

Commentary: How Data Architects Are Crafting Equitable Housing Policy Research

Matthew Murphy
NYU Furman Center

Nationwide, tools are surfacing to bridge knowledge gaps and deepen our comprehension of the effects of historical and recently changed housing policies. However, housing researchers, along with state and municipal governments, face a key and somewhat insurmountable challenge: insufficient data infrastructure to study housing policy comprehensively. To fill the gap, researchers are developing new methods to link together fragmented datasets to tell a more comprehensive story. Where data exist, researchers are refining the use of emerging technologies to avoid past pitfalls, such as racial biases. A focus has also been placed on pivotal housing justice concepts, like understanding the extent of the shift toward corporate ownership of America's rental properties, beyond public records.

I am excited about these kinds of advancements. The three articles in this series show the hoops researchers often go through to tell a comprehensive story about important housing policy considerations. The urgency to use data and new methods to adapt and meet the needs of residents and policymakers is evident in all housing-related work. Researchers require intricate data not just on cities and neighborhoods but on individual properties, their units, and their building systems. This kind of data has myriad applications. With escalating housing shortages, it is essential to grasp the full scope of available and potential housing inventory and what it might cost the public, or even the required investment from the market, to make new housing available. As natural disasters increase in frequency, understanding where people can seek refuge during and after these events becomes paramount. In a time of surging insurance premiums, robust asset management data will be needed to challenge broad assumptions made by insurers about low-income housing in particular. As we approach a preservation era, especially concerning low-income housing tax credit projects from the 1990s, in-depth property data will be crucial to ensuring the preservation and continuation of quality affordable housing. If local governments opt for social housing development, they must view these properties not just as monitoring points but as dynamic assets responding to market shifts. Finally, if local governments choose to use either a carrot or stick approach to force property owners' buildings to reduce carbon emissions, they will need detailed information about the systems in place tomorrow and what it will take to transform and recapitalize them for the next generation.

These particular issues influence me because I have seen them up close, having spent 8 years working in New York City government on affordable housing. I have sat in frustration, looking for data that should be much easier to find than they actually were. However, in my work as the executive director of the NYU Furman Center, I have come to realize that local authorities and advocates should perceive these challenges as opportunities.

The foundation for affordable housing policy discussions must be an ability to rely on high-quality and timely data. To fill this gap, researchers scrutinize whether they can rely on emerging technology, specifically regarding the concerns over racial bias. Linna Zhu, Michael Neal, and Caitlin Young's article explores automated valuation models (AVMs) that use advanced machine learning techniques. Their description of the possibilities of emerging technology provides a compelling case study on the potential technology has to address longstanding societal challenges. The study's findings reveal a higher percentage magnitude of AVM error in majority African-American neighborhoods, even after accounting for property condition and other variables, underscoring a troubling yet vital truth: although the allure of algorithms promises objectivity, they can inadvertently perpetuate biases present in historical data. It is especially concerning given the study's evidence that properties in poorer conditions or those in neighborhoods with a greater share of distressed sales are more susceptible to AVM inaccuracies. The use of the Light Gradient Boosting Machine, or LightGBM model, which demonstrated a 5.8-percentage-point improvement in model fit, showcases the potential of artificial intelligence in navigating the complexities and human biases assumed to be inherent in property valuation. However, the persistence of racial disparities in valuation, even with such advancements, is a stark reminder of the deeply entrenched effects of historical racism on property values and conditions. As we advance into the next generation of property valuation, a subjective field to begin with, it becomes crucial not just to adopt these sophisticated tools but to do so with an unwavering commitment to understanding and addressing the systemic racial biases they might reflect. At the same time, how do we come to rely on interior or even unit-level conditions at a more reliable scale? This article relies on a correlation between exterior conditions and interior ones. In practice, researchers or practitioners are extremely far away from having that knowledge available at scale. However, researchers should all strive to have the best possible data on housing conditions in their toolkits.

In a work that is near and dear to my heart, "Local Landscapes of Assisted Housing: Reconciling Layered and Imprecise Administrative Data for Research Purposes," researchers Shiloh Deitz, Will Payne, Eric Seymour, Kathe Newman, and Lauren Nolan of Rutgers University untangle the intricate data behind affordable housing programs. It is a funny but fitting name for an article—it reads as nearly satirical. However, the work is critically important for any state in the country. Their methodology involved meticulously cross-referencing disparate subsidy datasets, aligning them with property-level outcomes—a task of paramount importance in the realm of affordable housing, given the intricacies of layered subsidies. When I was a graduate student at New York University working at the Furman Center, I worked on the creation of a database of nearly all of New York City's affordable housing projects. Researchers still use the database today to inform preservation decisions and show the public which programs created what degree of affordable housing and where. I empathize with the researchers who took on this task, but I am so happy for their contribution.

A key revelation from their study was the identification of an overcounting of approximately 17,000 housing units. On paper, maybe 17,000 units does not sound like a lot. For example, about 3.5 million units are in New York City. However, picture a world in which a hurricane has displaced 5,000 households. A governor asks where people might be able to live—what is available tonight? You need to be a lot closer than off by 17,000. Rather, picture creating a 10-year strategy and needing to budget in the appropriate amount of money to rehabilitate all publicly subsidized units for 10 years. At \$100,000 per unit, an overcount of 17,000 units is a \$1.7 billion mistake. How can you get your arms around an issue when total population data are shaky to begin with? The research not only underscores the inherent challenges in handling multifaceted datasets but also indicates the potential discrepancies that can emerge without vigilant oversight. Something that stood out to me was that, despite leveraging sophisticated data integration techniques, it was indispensable for the research team to engage in manual data verification to ensure the accuracy of their findings. This confluence of advanced analytics and manual validation underscores the team's commitment to precision and integrity in research.

I do not think any housing planners can get away with a worthwhile project that does not involve some amount of manual data collection. Studying housing policies and programs is complicated, and some of the most critical data are often lacking. However, understanding housing programs and policies in detail *should* be complicated; it requires going to a physical place, whether it be an actual housing development or a file room at an under-resourced government agency. Studying housing requires an examination of the intersection of how people live (or want to live) with a huge number of considerations housing policy researchers account for, such as the state of physical structures, complicated financing structures and government subsidy programs at both the tenant and development level, land use regulations, property tax policies, patterns of neighborhood change and development, climate risk, and so much more. Skipping any of these considerations risks missing out on another important story that might be going on. It is the fun and frustration of being a housing policy researcher.

However, part of the frustration also has to do with why researchers must go through these hoops to begin with. Although the researchers' dedication here is laudable, it should prompt reflection on the broader institutional responsibilities. Should such intricate data reconciliation be the purview of academic researchers alone? It is imperative for government agencies to assume a more proactive role in ensuring the accuracy and comprehensiveness of housing data. This research underscores the pressing need for housing data linkage to be at the forefront of best management practices rather than a peripheral concern. However, I am confident that this research represents a starting place that governments can use and move forward. Having someone from the outside do it can be refreshing and invigorating—it is hard to pull off these projects within government, and often, government agencies are so under-resourced and lack needed technology, much less the data, so a project like this can help create the pressure necessary to shepherd the resources government employees require to up their game.

Speaking of hoops researchers must go through, even when publicly available information is quite good, the last article in the series, "Who Owns Our Homes? Methods to Group and Unmask Anonymous Corporate Owners" by Renz Torres, reads like an FBI forensic investigation into what

it takes to understand who owns property, and to what degree. Torres uses graph-based analysis to study single-family rental ownership in Jacksonville, a methodological innovation. I have learned that a graph-based approach provides a more holistic view of the interconnections between property owners, addresses, and business officers. By transforming tax parcel and business registry data into a network of relationships, the study manages to uncover nuanced patterns of ownership and investment strategies. For instance, the clustering of single-family nonowner-occupied housing (NOOH) properties in Jacksonville's urban core versus the pronounced activity of mega investors in the suburbs provides a detailed landscape of market dynamics. In an interesting spin, Torres also uses corporate disclosure forms to examine the accuracy of the methodology. I love the use of public disclosure data in the absence of clear transparency.

The results, such as the differential investment behaviors in areas populated by people of color, are noteworthy; the novelty lies in the methodology. However, it is essential to view this approach with a degree of caution. The methods used to overcome the inherent challenges of such an analysis, such as over-clustering due to data inconsistencies, again remind us of the complexities of real-world data. Furthermore, although Florida's data standardization aids this approach, its scalability in less data-rich environments would have to be further understood. Nevertheless, the article marks a promising stride in housing research, hinting at the potential of graph-based methodologies to revolutionize our understanding of property markets and inform policy decisions. The question of what to do to advance homeownership remains, especially for first-time homebuyers, and what the positive roles that single-family rentals, even corporate-owned ones, might be, especially in an America that is becoming more housing-starved. However, Torres has shown us that with commitment, fancy data work, the right questions, and some manual labor, researchers can link together information to tell important stories.

I learned a lot reading these three articles, and I hope any data-driven housing practitioner does, too. The articles mix methodological innovations with common-sense applications. As a researcher and former government policymaker, I appreciate the use cases, and I am excited that the markets examined are not New York City or San Francisco. I look forward to how this world evolves and the role researchers play as technology grows exponentially during the next few years. The tools that were on the horizon 5 years ago are now here—whether and how they can be used to transform housing policy for positive effects remains to be seen. Current research workarounds, although essential, only partially address the core data challenges impacting our understanding and improvement of housing policy. A more impactful approach would involve directing federal funds to enhance technological and data collection capabilities of local and state governments. This enhancement should focus on comprehensive monitoring of the housing market, including cost fluctuations, natural disaster risks, and the condition of rental properties over time. Simultaneously, private entities should collaborate to provide accessible, standardized rent data, aiding in the accurate assessment of housing costs and the effectiveness of various policies. In addition, government-sponsored enterprises and state housing finance agencies could significantly contribute by making more of their data publicly available, contingent upon receiving necessary resources. Amidst these potential advancements, the research community continues to innovate. The authors of these articles offered a hint of the possibilities, and it is exciting to think about where we can all push next.

Acknowledgments

The author extends heartfelt gratitude to all colleagues at the NYU Furman Center, both past and present. Special acknowledgment is given to Rich Froehlich, a titan in the field of affordable housing, who dedicated his life to improving New York City. Rich's passion for using data, analysis, and technology to ensure affordable housing for New Yorkers was unparalleled. His commitment to educating a new generation of affordable housing practitioners in the intricacies of housing policy and finance, including the author, has left an indelible impact on the community.

Author

Matthew Murphy is the executive director of the NYU Furman Center.