

Methodology

This study forecasts owner-occupied and renter household formation at the county level across Indiana over a five-year horizon (2025 to 2029), then translates those forecasts into estimates of housing supply need. The core modeling approach uses a rolling five-year window regression, trained on historical relationships between local economic and demographic conditions and subsequent household growth. The model is evaluated using leave-one-window-out cross-validation, where each test window is predicted using only data from earlier periods.

Data sources

The following data series are used as model inputs and for computing supply need components.

Household counts (owner and renter)

American Community Survey (ACS) 5-year estimates, retrieved via the Census Bureau API using the tidycensus R package. The base year is 2019 and the current year is 2024. Variables include owner-occupied units, renter-occupied units, vacant-for-sale and vacant-for-rent counts, and total housing units by structure type and year built.

Building permits

Single-family residential building permits by county, sourced from the Indiana Department of Workforce Development via HoosierData. Single-family permits are used throughout as a proxy for owner-occupied construction activity. This is a standard approximation; a small share of single-family homes are renter-occupied, and some owner-occupied units take other structural forms.

Employment

Annual average employment levels by county, from the Bureau of Labor Statistics Quarterly Census of Employment and Wages (QCEW), all-industry series. Year-over-year change in employment is used as a model input.

Migration

Domestic and international migration flows by county, sourced from the U.S. Census Bureau Current Population Estimates. Both five-year cumulative flow levels and rates (flows as a share of total households) are included as model inputs.

Home sales and prices

County-level home sales volume and mean sale price, used to capture local market conditions, sourced from Indiana Association of REALTORS® Housing Hub. Sales activity interacted with price level is the strongest positive predictor of owner household growth in the model.

Household formation forecast

Separate models are estimated for owner-occupied and renter household formation. Both use the same rolling window approach: for each historical five-year window, the model is fit on anchor-year conditions and asked to predict the change in households five years later. This structure mirrors the forecast task directly.

Owner model

The owner household model regresses five-year change in owner-occupied households on the following anchor-year predictors: mean home sale price; the interaction of sales volume with mean price; one-year change in owner households; one-year change in employment; five-year domestic migration level; five-year international migration level; domestic and international migration rates (flows as a share of total households); and interactions between current owner household count and each migration level term. The model specification in R is:

```
total_change ~ mean_price + total_sales:mean_price + owner_diff1 + employment_diff1 +  
dom_mig_5yr + intl_mig_5yr + dom_mig_rate + intl_mig_rate +  
owner_householdsE:dom_mig_5yr + owner_householdsE:intl_mig_5yr
```

Renter model

The renter model uses the same rolling window structure with a parallel set of renter-specific predictors, including renter vacancy rates, median rent interacted with sales volume, and renter-specific migration and trend terms.

Training and forecasting procedure

For validation, the model is trained on all five-year windows where the anchor year is earlier than the test window. No future data is ever used to train a model that predicts an earlier period. For the production forecast, all valid historical windows are used for training, and the model is applied to 2024 anchor-year conditions to project 2025 to 2029 change. Predicted change is applied purely from the regression; no trend blending or post-hoc adjustment is applied.

Driver contributions reported in the dashboard are computed as the term-level product of each predictor value and its estimated coefficient, then grouped into reader-friendly categories (Market conditions, Migration, Employment, Recent trend, Market size).

Model validation

Validation uses a leave-one-window-out procedure constrained so that the training period always precedes the test period. Five test windows are evaluated for the owner model (holdout anchor years 2015 through 2019) and four for the renter model (2016 through 2019), each covering all 92 Indiana counties, yielding 460 and 368 county-year observations respectively.

Error categories are defined as follows. A prediction is classified as “Nailed it” if the absolute error is 100 or fewer households, or the percentage error is 10 percent or less. “Pretty close” covers predictions within 500 households or 25 percent of actual. “Right direction, too far” and “Right direction, not far enough” describe predictions with the correct sign but incorrect magnitude. “Wrong direction” cases are those where the model predicts growth when decline occurred, or vice versa.

Owner model performance (460 county-year observations, 5 windows)

MAE: 510 households

Median absolute error: 316 households

RMSE: 889 households

Correlation (predicted vs. actual): 0.963

R²: 0.927

Error category breakdown:

Nailed it: 72 of 460 (16%)

Pretty close: 198 of 460 (43%)

Right direction, not far enough: 52 of 460 (11%)

Right direction, too far: 24 of 460 (5%)

Wrong direction (too high): 18 of 460 (4%)

Wrong direction (too low): 4 of 460 (1%)

Renter model performance (368 county-year observations, 4 windows)

MAE: 357 households

Median absolute error: 221 households

RMSE: 543 households

Correlation (predicted vs. actual): 0.850

R²: 0.722

Error category breakdown:

Nailed it: 77 of 368 (21%)

Pretty close: 134 of 368 (36%)

Right direction, not far enough: 32 of 368 (9%)

Right direction, too far: 7 of 368 (2%)

Wrong direction (too high): 8 of 368 (2%)

Wrong direction (too low): 18 of 368 (5%)

The owner model performs well, with nearly 60 percent of county-year predictions landing within 500 households or 25 percent of actual, and a correlation of 0.963 between predicted and observed change. Wrong-direction errors are rare (22 of 460, under 5 percent). The renter model shows lower but still meaningful predictive power, reflecting the greater volatility of the rental market and the sensitivity of renter household counts to short-term economic and migration shocks.

Supply need calculation

For each county and tenure type, the five-year supply need is calculated as:

Supply needed = Projected new households + Anticipated housing loss – Excess vacancy above equilibrium

Anticipated housing loss is estimated from observed historical loss rates applied to the current stock. Housing loss is estimated using the Census Bureau’s methodology for housing unit loss, which applies rates derived from the Components of Inventory Change (CINCH) supplement to the American Housing Survey, differentiated by structure type, age of unit, and Census region. Older homes and mobile homes carry substantially higher loss rates than newer single-family construction. Midwest regional rates are applied throughout.²

Excess vacancy is the difference between actual vacant-for-sale or vacant-for-rent units and the equilibrium count, defined as 1 percent of the owner market and 5 percent of the renter market. Where actual vacancy falls below equilibrium, this term is negative, adding to supply need. Where vacancy exceeds equilibrium, it reduces the net need.

Supply need is reported separately for owner-occupied and rental units. The analyses in this report focus on owner-occupied need, with rental need noted but not examined in detail.

Contact. Questions about methodology or data should be directed to the Indiana Association of REALTORS® (mnowlin@indianarealtors.com).