Affordability of Owner-Occupied Housing across US Cities

By REBECCA DIAMOND AND WILLIAM DIAMOND*

Housing is the single largest component of consumer spending and has enormous price dispersion across space. As a result, an accurate measure of spatial differences in housing costs is crucial to accurately measure differences in the affordability of places to live.

Measuring housing affordability across space requires comparing the prices of a single year of housing services. For the rented housing market, this is easily done by comparing rents for simple units in different cities. This is much more challenging for the owner-occupied housing stock, which makes up about two-thirds of the occupied housing stock. The challenge is that we do not directly see the rental value of owner-occupied housing.[1]

This paper applies a new method we developed in Diamond and Diamond (2024) (DD) to impute the rental value of owner-occupied housing across 15 CBSAs. This method uses a panel of housing units that switch between the owner-occupied and rental markets to more accurately estimate owner-occupied properties' rents. This allows us to include a property's value as a covariate to predict rents in addition to the standard hedonic controls. By including property value as a covariate, we can proxy for unobserved housing and neighborhood quality of the housing unit.

Our method predicts a 3.2 percent higher average rental value of owner-occupied housing than standard hedonic methods. This masks substantial spatial heterogeneity. We find hedonic methods overestimate the rental value of owner occupied housing in the most expensive and land-use regulated CBSAs and underestimate its rental value in the least expensive and regulated CBSAs. Our results imply a smaller difference in housing affordability between high and low cost CBSAs than previous work. For example, standard hedonic methods predict that owner-occupied housing is 225 percent more expensive in San Francisco CBSA than Detroit. The DD method implies that this difference gap is only 200 percent.

We provide suggestive evidence that our results are due to spatial misallocation of physical housing quality caused by land-use regulation. In the most regulated CBSAs, it is much harder to develop high-quality housing in the most desirable locations. As a result, the correlation between location desirability and housing quality is lower (or even negative) than in more regulated CBSAs.

In the most regulated CBSAs, hedonic methods should therefore underestimate the cost of the most desirable locations, where the housing stock is low quality. This can explain why our method reduces the cost estimates of the most regulated cities and increases it it in the least regulated cities.

I. Data

We use the public-use American Housing Survey (AHS) from 2015-2021. The AHS is a panel of properties that are surveyed every two years. The panel was completely resampled in 2015, so we focus on these later years with a consistent panel structure. Our analysis will focus on comparing the rental value of owner-occupied properties across US CBSAs. The public-use AHS only identifies the 15 largest CBSAs in the public-use files, so we restrict our analysis to these 15 CBSAs. Summary stats of key variables are listed in Appendix Table 1.

II. Measuring Owner-Occupied Housing Costs

As discussed in Poterba (1984), the purchase price of housing is not an accurate measure of the annual flow cost of housing consumption. Since housing is an asset, its purchase price represent the present discounted value of future expected rents. The ideal measure of current hous-

* R Diamond: Stanford GSB and NBER, diamondr@stanford.edu. W Diamond: Wharton, diamondw@wharton.upenn.edu. We thank Marie Kaul for excellent research assistance.

[1] In the construction of the Consumer Price Index for calculating inflation, imputing the rental value of owner-occupied housing is both the largest and most model-based step.

ing costs is the "spot market" price of a year of housing services, namely the rental cost. For properties in the rental market, this is directly observed in data. For owner-occupied properties, the annual rental costs must be estimated.

A standard approach to estimate the rental value of an owner-occupied housing unit is to use the rent of a "matched" rental property with similar observable characteristics (Davis, Lehnert and Martin, 2008; Demers and Eisfeldt, 2022). This is done with a hedonic regression:

$$r_{ijt} = \delta_t + \delta_j + \beta X_{it} + \varepsilon_{ijt}, \tag{1}$$

where i indexes properties, j CBSAs, and t years. r_{ijt} is the annual log rent of property i, X_{it} is a vector of property characteristics: property age, property age squared, dummies for single family homes and multi-family buildings (excluded category is mobile home), and the number of rooms in the unit.

We can estimate equation 1 by pooling our rental property data across CBSAs and years. To measure the difference in log rent across cities for the owner-occupied market, we predict rents for each CBSA using equation (1), with X_{it} equal to the nationwide owner-occupied average property characteristics. This method is used by the BEA to measure Regional Price Parities and by Diamond and Moretti (2021) and Moretti (2013) to measure local housing cost differences.[2]

One shortcoming of this method is that it ignores unobservable characteristics. If owner-occupied housing is higher quality than rental housing for reasons not captured by our observables, the hedonic method would underestimate the rental value of the owner-occupied housing stock. To account for unobservable quality differences between the rented and owner-occupied housing stocks, we apply new methods we developed in Diamond and Diamond (2024) (DD).

DD show that housing units that switch between the owner and rental markets over time, for which we observe both a rent and a sales price, can be used to account for the impact of unobserved housing quality. We refer to housing units that switch between the owner and rental market in consecutive survey waves of the AHS as switchers. DD show that their method accurately predicts the rents of switchers across the price distribution. They also show that hedonic methods overestimate the rent of low-price switchers (which have low unobserved quality) and underestimate the rent of high priced switchers (which have high unobserved quality).

To give some intuition about how switchers can improve rental cost estimation of owned units, we compare switchers' reported property values and rents to the overall property value and rental distributions. Housing units that switch between the rental and owner markets rent for 17% more than the average rental unit, but are valued 30% less than the average owned property. This shows that is significant segmentation between the owner-occupied and rental markets. If this segmentation is not fully captured by the observable differences between the owner-occupied and rented properties, hedonic rent estimation using equation (1) could be biased.

We briefly describe the DD method here; see DD for details. Since housing values and rents for switchers are measured two years apart, we transform rents and value dollar amounts to rent and value *ranks* within the nationwide AHS distribution for the year in which these prices are observed. Define $S_{R_{ijt}}$ and $S_{V_{ijt}}$ as housing unit $i's$ rank in the nationwide rental and value distribution in year t.[3] For any given unit in a given year, we either observed $S_{R_{ijt}}$ or $S_{V_{ijt}}$. We estimate the following quantile regression at quantile q using housing units that have switched between the rental and owner markets:

$$S^q_{R_{ijt'}} = \delta^q_{t'} + \delta^q_j + \beta^q X_{it'} + \gamma^q_1 Q_{V_{ijt^*}} + \gamma^q_2 Q^2_{V_{ijt^*}} + \varepsilon^q_{ijt'}, \tag{2}$$

where t' is the year a switcher house is observed in the rental market and year t^* is the year of the consecutive survey wave (either before or after) where house is observed in the owner market.

[2] Another common approach is to multiply the purchase value of the property by a fixed percentage, often 0.0785, following Peiser and Smith (1985) to impute rent. This method is used by Gyourko and Tracy (1991), Albouy (2016), and Diamond (2016).

[3] Because rent-to-price ratios fluctuate over the business cycle (Diamond, Landvoigt and Sanchez, 2023), the relationship between rents and prices is more stable over time in ranks than in levels.

We estimate this quantile regression at quantiles 15.5, 39, 59, 79, 92, 97 and 99. We then predict the distribution of rental market ranks for each CBSA in each year using equation (2) where we where we set X_{it} equal to the nationwide owner-occupied average property characteristics and $Q_{V_{ijt^*}}, Q^2_{V_{ijt^*}}$ to the average property value rank (and its square) within each CBSA. $\hat{S}^q_{R_{jt'}} = \delta^q_{t'} + \delta^q_j + \beta^q \bar{X} + \gamma^q_1 \bar{Q}_{V_{j^*}} + \gamma^q_2 \bar{Q}^2_{V_{j^*}}$. We then transform these predicted ranks in the rental distribution, $\hat{S}^q_{R_{jt'}}$, to their corresponding log annual rents according to the distribution of rents observed in the AHS in year t, F_t. $\hat{r}^q_{jt} = F_t(\hat{S}^q_{R_{jt'}})$. To get the expected annual log rent in CBSA j in year t we then integrate over this distribution of log rents. $E(r_{jt}) = \int_0^1 \hat{r}^q_{jt} dq$. We approximate this integral using the midpoint rule with the midpoint set at the quantiles mentioned above.

A few studies have analyzed housing units that switch between the owner-occupied and rental market to estimate price-to-rent ratios (Bracke, 2015; Eichholtz et al., 2021). However, our method utilizes these switchers as a way to extrapolate rental cost estimates to the entire stock of owner-occupied housing.

III. Results

On average, we find that our switcher methods predicts expected log rents for the owner occupied housing stock that are 3.2 percent higher than those predicted by hedonic methods (equation (1). We find substantial heterogeneity across CBSAs in the difference between rents predicted by the two methods. In Detroit, the switcher method predicts rent are 12.3% higher than the hedonic estimates. On the other end of the distribution, we find the switcher method estimates New York rents to be 3.6% lower than suggested by the hedonic method.

We find the difference in predicted owner-occupied rents between the switcher and hedonic method is strongly correlated with the housing property (asset) values across CBAS.[4] Figure 1a shows a 1 percent increase in a CBSA's property value index is associated with a 0.08 percent decrease in the difference between rental predictions from the switcher and hedonic estimation methods. Hedonic methods overstate differences in the rental cost of owner-occupied housing across CBSAs.

To highlight a salient example, we compare the CBSAs of Detroit and San Francisco. According to the hedonic property value estimates, the difference in house asset values of owner-occupied properties in these CBSAs is 440 percent. Asset value of housing is a poor measure of the per-period cost of housing consumption. Higher purchase prices are associated with higher subsequent house price appreciation, confounding this a measure of cost (Demers and Eisfeldt, 2022). Looking at hedonic rental differences for owner-occupied properties, San Francisco is 225 percent more expensive than Detroit. Using the switcher estimation method, San Francisco is 200 percent more expensive than Detroit.

IV. Mechanisms: The Role of Land-use Regulations

A full analysis of why hedonic methods disproportionately under predict rental values for owned homes in low price cities is beyond the scope of this paper. We investigate one possible theory: geographic variation in land-use regulation. Prior work has shown the growing importance of land-use regulation over the past 70+ years in restricting housing supply and inflating the cost of housing in highly regulated cities. See Gyourko and Molloy (2015) for a review of this literature.

Land-use regulations restricts the choice-set of what types of housing can be built in different locations. Since most of the present laws were enacted prior to 1970, many areas that have highly restrictive land-use laws today have their housing stocks "frozen in time" with what was built in or before the 1970s. This leads to spatial misallocation in the characteristics of the housing stock. For example, Cambridge, MA has a large stock of 2-3 unit multi-family properties that were built in the 19th century and cannot be replaced with modern housing by regulation. Despite their low physical house quality, they rent for a high price due to their central, desirable location. Compare this to sub-

[4] We estimate the property value price index by using equation 1, but change the dependent variable to the house value (as opposed to rent) reported for each owner-occupied house in the AHS. We then predict CBSA property values holding fixed observable differences in housing quality.

urban North Andover, MA. According to the town-level Wharton Land-Use Regulation Index, North Andover's land use regulation is 0.5 standard deviations *lower* than the nationwide mean, while the overall Boston CBSA land-use level is 0.4 standard deviations higher. North Andover has substantial new construction and affordable single family homes. Despite the housing quality being higher in North Andover, the location is less central and does not command the high land values of Cambridge. This misallocation creates a negative correlation between location quality and housing quality in CBSAs that are highly regulated: new construction can only be built in less desirable parts of the CBSA with less land-use regulation.

In contrast, CBSAs like Detroit have very little land-use regulation and allow construction of most housing types in most areas. Since higher income households are likely willing to pay for high quality housing in high quality neighborhoods, one would expect there there be strong income sorting on housing quality in CBSAs that have lower land-use regulation. In highly regulated CBSAs, the high quality housing will be misallocated to lower quality neighborhoods more often. This low quality housing stock in desirable neighborhoods is also more likely to be rented. In these high regulation CBSAs, high income households are more likely to have to choose either a high quality neighborhood with inferior house quality, or a lower quality neighborhood with higher house quality.

What does this imply about the rental value of owner-occupied housing? Since high-income households are much more likely to be home owners vs. renters, the owned housing stock will be selected towards to the high-end of the housing market, both in terms of housing quality and location quality. However, most hedonic rent imputations, such as equation (1), only adjust for quality differences in the housing structure and not neighborhood differences. When land-use is not restricted, this suggests that for a given level of structure quality, owner-occupied houses tend to be in higher quality neighborhoods than equivalent rental houses. Thus, hedonic rental predictions would be biased downward for low regulation CBSAs. In high regulation CBSAs, the sorting of high quality houses to high quality neighborhoods is weakened and could even be reversed, as suggested by the Boston example above. If this is the case, the hedonic rent imputations can over predict rents, since rental housing units for a given physical quality could be in higher quality neighborhoods than similar owner-occupied units.

To test this theory, we first regress the difference in predicted owner-occupied rents between the switcher and hedonic method on the 2018 Wharton Real Estate Land-Use Regulation Index (Gyourko, Hartley and Krimmel, 2021).[5] Figure (1b) shows a one standard deviation increase in the WRLURI is associated with a 0.08 drop in the difference in log rental value predictions between the hedonic and switcher methods. This shows that in low regulation CBSAs, such as Detroit, the hedonic rents are below the switcher rents, while in high regulation places, such as New York, hedonic rents are above switcher rents.

To further probe our hypothesis that land-use regulation is causing misallocation of housing quality across neighborhoods, we estimate the level of income sorting across housing quality for each CBSA. To do this, we use our hedonic rent equation (1) to predict a housing quality index for each housing unit in the the AHS (including both rented and owner properties). Housing quality is simply: $\hat{\beta}X_{it}$. We then regress this housing quality on the log household income, $LnInc_{it}$ of the household living in the housing unit: $\hat{\beta}X_{it} = \gamma_j LnInc_{it} + \delta_{jt} + \varepsilon_{it}$. γ_j represents the elasticity of housing quality with respect to household income in CBSA j. When γ_j is high, this means households are strongly sort on physical housing quality across the income distribution.

Our theory of housing misallocation in regulated CBSAs would predict a lower level of income sorting in highly regulated CBSAs, and a higher level in low regulation CBSAs. Figure (1c) regresses γ_j on the level of land-use regulation. We find a 1 standard deviation in land-use regualtion is associated with a 0.02 lower elasticity of housing quality with respect to household income. This is a substantial effect. Detroit's elasticity of housing quality with respect to household income is 0.078. Moving Detroit's level of regulation to San Francisco's level

[5] Following Gyourko, Hartley and Krimmel (2021), we measure a CBSA's land-use regulation by taking a simple mean WRLURI across municipalities in the CBSA.

would lower this elasticity to 0.048, a forty percent decline.

We next look at how γ_j relates to the the difference in rental predictions between our hedonic and switcher methods. Figure (1d) shows a 1 unit increase in the elasticity of housing quality with respect to household income is associated with a 2.81 percent increase is in the difference between the switcher and hedonic rental market predictions, consistent with our housing misallocation theory of land-use regulation.

V. Conclusion

We apply the method of Diamond and Diamond (2024) to estimate housing rental prices of owner-occupied housing across the 15 largest CBSAs in the US. We show that standard hedonic rent methods overstate the spatial variation in housing affordability. We find that hedonic rent methods systematically under predict rents for owner occupied units, but this is especially true in low land-use regulation CBSAs. In the most highly regulated CBSAs, hedonic rent methods over predict rents for owner-occupied units. This difference appears to be driven by the misallocation of housing quality across neighborhoods in regulated CBSAs.

REFERENCES

Albouy, David. 2016. "What are cities worth? Land rents, local productivity, and the total value of amenities." *Review of Economics and Statistics*, 98(3): 477–487.

Bracke, Philippe. 2015. "House prices and rents: Microevidence from a matched data set in Central London." *Real Estate Economics*, 43(2): 403–431.

Davis, Morris A, Andreas Lehnert, and Robert F Martin. 2008. "The Rent-price ratio for the aggregate stock of owner-occupied housing." *Review of Income and Wealth*, 54(2): 279–284.

Demers, Andrew, and Andrea L Eisfeldt. 2022. "Total returns to single-family rentals." *Real Estate Economics*, 50(1): 7–32.

Diamond, Rebecca. 2016. "The determinants and welfare implications of US workers' diverging location choices by skill: 1980–2000." *American Economic Review*, 106(3): 479–524.

Diamond, Rebecca, and Enrico Moretti. 2021. "Where is standard of living the highest? Local prices and the geography of consumption." National Bureau of Economic Research.

Diamond, Rebecca, and William Diamond. 2024. "Racial Differences in the Total Rate of Return on Owner-Occupi." *Stanford GSB Mimeo*.

Diamond, William, Tim Landvoigt, and German Sanchez Sanchez. 2023. "Printing Away the Mortgages: Fiscal Inflation and the Post-Covid Housing Boom."

Eichholtz, Piet, Matthijs Korevaar, Thies Lindenthal, and Ronan Tallec. 2021. "The total return and risk to residential real estate." *The Review of Financial Studies*, 34(8): 3608–3646.

Gyourko, Joseph, and Joseph Tracy. 1991. "The structure of local public finance and the quality of life." *Journal of political economy*, 99(4): 774–806.

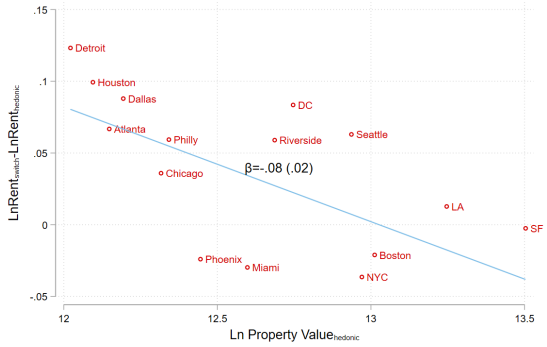
Gyourko, Joseph, and Raven Molloy. 2015. "Regulation and housing supply." In *Handbook of regional and urban economics*. Vol. 5, 1289–1337. Elsevier.

Gyourko, Joseph, Jonathan S Hartley, and Jacob Krimmel. 2021. "The local residential land use regulatory environment across US housing markets: Evidence from a new Wharton index." *Journal of Urban Economics*, 124: 103337.

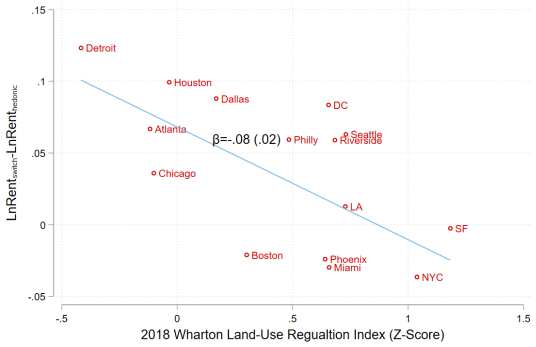
Moretti, Enrico. 2013. "Real Wage Inequality." *American Economic Journal: Applied Economics*, 5(1): 65–103.

Peiser, Richard B., and Lawrence B. Smith. 1985. "Homeownership Returns, Tenure Choice and Inflation." *Real Estate Economics*, 13(4): 343–360.

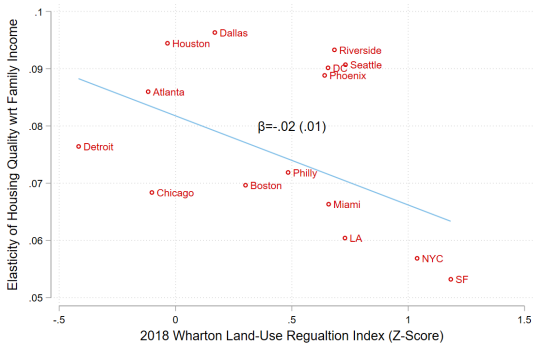
Poterba, James M. 1984. "Tax subsidies to owner-occupied housing: an asset-market approach." *The Quarterly Journal of Economics*, 99(4): 729–752.



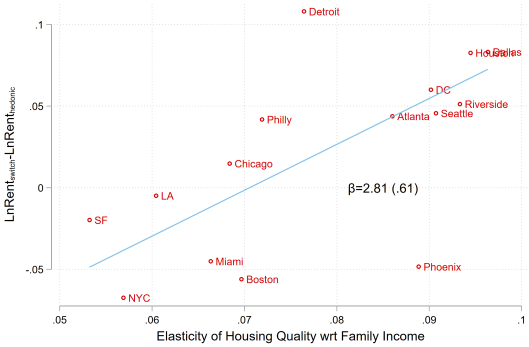
(a) Δ Log Rent vs Property Value Price Index



(b) Δ Log Rent vs WRLURI



(c) Elasticity of Housing Quality wrt Income vs WRLURI



(d) Δ Log Rent vs Elasticity of Housing Quality wrt Income

Figure 1.

Note: Panels a-c's dependent variable is the difference in the switcher method estimate of the owner-occupied rent index and the hedonic method estimate of the owner-occupied rent index. WRLURI stands for the Wharton Real-Estate Land-Use Regulation Index,.

Table 1—: Summary Statistics

	(1)		
	mean	sd	Obsevations
Ln Annual Rent	9.49	0.81	52202
Ln Property Value	12.6	0.94	77298
Owner Occupied	0.58	0.49	128419
Number of Rooms	5.56	1.63	128419
Property Age	47.4	25.2	128419
Single Fmaily Home	0.63	0.48	128419
Multi Family Property	0.35	0.48	128419
Ln Family Annual Income	10.9	1.22	124958
Switched btwn Owner and Renter Markets	0.035	0.18	128419
Year	2018.1	2.24	128419

Note: Data come for the the 2015-2021 American Housing Survey covering the 15 largest CBSAs, as identified in the public-use ASH data.