

The Cost of Living: A Zillow Housing Forecast

Patrick Beal
Tiara Eddington
Madelyn Vines

Math 404 Data Project Presentation

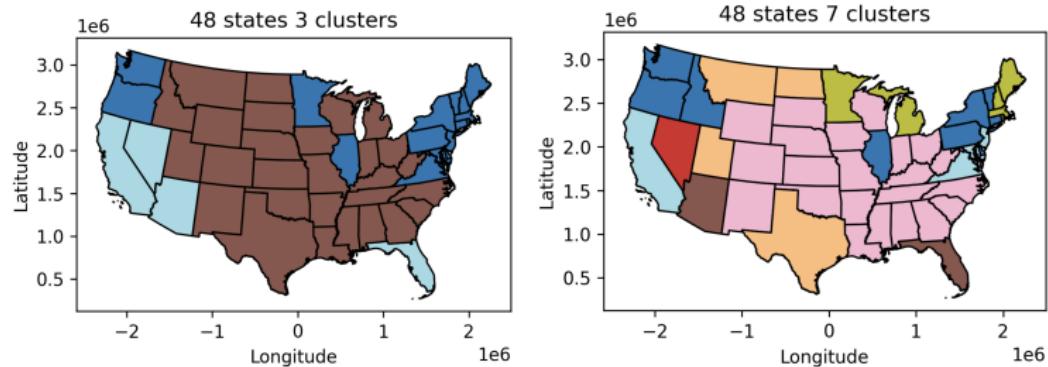
Motivation & Goals

- Housing prices shape wealth, policy, and investment.
- We aim to:
 - Identify similar housing markets (clustering).
 - Forecast housing prices using multiple models.
 - Identify correlations between housing markets and demographic features.
- Focus on interpretable and regional trends.

Data Overview

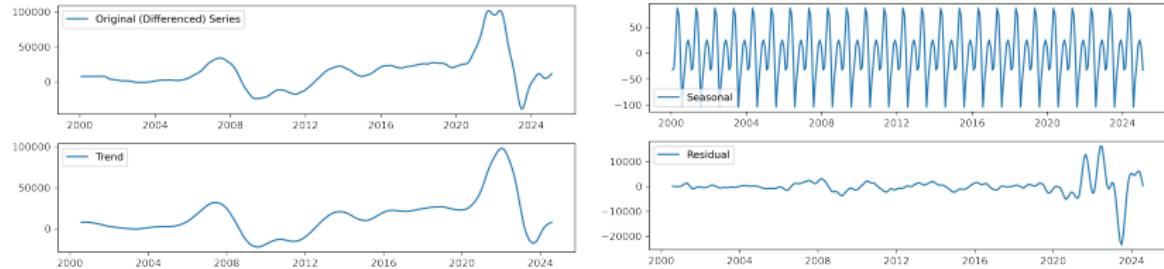
- **Zillow HPI:** State-level median home prices (monthly, 2000–2020).
- **CPS/IPUMS:** Demographics, income, and tax (annual, interpolated).
- Filtered to pre-COVID period.

Clustering States by Price Trends



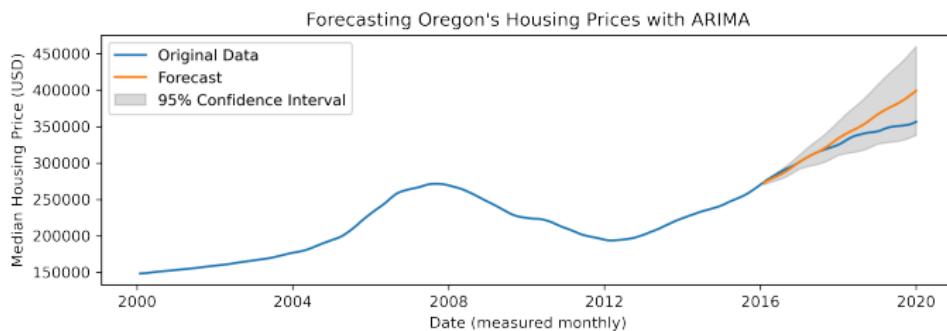
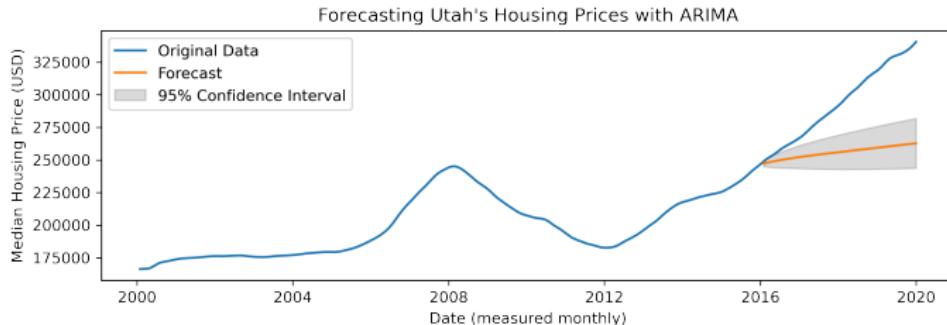
- K-means on housing price changes \Rightarrow groups with similar growth.
- Coastal states form distinct clusters.

Classical Decomposition of Housing Prices



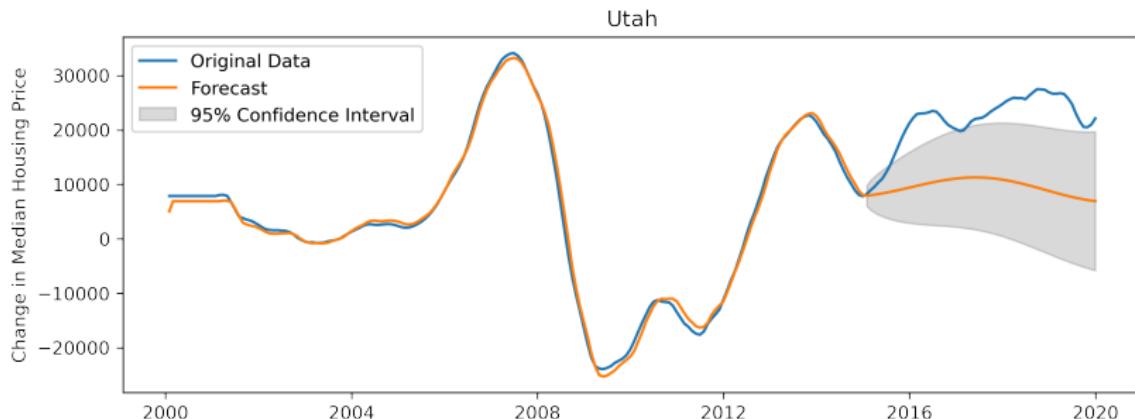
- Classical decomposition \Rightarrow double-peaked seasonal trends.
- Residuals spike after 2020 \Rightarrow excluded post-2020.

Univariate ARIMA Forecasting



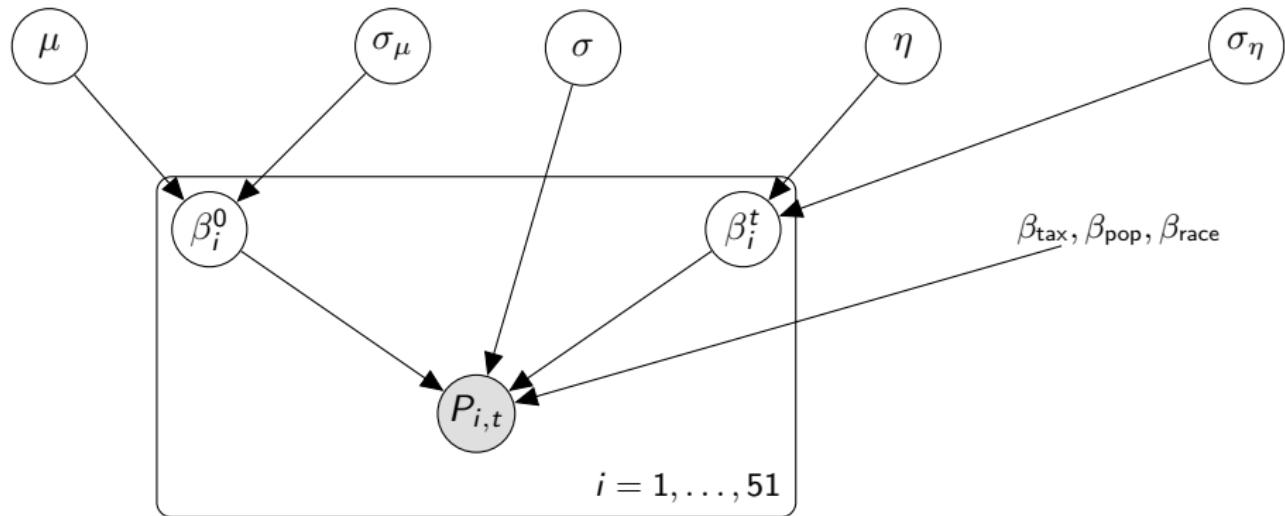
- Best ARIMA model: $(4,1,0)$ – approximately linear forecasts
- Limited accuracy: large confidence intervals, sensitivity to shocks

VARMAX Forecasting



- Cluster-based VARMAX \Rightarrow improved trend modeling
- Not applicable to states in singleton clusters
- Using demographic data led to convergence issues

Bayesian Hierarchical Model: Plate Diagram



Bayesian Hierarchical Model Setup

Model Equation

$$P_{i,t} = \beta_i^0 + \beta_i^t t + \beta_{\text{tax}} \text{tax}_i + \beta_{\text{pop}} \text{pop}_i + \beta_{\text{native}} \alpha_{\text{native},i} \\ + \beta_{\text{asian}} \alpha_{\text{asian},i} + \beta_{\text{black}} \alpha_{\text{black},i} + \beta_{\text{white}} \alpha_{\text{white},i} + \epsilon$$

- Captures variation across states via hierarchical priors.
- Standardized predictors allow coefficient comparisons.
- Fit using NUTS algorithm (via PyMC3).

Bayesian Hierarchical Prior Distributions

$$\beta_i^0 \sim \mathcal{N}(\mu, \sigma_\mu)$$

$$\mu \sim \mathcal{N}(200000, 100000)$$

$$\sigma_\mu \sim \text{Exp}\left(\frac{1}{50000}\right)$$

$$\beta_{\text{tax}} \sim \mathcal{N}(0, 20)$$

$$\beta_{\text{race}} \sim \mathcal{N}(0, 10000)$$

$$\sigma \sim \text{HalfCauchy}(100)$$

$$\beta_i^t \sim \mathcal{N}(\eta, \sigma_\eta)$$

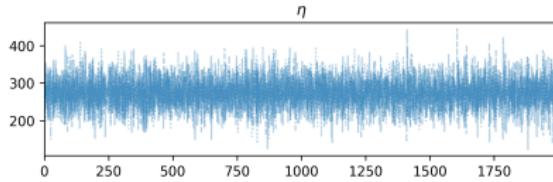
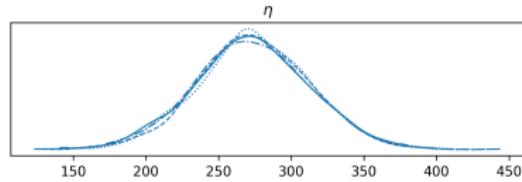
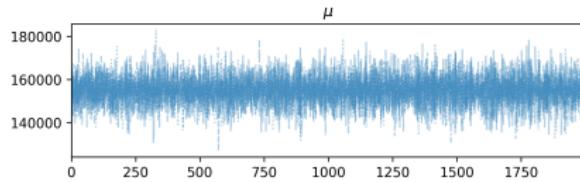
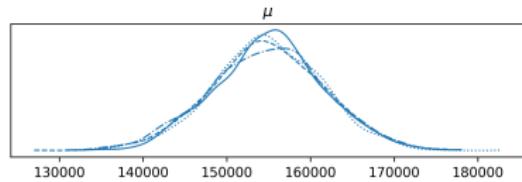
$$\eta \sim \mathcal{N}(1000, 1000)$$

$$\sigma_\eta \sim \text{Exp}\left(\frac{1}{5000}\right)$$

$$\beta_{\text{pop}} \sim \mathcal{N}(0, 1)$$

$$\epsilon \sim \mathcal{N}(0, \sigma)$$

Bayesian Hierarchical Modeling



- State-level intercepts/slopes + global covariates.
- Accounts for uncertainty and data imbalance.

Bayesian Hierarchical Modeling



Key Findings

- Central states show more homogeneous trends.
- ARIMA and VARMAX useful, but limited by volatility.
- Bayesian model provides interpretable insights:
 - Native/Asian/White proportion \Rightarrow lower prices.
 - Black proportion \Rightarrow rising prices.

Ethical Considerations

- Demographics **should not** guide individual decisions.
- Models risk perpetuating bias.
- Better use: *Policy insights and funding allocation.*

Conclusion

- Forecasting housing prices is difficult due to shocks.
- Combined methods yield the best insights.
- Future work: Combine Bayesian models with ARIMA/Kalman filters.

Resources

- GitHub Repo: github.com/jpatrickb/vol3_housing_project
- Data from Zillow + IPUMS/CPS

Questions?

Thank you!