

Introduction to the Special Issue

Property Value Analysis Using ZTRAX: Applications under the Approaching Sunset

Daniel J. Phaneuf *Henry C. Taylor Professor, Department of Agricultural and Applied Economics, University of Wisconsin–Madison; dphaneuf@wisc.edu*

Jeffrey Zabel *Professor, Department of Economics, Tufts University, Medford, Massachusetts; jeff.zabel@tufts.edu*

Andy Krause *Senior Manager, Applied Science, Zillow, Seattle, Washington; andy@andykrause.com*

Introduction (Daniel J. Phaneuf and Jeffrey Zabel)

We are happy to introduce this joint special issue of *Land Economics* (*LE*) and the *Journal of Housing Economics* (*JHE*), focused on property value applications using Zillow’s ZTRAX data. ZTRAX is a real estate–focused database with over 400 million public records spanning nearly all U.S. counties across more than 30 years. The spatially explicit data on deed transfers, sale prices, and property characteristics have provided the basis for analyses on themes such as disaster risks, including wildfires, flooding, and chemical accidents; natural resources, including water quality, farmland, and coal; and land uses, such as open space, national parks, and critical habitat designation, among others.

Beginning in 2017, Zillow made its ZTRAX database available to qualified researchers free of charge. This facilitated an impressive wave of research in environmental, housing, regional, public, urban, and related fields in economics. An especially valuable feature of access to the full database is that it enabled researchers to work at broad temporal and spatial scales (state, regional, and even nationwide analyses). It also helped investigators with small research budgets access

information to support research agendas that otherwise would not have existed. The influence has been enormous. Over the past seven years, nearly 300 academic, government, and nonprofit institutions signed agreements with Zillow to use its data. As of September 2023, a simple Google Scholar search of ZTRAX lists more than 475 scholarly articles that are the result of these agreements.

In 2021, Zillow announced that it would discontinue the ZTRAX access program. This was due in part to the program’s success—Zillow could no longer adequately service its large and growing list of research partners. The sunset for the program was September 2023, at which point all raw data needed to be deleted. This deadline provided the motivation for this project. Our sense was that coordinated special journal issues with evaluation timelines that recognized the approaching sunset would be helpful for researchers working on finishing up projects. We also thought that a collection of articles spanning several topics across two journals would highlight the value of access to a single and comprehensive database for the research community and help spur efforts to find substitutes for maintaining the productive research agendas showcased here.

Our role in the project was to serve as handling editors for the contributed articles at *LE* (Phaneuf) and *JHE* (Zabel). Kevin Boyle and Christoph Nolte rounded out the organizing committee, and both deserve recognition for the key roles they played. Andy Krause from Zillow was an enthusiastic supporter, and he shares a few remarks in the next section.

Land Economics • February 2024 • 100 (1): 1–5
DOI:10.3368/le.100.1.intro
ISSN 0023-7639; E-ISSN 1543-8325
© 2024 by the Board of Regents of the University of Wisconsin System

Our process began in May 2022 with nearly 50 expressions of interest in response to our call. To triage this enthusiastic response, we encouraged participation from authors who were pursuing large-scale applications (state, regional, and national) that would highlight the unique advantages (and challenges) of working with the ZTRAX data. Early versions of the articles were presented during an online workshop in August 2022, and manuscripts were submitted for review in October. We largely tracked articles that had an environmental theme to *LE* and those with urban and regional themes to *JHE*, although this division is not strict, and we encourage readers to examine both collections. Manuscripts were subject to standard peer review. The 10 articles in *LE* and seven in *JHE*, accepted between May and September 2023, are the result of this process.

We believe these articles do an outstanding job of illustrating the wide range of research questions that can be addressed with large-scale real estate databases like ZTRAX. Their value is more than the sum of their parts. In working from a common database, the author teams needed to complete similar data-processing steps. For several, including selecting arm's-length transactions, spatially relating local land uses and amenities, and addressing missing property characteristics, we asked authors to include in their appendixes a narrative of how these decisions were made. In addition, authors completed a column in a Google Sheet that displays a matrix of how the data-processing decisions were completed for each article. A link to this document can be found as an [Appendix](#) to this Introduction. We believe that this accumulated experience will be a valuable resource for future researchers working with large-scale property value databases.

We are grateful to the many people who helped create these special issues. This includes the many anonymous reviewers, the authors who responded quickly to revision requests, and the production staff at both journals. We are confident that the collected articles will further our understanding of large-scale property value analysis and illustrate the knowledge payoff from a coordinated and broadly accessible database.

Comments from Zillow (Andy Krause)

At Zillow, we know firsthand the critical role that data play in running an information-centered business. High-quality data, in addition to fueling the knowledge economy, are a crucial input to modern governance, scientific research, and higher education. Much like key commodities of prior ages—iron, wheat, timber, oil—in its raw form, these data are often unfit for direct consumption. They must be refined before they can be used. Scarcely is there anywhere this is more true than in the realm of the built environment and property markets.

Property market data—the information about land and the built assets on it as well as the market that trades in it—is observational not experimental. It varies greatly over time and space in measures of data quantity and quality. It is often poorly standardized, if at all. And, importantly, most data about property and its transaction are collected and recorded somewhere, somehow by a human hand. Collectively, this means that, although a national dataset like ZTRAX is “big” in the sense that it can require novel methods of storage and computation, these data also have all the nuances and problems often associated with smaller, human-scale data such as those found in epidemiological, paleontological, and policy evaluation disciplines.

As a result, to extract the full benefit from property data both of the following are required: extensive cleaning and preparation and deep contextual knowledge. Preparing and understanding data is what allows practitioners and researchers to refine it, to turn raw data into useful information. That information can then be analyzed and modeled to generate knowledge on which better decisions—policy, personal, and commercial—can be based.

The collection of work in this special issue highlights the importance of human actions—the preparation and application of context—to generating knowledge and policy from property data. It is a pleasure to see ZTRAX data support important work quantifying the market impact of a wide variety of factors, ranging from environmental features

and resource policy related to flood events and industry emissions to land policy and taxation to flood events and major locational demand drivers like the rise of remote work. The property market is highly complex; this work helps greatly in unpacking some of those multifaceted relationships driving prices.

Zillow was founded on the idea that consumers want and should freely have more information about the housing market. To expand market transparency, we have had to ideate, innovate, and sometimes challenge entrenched positions. Our efforts in this direction have relied heavily on research coming from academic institutions, research publications, and open-source software.

The research highlighted in these issues remains true to this goal. As such, we are thrilled to have been able to play a role—limited as it was—in helping bring even more enlightening information to consumers, policy makers, industry professionals, and other researchers. Finally, we look forward to evaluating the potential of the information developed and the methods used in this work to augment our own suite of valuation models and housing metrics.

Land Economics (Daniel J. Phaneuf)

The articles in this issue of *LE* focus on a range of avenues through which home prices interact with place-based environmental resources. This begins with the first two articles, which both look at the property value impacts of wildfires. Ma and colleagues are interested in how home prices respond to ex ante wildfire risk. Their challenge in doing so is twofold: wildfire risks are correlated with woodland amenities, and homebuyers may not be informed about risks they face. The solution is to exploit a program in California with administrative rules that generates risk disclosure and nondisclosure zones in proximity. The authors use the disclosure rules as an information shock and the neighboring disclosure/nondisclosure zones to implement a boundary discontinuity design (in which woodland amenities vary continuously across the boundary) that enables estimation of the

wildfire risk discount in disclosure zones relative to nondisclosure zones.

Huang and Skidmore are interested in how home prices respond to ex post wildfire events. They study the direct property value effects of nearby past wildfire events and the indirect effect operating through air pollution. The scale of the study is ambitious: the authors assemble data on all wildfire occurrences in the continental United States between 1992 and 2018 and link these spatially and temporally to home sales between 2010 and 2018. Using a repeat sales model, the authors report a price discount for nearby upwind fire events (the direct effect) and a second discount related to particulate matter impacts of wildfires (the indirect effect). The latter is identified using an IV strategy in which distant upwind wildfires are an instrument for local particulate matter concentration.

The next two articles are related through their emphasis on exposure to toxic pollution as a mechanism affecting home prices. Fraenkel, Graff Zivin, and Krumholz study the health and property value impacts of the phase-out of coal-fired electricity generation in the United States. As natural gas generation has displaced coal generation in the past two decades, more than 30% of coal fired plants have been fully or partially closed. The authors first examine the direct health effects of these closures and find that counties with population centers within 30 miles of plants that close at least one unit experience a decrease in mortality. There is a distance gradient for this: cardiovascular mortality decreases 2.77% for areas 15 miles or closer to a closure and 1.47% for areas out to 30 miles distant. They then test if these health benefits capitalize into home prices. They find incomplete capitalization in that prices are impacted by full closures for homes less than 15 miles distant but not by partial or more distant closures. This almost certainly represents an information effect—fatalities can decrease without people knowing the cause, while home prices only respond to what is known.

Moulton, Sanders, and Wentland consider a context exactly opposite to Fraenkel, Graff Zivin, and Krumholz—an information shock in the absence of any change in environmental outcomes. The authors study the year 2000

changes in toxic release reporting requirements that vastly expanded the industries that reported. With this change, newly reporting facilities contributed to toxic release inventories that were substantially larger than previous reporting periods, even though actual releases were largely unchanged. The authors use this to study the property value impacts of a pure information treatment at the national scale. Their findings confirm the existence of highly localized price discounts following the information release for homes near the largest polluters. This is consistent with the finding in Ma et al. in a very different context, and together the articles show that reducing uncertainties at a large spatial scale can help markets better price environmental risks.

The next three articles focus on water but from very different perspectives. Swedberg and colleagues consider inland lake water quality, Chen and Towe examine coastal amenities and flood hazard, and Chaudhry, Fairbanks, and Nolte look at water markets for agriculture. The starting point for Swedberg et al. is recent studies (e.g., Guignet et al. 2022; Mamun et al. 2023) reporting national average willingness-to-pay estimates for inland lake water quality. These national averages likely mask considerable heterogeneity based on variation in ecological and local market conditions. The authors systematically quantify this heterogeneity by running models at five different spatial scales; they also vary some methodological dimensions to compare how spatial scale and other investigator decisions generate variation in findings. The confirmation of substantial heterogeneity raises important questions on how to correctly match the spatial scale of analysis to the intended use of the findings.

Chen and Towe address the challenge of disentangling the price effects of correlated amenities and flood risk in coastal areas. This is like the challenges faced by Ma et al. and Huang and Skidmore, who also sought to identify a risk impact that covaries with an amenity. For Chen and Towe, the solution is data driven. In their application to coastal New England, the authors use GIS and remote sensing tools to assign a unique ocean viewshed measure to almost 400,000 properties in Connecticut and Massachusetts. To account

for amenity impacts they include the viewshed measure and a waterfront indicator as controls and then estimate the price gradient for a home lying in a special flood hazard area. The estimates illustrate the important role that accounting for amenities plays in estimating flood risk price discounts, and the study itself provides an example of a scalable method for computing property-specific coastal amenity scores.

Chaudhry, Fairbanks, and Nolte study farm prices in California to estimate how participation in water markets capitalizes into land prices. They focus on sales in the Sacramento River valley—a comparatively water-abundant region (in terms of snowmelt and groundwater) that hosts irrigated agriculture and exports water to the rest of California during drought years. Division of the region into water districts determines access to surface water and the ability to benefit from water transfers. The authors examine heterogeneity in how in water transfers affect land prices, with an emphasis on distributional aspects. They find that farms in water districts with surface water rights and ready access to groundwater as a substitute for surface water appreciate following transfers, whereas farms outside of districts that are constrained in their use of groundwater may suffer lower prices. The results demonstrate that efficiency gains from water markets will be unequally distributed, with historic water right allocations and groundwater use governance interacting to determine winners and losers from transfers.

The next two articles relate to surrounding land use impacts. Mamun, Nelson, and Nolte study a feature of the Endangered Species Act that requires designation of some private lands as critical habitat for listed species. The authors measure the price effect of designation on developed and undeveloped parcels, with the goal of teasing out potentially counteracting impacts through improved amenities and reduced development options. Using a boundary design that compares nearby parcels inside and outside of critical habitat zones, they report mixed findings, with the sign and significance of estimates varying with the spatial scale of analysis. In general, null estimates are generated from national models, while geographically focused models

for specific species generate economically significant price effects. These mixed findings complement Swedberg et al. by highlighting how important heterogeneity can be masked by aggregate, national level analyses.

Zabel, Nolte, and Paterson consider the price effects of proximity to national park units. The motivation is the National Park Service's need to calculate the economic benefits from their units, and the authors examine how local amenity values contribute to these impacts. A unique feature of the study is its consideration of many separate units, which range from iconic national parks to urban national historical sites. Five of the units were formed during the period of data availability, so the authors have observations that bracket the event, allowing them to estimate how the proximity premium adjusted following designation. For the other 13 units, they estimate standard proximity premia that are heterogeneous across unit types and locations. A key takeaway from their research is that local context matters in terms of both interpreting effects and understanding threats to identification.

The final article in the collection is different from the others. Rather than a specific

application, Nolte and colleagues present a systematic overview of best practices for using nationwide property value data in the context of valuing environmental amenities and hazards. The insights in this article were codeveloped with several of the applications presented in the *LE* and *JHE* special issues and tie together the common data processing themes the different author teams faced. More important, it provides a template for future authors using large-scale property value databases to evaluate their decisions and compare these with the best practice advice provided by the authors.

References

- Guignet, D., M. T. Heberling, M. Papenfus, and O. Griot. 2022. "Property Values, Water Quality, and Benefit Transfer: A Nationwide Meta-analysis." *Land Economics* 98 (2): 191–218. <https://doi.org/10.3368/le.98.2.050120-0062R1>.
- Mamun, S., A. Castillo-Castillo, K. Swedberg, J. Zhang, K. J. Boyle, D. Cardoso, C. L. Kling, et al. 2023. "Valuing Water Quality in the United States Using a National Dataset on Property Values." *Proceedings of the National Academy of Sciences* 120 (15): e2210417120. <https://doi.org/10.1073/pnas.2210417120>.