Forecasting Home Prices

Jeffrey Camp
Cristian Gavrus
Eric Munsing
Mentor: Kevin Teh

github.com/kkwteh/cdips_hpi_forecast

Opendoor

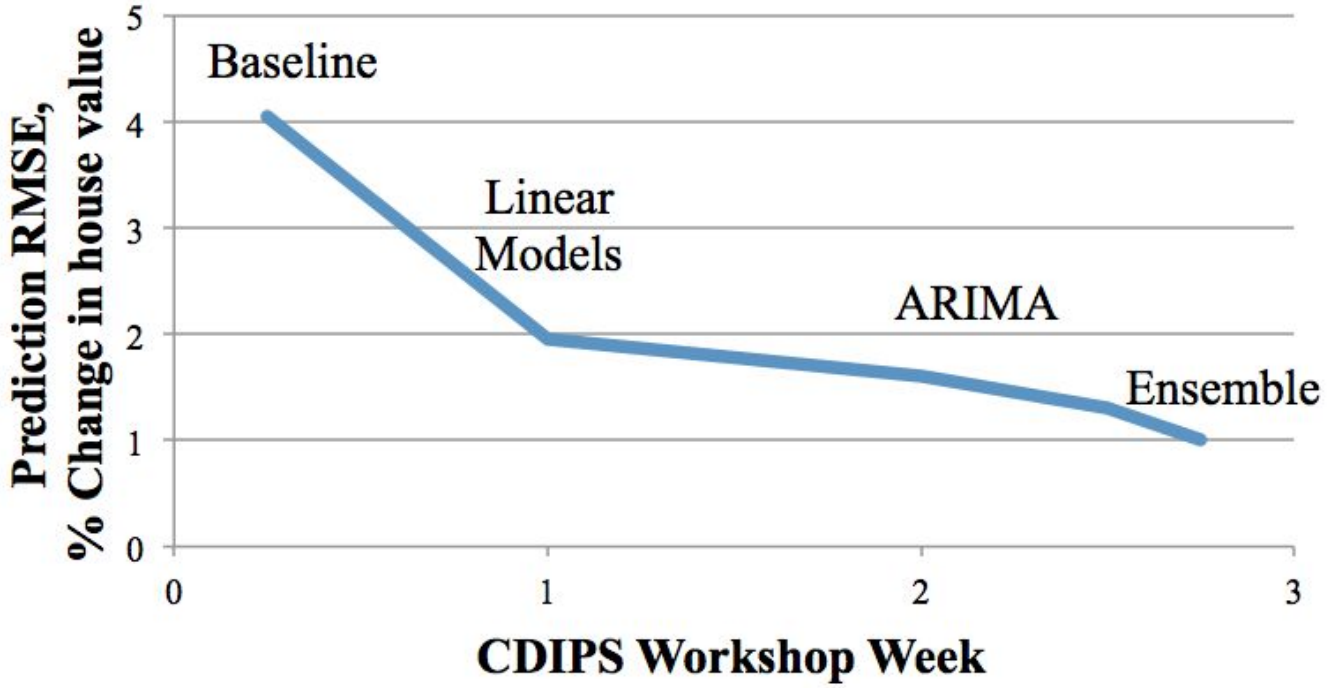
Opendoor buys houses and holds them 1-12 months until they find a seller

- Exposed to market volatility
- Price-sensitive sellers
- *Price must be competitive but factor in market risk*

CDIPS challenge:

- Forecast 6-month change in home price
- Metric: % change in housing price index

CDIPS project: Predict 6-month housing prices



Improving housing price prediction by 3% = reduce risk by $6000 per house!

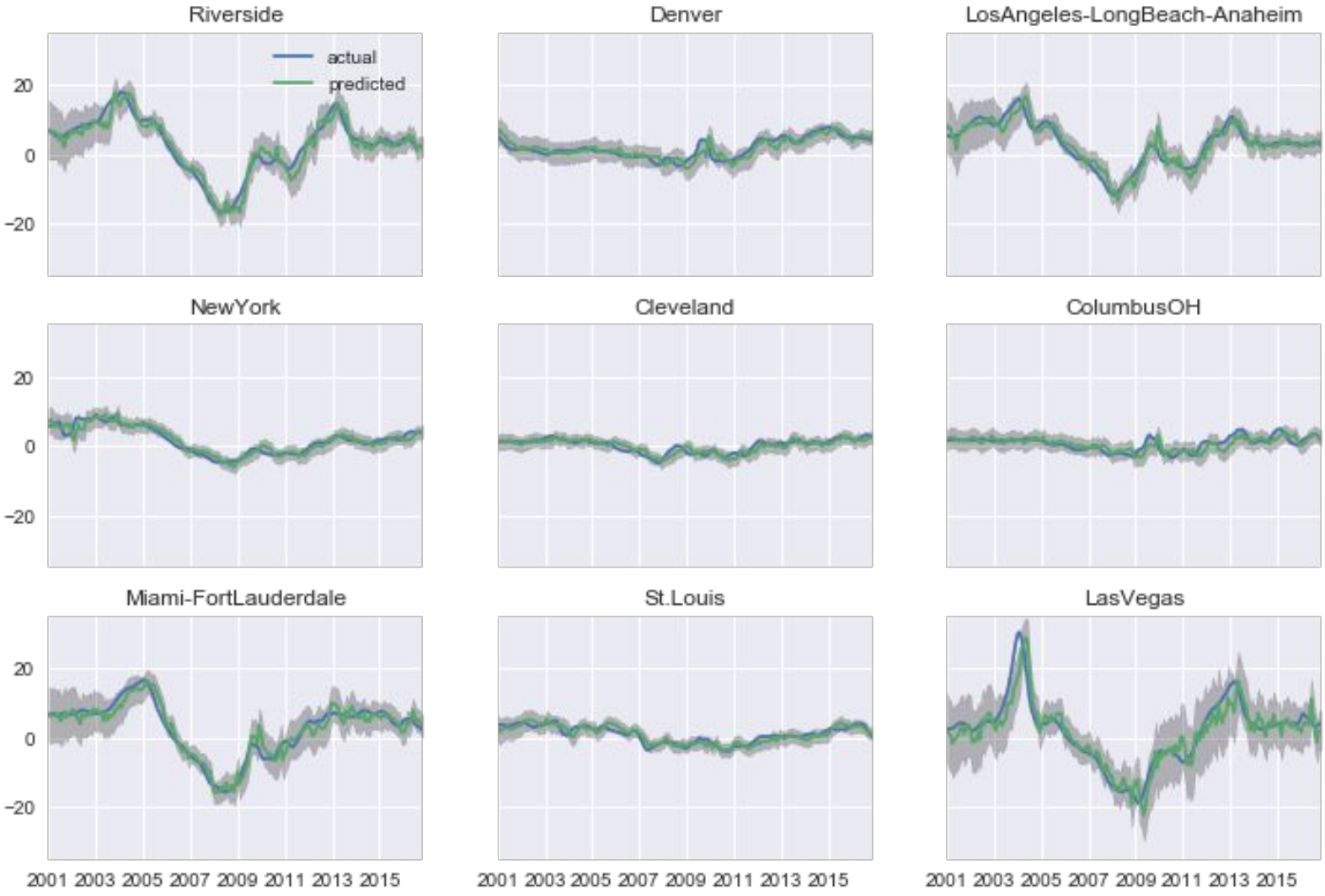
Model Details

- $ARMA(p, q)$ is a model for a time series data X_t indexed by time t :

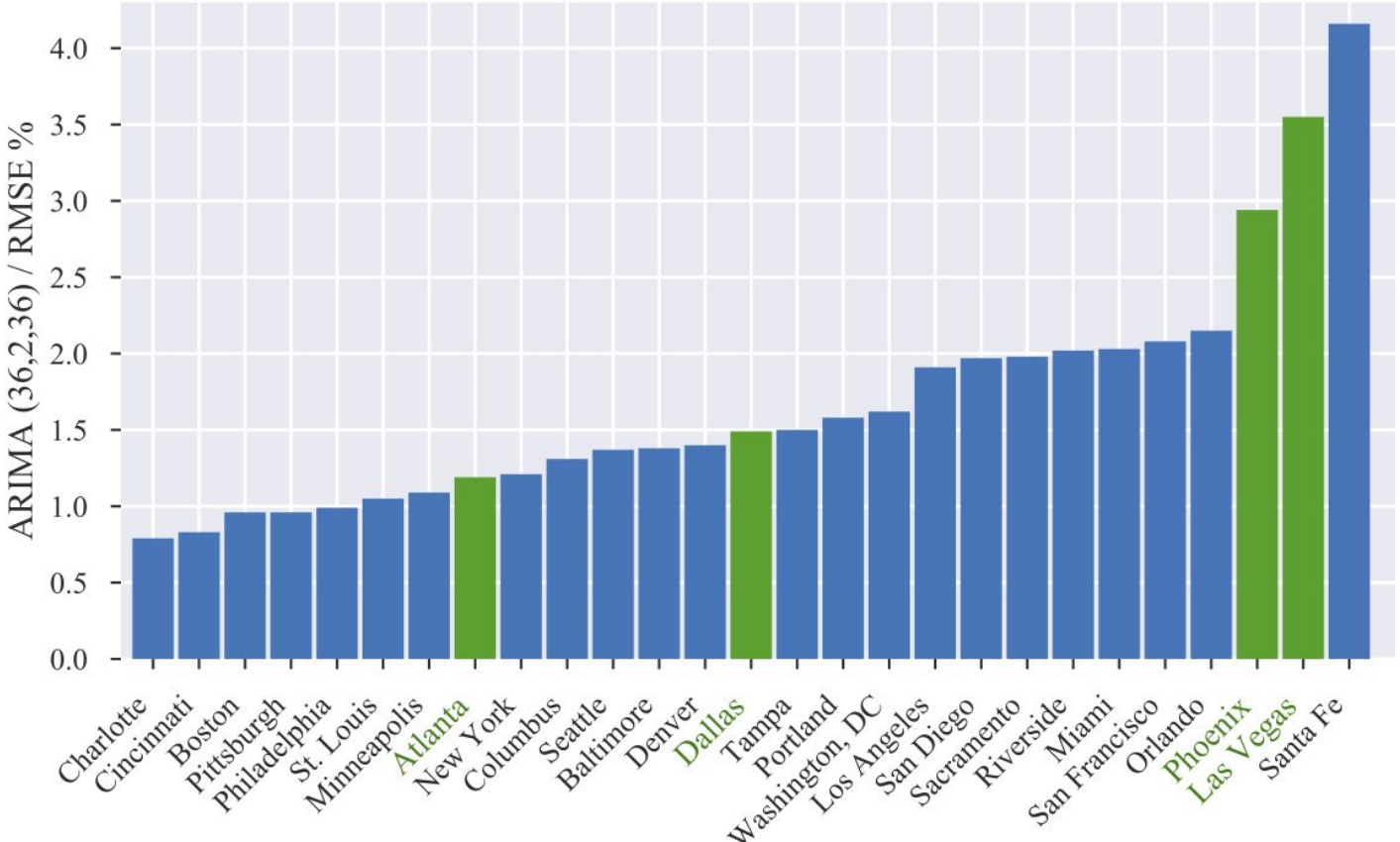
$$X_t - \alpha_1 X_{t-1} - \cdots - \alpha_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots \theta_q \varepsilon_{t-q}$$

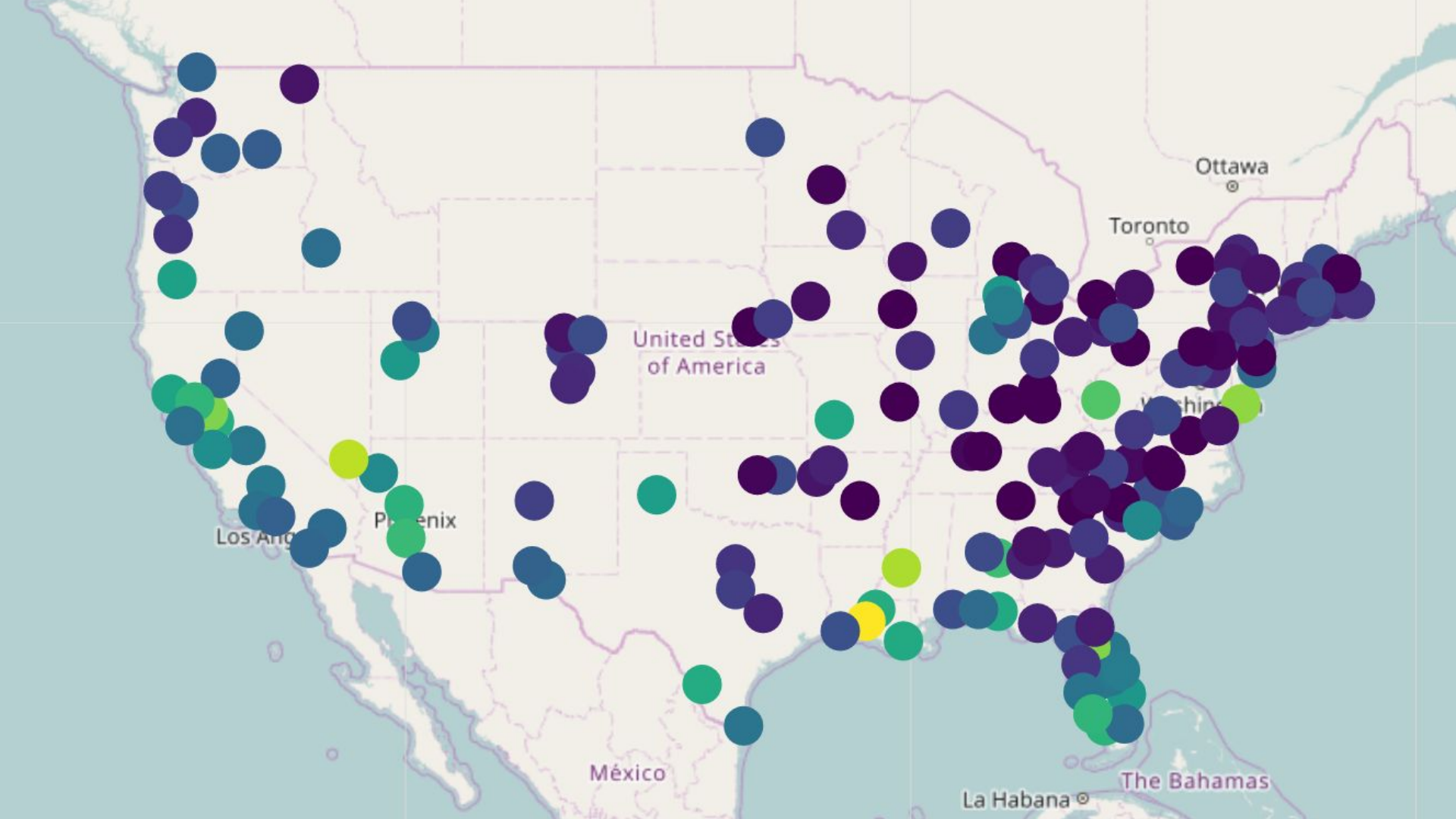
where α_i, θ_i are the parameters of the regression and ε_t are error terms.

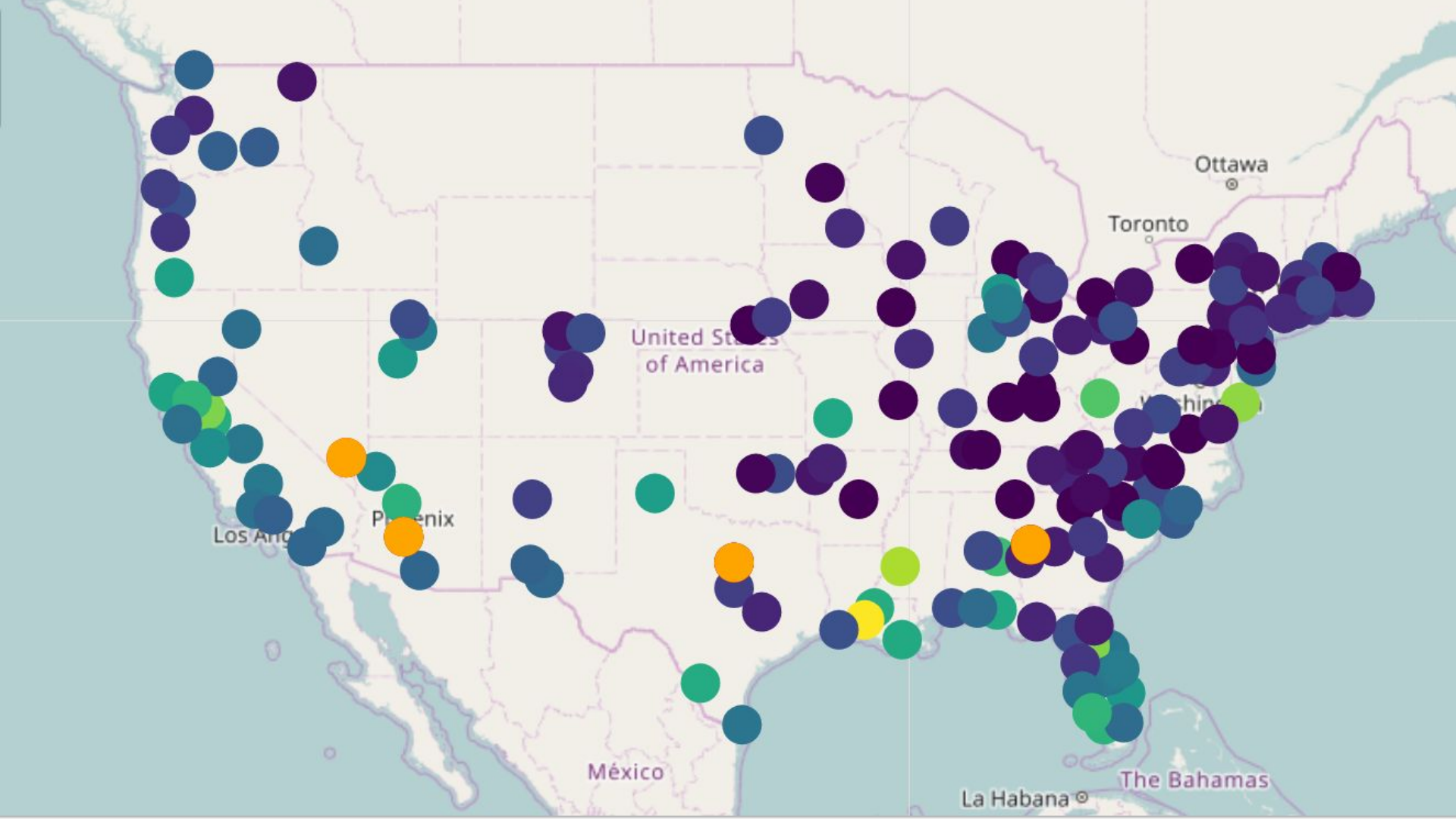
- We use second-order seasonally-adjusted data: tuning p,q in $ARIMA(p, 2, q)$
- Boost model by training a Random Forest Regressor on ARIMA residuals and exogenous data


Riverside
Denver
LosAngeles-LongBeach-Anaheim
NewYork
Cleveland
ColumbusOH
Miami-FortLauderdale
St.Louis
LasVegas
actual
predicted
20
0
−20
2001 2003 2005 2007 2009 2011 2013 2015

What markets should OpenDoor explore next?

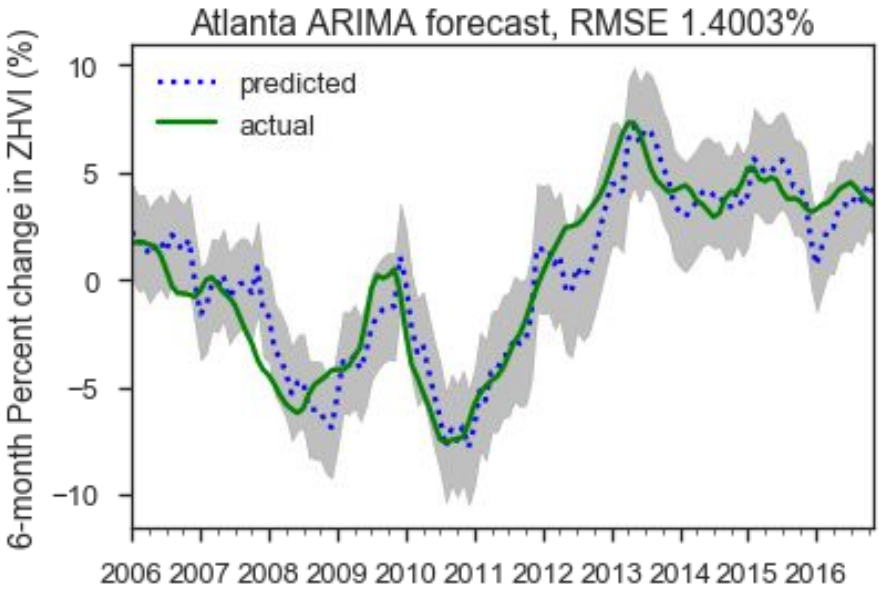


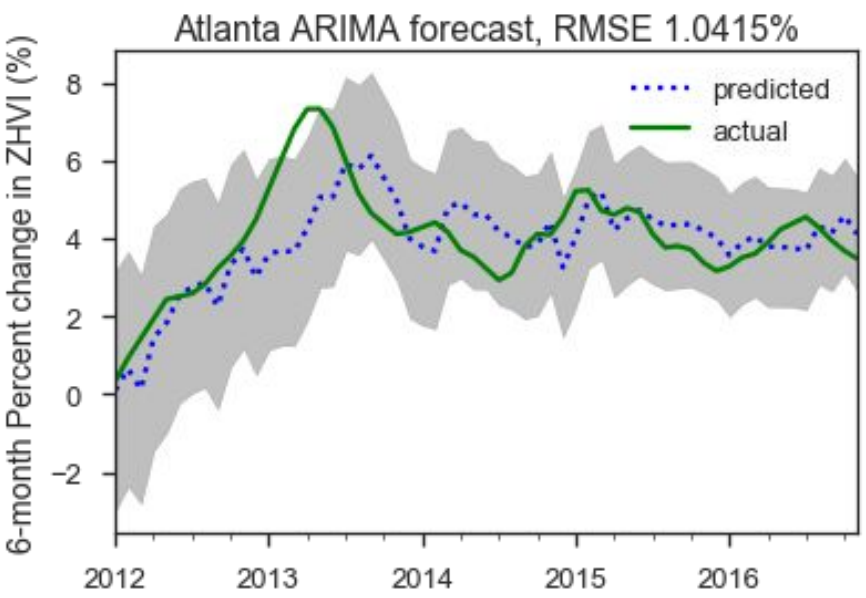

Ottawa
Toronto
United States of America
Phoenix
México
La Habana
The Bahamas


Ottawa
Toronto
United States
of America
Phoenix
Los Angeles
México
The Bahamas
La Habana

How does the Recession affect our predictions?

It's much easier to forecast post-recession… RMSE improves by 30%





Can we predict the next recession?

No. However, we can help Opendoor be prepared for the next recession:

- Predict increasing volatility
- Provide confidence intervals

Thanks!

Eric Munsing - Civil Engineering - e.munsing@berkeley.edu
Cristian Gavrus - Mathematics - cristian@berkeley.edu
Jeffrey Camps - LBNL - jeffcamp@lbl.gov

Additional notes for Opendoor: Key learnings

ARIMA models: surprisingly powerful

Housing market has long time lags- ARIMA(36,2,36) minimizes cross-validation error

Zillow market data (median number of days on market, rent cost, etc), provides significant improvement in prediction (~0.4%) but not available for recession backtesting

Smaller markets are generally less predictable, but there is significant variance in large markets (e.g. Las Vegas vs. Atlanta)

Additional notes for Opendoor: Future Steps

If we had more time:

- Collect and test more datasets to build a better second-stage regressor:
  - Mortgage rates
  - Earnings of Fortune 500 companies in each metro area
  - REIT and homebuilder stock prices
  - Weekly jobless claims
- Explore vector auto-regressive models using housing market data
- Explore recurrent neural networks instead of ARIMA predictions
  - Challenge: limited training data