August 2022

New Tools to Nowcast Rents

By Michael Clawar, Eric Draeger, Rachel Levy, and Gleb Nechayev[1]

Using data science to understand how underlying market trends impact performance can help investors identify underperformers as well as locations with high potential rents, thus allowing for faster decision-making and better investment returns.

Real estate investing is often referred to as a "long-distance race" as owners typically hold assets over several years. However, a property's rent levels at acquisition and right after (the next three to six months) can be as critical to investment returns as rent growth over the rest of the holding period. Consider a hypothetical acquisition of an apartment property with rents 15% or more below market: over a five-year period, it could generate the same investment returns as another property charging fair market rents due to much faster rent growth. There is good reason for the expression that one usually makes, or loses, money on a buy rather than a sale.

Despite recent advancements in technology and available data, accurately estimating "fair" current rent levels by measuring and explaining market shifts on a typical acquisition timeline remains more art than science. That said, a good model or algorithm can go a long way in taking the guesswork out of underwriting both current rent levels and near-term changes. With greater availability of both high-frequency data at the micro-location level and modeling techniques, data science can enable more precise measuring, predicting, and forecasting of rents. These models tie together property characteristics, economic and demographic factors to "nowcast" and forecast rent growth for specific asset classes or even hyper-local geographies. With explainable AI (XAI), models formerly considered black boxes can give interpretable results to investors, helping explain why certain properties or markets have stronger growth potential.[2]

---

[1] Michael Clawar is Vice President, Data Science at Altus Group, Eric Draeger is Chief Investment Officer at Berkshire Residential Investments, Rachel Levy is Data Scientist at Altus Group, Gleb Nechayev is Head of Research at Berkshire Residential Investments.

[2] Explainable AI (XAI) is artificial intelligence with interpretable results. For a recent example of interpretable machine learning in real estate, see Lorenz, Felix, Jonas Willwersch, Marcelo Cajias and Franz Fuerst (2022). Interpretable machine learning for real estate market analysis. *Real Estate Economics,* May 31.

Investors armed with data science tools can be more confident in projecting asset pricing, revenues and returns over the holding period. Moreover, they can adjust forecasts in real time rather than waiting for third-party data providers to catch up. The key is that these modeling tools allow investors to constantly recalibrate as both the market environment and available data evolve. As facts on the ground change, these models can reveal underlying trends faster than it takes for market consensus to catch up.

In this article, we explore the results of one promising data science model called random forest to nowcast 2022 rents on the zip-code level and identify areas with higher growth potential in the residential market. We found that this approach highlighted performance opportunities and provided a greater understanding that while home values are a strong predictor of near-term rent growth, incorporating economic and demographic fundamentals significantly improve accuracy when nowcasting fair market rent levels. This modeling approach could potentially be applied to additional asset types.

Random forest model for nowcasting

For our nowcasting approach, we utilized a random forest model to predict the latest 2022 Zillow Observed Rent Index (ZORI). ZORI, a smoothed, repeat-rent index measuring the typical observed market rent rate across a given region, is computed using the mean of listed rents that fall into the 40th to 60th percentile range for all homes and apartments within a given zip code.[3] For our study, all US zip codes with complete data between 2015-2022 were chosen.

A random forest model pools predictions from multiple tree-based models and considers the diversity of those outcomes to make a final informed decision.[4] In addition to the Zillow Home Value Index (ZHVI),[5] we enhanced the model with several demographic and economic variables, including median renter income, net migration rate, rental affordability (a household income to rent ratio), economic employment diversity and labor market health.[6] Although other types of models could have been utilized to predict rent levels, we chose

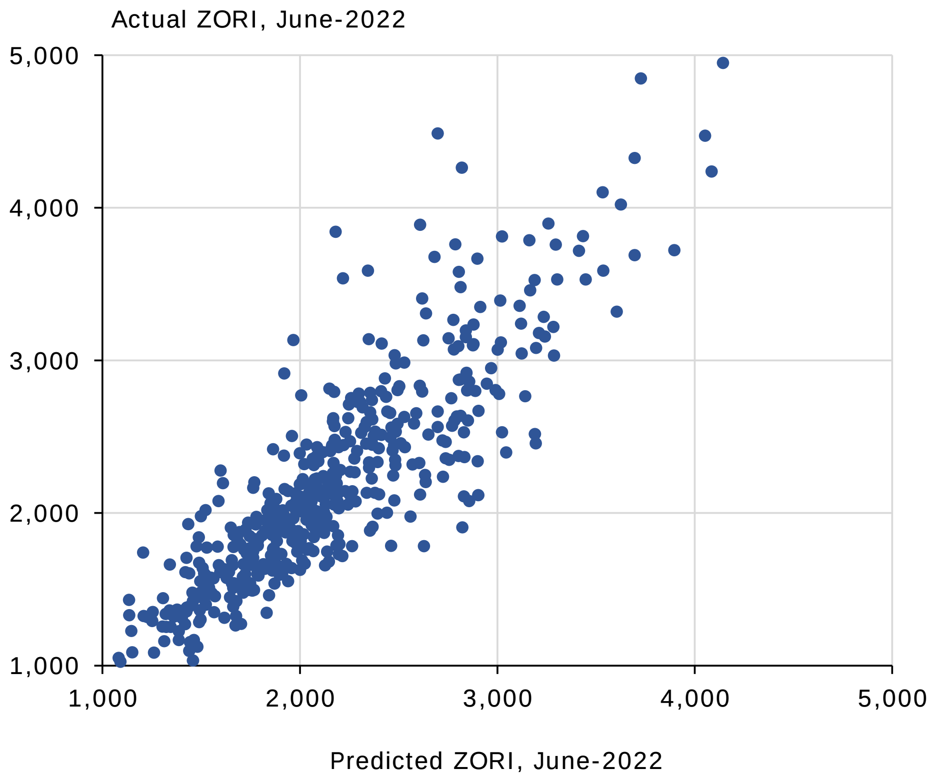
---

[3] Methodology: Zillow Observed Rent Index (ZORI). https://www.zillow.com/research/methodology-zori-repeat-rent-27092/.
[4] Specifically, each 'tree' is based on a random subset of the available ZIP codes and modeling features. The 'forest' combines these relatively weak individual predictors into what is generally found to be a strong and robust model less prone to data mining.
[5] Zillow Home Value Index Methodology, 2019 Revision: Getting Under the Hood. https://www.zillow.com/research/zhvi-methodology-2019-deep-26226/.
[6] ZIP code level data from Stratodem Analytics/Altus Group.

the random forest algorithm for this nowcasting exercise as it can handle both continuous and categorical variables.

The results of the random forest model are represented in the scatterplot in Exhibit 1, reflecting zip codes and their actual rents versus predicted value of the ZORI. On a hold-out set of 476 zip codes, the model predicted the ZORI with a high degree of accuracy, representing an R^2 score (a model performance metric ranging from 0-1) of 0.77 (Exhibit 2).[7] While home values were the most significant predictor of zip code rent levels, additional economic and demographic factors significantly improved model performance from an R^2 of 0.64. With the inclusion of economic and demographic factors, the median absolute error decreased from 12.8% to 8.9%.

Predicted vs. Actual Rents by ZIP Code



Sources: Altus Group, Berkshire Residential Investments, Zillow Group

[7] The hold-out set on which the model is evaluated is also referred as the 'test' set, while the model is estimated on the 'training' part of the full data. This separation helps identify whether the model is overfitted.

Random Forest Model Results

	ZHVI as the only factor		ZHVI plus other factors	
	Training set	Test set	Training set	Test set
Observations (number of zip codes)	1,900	476	1,900	476
R^2	0.67	0.64	0.93	0.77
Median absolute error (%)	10.8%	12.8%	5.3%	8.9%

Sources: Altus Group, Berkshire Residential Investments

Utilizing nowcast results

With the nowcast approach, the goal is not to achieve perfect accuracy for the random forest model as this would imply that all market rents are already accurately priced. Instead, a model that has a thorough understanding of how underlying market trends impact property performance can help investors gain insight on how certain areas will play out in the long run and identify both underperformers and, importantly, locations with high potential rents.

Backtesting the same factors as predictors on 2015 ZORI data, we find that rent growth for "underpriced" zip codes outperformed 'overpriced' zip codes by 14.2% over the next five-year period (Exhibit 3). For the zip codes that the model predicted 2015 rent levels to be 15% higher than its actual value, suggesting that rents were underpriced by the model's standards, actual rent growth by 2019 averaged 30.7%. For the zip codes where the model predicted rent levels to be 15% lower than its actual value, suggesting rents may be overpriced, the 5-year growth rate turned out to be only 16.5%.

Backtest results found underpriced zip codes outperforming overpriced zip codes

2015 Nowcast Model *5-Year Rent Growth by Residual Group*	
Residual *Actual minus Predicted*	**Average Rent Growth** *Jan 2015-Dec 2019*
-15% or less	30.7%
-15% to –5%	25.2%
-5% to 5%	22.4%
5% to 15%	18.1%
15% or more	16.5%

Sources: Altus Group, Berkshire Residential Investments

Assessing residuals from the June 2022 prediction, we can similarly gain insight into which zip codes have potential for higher rent levels and a strong propensity to outperform widely expected revenue growth within the next year. Exhibit 4 shows the top five zip codes where the actual rent levels are tangibly lower relative to where the model expects them to be given their specific location's home price levels and economic and demographic features. The model can assess thousands of locations around the nation to identify areas with the most upside potential for rents as well as downside for risk. This can help guide both site selection and underwriting, in this case, going well beyond what traditional forecasting platforms offer now.

Top underpriced zip codes based on the nowcast model

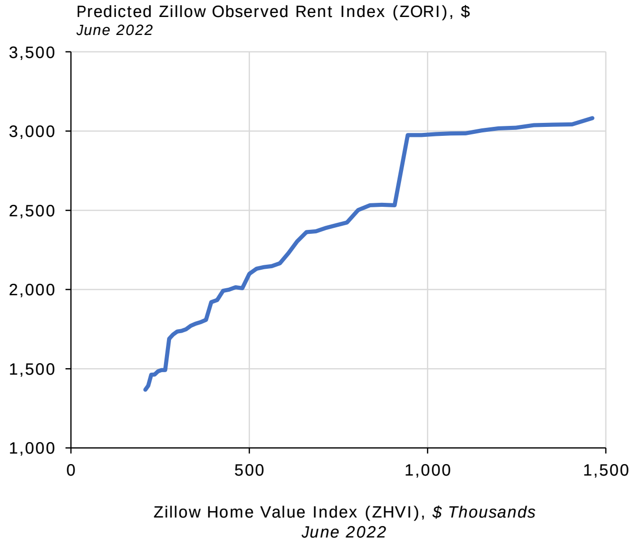
2022 nowcast model: Five lowest residuals by zip code				
ZIP	Market	2022 ZORI, $		Residual
		Actual	Predicted	
75201	Dallas-Fort Worth, TX	$2,105	$2,657	-20.8%
34683	Tampa, FL	$1,767	$2,211	-20.1%
78746	Austin, TX	$1,993	$2,460	-19.0%
92663	Los Angeles-Long Beach-Anaheim, CA	$2,397	$2,910	-17.6%
20015	Washington, DC	$2,297	$2,740	-16.2%

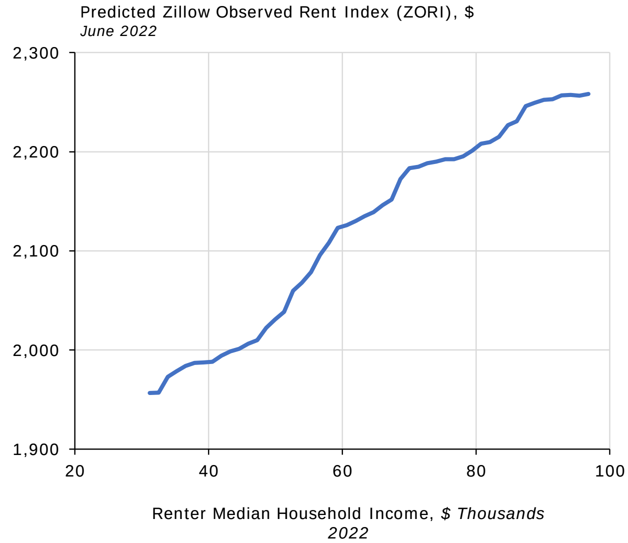
Sources: Altus Group, Berkshire Residential Investments

Interpreting the model's predictive impact

Many machine learning models are black boxes, consuming information and creating outputs through processes too complex for human understanding. Explainable AI (XAI), or interpretable machine learning models, is a critical approach for commercial real estate investors to understand the factors contributing to predictions.

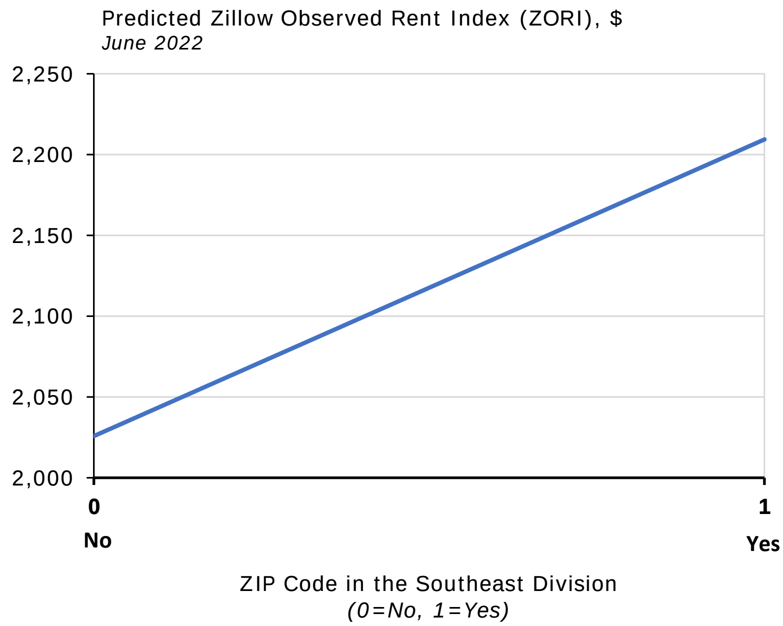
One simple example comes from partial dependence plots (PDPs) which show the relationship between the target variable (in our case, ZORI), and one or more input features (e.g., ZHVI) controlling for all the other factors. Exhibits 5 through 7 are three PDPs for the strongest predictors in our nowcast model based on relative impact to rents: ZVHI, median renter household income in the zip code and a dummy variable for whether the zip code is in the Southeast division of the US. We see strong, positive, non-linear relationships for home values and renter income. With ZHVI, for example, once typical home values in the zip code exceed $925,000, predicted ZORI jumps noticeably from $2,500 to $3,000. As a result, two zip codes on either side of this cutoff may have very different pricing dynamics.





Sources: Altus Group, Berkshire Residential Investments

The Southeast division flag has a more straightforward interpretation: after controlling for economic and demographic factors, the Southeast division today has $184 higher rents than other divisions.



Sources: Altus Group, Berkshire Residential Investments

Conclusion

With the improvement in high-frequency data widely available today, data science models and analytical tools can be used to assist with the challenge of assessing fair market rent values and higher-growth opportunities. Investors need confidence in pricing and forecasting models, supported by backtesting and explainable AI. While we find that underpriced zip codes from 2015 significantly outperformed over the following five-year period, other work in time series machine learning supports separating near-term and longer-term forecasting.

Although home price data is a principal factor in these models, it is not sufficient to identify true investment opportunities in today's market. We also find that economic and demographic fundamentals materially impact model and hypothetical investment performance. As a result, investors will gain insight on underlying market trends and the impact on rent levels before the market catches up and comes to a consensus on "fair" rent levels, allowing for faster decision-making and the potential for better investment returns.

Disclosures

The opinions expressed herein represent the current, good faith views of Berkshire Residential Investments (Berkshire) at the time of publication and are provided for limited purposes. The information presented in this article has been developed internally and/or obtained from sources believed to be reliable; however, Berkshire does not guarantee the accuracy, adequacy, or completeness of such information. Predictions, opinions, classifications, use of common industry terms, and other information contained in this article are subject to change continually and without notice of any kind and may no longer be true after the date indicated.

Any forward-looking statements speak only as of the date they are made, and Berkshire assumes no duty to and does not undertake to update forward-looking statements. Forward-looking statements are subject to numerous assumptions, risks, and uncertainties, which change over time. Actual results could differ materially from those anticipated in forward-looking statements. This material is for informational purposes only. It is not intended to, and does not constitute financial advice, investment management services, an offer of financial products or to enter into any contract or investment agreement in respect to any product offered by Berkshire and shall not be considered as an offer or solicitation with respect to any product, security, or service in any jurisdiction or in any circumstances in which such offer or solicitation is unlawful or unauthorized or otherwise restricted or prohibited.

Berkshire provides investment management services to advisory clients that invest in the multifamily housing sector. In respect of its investment management services, Berkshire may receive performance-based compensation from such advisory clients. Accordingly, Berkshire may financially benefit from the appreciation of multifamily housing units.