Toxic Assets: How the Housing Market Responds to Environmental Information Shocks*

Jeremy G. Moulton¹  Nicholas J. Sanders²  Scott A. Wentland³

June 22, 2023

Abstract

Employing national microdata from Zillow, we examine how United States housing markets respond to expanded information on local pollution stemming from a 1998 reporting change to the Toxics Release Inventory (TRI). Using both a difference-in-differences and a regression discontinuity in time design, we find news coverage of the new TRI data lowered sales prices of homes near the largest reporting polluters, but only within a tight geographic distance. Effects are isolated to homes within 0.5 miles of facilities reporting the largest amount of emissions (>100 tons). This price capitalization implies public information on local polluters shifted private market behavior, suggesting a role for government as provider of information.

Keywords: pollution, information disclosure, residential housing, home prices, spillovers, externality, toxics release inventory, regression discontinuity

JEL Classifications: D62, Q50, Q53, H23, G14, R32, R38

Disclaimer: Any views expressed here are those of the authors and not necessarily those of the Bureau of Economic Analysis or the U.S. Department of Commerce. Data provided by Zillow through the Zillow Transaction and Assessment Dataset (ZTRAX). More information on accessing the data can be found at http://www.zillow.com/ztrax. The results and opinions are those of the author(s) and do not reflect the position of Zillow Group.

* Thanks to Caroline Hoxby, Hilary Hoynes, Christopher Knittel, Jed Richardson, Gary Cornwall, Susanna Berkouwer, participants in the Stanford Environment and Energy Policy Analysis Center seminar, the Stanford Public/Environmental Seminar, the University of California, Davis Public/Labor Seminar, the Stanford Institute for Theoretical Economics summer meetings, the 2018 Urban Economics Association Meeting, 2019 American Real Estate Society Annual Meeting, 2020 ASSA meetings, and participants in the EPIC Junior Scholars meeting. Thanks also to Tom Blake, who helped substantially with data for an earlier version of the paper, and to Maggie Goulder for excellent research assistance.

¹ Department of Public Policy, University of North Carolina — Chapel Hill, moulton@email.unc.edu
² Brooks School of Public Policy and Department of Economics, Cornell University, njsanders@cornell.edu
³ Office of the Chief Economist, U.S. Bureau of Economic Analysis, scott.wentland@bea.gov
“It’s not news that they’re polluting, but it is news to the extent that they are polluting.” (John F. Sheehan of the Adirondack Council4)

1 Introduction

Since 1987, the Environmental Protection Agency (EPA) has required several polluting industries report their toxic releases via a data set known as the Toxics Release Inventory (TRI). The EPA gathers and aggregates these data, making them publicly available in an annual report. In 1998, the EPA expanded the industries required to report, including some of the most intense polluters, coal and oil power plants. The new industries’ data were first disclosed to the public in the spring of 2000. Their addition drastically shifted publicly reported pollution levels, with some local areas seeing increases in reported toxic releases in the range of 800%. Using data from Zillow covering millions of home sales across the United States, we find updated TRI information reduced housing prices near the largest reporting polluters. Overall, the results show information for that year’s data drew substantial news attention, placing focus on the largest polluters. This information was capitalized into home prices, resulting in a decrease in values of about 11%, but only in a highly localized space (<0.5 mile) near the highest emitting facilities (>100 tons of air emissions).

This was a shift in reported releases, not a shift in emissions per se, and thus the change in information on local (dis)amenities was independent of a change in the level of (dis)amenities themselves. This provides a unique opportunity to investigate several important policy questions. First, we study the effect of government as provider of information on only partially observable local amenities. The covered industries were almost exclusively obvious polluters, allowing households to establish priors of daily exposure to localized “bads”. The modified TRI serves as a

---

4 Hu (May 12, 2000).
government-provided source of information by which households updated perceptions. For example, homeowners located near a coal power plant (which was newly added to the TRI database) or a steel plant (which had reported in past reports) are likely aware these facilities emit pollution, but homeowners near the former largely did not know the absolute magnitude of emissions before 2000, nor did homeowners near the latter know the magnitude of a steel plant relative to something like a coal plant. It remains an empirical question whether markets would price such new information regarding the absolute or relative magnitude of emissions into nearby home transactions.

Second, we demonstrate a potential complication in using market mechanisms to rectify the inefficiencies negative environmental externalities generate. Several models in economics support prices as a sorting guide to efficiently price the costs and benefits of related non-market goods. But for such models to operate properly, markets must accurately assess environmental conditions, and imperfect information can result in a socially inefficient equilibrium. When direct observation and measurement are difficult, public disclosures like the TRI play an important role for both market efficiency and avoidance and mitigation behavior in the face of environmental dangers.

The right data are essential to investigate this issue. Prior research on the role of TRI information has been largely limited to individual cities (Bui and Mayer 2003, Oberholzer-Gee

---

5 Tiebout (1956) proposes a model where individuals sort in communities with their optimal combination of taxes and amenities. This “voting with your feet” can be applied to an environmental context where, rather than government establishing constraints and regulations, firms are allowed to pollute and households sort based on their preferences for environmental quality. Coase (1960) proposes an alternate solution via private bargaining. Property rights are assigned, and households and firms engage in market transactions to find an agreed-upon level of pollution.

6 When households have information regarding environmental hazards, they adjust behavior in ways that can help offset potential for health consequences. For example, Graff Zivin et al. (2011) find notification of water quality violations leads households to shift consumption from tap to bottled water, and Neidell (2004), Neidell (2009), and Moretti and Neidell (2011) find people adjust behavior to avoid spending time outside on days with dangerous levels of ambient ozone.
and Mitsunari 2006, Mastromonaco 2015), and individual local markets may not be nationally representative. With new industry information in the 2000 release of the TRI oriented toward high-emitting facilities, individual markets may not provide sufficient coverage of transactions near the subset of facilities that garnered the most pointed information updates. Crucially, the national data provided in Zillow’s ZTRAX database allows us to observe thousands of home transactions near high emitting facilities across the U.S. (along with those near the zero/lower-emitters as counterfactuals) while providing the spatial granularity to test how the effect on home prices decays with distance from the polluters.

The remainder of this paper is organized as follows. Section 2 describes the TRI in detail, as well as the relevant policy changes used for identification. Section 3 discusses prior findings on TRI information and home prices within the context of the larger TRI literature. Section 4 describes the data used in the analysis. Section 5 describes possible identification challenges when using the TRI as a measure of pollution and presents our different estimation methods. Section 6 presents our primary results and explores various robustness checks. Sections 7 and 8 provide discussion and concluding remarks. We also present a discussion of the TRI in the media and media exposure as a potential vector for new information in the Appendix.

2 The Toxics Release Inventory – Background

The TRI began reporting on establishment-level pollution and toxics in two major phases. Its initial phase, which extended through 1997, was first established by Public Law 99-499 (the “Superfund Amendments and Reauthorization Act of 1986”). It amended the Comprehensive Environmental Response, Compensation, and Liability Act (hereafter “the Act”) of 1980 and created the TRI. The Act required:
“The owner or operator of a facility subject to the requirements of this section shall complete a toxic chemical release form as published under subsection (g) for each toxic chemical listed under subsection (c) that was manufactured, processed, or otherwise used in quantities exceeding the toxic chemical threshold quantity established in subsection (f) during the preceding calendar year at such facility.”
(Public Law 99-499)

The Act applied to facilities with 10 or more full-time employees, carrying Standard Industrial Classification (SIC) codes ranging from 2000 through 3999, and producing/releasing over a threshold level of specifically noted toxics per year. The original SIC range corresponded largely to manufacturing and industrial applications, but notably did not include transportation and energy production, two large sources of toxic emissions. Firms self-report data in the TRI to the EPA at the end of each calendar year. The EPA then releases these data to the public in the form of the Toxics Release Inventory Report. Due to lags between when firms generate emissions and report data and when the EPA ultimately releases it to the public, the full TRI data for any given year become public around 12-15 months after the end of the relevant reporting year.

The next phase of reporting began in 1998, when the EPA added seven industries to the list of those required to report emissions. They were: electricity production via coal and oil burning (SIC codes 4911, 4931, and 4939); metal and coal mining (SIC codes 10 and 12); solvent recyclers (SIC code 7389); hazardous waste treatment and disposal facilities (SIC code 4953); chemical distributors (SIC code 5169); and petroleum bulk terminals (SIC code 5171). These previously non-reporting industries represented a large share of total toxic releases, particularly the electricity

---

7 In the first reporting year, this threshold was set at 75,000 pounds. This was lowered to 50,000 pounds in the second reporting year, and 25,000 pounds in the third reporting year, and then stabilized for some time. The initial listing of chemicals required to report was a combination of two pre-existing lists of hazardous toxics, the New Jersey Environmental Hazardous Substance List and the Maryland Chemical Inventory Report List. In 1993, the EPA added 23 additional chemicals to the reporting list, with 286 more added in 1994 as the list of who was to report expanded to include all Federal facilities.
production sector. A public statement by then EPA administrator Carol M. Browner upon the release of the new information noted (emphasis added in bold):

The new results, when added to the manufacturing sector already reporting, bring the total releases of toxic chemicals reported nationally to 7.3 billion pounds — **nearly triple the previous number**. Americans now will have the best picture ever of the actual amounts of toxic pollution being emitted by industry into local communities [...] **For the record, between 1997 and 1998, total releases of toxic pollution for the manufacturing sector continued to decline** — this time by 90 million pounds. Next year, we’ll be able to see how all of the combined sectors will “trend” in terms of total emissions and individually [...] You have been given press kits today similar to previous years. This time, however, **as a result of the new data being presented, you will notice lists of states and facilities in eight different categories. The categories are the traditional manufacturing sector and the seven new sectors.**

(Remarks Prepared for Delivery, TRI Announcement, May 11 2000)

In investigating the impact of the policy change, we focus on airborne releases, as they are by far the largest changes due to the policy. This also makes our results more comparable to prior findings using the TRI: Bui and Mayer (2003) and Oberholzer-Gee and Mitsunari (2006), for example, focus on airborne releases, as do many of the studies on health using the TRI (Currie and Schmieder, 2009; Currie, 2011; Currie et al., 2015). From this point forward, unless otherwise noted, the term “new releases” refers to said airborne releases.

We limit analysis to periods before the data release in 2002 to avoid another policy change in the TRI. In reporting year 2000 (data released in 2002), the EPA again expanded the toxics on the reporting list, adding new persistent bioaccumulative toxic (PBT) chemicals and lowering the reporting threshold for certain toxics already on the list, including metals such as lead (100-pound threshold) and mercury (10-pound threshold). Certain dioxins were given low reporting thresholds

---

8 The full (national) TRI release announcement from Carol M. Browner is available at the following URL: https://www.epa.gov/archive/epapages/newsroom_archive/speeches/83c9dac72c1425068525701a0052e3dd.html
of anything greater than 0.1 gram of releases. This policy change impacts many of the same industries. For example, power plants are a large source of both lead and mercury.

Figure 1 shows reported releases, in thousands of tons, from 1988 (the year of first release) through 2002, the final year in our analysis. It splits emissions by SIC codes covered in the initial Act (“Original Reporting SICs”, solid line) and SIC codes included in the 1998 law change (“Newly Added SICs”, dashed line). As expected, reported releases for the Newly Added SICs are effectively zero prior to the 1998 policy change. After the change, releases from these seven SIC are close to the magnitude of all the Original Reporting SICs combined. Figure A1 in the Online Appendix shows the number of reporting plants by Newly Added SICs (dashed line) and Original Reporting SICs (solid line). The number of Original Reporting SIC firms far outnumbers Newly Added SIC firms, illustrating the large shift in reported toxics was due to a smaller number of heavily polluting locations.

[Figure 1 about here]

3 Environmental Hedonic Pricing and Prior Evidence from the TRI

Changes in the value of homes is a common hedonic measure of how households value environmental amenities, but specific TRI-related hedonic work is less common. Bui and Mayer

---

9 For an in depth list of the PBT listing and threshold changes in reporting year 2000, see Chapter 3 of the 2001 Toxics Release Inventory Public Data Release.

10 For example, Greenstone and Gallagher (2008) examine housing prices near Superfund sites both before and after cleanup, and Gamper-Rabindran et al. (2011) consider how the results in Greenstone and Gallagher (2008) vary with levels of geographic aggregation. Chay and Greenstone (2005) use changes in pollution resulting from the Clean Air Act to show improved air quality was associated with increases in home prices in the impacted regions, and Bento et al. (2011), using the more recent Clean Air Act Amendments, show impacts of air quality improvement vary by spatial aggregation as well. Leggett and Bockstael (2000), using variation in water quality in the Chesapeake Bay, show a positive willingness to pay to avoid exposure to fecal coliform. Studies specifically investigating how power plants influence housing prices include Blomquist (1974), Nelson (1981), Gamble and Downing (1982), and Davis (2011). Banzhaf and Walsh (2008) find people “vote with their feet” for environmental quality, using air releases from the TRI as a measure of toxic levels. Davis (2004) shows that the proliferation of information on elevated cancer rates
(2003), Oberholzer-Gee and Mitsunari (2006), and Mastromonaco (2015) examine the impact of toxics reporting on home prices using the TRI, and all three differ from our analysis in both time and geographic scope. Oberholzer-Gee and Mitsunari (2006) use sales records from homes across five Philadelphia counties using the first-ever release of TRI data in 1989. They find home prices decreased across the time of the data release and interpret this change as a revision of the risk expectations of households who, prior to the TRI data release, had underestimated true toxic exposure, though they do not control for background trends in home prices independent of the TRI period. They find results are highly sensitive to distance from a TRI facility, with perceptions being revised only in households a quarter to a half-mile away, and zero effect for homes closer to or further from TRI sites.

Bui and Mayer (2003) use 231 zip codes in Massachusetts and examine both the impact of the initial data release as well as short-run changes in reported toxics in the years that follow. In both cases they find no detectable impact on home prices, even in communities with high newspaper readership (as measured by the Audit Bureau of Circulations) taken as a proxy for access to information. While they find reported releases declined substantially after the initial reporting years, the declines did not seem related to political economy, neighborhood influence, or price changes. Mastromonaco (2015) shows the effect of the later TRI policy change in 2002, with a shift in reporting thresholds for lead and PBT chemicals lowering housing prices in some California cities. Work by Janet Currie and coauthors is similarly related to the TRI but varies in

---

in a Nevada county caused a decrease in home prices of almost 16 percent, and Gayer et al. (2000) find the release of risk information about Superfund sites caused households to revise their expected cancer risks.

11 Other work uses the TRI to understand how the stock market capitalizes information on firm toxic emissions. Hamilton (1995) found stock losses for polluters in the days directly following the initial data release, and Konar and Cohen (2006), using 1988 TRI data, find both toxic chemical releases and environmental lawsuits to be associated.
either outcome or treatment. Using the 2000 policy change, Currie (2011) finds compositional changes in the characteristics of mothers living nearby TRI factories when the EPA report additional information, while Currie et al. (2015) use the opening and closing of TRI facilities rather than changes in reporting behavior.

Our analysis provides two unique contributions to the literature on the TRI and market responses. First, we use microdata on individual housing transactions, with detailed information on homes across over 4,300 zip codes, making it the largest geographic span studying the impact of the information in the TRI in economics to date. Large literatures in real estate and urban economics using state-specific datasets have shown that home prices capitalize other policy changes (e.g., changes to property taxes) heterogeneously across states (Sirmans et al. 2008). Gindelsky et al. (2023) find idiosyncratic household characteristics and geographic-specific factors contribute to this heterogeneity, making extrapolations and generalizations about price capitalization from a single state’s market particularly problematic. Further, the large national dataset we use (which we describe in more detail in Section 4 below) allows for analysis of finer spatial bins around TRI facilities, as a single state may not have sufficient observations within close distance of a high-emitting TRI facility, making state-specific estimates statistically noisier and underpowered relative to a national dataset.

Second, markets may react differently to information about heavy polluters as compared to more typical facilities, given that newspaper coverage was disproportionately focused on the

with negative stock returns. Khanna et al. (1998) found repeated release data had lasting effects on firms already known to be large polluters.

12 Sirmans et al. (2008) review and summarize decades of property tax capitalization research in real estate and urban economics, showing results ranging from no capitalization, to partial capitalization, to full capitalization from dozens of studies, where the vast majority of these studies were each contained to a single state/locality.
highest emitters. In 2000, TRI newspaper coverage often began with headline grabbing superlatives regarding the top or worst polluters in the state/locality, describing the nature of the newly added facilities to the TRI and rankings among the highest emitting facilities. These articles also listed high emitting facilities who reported in prior releases but still neared the top of the list. Prior to 2000, the public had little, if any, information about the absolute magnitude of emissions from power plants or the relative emissions of existing reporting plants. In Appendix B, we highlight local news articles that point to power plants topping the list of highest emitters, but subsequently lump in high emitting facilities for comparison, many of which had reported before. One article from Albany, NY, for example, points to two newly reporting power plants topping their list, followed by two prior reporting manufacturers (a camera manufacturer and a paper mill). In an article from Savannah, GA, the headline and opening paragraph both focus on a newly reporting electric plant, but then list other prior high emitters, with the plant ranking far behind a paper mill (by a factor of 14). Nevertheless, if news coverage is oriented toward the highest emitting facilities, this raises the empirical question as to whether nonlinearities matter to the market, and whether (or to what extent) market prices measurably capitalize this information for large polluters in particular, as opposed to capitalization that varies linearly with the intensity of emissions.

4 Data

4.1 Home Prices

The information shock we examine is national, but likely with very localized or idiosyncratic effects, depending on how far one lives from a major polluter reporting new data. Accordingly, we use national microdata from ZTRAX, a large dataset initially compiled by Zillow that contains
transaction data as well as rich individual property characteristics for sales recorded from local tax assessment data. A key advantage of these data is that it allows for both fine-grained analysis of localized effects and large national coverage. The data are representative of the United States’ national housing market, initially containing 374 million detailed records of transactions across more than 2,750 counties. They include information on each home’s sale price, sale date, mortgage information, foreclosure status, and other information commonly disclosed by a local tax assessor’s office. We link these data with each home’s property characteristics Zillow obtained from the local assessors’ offices, which typically includes the size of the home (in square feet), number of bedrooms and bathrooms, year built, and other such characteristics of the home.

We scrutinized missing data and extreme values as part of our data cleaning and initial culling of outliers following similar approaches to Nolte et al. (2023). For example, we omit foreclosures...

---

13 Data provided by Zillow through the Zillow Transaction and Assessment Dataset (ZTRAX). More information on accessing the data can be found at http://www.zillow.com/ztrax. The results and opinions are those of the author(s) and do not reflect the position of Zillow Group. Nonproprietary code used to generate the results for this paper is available upon request of the authors. This data vintage was downloaded in Q4 of 2016.

14 The ZTRAX database often contains multiple dates for a given transaction, but in some cases only a single date, which is a convention that varies from county to county. We follow Zillow’s FAQ in assigning our sale date: “Generally, we recommend using the DocumentDate. If missing, then Signature Date, if missing, then recording date” (https://www.zillow.com/research/ztrax/ztrax-faqs/).

15 Some states do not require mandatory disclosure of the sale price, so we have limited data for the following states and remove them for our analysis: Idaho, Indiana, Kansas, Mississippi, Missouri, Montana, New Mexico, North Dakota, South Dakota, Texas, Utah, and Wyoming. Even for states that do disclose sale price, there are some limitations in coverage, as coverage begins at different times depending on the state/county. Zillow notes in their Frequently Asked Questions site that, “In Zillow’s work analyzing these data sets to create home value indices, we find sample sizes are sufficient for most areas going back to April 1996,” (https://www.zillow.com/research/ztrax/ztrax-faqs/) and Nolte et al (2023) find more recent data is more reliable. The bulk of our analysis centers on 2000.

16 Zillow’s ZTRAX data contains separate transaction and assessment files by state, where all transactions need to be linked to corresponding assessment records. With guidance from Zillow, we were able to merge the bulk of the data, but not without some data loss (which figures into the size of our final sample). The assessment files from our data vintage generally come from 2014-2016, which means we must assume observable property characteristics in our dataset are static over time. This likely adds to measurement error as some properties have largescale renovations that might change the number of bedrooms or bathrooms, for example.
or sales that appear to be non-arm’s length using several document code categories.\textsuperscript{17} The raw data included sales of empty plots of land, commercial/industrial property transactions, agricultural sales, and a host of types of properties not relevant to the scope of our paper, so we limited the sample to single family homes.\textsuperscript{18} We cull residences larger than 2.5 acres and below the 2.5\textsuperscript{th} percentile by state, and cull or winsorize outlier homes that are in the upper tail of the distribution of other characteristics (i.e., winsorize homes with more than five bedrooms or more than four bathrooms or omit homes larger than 5,000 square feet). We remove homes at the extreme ends of the price distribution for our analysis, which were homes that had a reported sale price of less than $15,000 or greater than $1 million. We cull homes reported as built prior to the year 1900 or after 2003 given the timing of our analysis. While the Zillow data set contains a vast number of property characteristics, we primarily rely on the variables above that have the most coverage nationally to limit data loss due to missing variables.\textsuperscript{19} We also include missing flags for some control variables to allow for the inclusion of a slightly broader range of variables, as some states/municipalities do not collect all characteristics.

From these data we construct the following three different samples for our analyses. Our first sample includes 5,383,770 transactions that took place within a county containing any TRI facility within approximately two years (pre/post) years of the 2000 release (from 1998 through 2002), which we use for a difference-in-differences analysis. In our second sample, for comparison with

\textsuperscript{17} We limit transactions to arm’s-length transactions in a way generally consistent with those outlined by Nolte et al (2023) and other papers using ZTRAX data. Our code is available upon request.

\textsuperscript{18} We limited our sample based on the single-family residential categories used by Wentland et al. (2020), which shows the land use codes in their Online Appendix (Table A20) associated with SFR (suburban and rural) land use types.

\textsuperscript{19} We conducted a sensitivity analysis in untabulated regressions that incorporates property characteristics to determine whether the results are sensitive to omitted variables for which we can control. Our results are largely robust to omitting variables that have more limited coverage.
a prior version of this paper using zip code level data scraped from Zillow’s website, we aggregate
data to the zip code level for monthly zip code average and median prices respectively, over the
same sample period (N=223,120). Third, to explore more immediate effects of the information
shock using a regression discontinuity in time design, we use smaller subsets of the microdata to
look at transactions within short time windows around the release period (+/- 210 days in our
default specification) and within 0.5 miles of the facility. Because this analysis centers on only the
transactions that occurred during these windows around the information shock, this third sample
is much smaller than our initial data set. We return to this point when we describe our methodology.

Table 1 shows the summary statistics for the variables used in our main sets of analyses,
tabulating the means and standard deviations between properties within 0.5 mile and with 5 miles
of a TRI facility. In Table 1, we also split apart properties nearest to High Emitters (TRI
emissions>100 tons) from those near Low Emitters (<100 tons) within 0.5 miles. The summary
statistics suggest homes near the highest emitters have slightly less desirable property
characteristics (e.g., smaller and older, on average) and have a lower average price. In Appendix
Table A2, we also compare observable household demographics at the zip code level from the
2000 census. Zip codes with emitters in the higher levels tend to have higher vacancy rates, be of
lower income, and have a higher share of black residents. These baseline differences motivate our
use of property-specific controls in subsequent analysis. Moreover, these differences also help
motivate use of quasi-experimental methods like difference-in-differences and regression
discontinuity in time (RDiT) research designs.

[Table 1 about here]
4.2 TRI – Toxic Release Inventory and Data Imperfections

Toxic data are from the TRI Basic Data Files on the EPA website, annually aggregated by facility and emission type, with information on facility name and location, toxics released, and on- and off-site releases. All emissions data in our analysis are recorded in pounds, which we convert into tons.\(^{20}\) Also included are the SIC classification codes for each reporting producer, which we use to identify polluters impacted by the policy change.\(^{21}\) Prior work on the TRI and home values has separated toxics by categories of potential health damage to test for differences across assessed health risk, and found none (Bui and Mayer, 2003), so our models do not separate by toxicity.

Imperfections in data collection mean the TRI is an imprecise measure of released toxics. Firms appear and disappear due to openings/closings, failure to produce above the cutoff amount of toxics required to report, etc. This causes year-to-year changes in both the number of firms reporting and total emissions. Reported data are often estimates based on production levels rather than directly measured emissions, and while the EPA does enforce reporting, there is no regular verification of reported versus true toxic releases.\(^{22}\) Such problems mean the TRI data may be an unreliable measure of exact toxic exposure, which led Currie et al. (2015) to develop an instrumental variable strategy using firm openings and closings.\(^{23}\) We address this issue by focusing on the addition of

\(^{20}\) After 2000, most of the raw data remain in pounds, though dioxins are reported in grams.

\(^{21}\) In later years, the TRI reports NAICS codes rather than SIC codes.

\(^{22}\) de Marchi and Hamilton (2006), for example, show that when pollution monitors can be used to examine ambient toxic levels, drops in emissions reported in the TRI are often smaller than those measured by nearby monitors. They further show the distribution of certain reported emissions fails the “Bedford’s Law” test for a distribution of “true” data, and in some cases reported numbers appear to suggest “rule of thumb” uses for reporting rather than direct production numbers. Bedford’s Law states that in a distribution of data, the first digit of all values is unevenly distributed across the 1-9 spectrum similar to a logarithmic scale, with 1 being represented approximately 30% of the time and each larger number appearing less and less frequently.

\(^{23}\) When investigating the impact of toxics on infant health, they find no significant effects with OLS and large, significant effects with IV, suggesting measurement error in TRI data is a problem.
large-scale newly recorded releases rather than smaller year-to-year marginal effects, and by focusing on large vs. small emitters rather than linear emission effects.

There is also the concern that the general public is unaware of the existence of the TRI, and thus cannot benefit from any expansions of the data available. We provide examples in Appendix B that illustrate how the media focused on TRI-related stories at the time of each new data release, particularly around the releases impacted by the 1998 policy change. It need not be that households actively sought TRI data, but instead responded to the coverage of the data provided by the media, or even learned from neighbors or real estate agents who had learned from media coverage, etc. If households neither sought out this information nor remembered it from news coverage, we would expect little, if any, capitalization of any kind. With null capitalization, we cannot, however, disentangle whether households are not informed from households who are informed but do not care/act. We discuss further information challenges and return to this point in our discussion of the results.

5 Methodology

5.1 Identification challenges – pollution effects vs. information effects

Pollution changes and information changes often move together: a toxic event bringing firms to public attention, such as the incident at Three Mile Island (Nelson, 1981; Gamble and Downing, 1982), or newly constructed power plants or TRI facilities moving into neighborhoods (Davis, 2011, Currie et al. 2015). Our design avoids potential contamination from other factors that move with changes in environmental quality, such as plants openings/closings, economic development, migration patterns, and emissions regulation. Still, interpreting price changes around the time of the TRI release as the result of information requires no other factors correlated with treatment
changed due to the policy. For example, if firms that are newly required to report adjust production or employment as a result, there could be economic impacts that, in turn, influence housing prices. Similarly, if firms actively reduce pollution as a result of the policy, information and true pollution levels change simultaneously, making it difficult to separate specific impacts of information disclosure and the willingness to pay to avoid toxics.\textsuperscript{24} A benefit of using the TRI is there is a substantial lag between when firms produce toxics, when they report data, and when the TRI make the data publicly available. This helps us separate the impact of the information shock from any changes caused within the firms in response to the new reporting regulations. That is, if the policy change itself impacts home prices, changes should occur during or just following the year of toxic production (1998) rather than in 2000 when the data are made public.

Another concern is that information from newly reporting facilities coincides with already existing dirty locations, and any effects we detect are about prior emissions trends. However, we show the location of newly reporting industries has almost no relationship to prior levels of TRI-reported pollution. We find little correlation between the levels of reported toxics in 1997 (based on the prior year’s TRI report) and the level of newly reported toxics in 1998. Appendix Figure A2 shows a scatter plot of toxics from new industries in 1998 against 1997 reported toxic levels, along with a fitted line. The correlation coefficient is 0.003, and the bivariate regression fitted line yields a coefficient of 0.002.

\textbf{5.2 Media coverage and determining the treatment}

\textsuperscript{24} Active attempts were made by some firms to reduce emissions after the initial TRI release in the form of the “33/50” plan, where a number of producers aimed to reduce toxic emissions by 33\% in 1992 and 50\% in 1995 (Environmental Protection Agency, 1999).
The first TRI data release focused on emissions from the manufacturing sector. The legislative change in 1998 meant the updated TRI now included additional heavy polluters, but media coverage was not exclusively focused on the new group. News coverage suggests an asymmetric emphasis and attention on the largest polluters reporting in the TRI. Headlines would initially point to the new types of firms in the TRI (e.g., “POWER PLANT NUMBERS NOW A PART OF POLLUTION REPORT” from the Savannah Morning News, or “COAL-BURNING PLANTS TOP EPA LIST MOST TOXIC CHEMICALS RELEASED FROM POWER PLANTS, ACCORDING TO AGENCY. FIRSTENERGY HAS TWO IN TOP 10” from the Akron Beacon Journal), but the coverage would then list other high emitting facilities from manufacturing or other sector who had reported in prior years, e.g., paper mills in Georgia, or steel manufacturers in Ohio. Because the news coverage focused on the highest polluters, we first examine the hypothesis that home prices near the highest emitting facilities capitalized this information. We later separately estimate whether newly reporting high emitters had a different effect than prior reporting facilities, although there are data limitations to this approach.

We first match TRI facilities to individual homes based on location data, focusing on reported air releases as the most salient variable from the release. To illustrate the benefit of using individual transaction data, we also conduct analysis contrasting aggregated zip code-level sales information with property-level data. In subsequent analysis we leverage finer location-by-location matching

---

25 We use latitude and longitude coordinates in ZTRAX to calculate the distance between each property and each TRI facility in Stata, identifying the nearest TRI facility for each property. Based on its nearest facility, we then assign each property a discrete distance bin and its corresponding TRI emissions intensity bin. According to Zillow, the ZTRAX database uses enhanced Tiger coordinates, where its geocoding is interpolated on block segments by address. They note in their Frequently Asked Questions about ZTRAX that, “these are more useful for measuring distance between properties, [and are] not as useful in rendering on a map” (https://www.zillow.com/research/ztrax/ztrax-faqs/). As an additional quality check, we manually checked a random sample of properties’ distances to TRI facilities using Google Maps and other sources, finding the distances and coordinates to be accurate in our sample.
by distance from TRI facilities. We assign each home the closest large-emitter, either by zip code or by distance depending on the specification. When multiple locations exist within a given distance range, we select the largest emitter, as the main mechanism for the information shock (media coverage) focused overwhelmingly on the largest emitters.

We use two common empirical approaches in the literature that leverage the timing and spatial dimension of the information shock for identification: difference-in-differences (DiD) and regression discontinuity in time (RDiT). Recent literature (e.g., Cheng and Long 2022) uses these methods in tandem to explore longer-run effects and shorter-run effects, respectively, as we employ the DiD method for our analysis of a longer (4 year: 2 years pre/post) sample and a RDiT for an assessment of the immediate impact of the information shock over a shorter window (+/- 210 days or 30 weeks in our default specification). We describe each method in sections 5.1 and 5.2.

Using time as part of our estimation process requires determining treatment timing. Unfortunately, it is impossible to know the exact time at which individual buyers/sellers became aware of the 1998 TRI information. The official release of the national 1998 data was May 11th of 2000. However, searching through news archives we found many local and national printed news sources provided more localized information as early as January of 2000, which likely came from other data sources such as state-level environmental agencies releasing data from their own state to the public and news media. Using a news article search, we assign treatment using the first news article we could find on a state-by-state basis. When we were unable to find specific dates for a given state, we used March 1st 2000 as our definition of the treatment period, which is approximately the average or midpoint of the sample of articles we found. In Appendix Table

26 The average in Appendix Table B1 is February 27, but given that the DiD design uses monthly data, we round to March 1 for the default treatment date among states not found in our news search.
B1, we tabulate newspaper coverage by state which we observe through Lexis-Nexis, Newsbank, and other news archives. We also investigate the sensitivity of our results to this measurement of the treatment date.

5.1. Difference-in-Differences (DiD) Design

Hedonic regression analysis has been a commonly used methodology for valuation of nonmarket amenities in the housing literature since Rosen (1974). More recently, researchers couple this approach with a quasi-experimental research design (for a review, see Parmeter and Pope, 2013). We begin with a standard hedonic model that accounts for portion of a home’s (logged) price that can be explained by the idiosyncratic characteristics the home ($HC_h$), allowing the remaining price variation to be modeled in a difference-in-differences framework as follows:

$$\log(price)_{hmz} = \alpha_z + \beta Post_m \times TRI_z + Month_m + ZipCode_z + \theta HC_h + \varepsilon_{zd}$$

(1)

where $Month$ is a time fixed effect for calendar month $m$, $ZipCode$ is a location fixed effect for zip code $z$, $Post$ is an indicator for a home $h$ transacting in the window following the TRI public release of emissions data that would have been reported on in local news outlets, and $TRI_z$ is an indicator for whether a home shares a zip code with a TRI facility (where we take the maximum facility’s emissions in a given zip code and bin them as described below). $HC_h$ represent the following controls common to hedonic price regressions, which account for many of the most prominent observable characteristics of heterogeneous properties: square footage, number of bedrooms and

---

27 The data is more suited to a hedonic approach (as opposed to a repeat sales model) for a couple reasons. First, given that homes are sold infrequently, to have a sufficiently large repeat sales sample we would need a larger proportion of the national sample to extend back much further into the 1990s, which is a limitation of the current ZTRAX sample. Second, the hedonic approach utilizes the rich availability of property characteristics in the data as controls for explaining variation in home prices, without having to throw away observations where we do not observe a prior sale in the data. A repeat sales method, however, may be more appropriate to analyze more recent time periods with ZTRAX.

28 A list of the states and corresponding release date assignments are available in Appendix B1.
bathrooms, garage, logged acreage, year the home was built, whether the home is a single-story ranch, has a pool, and missing flag indicators. We cluster standard errors by zip code.

We divide properties nearest to TRI facilities into five reported emissions buckets that roughly correspond to quartiles of positive emissions (just over 0-10, 10-30, 30-100, and 100+ tons of air emissions) and a category for zero emissions, and a category for no reports within the county. This difference-in-differences setup allows us to contrast the home price effect of being near the “most treated” highest emitters (TRI emissions >100) against the lower categories of emitters before versus after the release of this new information. This categorical approach also allows for potential nonlinear effects of emissions on home prices, allowing coefficients to vary by each bucket. To illustrate the benefits of property-specific transaction data, we vary this basic DiD setup to explore the differences between results using individual property-level data and the data aggregated to the zip code level (without property-specific controls). We discuss the variations to this DiD setup in more detail when we report the results in section 6 below.

5.2. Regression Discontinuity in Time (RDiT) Design

As an alternate estimation approach, we follow a regression discontinuity design similar to Moulton, Waller, and Wentland (2018), which is broadly consistent with the event study literature in finance and applied microeconomics. This approach combines a hedonic sale price model and a standard dual linear spline RD model, using the date of sale as the running variable. Equation (2) shows the estimation setup.\(^{29}\) We focus on logged observed sale price as the outcome of interest over the 210 days (30 weeks) prior to and after the estimated disclosure event in our default

\(^{29}\) Given our running variable is time, we acknowledge this is also referred to as comparative interrupted time series (CITS). See also Hausman and Rapson (2017) who discuss RD using a time running variable more generally.
We selected this window as our default based on the data-driven approach described by Calonico, Cattaneo, and Titiunik (2014) for selecting optimal bandwidth. The controls (HC\textsubscript{hd}) for home \( h \) sold on day \( d \) follow those in our difference-in-difference estimates:

\[
\log(\text{price})_{hd} = \alpha_s + \beta(\text{SaleDate}_{hd} \geq E) \times \text{TRI}_z + \delta_1(\text{SaleDate}_{hd} - E) \times \text{TRI}_z + \delta_2(\text{SaleDate}_{hd} \geq E) \times \text{TRI}_z + \theta HC_{hd} + \epsilon_{hd}
\]  

(2)

The coefficient \( \delta_1 \) captures the sale price time trend prior to the event \( E \) cutoff. We also include this same re-centered trend interacted with an indicator variable equal to one when the sale date was at or after the policy change cutoff for a home in a given state (see Appendix Table B1 for a listing of cutoff dates by state). The coefficient \( \delta_2 \) represents the change in the post-cutoff price time trend, which determines if any price change following the announcement dissipates or grows over the post-cutoff window. In our default specification, we interact trends with county level fixed effects to allow background trends to vary by locale. The \( \beta \) coefficient associated with an indicator variable (equal to one when the sale date is after the policy change) estimates the difference in the pre- and post-cutoff trends’ intercepts at the cutoff. Thus, the estimated intercept difference in this design can be interpreted as the treatment effect of the coverage of the TRI release, which is the key coefficient of interest and is referred to and labeled as “Discontinuity” or “D” in our results tables.

We also investigate the extent to which distance from the (dis)amenity (i.e., the relevant TRI facility) matters. Consistent with prior literature on environmental (dis)amenities, we expect homes

---

\( ^{30} \) Regarding bandwidth selection, the default for \texttt{rrobust} package in Stata uses one common mean squared error (MSE)-optimal bandwidth selector for the RD treatment-effect estimator, which follows from Calonico et al. (2014). We estimate a nonparametric local linear variation of this regression discontinuity design, which we describe in our Results section below. The default bandwidth selected using this approach is 216 days (30.85 weeks). We round to 30 weeks for the remaining specifications as our default; but, for robustness we explore a number of alternative window sizes later in the paper.
closest to the polluting plants will be most affected by the policy. We hypothesize the estimated effect should be larger for homes nearer higher emitting plants, as the highest polluting plants tended to be most noticeable by their neighbors and tended to receive the most news coverage during the time period around the information release. In subsequent analysis, we explore how the effect decays with distance. In our default specification, we focus on properties within 0.5 mile of the nearest TRI facility, based on prior findings in the TRI literature (Currie et al., 2015). We later consider alternative distances that measure the effect on the housing market up to 5 miles away from a TRI facility.

Methodologically, the controls in both models serve several purposes. We examine a large cross-section of homes over time, and these homes are heterogeneous along numerous key dimensions. While aggregation across a large national data set may allay compositional concerns, controlling for arguably the most important determinants of a home’s price (size, bedrooms, bathrooms, age, location, etc.) allows for a more straightforward apples-to-apples comparison of a cross-section of homes within a given period. A handful of characteristics and location explain most of the time-invariant variation in home prices. Hence, the estimated emission announcement effect comes from the variation in price not explained by these factors, just as the “event” in the finance literature explains the variation in a firm’s equity price or return not already explained by quantifiable firm-specific factors that make up the fundamentals of its valuation. For robustness, we employ a residual approach, first estimating the hedonic elements in equation (2) (property characteristics, county and month fixed effects, and county specific linear trends) over a longer sample (1999 through 2001). This uses more of the data to estimate the coefficients on the property characteristics and fixed effects for a more precise model fit; and, because we use month fixed
effects over multiple years, also accounts for seasonal variation in home prices. We then estimate the RDiT elements ($\beta, \delta_1, \delta_2$) from equation (2) on these residuals in a second stage, which simplifies the RD specification to a more classic setup without covariates (as the dependent variable is already residualized).

One advantage of a short-window analysis is that, like an event study, we can be more confident that the information shock is properly identified as compared to longer windows where subsequent shocks have greater likelihood of playing a role. Similar to the DiD design, we also interact the discontinuity with a TRI treatment indicator equal to one when the facility has greater than 100 tons of emissions and zero if less than 100 tons to separately estimate the difference in discontinuity relative to the reference group of properties near a TRI facility with lower emissions. Our figures adapt equation (2) to stratify the sample by each TRI category, where the $\beta$ coefficient instead represents the intercept shift or break from its pre-event path. The former may be termed a “difference in regression discontinuity” design, where the latter is a classic RDiT design (as we present in equation (2) above).

We employ multiple approaches, variations of both DiD and RDiT designs, for two reasons. First, prior studies rely on coarser, more aggregated data (like zip-code level home values) using a DiD approach that we replicate, which we then contrast with a DiD approach using microdata in the next section. This helps illustrate the utility of microdata like ZTRAX in assessing research questions like those we examine in this study. Second, the classic DiD and RDiT designs have distinct assumptions and differ regarding the nature of the counterfactual. The classic DiD design assumes the treatment group would have evolved on a similar path as the control group; hence, parallel trends prior to the treatment date is a critical assumption for DiD designs. In contrast, a
classic RDiT design assumes the treatment group evolves on a path following its own prior trend, where a break or discontinuity is measured from the extrapolation from its own trend. There are, of course, variations of these designs, including a difference in discontinuity design that combines elements of both as we describe above. While differences in these types of designs can be subtle, it is important to note that if we are interested in an effect over a shorter horizon versus a longer horizon, this distinction matters more. For longer horizons or broader windows of analysis, subsequent contemporaneous shocks become more of a threat to identification in RDiT or event studies, which is why these designs are often tailored toward relatively narrow windows of time. When a control group in a DiD design is also affected by subsequent shocks, it is plausible (given parallel trends) that it continues to serve as a valid counterfactual over a longer period of time.  

Hence, we report results for both empirical approaches for robustness across these counterfactuals or modeling assumptions. Moreover, given these differences and subtly different interpretations, we caution at the outset that we do not expect the coefficient estimates to be identical across approaches. A similar directional result, however, provides more confidence a finding is not sensitive to particular modeling assumptions.

6 Results
6.1 Aggregate vs. Disaggregated/Property-level Estimates – DiD Results

Without large-scale ZTRAX home-level data, one way to estimate the effect of TRI facilities on home prices is to estimate a coarser variation of the DiD in equation (1), where home prices are measured as the mean or median home price in a given zip code each month. For example, Sanders (2012) used data scraped from Zillow’s website on median zip code-level prices, matching to zip

---

31 There are a number of other differences between our DiD and RDiT approaches, including accounting for local (county-level) trends.
codes containing corresponding TRI facilities. To provide contrast with the approach outlined in equation (1), we first conduct a more aggregated DiD analysis at the zip code level. Appendix Table A1 shows no statistically significant difference in property price trends (prior to the TRI release) across zip codes with varying emissions, which is consistent with the illustration of price trends in Appendix Figure A3 by emissions category.\textsuperscript{32} The common pre-trend helps support the use of DiD.

The first two columns of Table 2 show the results when we aggregate to (logged) mean and median zip code-level prices by month, respectively. In this case, TRI facilities are split among the emissions categories discussed above that correspond to the maximum emitting facility in the zip code: 0 to <10, 10 to <30, 30 to < 100, and 100+ tons of reported air emissions. In this more aggregated analysis, we have 4,360 zip codes containing TRI facilities, amounting to 223,120 zip code-by-month observations over the 1998-2002 window. Using this simplified approach, we see a negative impact on home prices of just over 3\% for zip codes with the highest emitting facilities after the TRI release in 2000 relative to homes near TRI facilities with zero air emissions. We do observe a small, statistically significant effect (nearly 2\%) for homes near facilities of the next highest emitter category (30 to <100) in columns 1 and 2. However, note that because we are aggregating in this analysis, we have no controls for property characteristics in these specifications.

If we instead estimate the DiD specification in equation (1) using over 5 million transactions in the ZTRAX microdata that occur in TRI facility zip codes, we observe more pronounced price

---

\textsuperscript{32} Trend differences between the omitted TRI disclosure group (TRI = 0) and the higher emissions group are economically small and, in all cases but column 3 in Table A1, statistically insignificant. When we shorten the window of analysis in column 3 to a year prior to the treatment (instead of two) in column 4, the pre-trends become statistically indistinguishable. In untabulated results, we observe a similar significant negative effect for the highest emitters when we alter the DiD windows in Table 2, including a +/- one year window from the treatment date. The seasonal variation in prices, as visible in Figure A3, is a point we return to in section 6.3.1.
effects among the highest emitters. Properties in zip codes with the highest emitting polluters fell by about 4.9% in value relative to similar properties near TRI facilities with no air emissions, while the next highest category (30 to <100) is no longer statistically significant. The results from the microdata in column (3) show a more pronounced negative effect on home prices, potentially due to compositional differences of homes sold on the market (for which the hedonic model accounts), among other reasons.

[Table 2 about here]

The specification in column (3) leverages the more granular scope of the property-level data but does not fully utilize this granularity for the assignment of the treatment group. Assigning treatment by zip code likely contributes to further measurement error because zip codes in the US are irregularly shaped. Some homes may be relatively close to TRI facilities but not share the same zip code, while others may be farther away but share the same zip code. Thus, a zip code-based analysis (which was the finest high-frequency geography available publicly prior to ZTRAX) may suffer from contaminated treatment and control groups, likely biasing the observed effect toward zero if the effect on home values is tied to distance. This leads us to our alternate approach of the RDiT.

6.2 Regression Discontinuity in Time (RDiT) Results

The RDiT analysis looks at a short temporal window around the “event” of the TRI release in the spring of 2000, estimating the model in equation (2) of Section 5. Rather than determining proximity by zip code, we assign a home sale to a treatment group in our initial analysis if it falls within 0.5-mile distance range of a TRI facility. Given our prior analysis only finds effects for the highest emitters (TRI>100), we collapse lower emitters into a single reference category to simplify
presentation of the results in Table 3 (Panel A). In untabulated tests, results are similar maintaining the same categories as Table 2.

Table 3, Panel A shows home prices near (<0.5 mile) the highest emitting TRI facilities (>100 tons) fell in value directly following the TRI information release in the spring of 2000. This is robust across numerous alternative specifications. In column (1), we estimate a simplified or naïve version of the RDiT in equation (2) with national (common) time trends, but no other controls or fixed effects, finding a large negative impact on home prices (-17%) for homes near the highest emitting facilities (D * High Emitter). Given the likely compositional differences in property characteristics highlighted in Table 1, column (2) incorporates property characteristic controls, which attenuates some of this large effect in column (1). Given housing markets are heterogeneous and spatially localized, our default specification in column (3) estimates equation (2) with the full set of controls, county fixed effects, and county-time trend interactions, to account for locale differences and market dynamics. Accounting for these factors reduces the estimate size to 11%, and it is statistically significant at the 1% level. For comparison, the size of this effect is similar to Mastromonaco (2015), who finds houses within one mile of facilities added under the TRI’s 2000 lead reporting expansion decreased in value by 11%.33

33 As another point of comparison, Currie et al. (2015) find that effects of TRI plant openings within 0.5 miles reduce home prices by about 11%, but the results are not directly comparable due to the magnitude and method of their treatment. First, they estimate the effect of a plant opening, while we estimate the effect of an information update on an existing plant. Second, the mean annual toxic emissions for the new plants is around 23,000 pounds, while our treatment category is reported releases over 100 tons.
A), we estimate a few variations of the RDiT design using residuals to explore this further. Recall from section 5.2 that we regress the hedonic elements from the default specification on multiple years of data (1999-2001), but also include month fixed effects to account for seasonal, month-specific market dynamics. When we do this, the discontinuity (D) for homes near low emitters is cut to 2.9%, while our main discontinuity effect for home values near high emitters is unchanged around 11%. When we restrict the sample to only properties near (<0.5 mile) the highest emitters and estimate a classic RDiT that is “within treatment group,” we find similar results as the interacted difference-in-discontinuity approach in the first four columns of Table 3. In particular, we estimate an RDiT on the residuals using a dual linear spline (as in equation (2)) and a nonparametric local linear approach in columns (5) and (6), respectively.\textsuperscript{34} Overall, while the modeling assumptions and interpretations are somewhat different across specifications, the collective evidence in Tables 2 and 3 suggest that housing markets near the highest emitting facilities capitalized the information from the news coverage of the TRI release in 2000.

Figure 2 provides a graphical illustration of our results from Table 3 (Panel A), showing binned residuals (by week) from our RDiT hedonic model stratified by TRI category over the 210-day period before and after the TRI release (day 0 is each state’s re-centered initial news coverage date). Specifically, Figure 2 illustrates a general upward trend in prices prior to the announcement for properties within 0.5 miles of facilities in the highest (>100) category, but a clear downward shift or break in home prices immediately after the TRI information release for these “most treated” properties. The bottom panel illustrates the home price trends for properties near counterfactual

\textsuperscript{34} In untabulated results, we also excluded California, given there was a contemporaneous stock market crash around this time period, which may have affected home values in a variety of ways. Overall, we find the results were very similar to excluding this state. We also exclude states with very few observations near TRI facilities to avoid issues with estimating fixed effects.
facilities and lower emitting ("less treated") facilities, which received relatively less media coverage or were favorably compared to the highest polluters. In contrast to the high polluting "treated" facilities, we observe a small but slightly positive effect for these "less treated" facilities.

[Figure 2 about here]

6.3 Robustness – Alternative Distances, Time Windows, and Model Specifications

In this subsection, we briefly discuss a variety of robustness tests and alternative specifications that provide different insights into the nature of the main finding and how it holds up to alternative specifications and placebo tests common in the RD literature.

6.3.1 Placebo and Falsification Tests

To further rule out seasonal concerns and other alternative explanations, we conducted placebo tests by re-estimating all specifications in Table 3, Panel A for the prior year, 1999, for Panel B. While there had been news coverage of the TRI for years, the nature of the coverage was different as it had not included many of the highest polluting facilities in the US that would subsequently report in 2000. Local polluting manufacturers might be named in newspaper coverage, but they had been reporting for years and there was little, if any, data that could equate their level of pollution to power plants and other large polluters who were not yet reporting. Table 3, Panel B shows no significant capitalization for homes near the highest emitters in the year prior to the 2000 release. We find some evidence of a positive effect for properties in the control group (lower emitting facilities) in columns (2) and (3), but when we estimate the RDiT after accounting for month fixed effects in column (4), this effect all but vanishes, again suggesting some small seasonal component for markets near control facilities. One interpretation of these findings is that the absolute level of information had already been capitalized among the highest emitting facilities.
prior to 2000; however, the 2000 release provided the market more perspective about the relative magnitude of these emissions (i.e., that some locations are as toxic, in terms of emissions, as some power plants). One local newspaper in West Virginia (see Appendix B) had said as much: “Even worse, this pollution has been there all along, though until publication of the latest TRI report, nobody had a handle on it.”

We next vary the bandwidth of the RDiT specification to ensure results are not sensitive to a single window of analysis. We adjust window size from our default of 210 days on both sides of the cutoff at 15-day intervals from 60 through 360 days. Moulton, Waller, Wentland (2018) find that extending time windows to cover more of the year makes seasonality and nonlinearities become more of a concern when modeling housing market trends. Hence, in Panel A of Figure 3 we show the default specification using a dual linear spline approach, while Panel B uses a quadratic spline. The main finding is generally unchanged across bandwidth specifications, as we observe a large negative price effect for homes nearest to the highest polluting facilities directly following the 2000 TRI release across all bandwidths. In both panels, the 95 percent confidence interval straddles zero in some specifications, but for reasons outlined by Moulton, Waller, and Wentland (2018), the quadratic generally has a better fit for longer windows and the linear has a better fit for shorter windows. We observe the same phenomena in this setting.

[Figure 3 A & B about here]

In an additional falsification test, we examine whether there are other structural breaks in the data that occur contemporaneously with our discontinuity. Specifically, we test whether our control variables exhibit a significant discontinuity in a way that might provide an alternative explanation for our findings, estimating the same RDiT as in our default specification (equation (2)), but with
each property characteristic as a dependent variable. Table A3 in the Online Appendix tabulates our results. We observe no alarming evidence of an alternative break in the data from these results, as all but one of the variables have a statistically insignificant noisy effect for the discontinuity of interest. Bathrooms is statistically significant discontinuity for high emitters, but is economically very small and cannot explain much of the price gap in the main results.\textsuperscript{35}

6.3.2 Alternative Distances

We next explore the extent to which this TRI house price effect decays spatially by estimating the effect on properties further away. In Figure 4, panels A and B, we estimate our default specification for homes within 0.5 mile increments up to 5 miles away from a high-emitting (>100 tons air emissions, annually) TRI facility. We only observe a negative home price effect for homes nearest (0-0.5 mile) to the highest polluting facilities, with no statistically significant effect for the highest polluting facilities in the next distance bin (0.5-1 miles) or any of the subsequent further facility interactions. This spatial decay is generally similar to Currie et al. (2015) results on plant openings. Results of this analysis also reveal what is going on “under the hood” of the more aggregated zip code level effects. The moderate home price decline of 3-5\% we observe in the DiD analysis for the highest polluters likely averages large effects closest to the facility with small or null effects further away. This highlights a key benefit of the fine-grained ZTRAX data, showing capitalization of this environmental information shock is highly localized and highly specific to homes near the largest polluters.\textsuperscript{36}

\textsuperscript{35} Statistically speaking, there is a nontrivial likelihood for a spurious correlation when we regress our RDiT on a large number of dependent variables, given a 95\% confidence threshold.

\textsuperscript{36} When we employ less granular fixed effects (like state fixed effects) and time trends by state in our equation where we generate residuals, the first few columns in Table A4 show a somewhat more pronounced effect. This is broadly consistent with the initial regressions from Table 3 (Panel A).
6.3.3 Alternative timing of the TRI release and measurement error

In the next set of placebo tests, we explore alternative “placebo cutoffs” by alternating the treatment date in both directions of the cutoff by 5 days at a time up to 380 days. Using our default RDiT specification, Appendix Figure A4 shows a large statistically significant effect for the treatment time that corresponds with the TRI release coverage (which we discuss in more detail in Appendix B). The figure also shows that arbitrarily altering these Post dates to cover other dates around the cutoff produces noisy, null effects.

To further assess whether this treatment time is spurious or whether deviations from this time are associated with measurement error, we test alternative approaches to setting this date. First, we ignore the individual newspaper coverage, and use our average March 1 as the date of the information release to assess whether the evidence points to symptoms of measurement error. For example, in Appendix Figure A5, we estimate our default RDiT specification using March 1 as the discontinuity date for all observations, and still find a negative but noisier effect of about -5%. As we iteratively modify states to use the more accurate date where we have identified specific dates associated with local newspaper coverage, the effect size becomes more negative and the estimate is more precise. That suggests more accurate information dates make estimates more pronounced and precise, as we would predict if the effect is properly identified and we are, in fact, reducing measurement error by using local newspaper coverage dates.

We also explored specifications that date the discontinuity from the news coverage of the national TRI release, which began on May 11, 2000, and incorporate a separate effect for properties near new reporting facilities. An issue with using May 11 as our default specification is that we
have ample evidence local news coverage occurred throughout the U.S. in the months prior. We address this using a donut RD design, where we cut sales from our dataset that occurred during the three months leading up to May 11. We estimate this donut RDiT that uses the winter months and prior (cutting 90 days prior to May 11) as the 210 pre-trend in columns (4) and (5) in Table A4 in the Online Appendix. We still find a negative effect similar to when we use our default discontinuity timing (-12%); but, it is noisier and not significant at conventional levels.

In columns 6 and 7 we include an interaction (or stratification) to test for differential effects for newly reporting facilities (vs. large facilities that had reported in prior years). This has a negative interactive effect, suggesting that homes near newly reporting facilities saw the largest price effects, but the interacted coefficient estimate (D * New Reporter) is not statistically significant. The final column in Table A4, in which we estimate effects using only houses near newly reporting large emitters, shows one practical reason why. We observe a large negative coefficient when we confine the sample to properties near newly reporting high emitters, but we only observe 213 sales in this specification which may be too few observations for statistical power.

A practical limitation of the data is that there are simply relatively few homes near large new reporters. Many of these facilities grabbed the headlines, but the bulk of the relevant households it reached were located near high emitting and prior reporting facilities, which were lumped in with the news coverage. Relatedly, the data show properties near high emitting facilities are less liquid more generally. Appendix Figure A6 shows that the flow of sales near these properties is both small and not significantly changing around the discontinuity date in the spring and throughout the year. Low emitters saw an uptick in sales around this time, but the sale quantities in Figure A6 are
not seasonally adjusted. As we mentioned in 6.2, this is further evidence of some seasonal component among the houses near control facilities that is not present among those near high emitters.

6.3.4 Identification Challenges and Caveats

While none of our robustness results override the broader evidence pointing to significant and robust capitalization of information from the 2000 TRI release in homes near the highest emitters, overall evidence of noisy effects in alternative specifications reflects limitations in the data, which we briefly summarize. First, newspaper archives have incomplete coverage of local news, making it impossible to get exact timing of local news shocks. If local newspaper coverage in news databases like Lexis-Nexis or Newsbank were comprehensive (i.e., transcribing all local newspaper coverage of TRI releases in all states), it would improve default discontinuity measurement of newspaper coverage timing. But even then, many people could get local news from television and radio. Transcribed coverage dates of other media outlets like local television news and radio news (and, ideally, their ratings), could generate a more precise estimate of information capitalization. Third, even with the ZTRAX data, there are limited properties directly adjacent to the largest TRI emitters. Additional sales (e.g., possibly from non-reporting states in the ZTRAX data like Texas), would help in separately estimating a “newly reporting” effect.

7 Discussion

7.1 Results in Context: The Financial Impact of Information Capitalization

To place our findings within the context of similar studies on environmental “bads”, it is useful to consider prior environmental hedonic estimates using housing values. Davis (2011) finds the construction of new natural gas power plants reduced home values within 2 miles of plants by 4.1-
7.1 percent.\textsuperscript{37} Chay and Greenstone (2005) find that Clean Air Act total suspended particulate reductions during the 1970s increased home values by 2-3.5 percent, and Bento et al. (2011) find similarly sized county-level results for the later 1990 Amendments. Gamper-Rabindran et al. (2011) find cleanup of Superfund sites raised highly localized housing values by up to 19 percent, though at a different level of aggregation Greenstone and Gallagher (2008) find cleanups to have no effect. More directly related to the dissemination of environmental information, Davis (2004) finds that the increased information on cancer clusters dropped home values by 14 percent, while Gayer et al. (2000) find increased information on Superfund hazardous waste risk shifted risk expectations downward but still led to a home price decrease of approximately 1 percent.\textsuperscript{38,39}

### 7.2 Relation to prior findings from TRI information releases and housing markets

Prior work by Bui and Mayer (2003) and Oberholzer-Gee and Mitsunari (2006) on TRI information and housing prices finds no consistent effects of newly reported TRI releases. There are several possible reasons our results could vary from earlier TRI work. First, there is a large difference in the size of the information shock regarding changes in reported emissions. Mean non-zero zip code level releases were around 200 tons in Bui and Mayer (2003). In 2000, however, D\textsuperscript{\textsuperscript{37}}

\begin{footnotesize}
\begin{enumerate}
\item Davis (2011) shows many new plants opened in 2000, but notes almost all new plants were natural gas plants, which are exempt from reporting to the TRI. For our results to be due to newly constructed power plants, new natural gas plants would have to have opened in the same areas as already existing impacted industries at the same time as the new TRI release. Davis (2011) also finds the effects of being close to a power plant fade within approximately 2 miles, meaning the probability of a treatment zip code in our analysis being close to a treatment area from that analysis is relatively small.
\item This is calculated using their reported price drop of $661 divided by the mean housing value of $74,176 (in 1996 dollars).
\item An example of the effect of information in non-environmental literature is Linden and Rockoff (2008) or Wentland, Waller, and Brastow (2014) who find that releasing information on sexual offenders in the neighborhood lowers home values by approximately 4 percent and 7 percent, respectively.
\end{enumerate}
\end{footnotesize}
reported aggregate zip-level emissions totals were almost twice that for zip codes where a new
treatment code was present. Second, our geographic variation is larger, covering millions of sales
across multiple states. Third, the world in which TRI data were released for the first time is
different from that in which TRI data are updated. The initial data release, for example, did not
have the advantage of the Internet, and households had to seek out hard copies of the TRI if they
wanted information. Data are now available online, news outlets have expanded both in number
and scope of coverage, and additional information is more readily available on the dangers of
environmental toxics. Communication was also more costly in the past, so dissemination of
information across households and neighborhoods is now higher.

A fourth consideration is that in the early years of TRI data reporting the housing market may
not have held solid priors on local emissions levels. It may have taken time before people knew
how to interpret toxic data. By the time of the 2000 data release, the TRI had been around for over
a decade. For those near newly reporting polluters, the updated data represented a substantial shift
from what they may have perceived as a previously accurate baseline. For those near prior
reporting large emitters, the data shock provided a greater context for the true nature of their
exposure to toxics. Both of these could lead to a larger response under the 2000 data update. We
should note, however, that our results do not directly answer whether the market response was full
(or under/over) capitalization of the net present value of the expected costs of pollution (akin to
the question asked in the property tax capitalization literature), as our research setting does not
measure a change in toxic pollution itself. Given the price drop in response to new information, it
may be reasonable to speculate that the risks may have been undercapitalized by the market near
high emitters prior to the 2000 data update, but without more detailed information about prior
expectations and expected risks, we leave further empirical examination of the question of full/under/over capitalization to future research.

8 Conclusion

We show how housing markets respond to increased information on local amenities, using changes in reporting policies as a shock to perception. The Toxics Release Inventory, a publicly released annual report of toxic emissions produced by many polluting sectors, served as a manner for households to assess local environmental amenities. Some of the largest emitters of toxics, including power plants, did not initially report emissions information in the TRI. A law change added such facilities to the TRI in 1998, substantially increased the amount of reported toxic emissions near some households, and provided greater context for emissions levels from existing reporting facilities. This large increase in publicly reported emissions information drew new attention to the largest polluters in the area and an increase in discussion of pollution sources in printed media.

We find the addition of the new data and the attention it created led to statistically and economically significant decreases in home sales prices that are non-linear in nature, showing up only for the largest sources of reported toxic emissions. Using data on archived local printed news sources, we show it was almost entirely these facilities that received news attention during the data release, with specific locations, noted by name, highlighted in news articles. Using the Zillow ZTRAX data, we first demonstrate how using spatially aggregated statistics to estimate home price effects mask substantial differences with even small distance variation. Using zip code level analysis, at higher emissions levels we see decreases in home sales price of approximately 5% in zip codes with large polluters. In a more spatially fine-tuned analysis using specific household...
traits and latitude and longitude for distance matching we find homes within 0.5 miles of the largest polluters saw price drops of from about 8 to 11%. Finally, we observe that this effect is highly localized, dropping to zero at larger distances beyond 0.5 mile from a TRI facility.

More broadly, our results speak to the role of information as a metric by which market forces can address the task of dealing with environmental externalities. Market mechanisms exist that, in theory, achieve socially efficient equilibria; but, they often require markets to be fully informed about the size of the externality. In the case of environmental toxics, our results show that household perceptions of the externality are likely imperfect, and credible information about pollution matters to (at least a subset of) the market. Pareto optimal free-market solutions are more likely when reliable, publicly provided information about the magnitude of these emissions exist, suggesting a role for government as a gatherer and provider of relevant information.
References


Environmental Protection Agency (1999): “33/50 Program: The Final Record,”.


## Tables and Figures

### Table 1
Summary Statistics for All Homes Sold in a 210 Day Window around the TRI Announcement

<table>
<thead>
<tr>
<th></th>
<th>TRI &gt; 100 (0 &lt; miles &lt; 0.5)</th>
<th>TRI &lt; 100 (0 &lt; miles &lt; 0.5)</th>
<th>Full Sample (0.5 &lt; miles &lt; 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Sales Price ($)</td>
<td>116,193</td>
<td>(95,896)</td>
<td>128,692</td>
</tr>
<tr>
<td>Residual</td>
<td>-0.02</td>
<td>(0.41)</td>
<td>0.00</td>
</tr>
<tr>
<td>Ranch (single story)</td>
<td>0.48</td>
<td>(0.50)</td>
<td>0.50</td>
</tr>
<tr>
<td>Year Built</td>
<td>1,954</td>
<td>(28)</td>
<td>1,963</td>
</tr>
<tr>
<td>MF Year Built</td>
<td>0.08</td>
<td>(0.27)</td>
<td>0.07</td>
</tr>
<tr>
<td>Garage</td>
<td>0.45</td>
<td>(0.50)</td>
<td>0.39</td>
</tr>
<tr>
<td>MF Garage</td>
<td>0.27</td>
<td>(0.44)</td>
<td>0.25</td>
</tr>
<tr>
<td>Pool</td>
<td>0.03</td>
<td>(0.16)</td>
<td>0.05</td>
</tr>
<tr>
<td>Acreage</td>
<td>0.21</td>
<td>(0.21)</td>
<td>0.25</td>
</tr>
<tr>
<td>Log(Acreage)</td>
<td>-1.70</td>
<td>(0.55)</td>
<td>-1.57</td>
</tr>
<tr>
<td>MF Acreage</td>
<td>0.14</td>
<td>(0.34)</td>
<td>0.12</td>
</tr>
<tr>
<td>Sqft</td>
<td>1,544</td>
<td>(680)</td>
<td>1,672</td>
</tr>
<tr>
<td>MF Sqft</td>
<td>0.06</td>
<td>(0.24)</td>
<td>0.05</td>
</tr>
<tr>
<td>Bedrooms</td>
<td>2.74</td>
<td>(1.05)</td>
<td>2.68</td>
</tr>
<tr>
<td>MF Bedrooms</td>
<td>0.13</td>
<td>(0.34)</td>
<td>0.17</td>
</tr>
<tr>
<td>Bathrooms</td>
<td>1.55</td>
<td>(0.71)</td>
<td>1.66</td>
</tr>
<tr>
<td>MF Bathrooms</td>
<td>0.11</td>
<td>(0.31)</td>
<td>0.12</td>
</tr>
</tbody>
</table>

| N                    | 2,775 | 15,902 | 2,217,758 |

Notes: Summary statistics for home information used in main analysis in Tables 3 and 4. Results are stratified by TRI emissions amounts (in tons) for homes within 0.5 miles of the facility and for the full sample for those between 0.5 and 5 miles. Data derived from Zillow ZTRAX.
### Table 2
TRI Facilities and Home Prices – Difference-in-Differences Results

<table>
<thead>
<tr>
<th>ZTRAX Aggregated Zip Code x Month</th>
<th>TRAX Sales log(price)</th>
<th>log(age(price))</th>
<th>log(med(price))</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post x (TRI &gt; 100)</td>
<td>-0.032***</td>
<td>-0.034***</td>
<td>-0.049***</td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Post x (30 &lt; TRI &lt; 100)</td>
<td>-0.018**</td>
<td>-0.019**</td>
<td>-0.012</td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Post x (10 &lt; TRI &lt; 30)</td>
<td>0.000</td>
<td>0.003</td>
<td>0.006</td>
<td>(0.007)</td>
<td>(0.008)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Post x (0 &lt; TRI &lt; 10)</td>
<td>0.004</td>
<td>0.004</td>
<td>0.003</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Post x (TRI = 0)</td>
<td>reference</td>
<td>reference</td>
<td>reference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>223,120</td>
<td>223,120</td>
<td>5,383,770</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: The table reports interaction coefficients from the Difference in Differences model for varying levels of TRI disclosure relative to 0 emissions. TRI levels (i.e., facility max) are determined at the zip code level. Post is equal to 1 if the sale month is in the month of the state’s first news date or thereafter. Includes sales from 1998 to 2002. Columns 1 and 2 include zip code by month average or median price data created using Zillow ZTRAX. Column 3 includes individual house sale data from ZTRAX and includes hedonic controls listed in the text. Standard errors are clustered on zip code.
### Table 3

#### Panel A: TRI Facilities and Home Prices – Difference in Regression Discontinuity Results

<table>
<thead>
<tr>
<th></th>
<th>log(price)</th>
<th>log(price)</th>
<th>log(price)</th>
<th>residual</th>
<th>residual</th>
<th>residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>D * (High Emitter)</td>
<td>-0.172***</td>
<td>-0.139***</td>
<td>-0.114***</td>
<td>-0.105***</td>
<td>-0.076**</td>
<td>-0.100***</td>
</tr>
<tr>
<td></td>
<td>(0.064)</td>
<td>(0.052)</td>
<td>(0.041)</td>
<td>(0.040)</td>
<td>(0.037)</td>
<td>(0.038)</td>
</tr>
<tr>
<td>D</td>
<td>0.055**</td>
<td>0.049**</td>
<td>0.052***</td>
<td>0.029**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.020)</td>
<td>(0.015)</td>
<td>(0.014)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIPS Pre-Trends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Only High Emitters & rdrobust

**Observations**

|                | 18,677 | 18,677 | 18,645 | 18,677 | 2,775 | 2,885 |

#### Panel B: Difference in Regression Discontinuity Results for Prior Year (1999)

<table>
<thead>
<tr>
<th></th>
<th>log(price)</th>
<th>log(price)</th>
<th>log(price)</th>
<th>residual</th>
<th>residual</th>
<th>residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>D * (High Emitter)</td>
<td>-0.047</td>
<td>-0.045</td>
<td>0.017</td>
<td>0.024</td>
<td>0.037</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.051)</td>
<td>(0.033)</td>
<td>(0.031)</td>
<td>(0.028)</td>
<td>(0.034)</td>
</tr>
<tr>
<td>D</td>
<td>0.040</td>
<td>0.056**</td>
<td>0.038**</td>
<td>0.013</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.023)</td>
<td>(0.016)</td>
<td>(0.015)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIPS Pre-Trends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Only High Emitters & rdrobust

**Observations**

|                | 17,656 | 17,656 | 17,619 | 17,656 | 2,666 | 2,650 |

**Notes:**
- Panel A of this table reports discontinuity coefficients from the Difference in Regression Discontinuity in Time model for TRI disclosures greater than 100 (relative to less than 100). All homes are located within half a mile of the TRI and are sold within a +/- 210 day window around the announcement. Columns 1-3 use logged price as the outcome, while columns 4-6 use the residualized logged price from a hedonic regression including the controls, calendar month fixed effects, zip code fixed effects, and zip code level linear trends. Column 6 uses the Stata rdrobust command and finds that the optimal bandwidth using the default specification is 216 days on both sides of the cutoff. Panel B estimates the same models as Panel A, except using a placebo treatment time by moving the cutoff to the prior year, 1999. Standard errors are clustered at the TRI facility level. Standard errors are clustered at the TRI facility level.
Figure 1

Total Toxic Releases Reported for Newly Added vs. Original Reporting Industries

Notes: Toxics are the sum of all land and air releases, in thousands of tons, across all toxics recorded as reported in the Toxics Release Inventory. “Newly Added” indicates firms classified under SIC codes 10, 12, 4911, 4931, 4939, 4953, 5169, 5171, and 7389 (see Section 2). “Original” includes all other industries, which had reported emissions in prior TRI releases. Data are from all available TRI locations and are not restricted to the zip codes used in the DiD analysis.
Figure 2
Regression in Discontinuity

Notes: This figure is the Regression Discontinuity in Time result for property prices near high emitting facilities (TRI > 100) in the top panel and low emitting facilities in the bottom panel (TRI < 100) for homes sold within half a mile of a TRI. This figure depicts binned residuals (by week) from the default hedonic regression specification.
Figure 3
Bandwidth Robustness –
High Emitter Effect on Home Prices by Different Bandwidths & Trends

Panel A: Linear Spline

Panel B: Quadratic Spline

Notes: These figures depict the discontinuities and 95% confidence intervals from the Difference in Regression Discontinuity in Time effect size for the model in Table 3, column 4 varying the bandwidth from 60 days on both sides of the cutoff to 360. Panel A uses a linear spline trend, while Panel B uses a quadratic spline.
Figure 4
Regression in Discontinuity for TRI > 100 by Distance Donut

Notes: Figure reports the discontinuity coefficient and 95% confidence intervals for the homes sold near a TRI > 100 from a model similar to Table 3, Column 4, but stratified by distance from the TRI. Panel A includes mutually exclusive distance donuts starting at 0 to 0.5 miles and increasing by 0.5 each iteration, up to 4.5 to 5 miles. Panel B repeats this exercise, but iteratively adds the homes within the next 0.5 miles to the sample.