tangible effect on housing market decisions. If there is little or no evidence of a housing impact, this could suggest the equally intriguing conclusion that values placed on potential contamination are low for consumers who have already made their choice about where to live.

II. Literature review

This project modifies the approach taken in much of the vast set of existing literature on pollution's effect on property values. Much of that literature takes the form of hedonic analysis, whereas the model used herein will rely largely on a difference-in-differences (DID), or more accurately, difference-in-trends, approach. The use of data from Zillow, which are available monthly and as far back as 2008, allow for this, and DID is used with the goal of avoiding much of the bias inherent in hedonic analysis.

The use of hedonics to estimate environmental effects—positive and negative—on property values is well-established. Analyses including Affuso et al (2010), Hanna (2007), and Hite et al (2000) rely heavily on this approach, using a basket of home characteristics as controls in order to single out the effect of negative environmental factors on home values. Affuso et al (2010), in studying the effect of lead contamination in an Alabama city on local property values, utilized a hedonic model with controls for various home characteristics, such as price, land area, and number of rooms, along with controls for distance to water and to a local

chemical center. The input of interest was lead contamination in the soil, which they found to be statistically significant in a two-stage least squares specification, at least within the hedonic framework. Similarly, Hite et al (2000) created a single-period hedonic model estimating the effect of landfills on property values, adding in an intriguing spatial component that measured distances to local landmarks and amenities. Hanna's (2007, p. 1) analysis "assesses the effect of new, polluting manufacturing facilities on both residential house values and neighborhood incomes." This comes closer to the type of analysis required for the situation in Michigan; Hanna's model allows for the introduction of polluting manufacturing facilities in order to attempt to assess the before-and-after effect of the sites.

However, these analyses all lack, to some degree, a true control for homes that are not near the contamination or environmental issue in question. They are restrained to single localities, so the potential for selection bias (and a potential omitted variable bias) in the location of contamination is readily apparent, despite the analyses' other merits. This lack of a true control leaves open the possibility that such analyses will overstate or understate the effect of contamination, which could be correlated with factors such as income. There is a danger of finding of a causal effect where there is none, or finding no causal effect when one does, in fact, exist.

One such source of potential bias is explored in depth by Michael et al (2000, p. 283); their analysis finds that, perhaps unsurprisingly, most studies "are conducted using whatever empirical measure(s) of environmental quality is (are) available as the environmental variable(s), and *these measures of environmental quality may not represent public perception of water quality*."[7] They scrutinize several measures of the hypothetical environmental variable, and find that "subjective choices by the investigator can affect the significance and magnitude

---

[7] Emphasis mine.

of implicit prices." They then focus on changes in the level of a suggested measure for contamination in Maine freshwater lakes, and explore how values changed as the contamination changed over time.[8]

Taking a different tack, and in one of the closer analogs to this project, Kiel (1995) extended existing literature around pollution from Superfund sites in the Boston area. Kiel (1995, p. 429) focuses on the effect of new information about pollution on the market, noting the possibility "that the adjustment process of the market in response to the incinerator is different from that in response to a toxic waste site due, in part, to the way that information is released to the market."

While the dataset does not contain enough data to discover a time trend, Kiel divides the data into several periods, representing time before information about pollution, the discovery period, a period of EPA announcements about contamination, and a period of announcements about cleaning the sites. Kiel (1995, p. 434) mentions odors and other signals at some sites, perhaps explaining the negative effect preceding EPA announcements, and finds that "information from sources including the EPA and local community groups does impact sales prices; thus at least in this case the EPA announcements seem to provide some additional information to the residents." In this way, Kiel's approach has a hedonic framework with a difference-in-differences expansion.

Kiel's model is based on summing the discounted future value of market clearing rents for houses, and implies that an expectation that a site being more contaminated than previously known would decrease price, while cleanup efforts would increase price. She uses the hedonic approach to estimate values, and finds statistically significant coefficients for distance to contamination once physical

---

[8] It does not deal with the selection issue of measurement, though it does comment on the effect—on hedonic analysis—of different measurement types.

indicators like odor were noticeable.[9] In keeping with other hedonic analyses, Kiel's does not have a proper control group.

Taylor et al (2016, p. 4) present an analysis that attempts to find the impacts of environmental contamination distinct from the effects of "undesirable" land use, addressing in explicit terms the omitted variable bias inherent in predominantly hedonic analyses.

> "We explicitly recognize that hazardous waste sites are commercial or industrial properties … that may be undesirable neighbors, irrespective of their environmental status. … Key to our estimation approach is the explicit consideration of uncontaminated commercial properties, which provide a benchmark against which we can compare any residual price impacts of remediated hazardous waste sites, and thus determine the degree to which stigma exists."

Using proximity to contaminated sites alongside uncontaminated (but otherwise similar) sites, Taylor et al construct a difference-in-differences model to dampen the potential biases of hedonics.

It is in this spirit that I will aim to quantify the effects of PFAS contamination on Michigan housing market decisions. Like Kiel (1995), this study will focus on the effect of new information about existing contamination; this will eliminate one of the land-use problems, namely that when new polluting locations are constructed, changes in perception are built on the economic impact of the new site as well as potential pollution. Given the widespread PFAS contamination across the state of Michigan, the study will further employ a difference-in-differences (difference-in-trends) model with the intent of avoiding the inherent problem of excluding non-contaminated sites from the analysis.

---

[9] The effect of EPA announcements was apparently dampened by the effect of these already-existent indicators.

III. Data sources

All data are current as of 9 November 2019, and are summarized below. They are divided into data on housing, data on PFAS contamination and knowledge about it, and economic indicators.

A. Housing data

Housing data are from Zillow's publicly available database. Zillow has average house prices, number of completed sales, average percentage of price cut, new monthly inventory, and more, all at the zip code level and all several years before the bulk of the release of new PFAS information.[10]

While Zillow claims the data are quite accurate, they mention that there can be a lag in full information being reported, noting this latency can be "anywhere from a few days to multiple months".[11] For this reason, the dataset used herein is limited to observations before and during August 2019. Their data are coded consistently over time, and each file contains the entire historical series for the variable in question.

Zillow has sporadic missing values for certain zip codes for each outcome variable mentioned below, though their data for turnover are largely complete since 2008. To avoid the missing values in each of their datasets, the data mapping each zip code to its respective county is from a third-party site.[12]

B. PFAS data

The data on PFAS, and specifically information about PFAS contamination, are compiled from the PFAS Project at the Social Science Environmental Health

---

[10] Zillow. 2008-2019. "Housing Data." https://www.zillow.com/research/data/ (accessed 9 November 2019).

[11] Zillow. 2019. "Methodology: Home Sales." https://www.zillow.com/research/home-sales-methodology-7733/ (accessed 13 November 2019).

[12] Zip Codes to Go. 2007-2018. "Zip Codes for the State of Michigan." https://www.zipcodestogo.com/Michigan/ (accessed 13 November 2019).

Research Institute (SSEHRI) at Northeastern University[13] and from the Michigan Department of Environment, Great Lakes, and Energy's (EGLE) Michigan PFAS Action Response Team (MPART) website.[14]

C. Economic data

A handful of economic indicators are pulled from various sources. (While some of the data are dated, they are used herein as static controls. The assumption is that these data have not significantly changed *relative to each other* since their publication.) Rurality data were compiled by the University of Michigan Population Studies Center.[15] Income data were pulled from Zip Atlas.[16] These are collinear with fixed effects at the zip code level, and will be used for inverse propensity weights, not regression analysis.

IV. Variables and methodology

This analysis utilizes numerous input and output variables, and a synthesis of several economic and statistical methods. Several assumptions need to be made in the use of these variables and methods, as detailed in Section IV.C.

A. Variables

Input variables are as follows. *Zip* is a dummy for zip code, varying only with zip code. *Year* is an integer (later used as a factor variable), varying with time. *Month* represents a group of 11 month dummies, varying with time. *PFAS* is a dummy for the presence of PFAS above 6ppt, varying with zip code. *Level* is the

---

[13] SSEHRI. 2019. "PFAS Contamination Site Tracker." https://pfasproject.com/pfas-contamination-site-tracker/ (accessed 9 November 2019).

[14] Michigan PFAS Action Response Team. 2019. "PFAS Sites." https://pfasproject.com/pfas-contamination-site-tracker/ (accessed 13 November 2019).

[15] University of Michigan Population Studies Center. 2003. "Measures of rurality for zip codes in the United States." https://www.psc.isr.umich.edu/dis/acs/faq/answer/1102 (accessed 23 November 2019).

[16] Zip Atlas. 2019. "Zip Codes with the Highest Median Household Income in Michigan." https://zipatlas.com/us/mi/zip-code-comparison/median-household-income.htm (accessed 16 November 2019).

level of PFAS contamination, ranging from 0 to 4, and varying with zip code. Those instances with *level* of 2 (corresponding to a level between 6ppt and 70ppt) or greater are where *PFAS*=1. *Post* is a dummy denoting whether public information about PFAS had yet been released, varying with zip code and time, and *county* represents a factor variable for the county, varying with zip code.

There are several output variables of interest. The obvious choice is perhaps median sale price (by zip code), and these data are available. However, given the likelihood of some level of stickiness in prices, a more interesting output variable may be turnover.[17] Other variables that may have a noticeable effect, if one exists, before a change in prices include average percentage price reduction, monthly listings, and new monthly listings.

As shown in Table 1, the contaminated and non-contaminated groups of zip codes do not appear to be a random selection; those zip codes with contamination tend to have more turnover and higher prices. This motivates the inclusion of several static, economic controls that could lend insight into the inherent differences between the two groups, and perhaps explain some of the variation found between the two groups.

These additional, static controls, all varying by zip code, are as follows. *Density* is the log of population density. *Population* is the population of the zip code. *Rurality* is a measure of the rurality of the zip code, with higher values denoting more rural areas. Likewise, *urban influence* rises with increasing rurality. *Income* is the median household income for the zip code, and *income rank* is the rank of those incomes, with higher values for *rank* corresponding to higher values of *income. University* is a dummy for the presence of a four-year public university in the same city.

---

[17] And, of course, if turnover is changing significantly, it too would have an effect on prices in the short run.

Some of these static controls are inherently correlated, such as log density, rurality, and urban influence. Since they are only being used to construct a predictor for the presence of PFAS, this will not diminish their utility in the analysis.

From the Zillow data, the output variables are as follows; all vary with zip code and time. *Turnover* is the number of sales per month. *Listings* is the number of active listings per month. *New listings* is the number of new listings per month. *List price* is the average list price for the zip code, while *sale price* is the average (actual) sale price. *Percent price reduction* is the average price reduction, as a percentage, on listings in that zip code.

Summary statistics for many of these variables appear in Table 1 below.

Table 1. Summary of Variables

	CONTAMINATED		NON-CONTAMINATED	
Variable	Mean (Std. Dev.)	Min/max	Mean (Std. Dev.)	Min/max
turnover	29.18 (22.82)	0/163	11.47 (17.83)	0/223
listings	148.79 (70.26)	22/427	99.65 (72.86)	11/563
new listings	40.75 (17.01)	11/123	42.51 (19.32)	11/144
list price	$183,571 ($93,539)	$12,400/ $529,018	$190,926 ($105,788)	$26,000/ $812,000
sale price	$169,383 ($70,580)	$40,200/ $444,900	$160,562 ($74,309)	$21,200/ $503,700[18]
percent price reduction	5.16% (2.34%)	0.14%/ 23.86%	5.54% (3.27%)	0.00%/ 48.28%
pfas	1 (0)	1/1	0 (0)	0/0
pfas level	3.20 (0.58)	2/4	0.00[19] (0.03)	0/1

---

[18] A lack of data, particularly in prices, leads to the apparently contradictory discrepancies between the various price measures.

[19] There are non-zero values for level in the non-contaminated group because for this analysis, contamination is defined to be existent where PFAS levels of over 6ppt have been found.

Variable	Mean (Std. Dev.)	Min/max	Mean (Std. Dev.)	Min/max
log density	6.08 (1.45)	3.21/8.81	5.00 (1.91)	-0.05/9.25
rurality	2.85 (1.84)	1/8	4.17 (2.65)	1/9
urban influence	3.10 (2.64)	1/12	4.75 (3.54)	1/12
population	22,584 (12,749)	2,155/ 49,086	9,620 (12,350)	21/67,012
household income	$44,614 ($11,746)	$22,146/ $77,274	$43,122 ($14,386)	$11,667/ $112,809
university	0.21 (0.41)	0/1	0.04 (0.21)	0/1

Notes: *turnover* through *percent price reduction* are Zillow data, 2008-2019 (accessed 9 November 2019). *pfas* and *pfas level* are from my compilation of PFAS Project (accessed 5 November 2019) data and MPART data (accessed 5 November 2019). The remainder are from the University of Michigan Center for Population Studies (accessed 23 November 2019). *Rurality* rises with rurality, and *urban influence* with urbanity. *University* is a dummy for presence of public four-year university. The left two columns are statistics for contaminated zips; the right two are for non-contaminated zips.

As discussed briefly in the literature review, this analysis is structured in the difference-in-differences (or difference-in-trends) framework. The known contaminated zip codes and the as-yet uncontaminated zip codes vary greatly in the behavior of the housing market. Between the contaminated and non-contaminated groups, the median sale price by zip code exhibits divergent behavior in the years following the Great Recession, though since at least 2014, these have tracked somewhat closely to each other. (See Figure 1 below.) This presents an opportunity for a DID analysis without differential trends.

Figure 1. Average Sales Price for Contaminated and Non-Contaminated Zip Codes
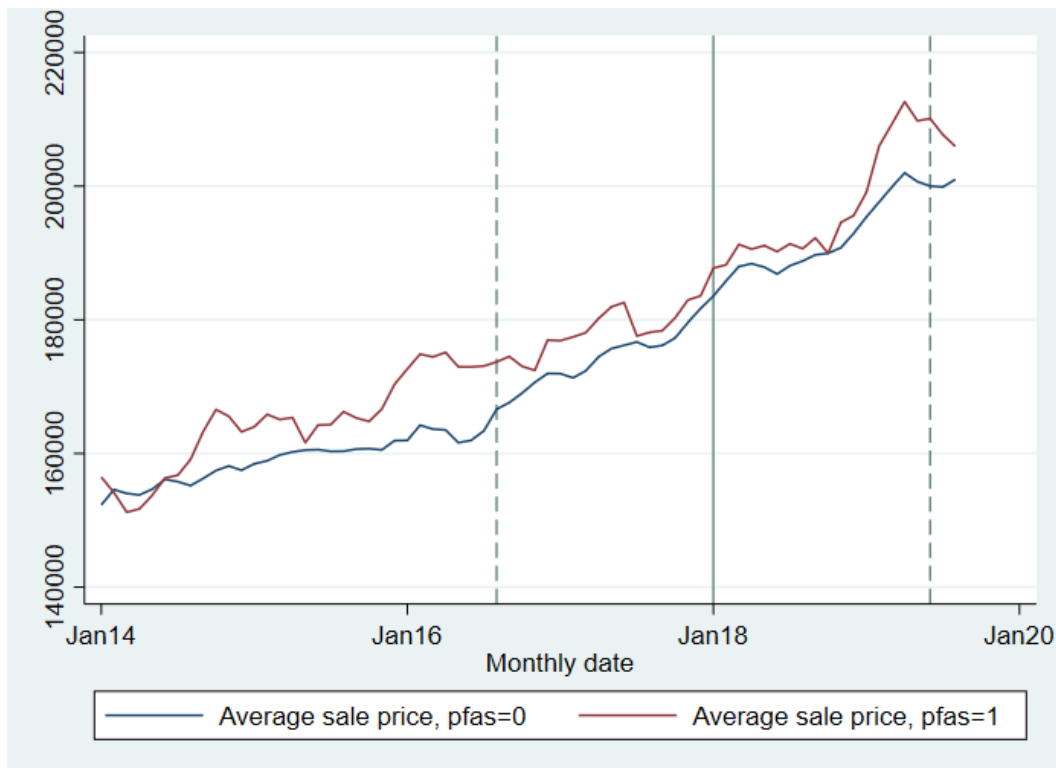
Note: *pfas*=0 is the average for non-contaminated zips, and *pfas*=1 is the average for contaminated zips. The vertical lines represent the mean date for post, and one standard deviation in either direction. Data are from Zillow (accessed 9 November 2019).

However, there is a clear change in the trends for the contaminated and non-contaminated groups when it comes to turnover within the zip code once PFAS information begins to spread.[20] (See Figure 2 below.) Further, the pre-trends are differential. The analysis therefore takes on the difference-in-trends framework to allow for (and quantify) the changes in trends, and must prioritize turnover, which if affected, would itself be confounding the effect on prices.

---

[20] Based on my analysis, as of 9 November 2019, 25% of current known PFAS sites were known by February 2018; 50% by July 2018; and 75% by October 2018.

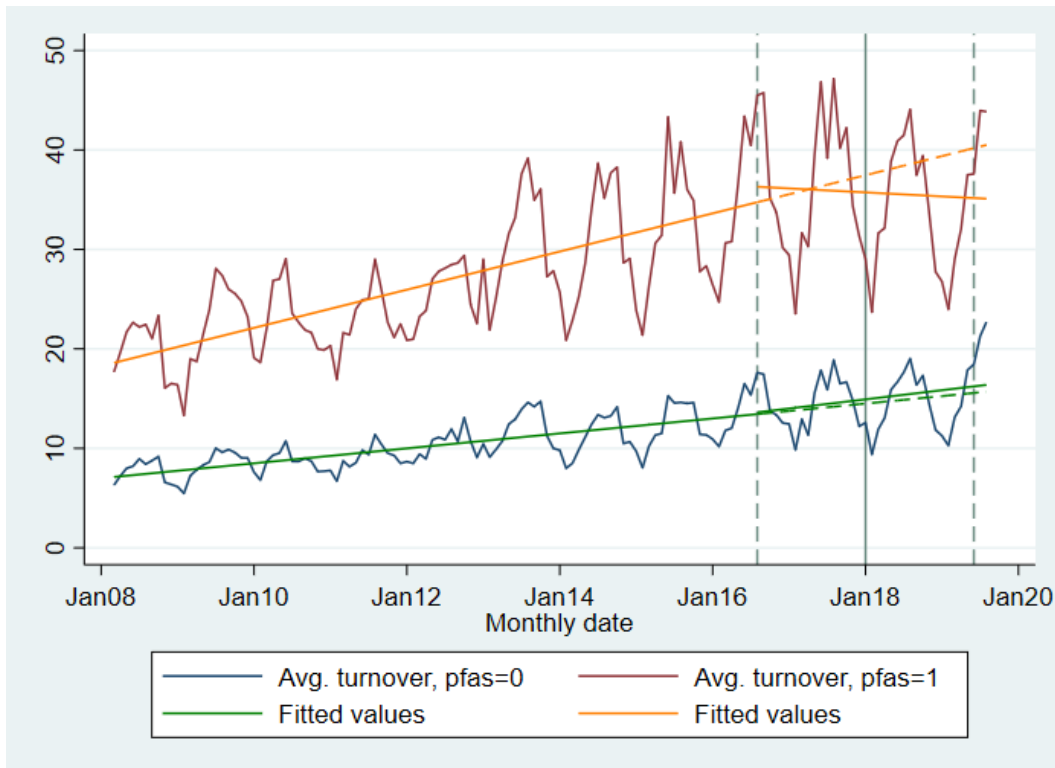**Figure 2. Average Turnover Rates for Contaminated and Non-Contaminated Zip Codes**

Note: *pfas*=0 is the average for non-contaminated zips, and *pfas*=1 is the average for contaminated zips. The vertical lines represent the mean date for *post*, and one standard deviation in either direction. Dashed lines for fitted values represent projection based on the time trend before one standard deviation before the *post* average; solid lines are actual. Housing are from Zillow (accessed 9 November 2019); *pfas* and *post* data are compiled from the PFAS Project (accessed 5 November 2019) and MPART (accessed 5 November 2019).

Making this change clearer, see below the difference in turnover trends between the two groups. Figure 3 demonstrates the clear shift in the difference between the two groups' turnover rates around 2017-2018, perhaps signaling changes due to PFAS contamination.

Figure 3. Difference in Average Turnover
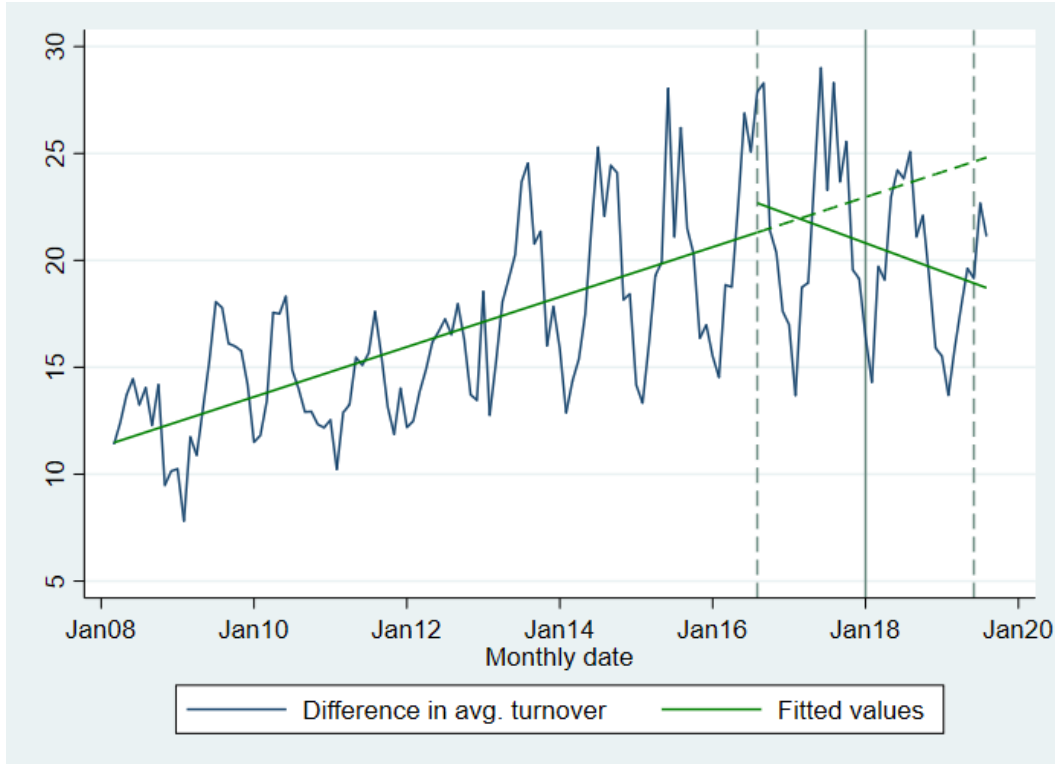
Note: this is the average turnover in contaminated zips minus the average turnover in non-contaminated zips. The vertical lines represent the mean date for post, and one standard deviation in either direction. The dashed line for fitted values represents projection based on the time trend before one standard deviation before the *post* average; solid line is fitted for actual. Data are from Zillow (accessed 9 November 2019).

B. Model construction

Taking into account the apparent shift in turnover, the model is constructed around the effect on turnover, turning to prices later. The primary model will aim to identify any difference in effects of the post period on the contaminated and non-contaminated groups. This model will take on the following form:

(1) $Y = \beta_0 + \beta_1 post_{zt} + \beta_2 post_{zt}*pfas_z + \beta_3 year_t + \beta_4 year_t*pfas_z + \beta_5 year_t*post_{zt} + \beta_6 year_t*post_{zt}*pfas_z + \beta_7 jan_t + \beta_8 feb_t + ... + \beta_{17} nov_t + \beta_{18z} zip_z + \varepsilon_{zt}$

The intuition behind these coefficients is as follows. β_0, β_1, and β_2 are intercepts: the coefficient for the earliest date in the non-contaminated group, the

earliest date for the trend-break line for the non-contaminated group, and the earliest date for the trend-break line for the contaminated group, respectively. β_3 represents the base slope of the pre-trend-break line, while β_4 is the additional slope on this line for the contaminated group. β_5 is the base slope for the post-trend-break line, while β_6 is the additional slope on this line for the contaminated group. β_7 through β_{17} are the coefficients for months, and the seasonal trend. β_{18} represents the many coefficients of the zip code dummies. The goal is to construct a framework like that of Figure 4 below, wherein the difference between trend breaks is observed.

Figure 4. Analytical Construction
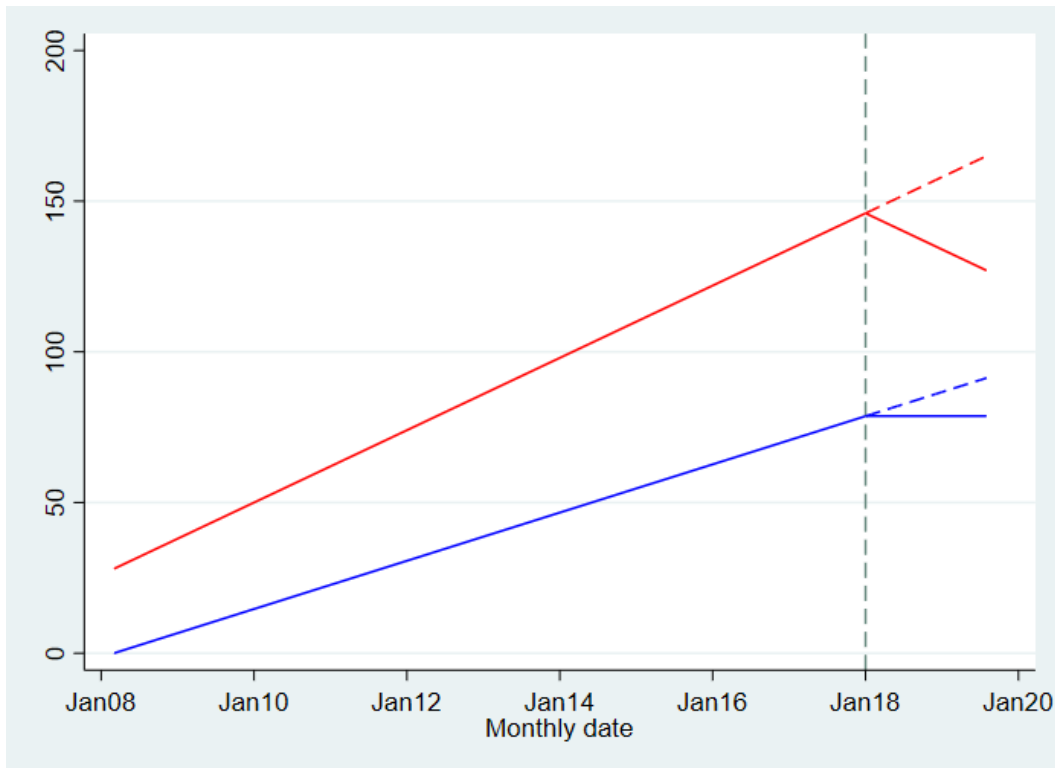
Note: In this example, there are two groups: red and blue. The dotted x-line represents the beginning of the *post* period. Dotted lines show the projection of the pre-trends. Of interest are the trend breaks, and in particular, the difference between their slopes.

16

In this framework, β_6 will be the coefficient of interest, representing the additional change in trend once PFAS information is public, or the only change, if β_5 is insignificant.

Following the main analysis, the study will examine the same model in an event-study framework, and later, adjust for differences in the propensity of zip codes to be contaminated with PFAS.

C. Assumptions, detailed

In coding the *post* variable, assumptions are needed regarding how to define *post* in the non-contaminated zip codes. The various choices for this are somewhat arbitrary, so in the interest of neutrality, the average date for PFAS information publicity (January 2018) is used as the *post*-date for the non-contaminated group. (An alternative would be the median, July 2018; see Table 2 below.) The analysis assumes that there is significant enough public information by this point that home buyers and sellers in non-contaminated zip codes are making their decisions with full information about the PFAS crisis in Michigan.

Table 2. Summary of *post* Variable Over All Zip Codes

Percentile	10%	25%	50%	75%	90%
post month	Jan. 2015	Oct. 2017	July 2018	Oct. 2018	May 2019

Notes: Data are from the PFAS Project (accessed 5 November 2019) and MPART (accessed 5 November 2019).

The other main assumption lies in how to determine which zip codes to include in the contaminated group, or more specifically, what level of PFAS contamination warrants inclusion in the contaminated group. This is again an arbitrary choice; the EPA limit of 70ppt is one option; the analysis could further include a handful of zip codes that are over 6ppt, based on the circulating studies that show PFAS to be harmful at much lower levels. This study assumes that levels over 6ppt would be sufficient for members of the public to make choices more in

line with the highly-contaminated zip codes than with uncontaminated (or unknown) zip codes. This means that *pfas*=1 when *level* is 2, 3, or 4, and *pfas*=0 otherwise.[21]

Finally, for economic controls, the study will assume that the relative levels of indicators, including for rurality and median household income, have not, relative to each other, changed sufficiently since their collection and during the period of this analysis to greatly change any inference that can be drawn from them. These controls are collinear with fixed effects at the zip code level, and are used only for the calculation of propensity to be contaminated with PFAS, not for the difference-in-trends regressions.

V. Results and analysis

Below are detailed the results of this study. First, results are presented for the regression analyses described in Section IV.B. Next, these regressions are re-figured for a corresponding event study analysis. Finally, probability weights are applied to the original models as a correction and robustness check.

A. Regression models

With *year* as an integer variable, the first analysis is based on the equation (1) above. (This model is hereafter called Model 1.) In this framework, the results indicate that there is an overall negative effect once the *post* period begins, with β_5 equal to -1.11, meaning that for each year into the *post* period, turnover is 1.11 below the pre-trend, on average. Additionally, the coefficient on the triple interaction, representing the additional downward influence in PFAS-contaminated

---

[21] In either case, the number of zip codes for which *level*=2—that is, PFAS levels are between 6 and 70 ppt—is relatively small compared to the population.

zip codes, is -0.826. (However, this is insignificant at the 10% level.) These results are summarized as in Table 3 below as Model 1.[22]

If, however, the assumption of completely linear trends is relaxed, and instead the base trend is modeled with dummies, all the interactions are significant at (at least) the 5% level. This model, replacing the *year* integer with *year* dummies, is hereafter called Model 2. In particular, in this model the base *post* trend (on *post*year*) is positive, so that turnover appears to be rising in comparison to the pre-trend in non-contaminated zip codes. The coefficient for the triple interaction, β_6, is equal to -3.772 at a 1% significance level, so that turnover in the contaminated zip codes is falling against the base trend, β_5. This indicates that for each year into the *post* period, turnover in contaminated zip codes is estimated to be 3.772 (per month) below what it would have been without new information about contamination, on average. Taken together, Model 2 estimates that the contaminated zip codes are falling by 0.631 per year in the *post* period against the base pre-trend,[23] or 3.772 below the base *post*-trend. These results are summarized as Model 2 in Table 3 below.

This brings up the issue of controls. It would be ideal to include at least some economic controls that vary over time at the zip code level in this analysis. However, given that much of the change in the real estate market occurred in 2018 or later, detailed information, such as that from the American Community Survey, is as yet unavailable. As an alternative, the next model includes an interaction control between *year* dummies and *county* dummies. This helps to control for much of the time-variant and regional economic changes over time, so that the zip codes are not viewed in isolation. This model is hereafter referred to as Model 3. Including this control, as shown in Table 3 below, still gives a result for the triple interaction

---

[22] In all tables, values for coefficients on zip code are omitted. There are 988 dummies for zip code in this dataset.
[23] Based on β_5 - β_6 = -0.631.

that is significant at the 1% level. (This is also true where the year variable is an integer and the county interaction is included.)

Table 3. Difference in Trends for Effect on Turnover

Variable	Model 1	Model 2	Model 3
year	0.595***	[factor]	[factor]
	(0.033)		
post	2239.595***	-6338.563**	-5137.925**
	(240.525)	(2652.438)	(2457.744)
post*pfas	1665.715	7611.033***	5775.185***
	(1225.114)	(2430.753)	(2233.405)
post*year	-1.110***	3.141**	2.546**
	(0.119)	(1.314)	(1.218)
pfas*year	0.939***	1.006***	0.748***
	(0.207)	(0.209)	(0.173)
pfas*post*year	-0.826	-3.772***	-2.861***
	(0.607)	(1.204)	(1.106)
county*year			[factor]
Adjusted R^2	0.878	0.879	0.890

Notes: [factor] indicates that this variable is included as a factor variable with many coefficients. Blank cells indicate that a variable was not included for the given model. Model 1 uses *year* as an integer, Model 2 uses *year* as factor variables, and Model 3 is the same as Model 2 but adds interactions for *county* and *year*, which are both factor variables. Data for *pfas* and *post* are from the PFAS Project (accessed 5 November 2019) and MPART (accessed 5 November 2019); turnover is from Zillow (accessed 9 November 2019).

Notably, the three models generally indicate insignificant results for the other measures of real estate market activity. The effects on listings and new listings are summarized below.[24]

---

[24] I have chosen to omit the results on prices, which are largely insignificant on sale price, due to the indication of significant effects on turnover, which would themselves affect list prices, sale prices, and price reductions.

Table 4: Difference in Trends for Effect on Listings

Variable	Model 1	Model 2	Model 3
year	-7.026***	[factor]	[factor]
	(0.415)		
post	-14456.71***	7081.011	-6163.27
	(1446.019)	(9663.202)	(9441.501)
post*pfas	10070.18***	-9813.326	1305.161
	(2909.684)	(7609.751)	(6107.415)
post*year	7.162***	-3.510	3.054
	(0.717)	(4.789)	(4.679)
pfas*year	-6.26***	-6.328***	-4.108***
	(1.254)	(1.241)	(1.225)
pfas*post*year	-4.983***	4.867	-0.645
	(1.443)	(3.771)	(3.026)
county*year			[factor]
Adjusted R^2	0.894	0.890	0.918

Notes: [factor] indicates that this variable is included as a factor variable with many coefficients. Blank cells indicate that a variable was not included for the given model. Model 1 uses *year* as an integer, Model 2 uses *year* as factor variables, and Model 3 is the same as Model 2 but adds interactions for *county* and *year*, which are both factor variables. Data for *pfas* and *post* are from the PFAS Project (accessed 5 November 2019) and MPART (accessed 5 November 2019); listings are from Zillow (accessed 9 November 2019).

Table 5. Difference in Trends for Effect on New Listings

Variable	Model 1	Model 2	Model 3
year	-1.973***	[factor]	[factor]
	(0.288)		
post	-2432.681*	1.668	1.367
	(1316.334)	(1.076)	(1.278)
post*pfas	2207.705	1726.3	2218.762
	(2151.043)	(1966.709)	(2934.496)
post*year	1.207*	[omitted]	[omitted]
	0.652		
pfas*year	0.696	0.440	1.157
	(0.537)	(0.493)	(1.170)
pfas*post*year	-1.095	-0.856	-1.099
	(1.066)	(0.975)	(1.454)
county*year			[factor]
Adjusted R²	0.756	0.756	0.757

Notes: [factor] indicates that this variable is included as a factor variable with many coefficients. Blank cells indicate that a variable was not included for the given model. Model 1 uses *year* as an integer, Model 2 uses *year* as factor variables, and Model 3 is the same as Model 2 but adds interactions for *county* and *year*, which are both factor variables. Data for *pfas* and *post* are from the PFAS Project (accessed 5 November 2019) and MPART (accessed 5 November 2019); listings are from Zillow (accessed 9 November 2019).

There are no consistently significant effects of PFAS on new listings, and the only significant effect on all listings (once controls are added and year is in factor form) comes in the coefficient on *pfas*year*, representing the different levels and slopes of listing trends in contaminated versus non-contaminated zip codes. This indicates that the effect of contamination on listings is likely insignificant beyond existing differences between those zip codes known to be contaminated and the rest. Most of the variation is accounted for by local and regional controls.

B. Event study

It is convenient to normalize the *post* dates around a single date in the model. In order to construct an event study, *post_t* is defined as being equal to the month new information arose in the contaminated zip codes, and equal to January

2018—the mean of the information dates—for the non-contaminated zip codes. T is then defined:

$$T = date - post\_t$$

This way, $T=0$ where the *post* dates begin for all zip codes. Y is defined as being the year from T, equalized with the lowest T having $Y=0$.[25] New interactions are therefore analogous to the original interactions. This model is much the same as the one presented above:

(2) $turnover = \beta_0 + \beta_1 post_{zT} + \beta_2 post_{zT}*pfas_z + \beta_3 Y_T + \beta_4 Y_T*pfas_z +$

$\beta_5 Y_T*post_{zT} + \beta_6 Y_T*post_{zT}*pfas_z + \beta_9 jan_t + \beta_{10} feb_t + ... + \beta_{19} nov_t +$

$\beta_{20z} zip_z + \varepsilon_{zT}$

The results of these regressions are summarized below, in Table 6. While some insight can be gained, such as a confirmation of the differential pre-trends, finding a significant value for β_6 is confounded by the problem that the control-group observations end after $T=19$. This lack of comparability is likely the culprit of not being able to identify significant effects for β_6 in the *post* period. There is a negative overall effect on *post*Y*, likely because the bulk of observations for positive T are those in the treatment group; this could indicate that the negative trend effect remains somewhat in this framework.

---

[25] Stata does not allow negative factor variables, and I need to use Y as a factor variable; else Y would be defined as negative where T is negative.

Table 6. DID Models for Effect on Turnover, Event Study

Variable	Model 1	Model 2	Model 3
Y	0.588***	[factor]	[factor]
	(0.033)		
post	11.474***	19.103***	20.750***
	(1.292)	(1.810)	(1.744)
post*pfas	8.227	0.130	-0.548
	(7.397)	(9.171)	(8.289)
post*Y	-1.110***	-1.865***	-2.02***
	(0.119)	(0.173)	(0.167)
pfas*Y	0.942***	0.928***	0.654***
	(0.207)	(0.206)	(0.163)
pfas*post*Y	-0.837	-0.151	0.039
	(0.608)	(0.808)	(0.725)
county*Y			[factor]
Adjusted R^2	0.878	0.878	0.890

Notes: [factor] indicates that this variable is included as a factor variable with many coefficients. Blank cells indicate that a variable was not included for the given model. Model 1 uses *Y* as an integer, Model 2 uses *Y* as factor variables, and Model 3 is the same as Model 2 but adds interactions for *county* and *Y*, which are both factor variables. *T* is the number of months before or after *post,* and *Y* is a year version of *T*. Data for *pfas* and *post* are from the PFAS Project (accessed 5 November 2019) and MPART (accessed 5 November 2019); turnover is from Zillow (accessed 9 November 2019).

It is intriguing to visualize the change that occurred in contaminated zip codes in an event study framework. From a strictly visual standpoint, it appears that independent of seasonality, the variance in turnover drastically increased in the *post* period for contaminated zip codes. (This also allows visualization of the lack of a comparison group after about *T*=19.) See Figure 5 below.

Figure 5. Event Study of Average Turnover in Contaminated and Non-Contaminated Zip Codes
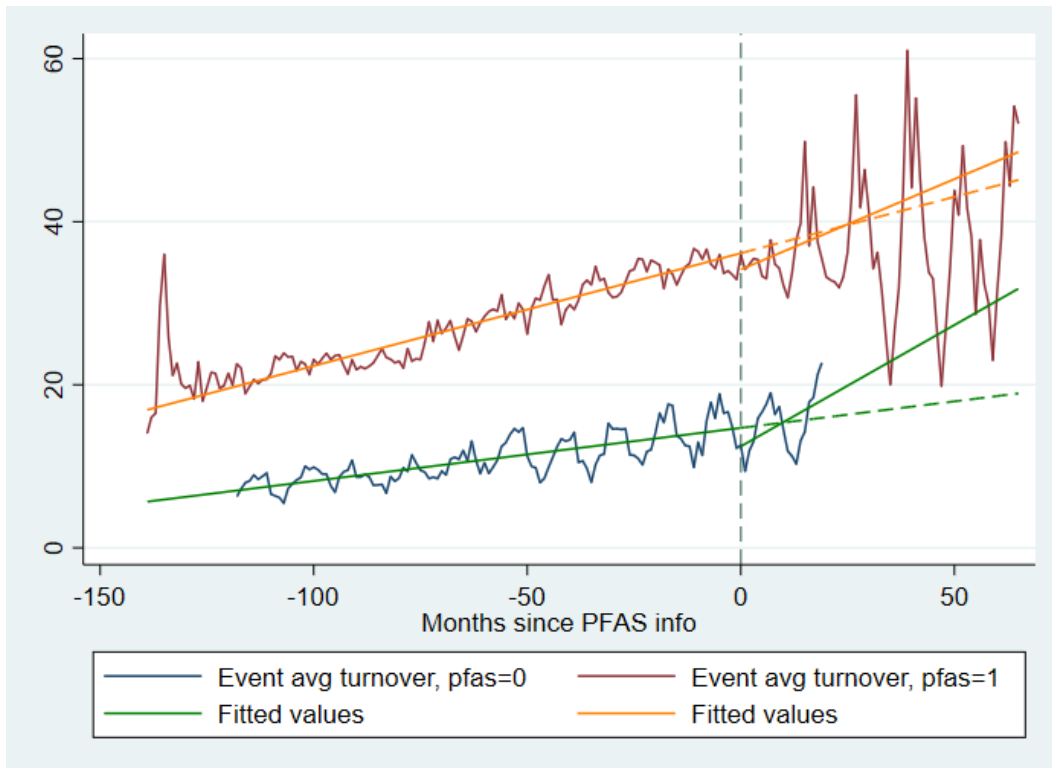
Notes: This is average turnover in contaminated zip codes (*pfas*=1) and non-contaminated zip codes (*pfas*=0), normalized so that *T*=0 is where the public knew about PFAS contamination, and *post*=1. The vertical line appears at *T*=0. The dashed lines for fitted values represent projection based on the time trend before *T*=0; solid line is actual. Housing data are from Zillow (accessed 9 November 2019); *pfas* and *post* data are from the PFAS Project (accessed 5 November 2019) and MPART (accessed 5 November 2019).

The appearance of an increase in variance in the contaminated group correlates with an apparent rise in home sale prices in the *post* period. (See Figure 6 below.) Given that the results for list and sale prices are insignificant, the most likely explanation is that the immediate effect of new PFAS contamination is primarily on turnover.

Figure 6. Event Study of Average Sales Price in Contaminated and Non-Contaminated Zip Codes

Note: This is average sale price in contaminated zip codes (*pfas*=1) and non-contaminated zip codes (*pfas*=0), normalized so that T=0 is where the public knew about PFAS contamination, and *post*=1. The vertical line appears at T=0. Fitted values omitted for clarity. Sales data are from Zillow (accessed 9 November 2019); *pfas* and *post* data are from the PFAS Project (accessed 5 November 2019) and MPART (accessed 5 November 2019).

C. Propensity to be contaminated with PFAS

Underlying the above analyses is the assumption that the contaminated and non-contaminated zip codes were different enough to warrant recognition of differential trends and levels, an assumption borne out in the significance of the different slopes for each group. The static economic indicators mentioned in Section V have varying levels of correlation with a tendency to be contaminated with PFAS; these correlations are summarized in Table 7 below.

Table 7. Correlation of Various Economic Indicators With PFAS

Contamination >6ppt

Variable	Correlation with *pfas*
Log density	0.131
Rurality	-0.116
Urban influence	-0.109
Population	0.236
Income rank	0.039
Median household income	0.024
University	0.170

Note: rurality and urban influence rise with rurality, and are heavily correlated. Income and income rank, where the rank is coded to be higher with higher incomes, are also heavily correlated. Rurality/population data are from the University of Michigan Center for Population Studies (accessed 23 November 2019); income data are from Zip Atlas (accessed 16 November 2019).

There is a high correlation between rurality and urban influence, as well as household income and income rank; however, the data are still independent and will lend precision to the probit models to predict PFAS contamination. Running a probit model on log density, rurality, population, median household income, the directions of the correlations mostly correspond to the coefficients on each predictor in the model. (See Table 8 below.) Also included is a linear regression of PFAS on the static variables to lend an interpretation of the results, and verify the signs on the coefficients.

Table 8. Results of Probit Models to Predict PFAS Contamination of >6ppt

Variable	Probit Without County Controls	Probit With County Controls	Linear Regression[26]
Log density	-0.119***	0.057***	1.78×10^{-3}***
	(0.006)	(0.009)	(6.16×10^{-4})
Rurality	-0.047***	4.032	0.002
	(0.007)	(164.34)	(0.005)
Urban influence	0.003	-4.787	-0.009**
	(0.005)	(167.592)	(0.004)
Population	3.48×10^{-5}***	3.97×10^{-5}***	5.27×10^{-6}***
	(6.31×10^{-7})	(7.75×10^{-7})	(7.01×10^{-8})
Median household income	-2.24×10^{-5}***	-2.42×10^{-5}***	-1.21×10^{-6}***
	(1.40×10^{-6})	(1.93×10^{-6})	(1.18×10^{-7})
Income rank	0.001***	0.001***	5.19×10^{-5}***
	(7.1×10^{-5})	(9.81×10^{-5})	(6.73×10^{-6})
University	0.285***	-0.072**	0.053***
	(0.024)	(0.033)	(0.003)
County		[factor]	[factor]
Constant	-0.967***	-2.537	0.089***
	(0.054)	(32.867)	(0.032)
Pseudo-R^2, R^2	0.124	0.214	0.184

Notes: *Rurality* and *urban influence* rise with rurality. *Income rank* rises with income. [factor] indicates that this variable is included as a factor variable with many coefficients. Blank cells indicate that a variable was not included for the given model. Rurality/population data are from the University of Michigan Center for Population Studies (accessed 23 November 2019); *pfas* data are from the PFAS Project (accessed 5 November 2019) and MPART (accessed 5 November 2019); income data are from Zip Atlas (accessed 16 November 2019).

The inclusion of county controls in the probit model presents both benefits and shortcomings. Clearly, the controls add a level of certainty, and can verify the accuracy of the signs on some the coefficients, though the high correlation of urban influence and rurality leads to opposing signs and astronomical standard errors for these two variables. Further, it presents a selection problem: since the state of

---

[26] With county controls.

Michigan is early in the discovery process of where PFAS is, there is no guarantee that those counties that are currently free of PFAS will remain so. However, by looking at all three models, there is verification that presence of PFAS is negatively correlated with household income and positively correlated with population. This means that PFAS contamination is generally more likely in more populated areas, as found in the correlation statistics, but unlike the correlation statistics, PFAS appears to be found in areas with lower median household incomes than average. While adding the county controls does confound the effect of rurality and urbanity measures, their inclusion allows for weighting down the observations in counties where PFAS has not yet been found, since it is an unknown factor in those locales.

The study will use these predictors to do propensity score matching and weighting of the regressions, using the process laid out in Bailey et al (2019). As seen below in Figure 7, the second probit model, which includes county controls, does have the requisite shape: observations could be trimmed where the predictor for PFAS is very close to zero and where the first line ends (about 0.8), and the line for the density of the predictor in contaminated zip codes quickly crosses the density line for non-contaminated zip codes and remains crossed, as desired. (The predictor variable is called *pfas_hat*.)

Figure 7. Density of contaminated and non-contaminated zip codes from
pfas_hat=0 **to** *pfas_hat*=1
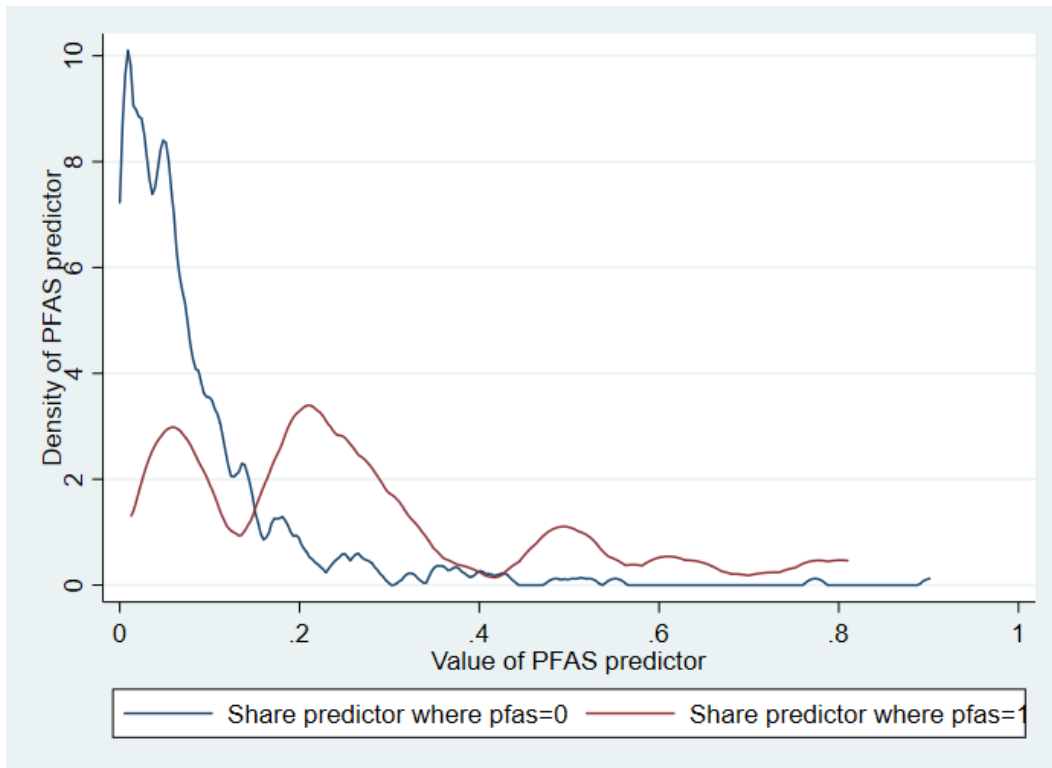
Note: The blue line follows the density of the PFAS predictor for non-contaminated zips, and the red line follows the density for contaminated zips. This is predicted using static economic and population data from Zip Atlas (accessed 16 November 2019) and the University of Michigan Center for Population Studies (accessed 23 November 2019), as well as county controls.

Probability weights are constructed as prescribed in Bailey et al (2019), so that the weight is equal to the probability of not being in the contaminated group over the probability of being in the contaminated group, multiplied by the proportion of the dataset in the contaminated group over the proportion in the non-contaminated group. This helps the problem of overstating the weight of those non-contaminated zip codes that are not expected to be contaminated anyway.

With the new weights, re-running the regression for Model 3 for the effect on turnover results in the findings for Model 3+ in Table 9 below.[27]

Table 9. Effect on Turnover with Probability Weights

Variable	Model 1	Model 2	Model 3	Model 3+[28]
year	0.595***	[factor]	[factor]	[factor]
	(0.033)			
post	2239.595***	-6338.563**	-5137.925**	[omitted]
	(240.525)	(2652.438)	(2457.744)	
*post*pfas*	1665.715	7611.033***	5775.185***	1,111.2
	(1225.114)	(2430.753)	(2233.405)	(1,531.2)
*post*year*	-1.110***	3.141**	2.546**	1.68×10^{-4}
	(0.119)	(1.314)	(1.218)	(2.32×10^{-4})
*pfas*year*	0.939***	1.006***	0.748***	0.632**
	(0.207)	(0.209)	(0.173)	(0.263)
*pfas*post*year*	-0.826	-3.772***	-2.861***	-0.551
	(0.607)	(1.204)	(1.106)	(0.759)
*county*year*			[factor]	[factor]
Adjusted R²	0.878	0.879	0.890	0.889

Notes: [factor] indicates that this variable is included as a factor variable with many coefficients. Blank cells indicate that a variable was not included for the given model. Model 1 uses *year* as an integer, Model 2 uses *year* as factor variables, and Model 3 is the same as Model 2 but adds interactions for *county* and *year*, which are both factor variables. Model 3+ adds probability weights based on static economic indicators from the University of Michigan Center for Population Studies (accessed 23 November 2019). Data for *pfas* and *post* are from the PFAS Project (accessed 5 November 2019) and MPART (accessed 5 November 2019); turnover is from Zillow (accessed 9 November 2019).

While some of the significance of the original results is lost, as is expected when adding probability weights, the basic intuition and direction of the results are the same. There is an insignificant trend in the *post* period for the non-contaminated group, and an additional (if no longer statistically significant) downward trend for

---

[27] The coefficient for *post* is excluded due to collinearity with *post*year* when weights are included.
[28] Adds probability weights.

the contaminated group, all while controlling for the differential trends present in the data.

VI. Conclusions

Using a variation on existing models, this analysis finds that there is a negative trend effect on home turnover in zip codes once they are known to be contaminated with PFAS at a level of greater than 6 ppt. It further finds that certain economic indicators can predict with some accuracy the probability that a zip code will be contaminated: namely, that more urban and poorer areas are more likely to be contaminated.

This study has compiled outcome housing data from Zillow, PFAS and *post* data from independent researchers, and static economic controls from different sources.

These results indicate that in addition to a statistically significant downturn in home turnover in contaminated zip codes once they are in the *post* period, there is an increase in slope in the non-contaminated zip codes in the *post* period. This indicates that the likelihood there is an overall, exogenous negative impact on turnover that is amplified in already-high-turnover, contaminated zips, is slim; rather, it appears that contaminated and non-contaminated zip codes diverged in the *post* period. Further, the opposite signs of the trend changes, with negative and positive changes in contaminated and non-contaminated zip codes, respectively, could indicate that some consumers are relocating to uncontaminated areas. (This is difficult to ascertain without corresponding population data.) Economically, however, these results indicate a turnover effect that may not be apparent to residents of contaminated areas; a fall of 0.5 to 3.8 (on average) in turnover per year is relatively small in comparison to 29 home sales monthly in contaminated areas.

As the PFAS story develops, I would like to expand the work in Section V.C to add a further layer of robustness to the existing difference-in-trends

32

framework constructed above. Once there is a complete picture of the PFAS situation in Michigan, and once American Community Survey (ACS) data are available for all *post* periods, a better predictor for PFAS contamination can likely be derived, and more economic controls can be added to the model. Finally, in the spirit of Taylor et al (2016), I would like to control for proximity to contamination sites, so that zip codes adjacent to contaminated zip codes are not treated the same way as distant zip codes that are nowhere near a contamination site. This would all help tell a more complete story of the full effects of new information about contamination on Michiganders' living situation decisions.

Works Cited

Affuso, Ermanno, et al. 2010. "The impact of hazardous waste on property values: The effect of lead pollution." *Urbani Izziv* Vol. 21, No. 2, pp. 117-126.

Bailey, Martha, et al. 2019. "Simple Strategies for Improving Inference with Linked Data: A Case Study of the 1850-1930 IMPUS Linked Representative Historical Samples." *Historical Methods: A Journal of Quantitative and Interdisciplinary History.* DOI: 10.1080/01615440.2019.1630343.

Ellison, Garret. 2018. "Michigan atop national PFAS site list." *MLive*, April 18. https://www.mlive.com/news/2018/04/ewg_2018_pfas_map_report.html, (accessed 5 November 2019).

Hanna, Bríd Gleeson. 2007. "House values, incomes, and industrial pollution." *Journal of Environmental Economics and Management* Vol. 54, pp. 100-112.

Hite, Diane, et al. 2000. "Property Value Impacts of an Environmental Disamenity: The Case of Landfills." *The Journal of Real Estate Finance and Economics*, Vol. 22, Issue 2-3, pp. 185-202.

Kiel, Katherine A. 1995. "Measure the Impact of the Discovery and Cleaning of Identified Hazardous Waste Sites on House Values." *Land Economics*, Vol. 71, No. 4, pp. 428-435.

Michael, Holly J., et al. 2000. "Does the Measurement of Environmental Quality Affect Implicit Prices Estimated from Hedonic Models?" *Land Economics*, Vol. 76, No. 2, pp. 283-298.

Michigan PFAS Action Response Team. 2019. "Michigan_PFAS_Sites_-_71_-_Table_669885_7." Michigan Department of Environment, Great

Lakes, and Energy. https://www.michigan.gov/pfasresponse/0,9038,7-365-86511_95645---,00.html (accessed 5 November 2019).

Michigan PFAS Action Response Team. 2019. "PFAS Sites." https://pfasproject.com/pfas-contamination-site-tracker/ (accessed 13 November 2019).

PFAS Project. 2019. "Public SSEHRI PFAS Contamination Site Tracker – last update 9.25.2019." Northeastern University Social Science Environmental Health Research Institute. https://pfasproject.com/pfas-contamination-site-tracker/ (accessed 5 November 2019).

SSEHRI. 2019. "PFAS Contamination Site Tracker." https://pfasproject.com/pfas-contamination-site-tracker/ (accessed 9 November 2019).

State of Michigan. 2019. "Michigan Public Four Year Universities." https://www.michigan.gov/documents/Public_4_Year_Map_151356_7.pdf (accessed 12 December 2019).

Taylor, Laura O., et al. 2016. "Disentangling Property Value Impacts of Environmental Contamination from Locally Undesirable Land Uses: Implications for Measuring Post-Cleanup Stigma." *Journal of Urban Economics*, Vol. 93, Issue C, 85-98.

Thorp, Ben. 2019. "Science advisory group recommends much stricter PFAS standards for Michigan." *Michigan Radio*, June 27. https://www.michiganradio.org/post/science-advisory-group-recommends-much-stricter-pfas-standards-michigan (accessed 5 November 2019).

University of Michigan Population Studies Center. 2003. "Measures of rurality for zip codes in the United States." https://www.psc.isr.umich.edu/dis/acs/faq/answer/1102 (accessed 23 November 2019).

Zillow. 2008-2019. "Housing Data." https://www.zillow.com/research/data/ (accessed 9 November 2019).

Zillow. 2019. "Methodology: Home Sales." https://www.zillow.com/research/home-sales-methodology-7733/ (accessed 13 November 2019).

Zip Atlas. 2019. "Zip Codes with the Highest Median Household Income in Michigan." https://zipatlas.com/us/mi/zip-code-comparison/median-household-income.htm (accessed 16 November 2019).

Zip Codes to Go. 2007-2018. "Zip Codes for the State of Michigan." https://www.zipcodestogo.com/Michigan/ (accessed 13 November 2019).

APPENDIX – STATA DO FILE

My .do file for this project is below appended.

```stata
1   *EFFECT OF PFAS CONTAMINATION ON PROPERTY VALUES
2   *Eric Fegan - created for Honors Thesis, UMich Econ, Class of 2019
3   *Created on 4 October 2019
4   *Last updated on 13 December 2019
5
6   *First dataset is from Zillow; "Sales_Counts_Zip" - this version downloaded 9 November 2019 -- will
    need to re-save all Excel docs as .xlsx
7   *Search "XYZ" for filenames
8
9   clear
10  cd "M:/" //XYZ
11  capture log close
12  log using PFAS, replace
13
14  *-------------------------------ENTRY/CLEANING-------------------------------*
15  import excel using "M:/Sale_Counts_zip.xlsx", firstrow //XYZ
16  tab StateName                                 //962 zip codes for Michigan
17  codebook StateName                            //0 missing values
18  gen stateus=1 if StateName=="Michigan"
19  tab stateus                                   //962 observations
20  rename RegionName zip
21  label variable zip "Zip Code"
22
23  drop if stateus!=1
24  drop StateName
25  drop stateus
26
27  *Renaming time variables to be usable
28  foreach v of varlist E F G H I J K L M N O P Q R S T U V W X Y Z AA AB AC AD AE AF AG AH AI AJ AK AL
    AM AN AO AP AQ AR AS AT AU AV AW AX AY AZ BA BB BC BD BE BF BG BH BI BJ BK BL BM BN BO BP BQ BR BS BT
     BU BV BW BX BY BZ CA CB CC CD CE CF CG CH CI CJ CK CL CM CN CO CP CQ CR CS CT CU CV CW CX CY CZ DA
    DB DC DD DE DF DG DH DI DJ DK DL DM DN DO DP DQ DR DS DT DU DV DW DX DY DZ EA EB EC ED EE EF EG EH EI
     EJ EK EL EM {
29  local lbl: variable label `v'
30  local lbl=subinstr("`lbl'","-","",1)
31  label var `v' `"turn`lbl'"'
32  local lbl2: variable label `v'
33  rename `v' `lbl2'
34  }
35
36  quietly reshape long turn, i(zip) j(date1)
37  tostring date1, replace
38  gen int year=real(substr(date1,1,4))
39  gen int month=real(substr(date1,5,6))
40  gen jan=1 if month==1
41  gen feb=1 if month==2
42  gen mar=1 if month==3
43  gen apr=1 if month==4
44  gen may=1 if month==5
45  gen jun=1 if month==6
46  gen jul=1 if month==7
47  gen aug=1 if month==8
48  gen sep=1 if month==9
49  gen oct=1 if month==10
50  gen nov=1 if month==11
51  gen dec=1 if month==12
52
53  gen date=ym(year,month)
54  format date %tmMonYY
55  drop date1
56
57  label variable turn "Monthly home turnover"
58  label variable year "Year"
```

```stata
 59    label variable month "Month"
 60    label variable jan "January"
 61    label variable feb "February"
 62    label variable mar "March"
 63    label variable apr "April"
 64    label variable may "May"
 65    label variable jun "June"
 66    label variable jul "July"
 67    label variable aug "August"
 68    label variable sep "September"
 69    label variable oct "October"
 70    label variable nov "November"
 71    label variable dec "December"
 72    label variable date "Monthly date"
 73
 74    drop RegionID SizeRank
 75
 76    **BRINGING IN OTHER DATA - MONTHLY LISTINGS, NEW MONTHLY LISTINGS, MEDIAN LISTING PRICES, MEDIAN
       PERCENTAGE OF PRICE REDUCTIONS, SALE PRICES**
 77    save pfas, replace
 78    clear
 79    import excel using "M:/MonthlyListings_NSA_AllHomes_Zip.xlsx", firstrow //XYZ
 80    gen stateus=1 if StateName=="MI"
 81    tab stateus                                     //549 observations
 82    rename RegionName zip
 83
 84    drop if stateus!=1
 85    drop StateName
 86    drop stateus
 87
 88    foreach v of varlist F G H I J K L M N O P Q R S T U V W X Y Z AA AB AC AD AE AF AG AH AI AJ AK AL AM
        AN AO AP AQ AR AS AT AU AV AW AX AY AZ BA BB BC BD BE BF BG BH BI BJ BK BL BM BN BO BP BQ BR BS BT
       BU BV BW BX BY BZ CA CB CC CD CE CF CG CH {
 89    local lbl: variable label `v'
 90    local lbl=subinstr("`lbl'","-","",1)
 91    label var `v' `"list`lbl'"'
 92    local lbl2: variable label `v'
 93    rename `v' `lbl2'
 94    }
 95
 96    quietly reshape long list, i(zip) j(date1)
 97    tostring date1, replace
 98    gen int year=real(substr(date1,1,4))
 99    gen int month=real(substr(date1,5,6))
100    gen jan=1 if month==1
101    gen feb=1 if month==2
102    gen mar=1 if month==3
103    gen apr=1 if month==4
104    gen may=1 if month==5
105    gen jun=1 if month==6
106    gen jul=1 if month==7
107    gen aug=1 if month==8
108    gen sep=1 if month==9
109    gen oct=1 if month==10
110    gen nov=1 if month==11
111    gen dec=1 if month==12
112
113    gen date=ym(year,month)
114    format date %tmMonYY
115    drop date1
116
117    drop RegionID SizeRank RegionType
118    save listings, replace
```

```
119
120     **
121     clear
122     import excel using "M:/NewMonthlyListings_NSA_AllHomes_Zip.xlsx", firstrow //XYZ
123     gen stateus=1 if StateName=="MI"
124     tab stateus                                    //328 observations
125     rename RegionName zip
126
127     drop if stateus!=1
128     drop StateName
129     drop stateus
130
131     foreach v of varlist F G H I J K L M N O P Q R S T U V W X Y Z AA AB AC AD AE AF AG AH AI AJ AK AL AM
         AN AO AP AQ AR AS AT AU AV AW AX AY AZ BA BB BC BD BE BF BG BH BI BJ BK BL BM BN BO BP BQ BR BS BT
        BU BV BW BX BY BZ CA CB CC CD CE CF CG CH {
132     local lbl: variable label `v'
133     local lbl=subinstr("`lbl'","-","",1)
134     label var `v' `"list_new`lbl'"'
135     local lbl2: variable label `v'
136     rename `v' `lbl2'
137     }
138
139     quietly reshape long list_new, i(zip) j(date1)
140     tostring date1, replace
141     gen int year=real(substr(date1,1,4))
142     gen int month=real(substr(date1,5,6))
143     gen jan=1 if month==1
144     gen feb=1 if month==2
145     gen mar=1 if month==3
146     gen apr=1 if month==4
147     gen may=1 if month==5
148     gen jun=1 if month==6
149     gen jul=1 if month==7
150     gen aug=1 if month==8
151     gen sep=1 if month==9
152     gen oct=1 if month==10
153     gen nov=1 if month==11
154     gen dec=1 if month==12
155
156     gen date=ym(year,month)
157     format date %tmMonYY
158     drop date1
159
160     drop RegionID SizeRank RegionType
161     save listings_new, replace
162
163     **
164     clear
165     import excel using "M:Zip_MedianListingPrice_AllHomes.xlsx", firstrow //XYZ
166     gen stateus=1 if State=="MI"
167     tab stateus                                    //375 observations
168     rename RegionName zip
169
170     drop if stateus!=1
171     drop State
172     drop stateus
173
174     foreach v of varlist G H I J K L M N O P Q R S T U V W X Y Z AA AB AC AD AE AF AG AH AI AJ AK AL AM
        AN AO AP AQ AR AS AT AU AV AW AX AY AZ BA BB BC BD BE BF BG BH BI BJ BK BL BM BN BO BP BQ BR BS BT BU
         BV BW BX BY BZ CA CB CC CD CE CF CG CH CI CJ CK CL CM CN CO CP CQ CR CS CT CU CV CW CX CY CZ DA DB
        DC DD DE DF DG DH DI DJ DK DL DM DN DO DP DQ DR DS {
175     local lbl: variable label `v'
176     local lbl=subinstr("`lbl'","-","",1)
```

```stata
177     label var `v' `"p_list`lbl'"'
178     local lbl2: variable label `v'
179     rename `v' `lbl2'
180     }
181
182     quietly reshape long p_list, i(zip) j(date1)
183     tostring date1, replace
184     gen int year=real(substr(date1,1,4))
185     gen int month=real(substr(date1,5,6))
186     gen jan=1 if month==1
187     gen feb=1 if month==2
188     gen mar=1 if month==3
189     gen apr=1 if month==4
190     gen may=1 if month==5
191     gen jun=1 if month==6
192     gen jul=1 if month==7
193     gen aug=1 if month==8
194     gen sep=1 if month==9
195     gen oct=1 if month==10
196     gen nov=1 if month==11
197     gen dec=1 if month==12
198
199     gen date=ym(year,month)
200     format date %tmMonYY
201     drop date1
202
203     drop City Metro CountyName SizeRank
204     save listprices, replace
205
206     **
207     clear
208     import excel using "M:Zip_MedianPctOfPriceReduction_AllHomes.xlsx", firstrow //XYZ
209
210     gen stateus=1 if State=="MI"
211     tab stateus                                    //301 observations
212     rename RegionName zip
213
214     drop if stateus!=1
215     drop State
216     drop stateus
217
218     foreach v of varlist H I J K L M N O P Q R S T U V W X Y Z AA AB AC AD AE AF AG AH AI AJ AK AL AM AN
        AO AP AQ AR AS AT AU AV AW AX AY AZ BA BB BC BD BE BF BG BH BI BJ BK BL BM BN BO BP BQ BR BS BT BU BV
         BW BX BY BZ CA CB CC CD CE CF CG CH CI CJ CK CL CM CN CO CP CQ CR CS CT CU CV CW CX CY CZ DA DB DC
        DD DE DF DG DH DI DJ DK DL {
219     local lbl: variable label `v'
220     local lbl=subinstr("`lbl'","-","",1)
221     label var `v' `"p_red`lbl'"'
222     local lbl2: variable label `v'
223     rename `v' `lbl2'
224     }
225
226     quietly reshape long p_red, i(zip) j(date1)
227     tostring date1, replace
228     gen int year=real(substr(date1,1,4))
229     gen int month=real(substr(date1,5,6))
230     gen jan=1 if month==1
231     gen feb=1 if month==2
232     gen mar=1 if month==3
233     gen apr=1 if month==4
234     gen may=1 if month==5
235     gen jun=1 if month==6
236     gen jul=1 if month==7
```

```stata
237     gen aug=1 if month==8
238     gen sep=1 if month==9
239     gen oct=1 if month==10
240     gen nov=1 if month==11
241     gen dec=1 if month==12
242
243     gen date=ym(year,month)
244     format date %tmMonYY
245     drop date1
246
247     drop RegionID SizeRank City County Metro
248     save pricereductions, replace
249
250     **sales
251     clear
252     import excel using "M:/Sale_Prices_Zip.xlsx", firstrow //XYZ
253     gen stateus=1 if StateName=="Michigan"
254     tab stateus                                  //220 observations
255     rename RegionName zip
256
257     drop if stateus!=1
258     drop StateName
259     drop stateus
260
261     foreach v of varlist E F G H I J K L M N O P Q R S T U V W X Y Z AA AB AC AD AE AF AG AH AI AJ AK AL
        AM AN AO AP AQ AR AS AT AU AV AW AX AY AZ BA BB BC BD BE BF BG BH BI BJ BK BL BM BN BO BP BQ BR BS BT
         BU BV BW BX BY BZ CA CB CC CD CE CF CG CH CI CJ CK CL CM CN CO CP CQ CR CS CT CU CV CW CX CY CZ DA
        DB DC DD DE DF DG DH DI DJ DK DL DM DN DO DP DQ DR DS DT DU DV DW DX DY DZ EA EB EC ED EE EF EG EH EI
         EJ EK EL EM {
262     local lbl: variable label `v'
263     local lbl=subinstr("`lbl'","-","",1)
264     label var `v' `"p_sale_sa`lbl'"'
265     local lbl2: variable label `v'
266     rename `v' `lbl2'
267     }
268
269     quietly reshape long p_sale_sa, i(zip) j(date1)
270     tostring date1, replace
271     gen int year=real(substr(date1,1,4))
272     gen int month=real(substr(date1,5,6))
273     gen jan=1 if month==1
274     gen feb=1 if month==2
275     gen mar=1 if month==3
276     gen apr=1 if month==4
277     gen may=1 if month==5
278     gen jun=1 if month==6
279     gen jul=1 if month==7
280     gen aug=1 if month==8
281     gen sep=1 if month==9
282     gen oct=1 if month==10
283     gen nov=1 if month==11
284     gen dec=1 if month==12
285
286     gen date=ym(year,month)
287     format date %tmMonYY
288     drop date1
289
290     drop RegionID SizeRank
291     save saleprices_sa, replace
292
293     ***MERGING ALL ZILLOW DATA***
294     clear
295     use pfas
```

```stata
296    merge 1:1 zip date using listings
297    drop _merge
298    merge 1:1 zip date using listings_new
299    drop _merge
300    merge 1:1 zip date using listprices
301    drop _merge
302    merge 1:1 zip date using pricereductions
303    drop _merge
304    merge 1:1 zip date using saleprices
305    drop _merge
306
307    foreach v of varlist jan feb mar apr may jun jul aug sep oct nov dec{
308    replace `v'=0 if `v'==.
309    }
310
311    label variable list "Monthly listings"
312    label variable list_new "New monthly listings"
313    label variable p_list "Median list price"
314    label variable p_red "Average price reduction %"
315    label variable p_sale "Median sale price"
316
317    **BRINGING IN POST DATA**
318    *Data are from an Excel file that I compiled using publicly available information about PFAS
       contamination*
319    save pfas, replace
320    clear
321    import excel using "M:/post dates.xlsx", firstrow //XYZ
322    gen level1=4 if level=="very high"
323    replace level1=3 if level=="high"
324    replace level1=2 if level=="mid"
325    replace level1=1 if level=="low"
326    drop if level1==.
327    drop Facility Address City source1025 Seehttpscdnknightlabcomli httpsdocsgooglecomspreadsh Military
328    drop level
329    rename level1 level
330    sort zip post
331    *Need to drop errant blanks and observations with duplicate information. In cases where there are
       multiple instances of testing in a zip code, the oldest testing information is kept.*
332    drop if zip==zip[_n-1]
333    gen m=month(post)
334    gen y=year(post)
335    gen post_m=ym(y,m)
336    format post_m %tmMonYY
337    drop post y m
338    label variable post_m "Public aware of PFAS"
339    label variable post_y "Alternate post measure"
340    label variable level "PFAS level"
341    save dates, replace
342    clear
343    use pfas
344    merge m:1 zip using dates
345
346    gen pfas=0 if level==.|level==1
347    replace pfas=1 if pfas==.
348    label variable pfas "PFAS present at >6ppt in zip"
349
350    *Only non-seasonally adjusted data is used (except sale prices), so the seasAdj variable is
       unnecessary.
351    drop seasAdj
352    drop _merge
353
354    gen post=date-post_m
355    replace post=1 if pfas==1 & post>0
```

```
356    replace post=0 if pfas==1 & post<=0        //These code post to be equal to one after public awareness
       in contaminated zip codes
357    codebook post_m                            //Identifies that 50th percentile for post date is 702
       (about July 2018); mean is 696 (about January 2018)
358    replace post=1 if pfas==0 & date>696
359    replace post=0 if pfas==0 & date<=696      //These code post to be equal to one after the average
       public awareness in non-contaminated zip codes
360
361    gen post2=year-post_y
362    replace post2=1 if post2>=0
363    replace post2=0 if post_y==.|post2<0
364
365    label variable post "Post-public awareness"
366    label variable post2 "Post-public awareness, alternate"
367
368    **IDENTIFYING ZIP CODES TO COUNTIES**
369    save pfas, replace
370    clear
371    import excel using "M:/zip-county.xlsx", firstrow //XYZ
372    rename ZipCode zip
373    rename City city
374    rename County county
375    drop ZipCodeMap
376    save counties, replace
377    clear
378    use pfas
379    merge m:m zip using counties
380    drop if _merge==2                           //These zip codes have no observations in the dataset; they
       may be PO box zip codes or similar
381    replace county="Calhoun" if _merge==1     //49037 seems to have been errantly omitted from zip code
       dataset; it may be a recent change. This corrects it to be in Calhoun County.
382    replace city="Battle Creek" if _merge==1 //See above.
383    drop _merge
384    encode(county), gen(_county)
385    label variable _county "County, destring"
386    encode(city), gen(_city)
387    label variable _city "City, destring"
388
389    **RURALITY DATA**
390    save pfas, replace
391    clear
392    import excel using "M:/t1101_ziprural.xlsx", sheet("Data") firstrow //XYZ
393    drop fipsstcou cnameabbr allocate lzpop ruca2 migmet zpop
394    label variable lzden "Log population density"
395    rename ru2003 rural
396    rename ui2003 ui
397    label variable rural "Rural-urban continuum code"
398    label variable ui "Urban influence code"
399    save rurality, replace
400
401    **INCOME DATA**
402    clear
403    import excel using "M:/zip-income.xlsx", sheet("income") firstrow //XYZ
404    drop location city nationalrank
405    rename A rank_mi
406    gen rank_up=963-rank_mi
407    label variable rank_up "MI income rank, higher/richer"
408    label variable rank_mi "Michigan income rank, higher/poorer"
409    label variable incomehh "Median household income"
410    format incomehh %8.0g
411    rename population pop
412    label variable pop "Population"
413    label variable zip "Zip Code"
```

```stata
414    save income, replace
415
416    clear
417    use pfas
418    merge m:m zip using rurality
419    drop if _merge==2 //This included all zip codes nationally.
420    rename _merge merge1
421    merge m:m zip using income
422    drop if _merge==2
423    replace incomehh=. if incomehh==0|incomehh==5000 //some errant values here.
424
425    gen inter_mm=merge1*_merge
426    codebook inter_mm //verifies that the 5,699 values missing for both merges are the same
427    drop _merge merge1 inter_mm
428
429    sort date zip
430
431    *adding in military bases either not tested or not tested positive XXX delete?
432    *replace mil=1 if zip==49707|zip==49012|zip==49037|zip==49091|zip==48045
433    *replace mil=0 if mil==.
434
435    *four-year, public university dummy
436    gen univ=1 if city=="Mount Pleasant"|city=="Ypsilanti"|city=="Big Rapids"|city=="Allendale"|city==
           "Sault Ste Marie"|city=="East Lansing"|city=="Houghton"|city=="Marquette"|city=="Rochester Hills"|
           city=="University Center"|city=="Ann Arbor"|city=="Dearborn"|city=="Flint"|city=="Detroit"|city==
           "Kalamazoo"
437    replace univ=0 if univ==.
438    label variable univ "Public univ. in city"
439
440    *dropping all observations from after August 2019 due to latency issues in Zillow dataset
441    drop if date==716
442    save pfas_clean, replace
443    *---------------------------END ENTRY/CLEANING-------------------------------*
444    use pfas_clean, clear
445
446    **GENERATING INTERACTION VARIABLES
447    gen inter_pc=pfas*post
448    label variable inter_pc "pfas*post"
449    gen inter_py=post*year
450    label variable inter_py "post*year"
451    gen inter_cy=pfas*year
452    label variable inter_cy "pfas*year"
453    gen inter_cpy=pfas*post*year
454    label variable inter_cpy "pfas*post*year"
455
456    **---SOME SUMMARY STATISTICS AND INFORMATIVE GRAPHICS---**
457
458    *TABLE 1 INFORMATION
459    replace level=0 if level==.
460    sum turn list list_new p_list p_sale p_red pfas level lzden rural ui pop incomehh univ if pfas==0
461    sum turn list list_new p_list p_sale p_red pfas level lzden rural ui pop incomehh univ if pfas==1
462
463    *INFORMATION ON POST
464    sum(post_m) //Identifies mean month for post as 696 (Jan 2018) and a standard deviation as 17 months.
465
466    *SALE PRICES
467    gen p1=p_sale if pfas==0
468    gen p2=p_sale if pfas==1
469    egen m_psale=mean(p1), by(date)
470    egen m_psale_c=mean(p2), by(date)
471    drop p1 p2
472    label variable m_psale "Average sale price, pfas=0"
473    label variable m_psale_c "Average sale price, pfas=1"
```

```
474
475    ttest m_psale=m_psale_c        //We reject the null here as well.
476    line m_psale m_psale_c date if date>=648, xline(679 713, lpattern(dash) lcolor(teal)) xline(696,
       lcolor(teal))  //Creates Figure 1
477    graph export "M:\Figure1.png", as(png) name("Graph") replace     //XYZ
478
479    *TURNOVER
480    gen turn1=turn if pfas==0
481    gen turn2=turn if pfas==1
482    egen m_turn=mean(turn1), by(date)
483    egen m_turn_c=mean(turn2), by(date)
484    drop turn1 turn2
485    label variable m_turn "Average turnover, pfas=0"
486    label variable m_turn_c "Average turnover, pfas=1"
487
488    ttest m_turn=m_turn_c        //Very easily reject the null here.
489
490    quietly regress m_turn date if date<=679
491    predict mt_hat_pre
492    quietly regress m_turn date if date>=679
493    predict mt_hat
494    quietly regress m_turn_c date if date<=679
495    predict mtc_hat_pre
496    quietly regress m_turn_c date if date>=679
497    predict mtc_hat
498    line m_turn m_turn_c date || line mt_hat date if date>=679, lcolor(green) || line mtc_hat date if
       date>=679, lcolor(orange) || line mt_hat_pre date if date>=679, lcolor(green) lpattern(dash) || line
       mt_hat_pre date if date<=679, lcolor(green) || line mtc_hat_pre date if date>=679, lcolor(orange)
       lpattern(dash) || line mtc_hat_pre date if date<=679, lcolor(orange) xline(679 713, lpattern(dash)
       lcolor(teal)) xline(696, lcolor(teal)) legend(order(1 "Avg. turnover, pfas=0" 2 "Avg. turnover,
       pfas=1" 3 "Fitted values" 4 "Fitted values"))       //Creates Figure 2
499    graph export "M:\Figure2.png", as(png) name("Graph") replace     //XYZ
500
501    gen turndiff=m_turn_c-m_turn
502    label variable turndiff "Difference in turnover"
503    quietly regress turndiff date if date<=679
504    predict mtd_hat_pre
505    quietly regress turndiff date if date>=679
506    predict mtd_hat
507    line turndiff date || line mtd_hat date if date>=679, lcolor(green) || line mtd_hat_pre date if date
       >=679, lcolor(green) lpattern(dash) || line mtd_hat_pre date if date<=679, lcolor(green) xline(679
       713, lpattern(dash) lcolor(teal)) xline(696, lcolor(teal)) legend(order(1 "Difference in avg.
       turnover" 2 "Fitted values"))  //Creates Figure 3
508    graph export "M:\Figure3.png", as(png) name("Graph") replace     //XYZ
509
510    *TABLE 2 INFORMATION
511    codebook post_m //In looking at when PFAS information arose, as of 9 November, 10% of total occurred
       by November 2016; 25% by February 2018; 50% by July 2018; 75% by October 2018; and 90% by May 2019.
512
513
514    **---ANALYSIS---**
515
516    *TABLE 3
517    areg turn jan feb mar apr may jun jul aug sep oct nov year post inter_pc inter_py inter_cy inter_cpy,
        a(zip) cluster(zip)
518
519    *TABLE 3 pt. 2
520    areg turn jan feb mar apr may jun jul aug sep oct nov i.year post inter_pc inter_py inter_cy
       inter_cpy, a(zip) cluster(zip)
521
522    *TABLE 3 pt. 3
523    areg turn jan feb mar apr may jun jul aug sep oct nov i.year year#_county post inter_pc inter_py
       inter_cy inter_cpy, a(zip) cluster(zip)
```

```
524
525    areg turn jan feb mar apr may jun jul aug sep oct nov year year#_county post inter_pc inter_py
       inter_cy inter_cpy, a(zip) cluster(zip)       //This gives the same result for at-issue coefficients
       as when year dummies are used.
526
527    *Models 1, 2, 3, with listings (output for Table 4):
528    areg list jan feb mar apr may jun jul aug sep oct nov year post inter_pc inter_py inter_cy inter_cpy,
        a(zip) cluster(zip)
529    areg list jan feb mar apr may jun jul aug sep oct nov i.year post inter_pc inter_py inter_cy
       inter_cpy, a(zip) cluster(zip)
530    areg list jan feb mar apr may jun jul aug sep oct nov i.year year#_county post inter_pc inter_py
       inter_cy inter_cpy, a(zip) cluster(zip)
531
532    *Models 1, 2, 3, with NEW listings (output for Table 5):
533    areg list_new jan feb mar apr may jun jul aug sep oct nov year post inter_pc inter_py inter_cy
       inter_cpy, a(zip) cluster(zip)
534    areg list_new jan feb mar apr may jun jul aug sep oct nov i.year post inter_pc inter_py inter_cy
       inter_cpy, a(zip) cluster(zip)
535    areg list_new jan feb mar apr may jun jul aug sep oct nov i.year year#_county post inter_pc inter_py
       inter_cy inter_cpy, a(zip) cluster(zip)
536
537
538    *EVENT STUDY
539    gen post_t=post_m
540    replace post_t=696 if pfas==0   //sets anywhere with PFAS of less than 6ppt to be at 'post' in the
       mean month, Jan 2018.
541    gen T=date-post_t
542    label variable post_t "post_m; equal to 696 if pfas=0"
543    label variable T "Months since PFAS info"
544
545    gen t1=turn if pfas==0
546    gen t2=turn if pfas==1
547    egen mt_t=mean(t1), by(T)
548    egen mtc_t=mean(t2), by(T)
549    drop t1 t2
550    label variable mt_t "Event avg turnover, pfas=0"
551    label variable mtc_t "Event avg turnover, pfas=1"
552    sort T zip
553    quietly regress mt_t T if T<=0
554    predict mt_t_hat_pre
555    quietly regress mt_t T if T>=0
556    predict mt_t_hat
557    quietly regress mtc_t T if T<=0
558    predict mtc_t_hat_pre
559    quietly regress mtc_t T if T>=0
560    predict mtc_t_hat
561    line mt_t mtc_t T || line mt_t_hat_pre T if T<=0, lcolor(green) || line mt_t_hat_pre T if T>=0,
       lcolor(green) lpattern(dash) || line mtc_t_hat_pre T if T<=0, lcolor(orange) || line mtc_t_hat_pre T
       if T>=0, lcolor(orange) lpattern(dash) || line mt_t_hat T if T>=0, lcolor(green) || line mtc_t_hat T
       if T>=0, lcolor(orange) xline(0, lpattern(dash) lcolor(teal)) legend(order(1 "Event avg turnover,
       pfas=0" 2 "Event avg turnover, pfas=1" 3 "Fitted values" 4 "Fitted values")) //Creates Figure 5.
       Will need to sort by T and zip before creating this figure.
562    graph export "M:\Figure5.png", as(png) name("Graph") replace    //XYZ
563
564    gen T2=T+140
565    egen Y1=cut(T2), at (1(12)220)
566    gen Y=(Y1-1)/12
567    drop Y1 T2
568    label variable Y "Event study year"
569
570    gen inter_pt=post*Y
571    gen inter_ct=pfas*Y
572    gen inter_cpt=pfas*post*Y
```

```
573     label variable inter_pt "post*Y"
574     label variable inter_ct "pfas*Y"
575     label variable inter_cpt "pfas*post*Y"
576
577     *Output for Table 6
578     areg turn jan feb mar apr may jun jul aug sep oct nov Y post inter_pc inter_pt inter_ct inter_cpt, a(
        zip) cluster(zip)
579     areg turn jan feb mar apr may jun jul aug sep oct nov i.Y post inter_pc inter_pt inter_ct inter_cpt,
        a(zip) cluster(zip)
580     areg turn jan feb mar apr may jun jul aug sep oct nov i.Y Y#_county post inter_pc inter_pt inter_ct
        inter_cpt, a(zip) cluster(zip)
581
582     gen p1=p_sale if pfas==0
583     gen p2=p_sale if pfas==1
584     egen mp_t=mean(p1), by(T)
585     egen mpc_t=mean(p2), by(T)
586     drop p1 p2
587     label variable mp_t "Event avg sale price, pfas=0"
588     label variable mpc_t "Event avg sale price, pfas=1"
589     quietly regress mp_t T if T<=0 & T>-50
590     predict mp_t_hat_pre
591     label variable mp_t_hat_pre "Fitted values, pfas=0"
592     quietly regress mp_t T if T>=0 & T<20
593     predict mp_t_hat
594     label variable mp_t_hat "Fitted values, pfas=0"
595     quietly regress mpc_t T if T<=0 & T>-50
596     predict mpc_t_hat_pre
597     label variable mpc_t_hat_pre "Fitted values, pfas=1"
598     quietly regress mpc_t T if T>=0 & T<20
599     predict mpc_t_hat
600     label variable mpc_t_hat "Fitted values, pfas=1"
601     line mp_t mpc_t T if T>-50 & T<20 || line mp_t_hat_pre T if T<=0 & T>-50, lcolor(green) || line
        mp_t_hat T if T>=0 & T<20, lcolor(green) || line mpc_t_hat_pre T if T<=0 & T>-50, lcolor(orange) ||
        line mpc_t_hat T if T>=0 & T<20, lcolor(orange) xline(0, lpattern(dash) lcolor(teal)) // fitted
        values don't tell us much
602     line mp_t mpc_t T if T>-50 & T<20, xline(0, lpattern(dash) lcolor(teal))
603     graph export "M:\Figure6.png", as(png) name("Graph") replace    //XYZ
604
605     *Below, I run the same regression for sale prices and list prices. All interactions are insignificant.
606     areg p_sale jan feb mar apr may jun jul aug sep oct nov i.Y Y#_county post inter_pc inter_pt inter_ct
         inter_cpt, a(zip) cluster(zip)
607     areg p_list jan feb mar apr may jun jul aug sep oct nov i.Y Y#_county post inter_pc inter_pt inter_ct
         inter_cpt, a(zip) cluster(zip)
608
609
610     *PROPENSITY TO BE CONTAMINATED WITH PFAS*
611     sort date zip
612
613     *Table 7 information
614     corr pfas lzden
615     corr pfas rural
616     corr pfas ui
617     corr pfas pop
618     corr pfas rank_up
619     corr pfas incomehh
620     corr pfas univ
621
622     corr rural ui //correlated at over 90%
623     corr incomehh rank_up //correlated at over 90%
624
625     *Table 8 information
626     probit pfas lzden rural ui pop incomehh rank_up univ
627     probit pfas i._county lzden rural ui pop incomehh rank_up univ
```

```
628    quietly predict pfas_hat
629    reg pfas i._county lzden rural ui pop incomehh rank_up univ
630
631    twoway (kdensity pfas_hat if pfas==0, legend(label(1 "Share predictor where pfas=0"))) (kdensity
       pfas_hat if pfas==1, legend(label(2 "Share predictor where pfas=1"))), ytitle("Density of PFAS
       predictor") xtitle("Value of PFAS predictor") //creates Figure 7
632    graph export "M:\Figure7.png", as(png) name("Graph") replace     //XYZ
633
634    codebook pfas
635    di 127698+7938 //There are 7,938 records with pfas==1, and 135,636 total
636    gen weight=(((1-pfas_hat)/(pfas_hat))*(7938/127698))
637
638    twoway (hist pfas_hat if pfas==0, color(blue) fintensity(inten20) width(.02)) (hist pfas_hat if pfas
       ==1, color(red) fintensity(inten20) width(.02))     //pfas=0 peak occurs before pfas_hat=.04; pfas=1
       peak occurs before pfas_hat=.06
639
640    areg turn jan feb mar apr may jun jul aug sep oct nov i.year year#_county inter_pc inter_py inter_cy
       inter_cpy [pw=weight] if pfas_hat>0.01 & pfas_hat<0.8, a(zip) cluster(zip) //Results for Table 9
641
642    gen y_1=(date-578)*(2/3)
643    gen y_2=(date-550)
644    gen y_3=78.7
645    gen y_4=842-date
646    line y_1 date if date<=696, lcolor(blue) || line y_1 date if date>=696, lcolor(blue) lpattern(dash)
       || line y_2 date if date<=696, lcolor(red) || line y_2 date if date>=696, lcolor(red) lpattern(dash)
       || line y_3 date if date>=696, lcolor(blue) || line y_4 date if date>=696, lcolor(red) legend(off)
       xline(696, lcolor(teal) lpattern(dash))
647    graph export "M:\Figure4.png", as(png) name("Graph") replace
648
649
650    save pfas_use, replace
651
652    close log
653    *--------------------*
```