A machine learning approach to House Price Indexes with the `hpiR` package

Andy Krause – Zillow Group Home Valuations

American Real Estate Society 2019 Meeting

Zillow

Outline

1. ‘hpiR’ package and data
2. Conceptual Model + Motivation
3. A Random Forest House Price Index
4. Performance Comparison
5. Discussion

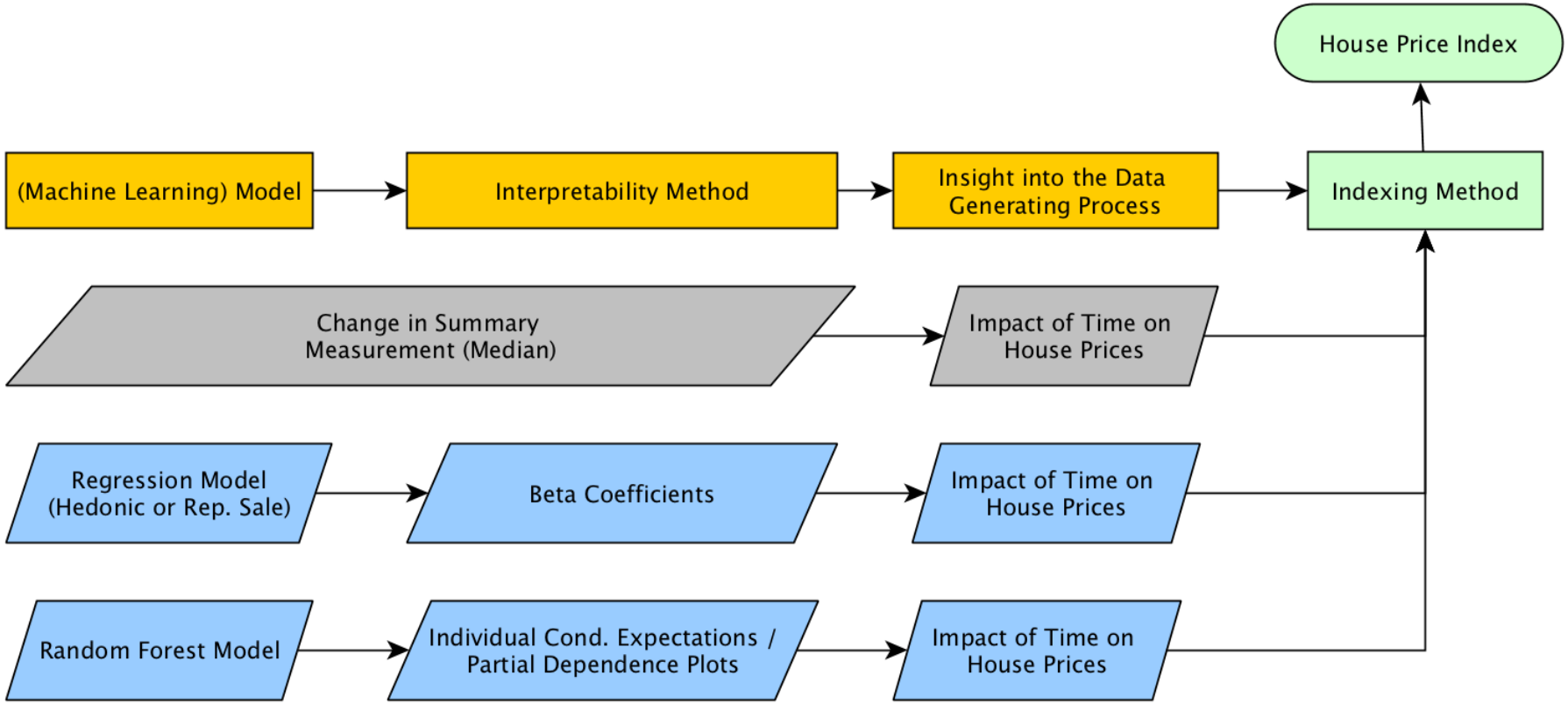
Zillow

‘hpiR’ package

- R package for house price indexes :https://cran.r-project.org/web/packages/hpiR/index.html
- Methods for generating ‘coefficients’ and converting to index
  - Hedonic + Repeat Sales
  - Random Forests (Github/development version only)
- Methods for index validation and comparison
  - Accuracy
  - Revision
  - Volatility
- Generic Functionality
  - Smoothing
  - Imputation
  - Series Creation
- Data Set
  - Seattle home sales from 2010 through 2016 (43k, 5k repeat)
  - Rich hedonic characteristics

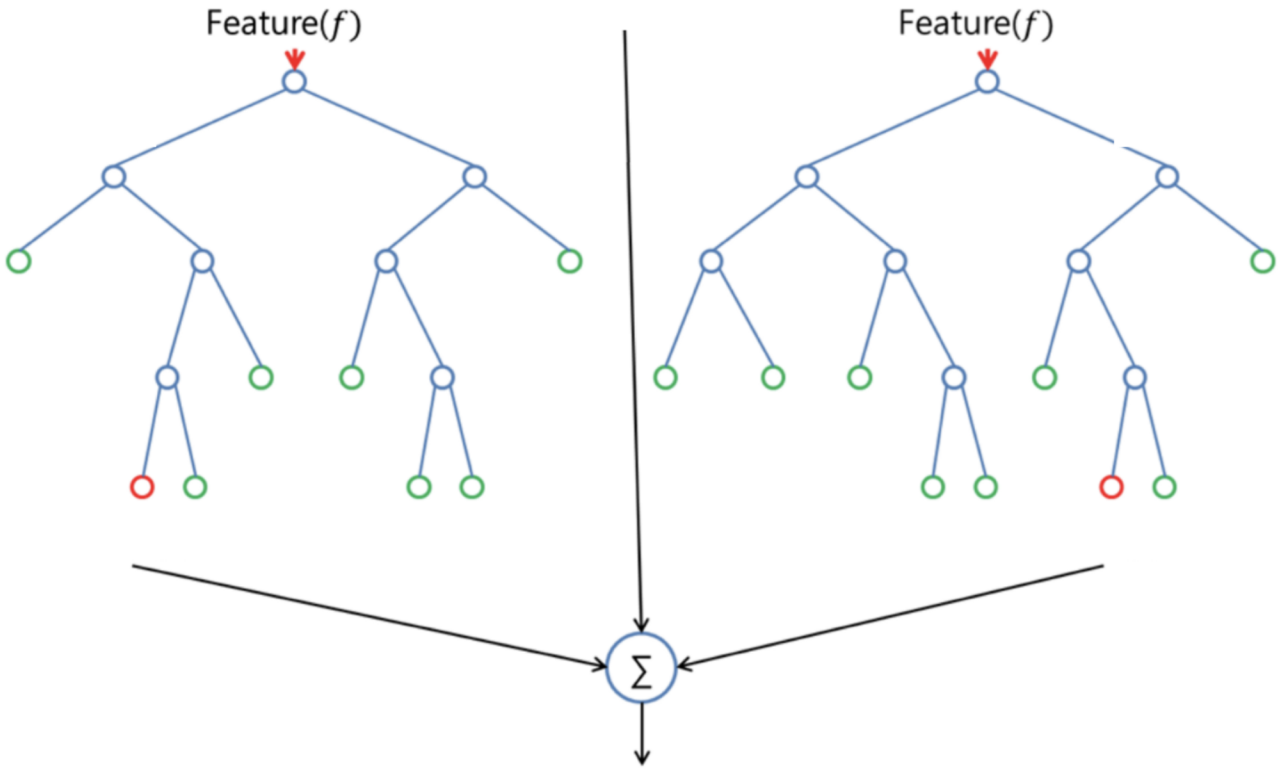
Zillow

Conceptual Model + Motivation



What is a Random Forest?

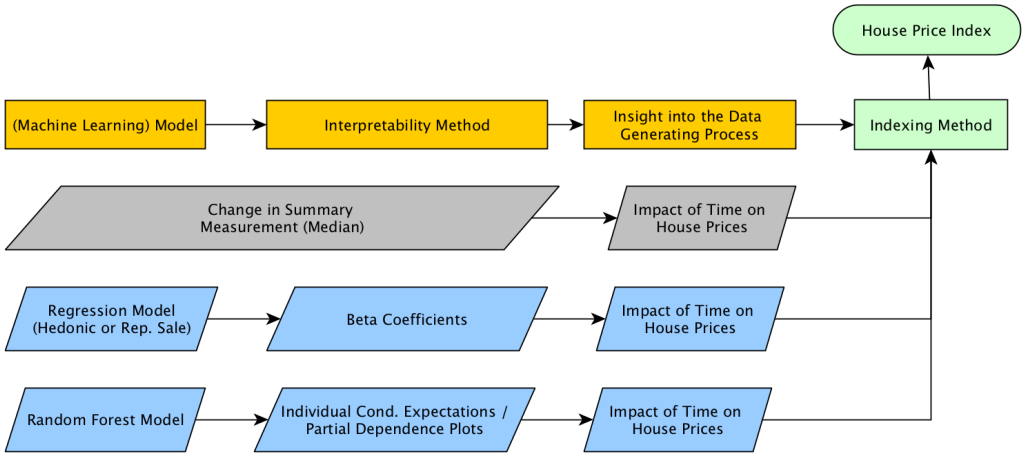
- Subsetting process
- Multiple decision trees trained on randomized sub-samples of data/features
  - Each split in tree is a single feature (e.g. Is Waterfront. Yes (left) / No (right)
- Prediction is (weighted) result of each tree



Zillow

A House Price Index via Random Forest

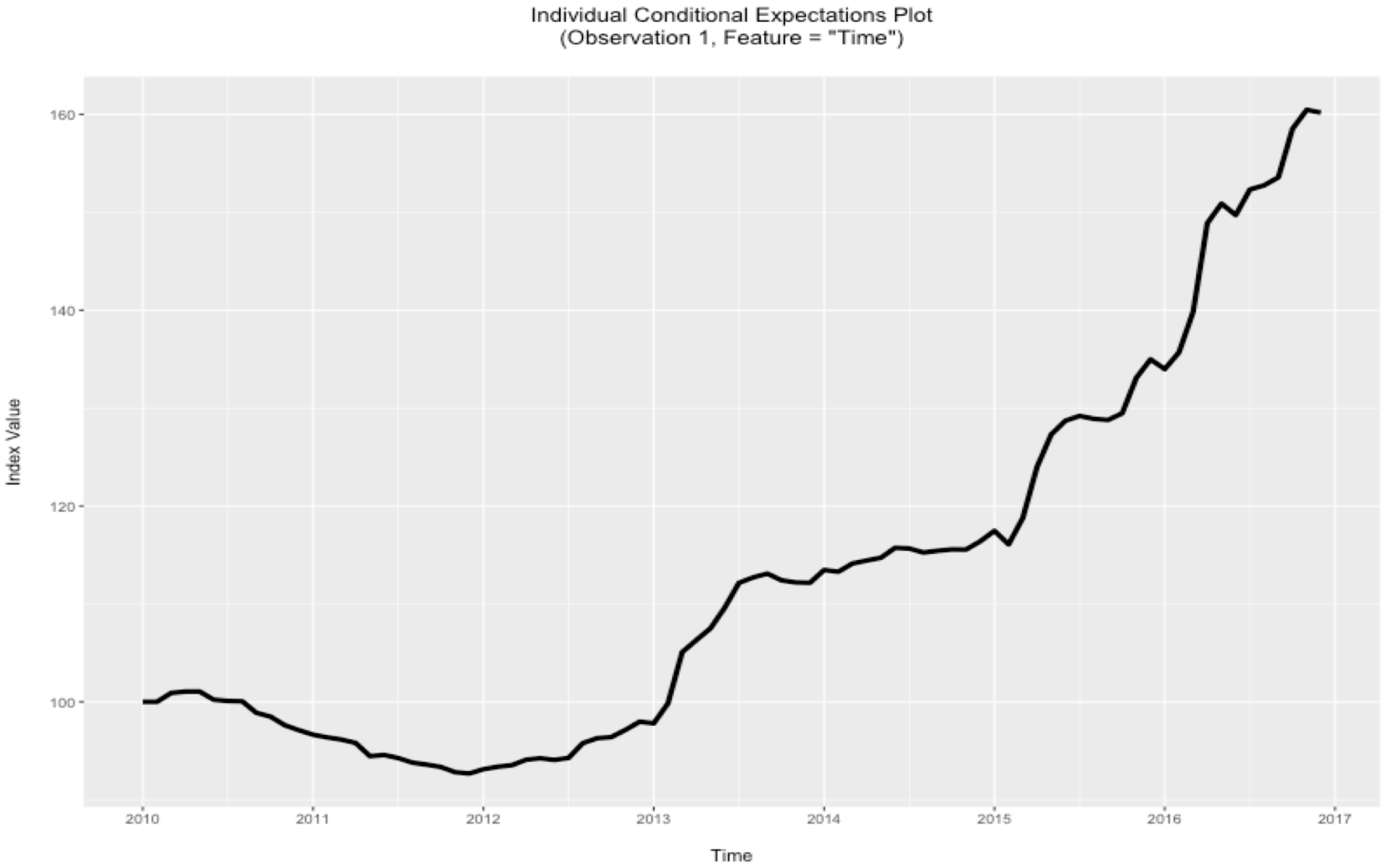
- House price indexes generation require an estimate of the impact of time on prices
- Random Forests do not produce readily interpretable coefficients
  - But we can simulate these with a Model Agnostic Interpretability Methods (Molner 2019)



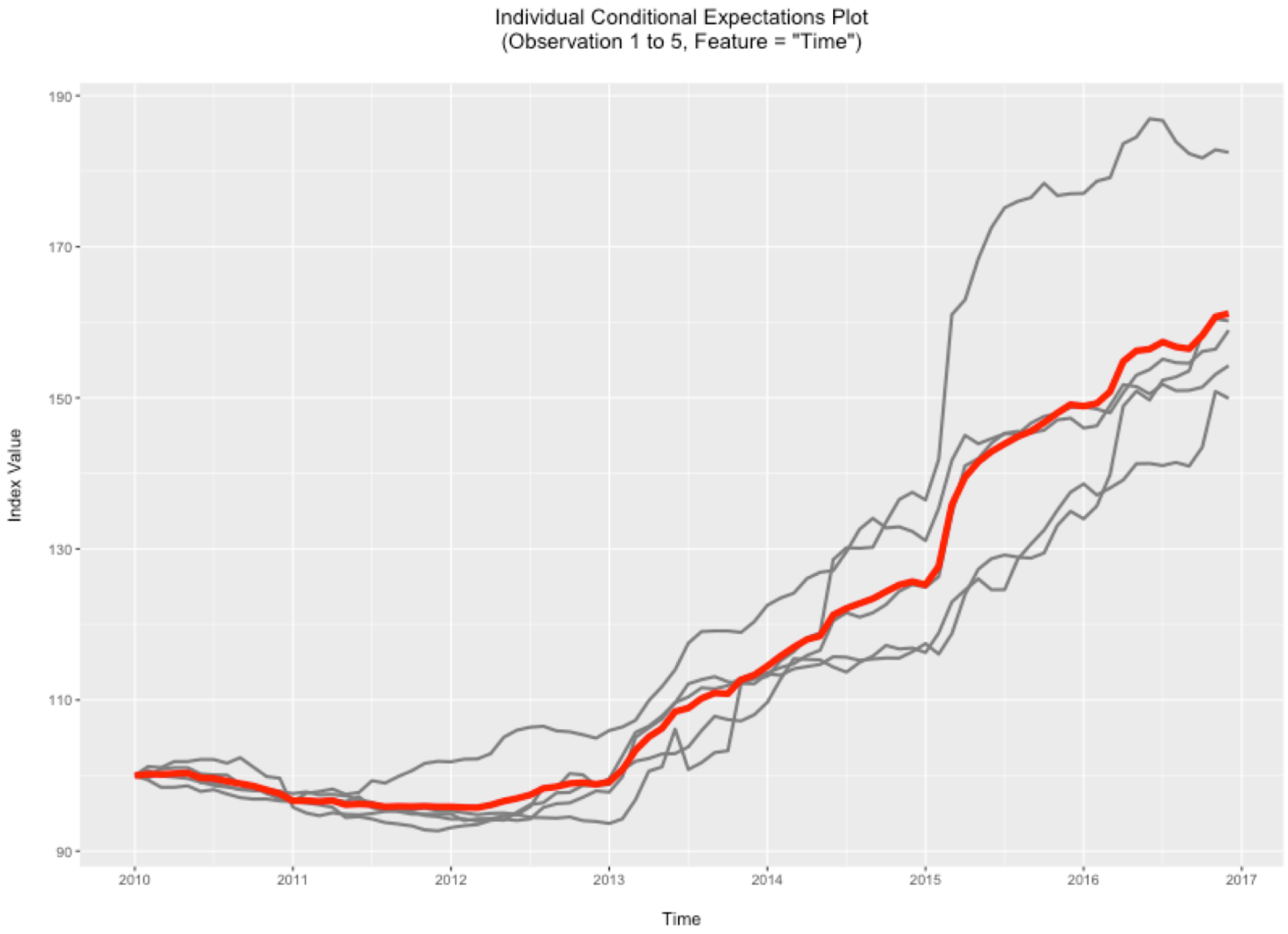
- Individual Conditional Expectations / Partial Dependence Plots
  - Build RF Model
  - (ICE) Value or simulate property Z with RF Model at period t, t+1, t+2, t+k
  - Convert this price trend to an index
  - (PDP) Apply the above to all instances/observations and average

Zillow

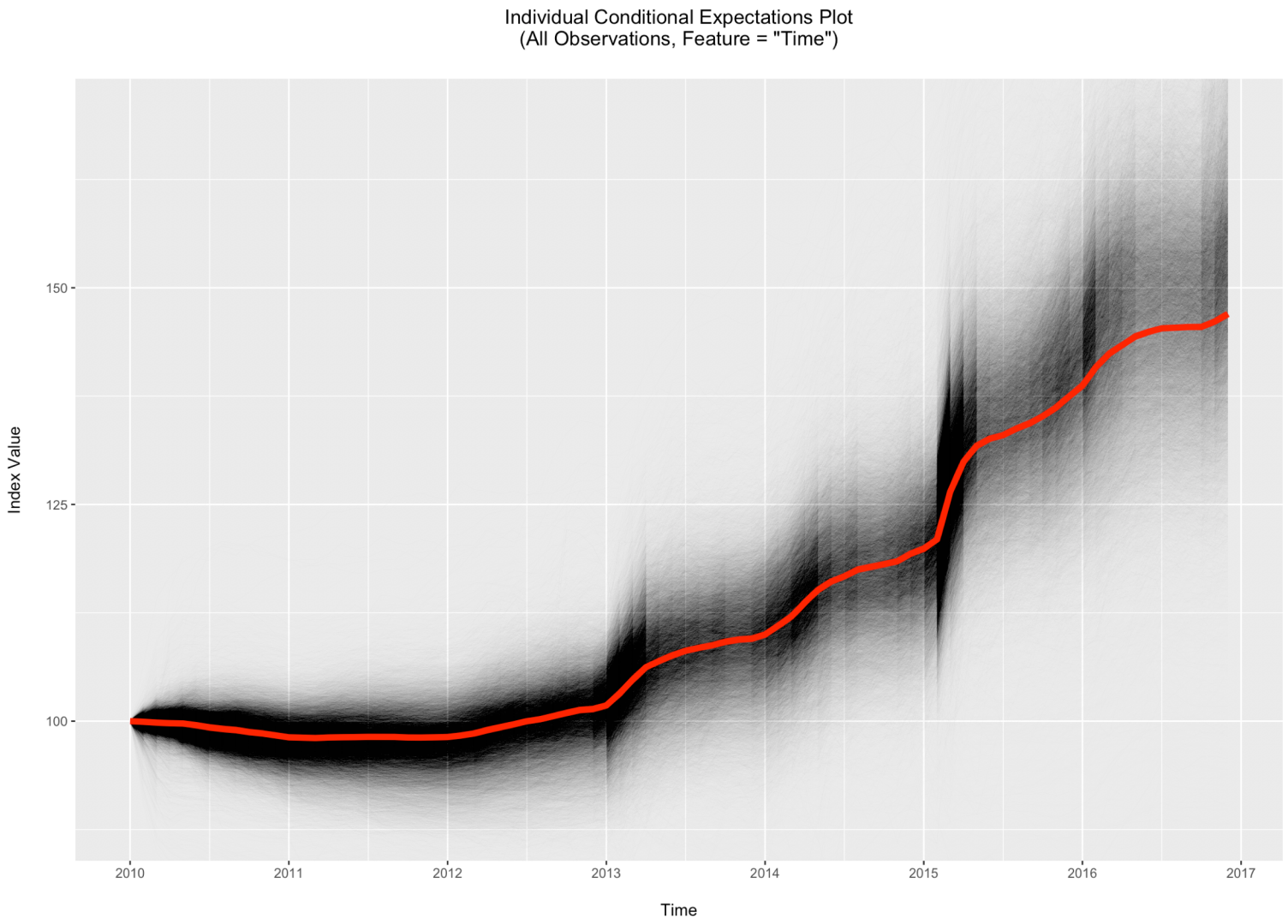
A Single ICE Example



5 ICEs + Average

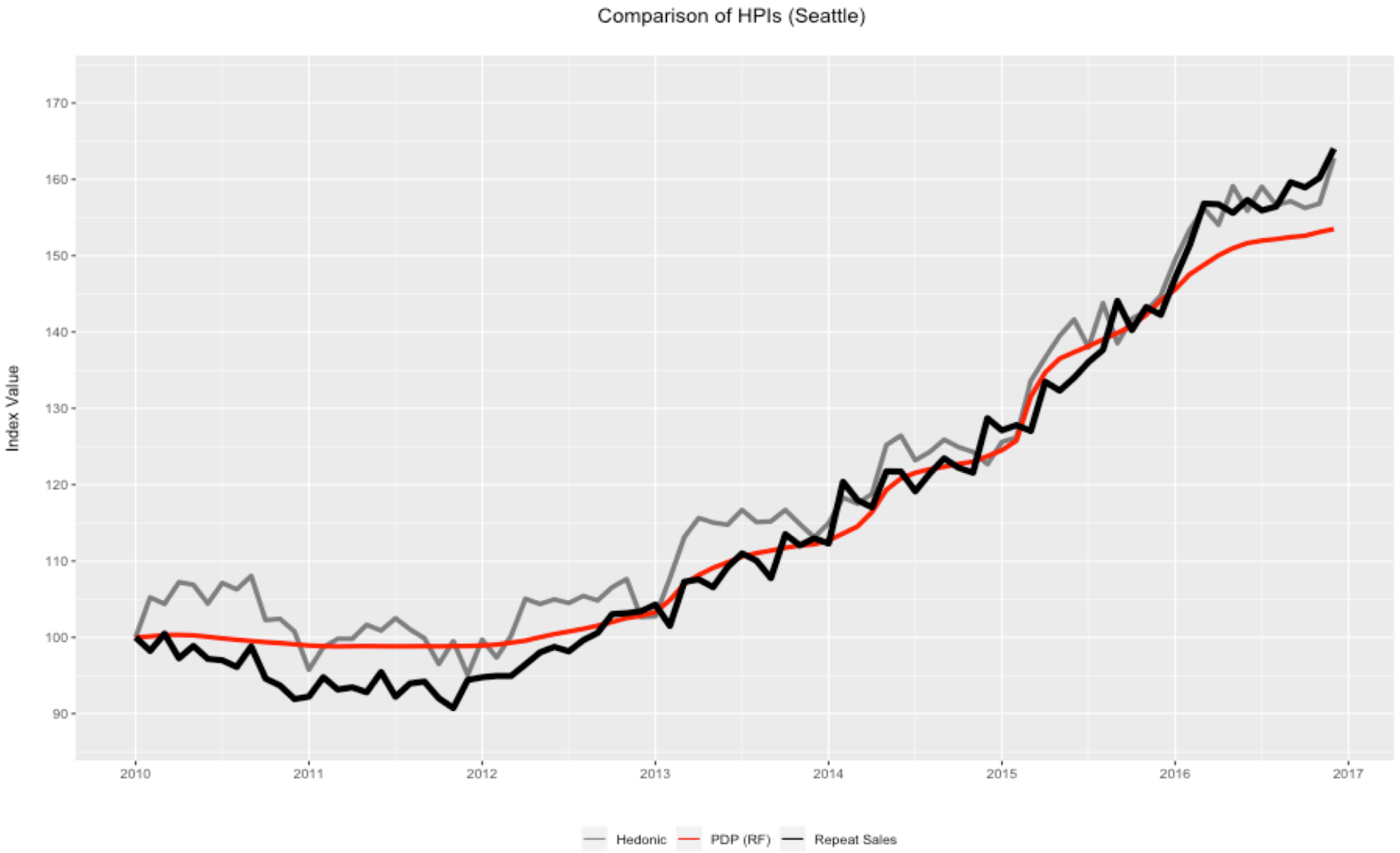


~43,000 ICEs -> PDP

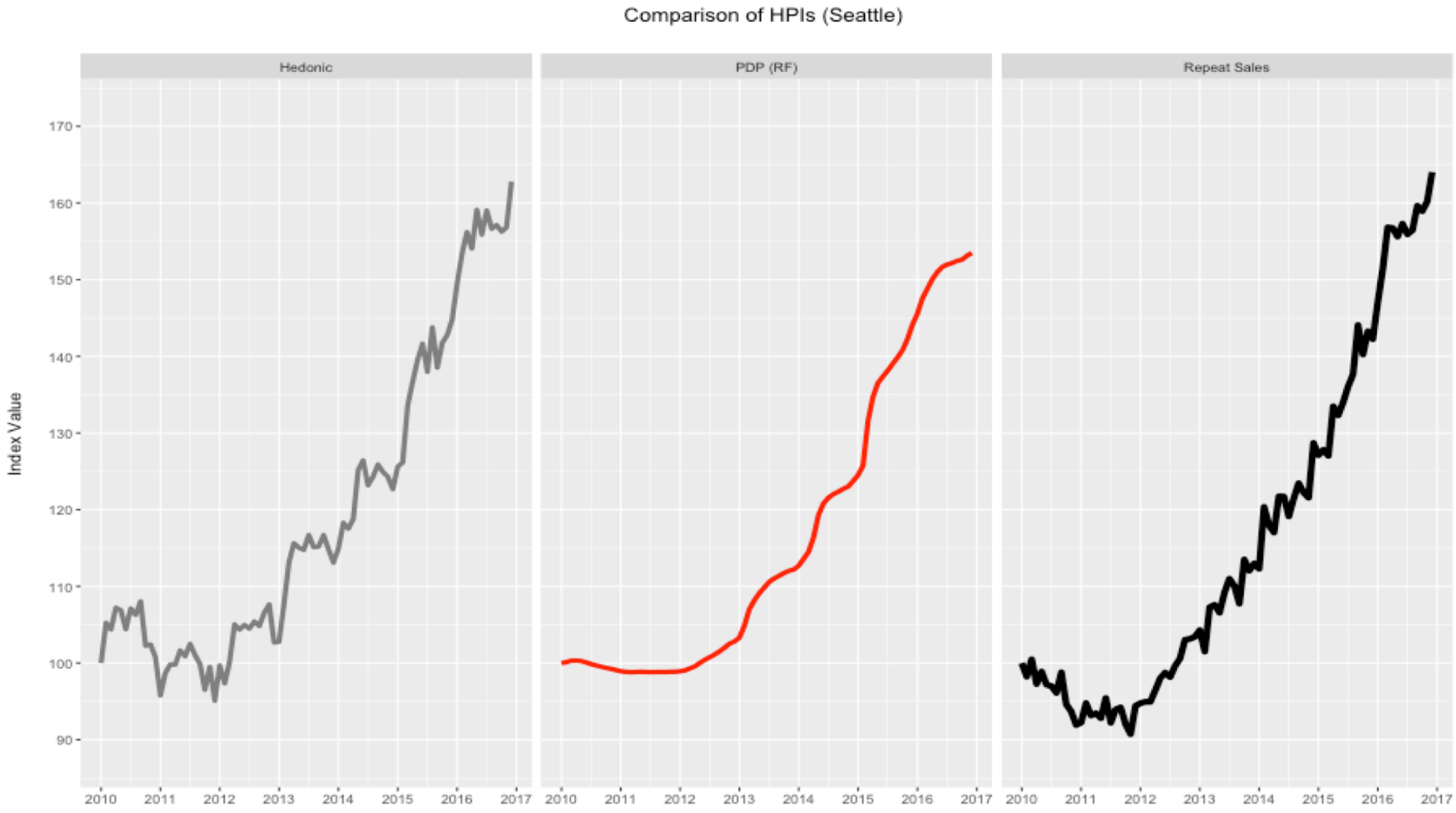


Zillow

Comparison


Comparison of HPIs (Seattle)
Index Value
170
160
150
140
130
120
110
100
90
2010
2011
2012
2013
2014
2015
2016
2017
Hedonic
PDP (RF)
Repeat Sales

Comparison


Comparison of HPIs (Seattle)
Hedonic
PDP (RF)
Repeat Sales
Index Value
170
160
150
140
130
120
110
100
90
2010
2011
2012
2013
2014
2015
2016
2017

Comparison to Existing Methods

- Looks competitive (and certainly smoother), but is it better?
- ‘hpiR’ allows comparison by:
- 1) Volatility – Standard Deviation of three month moving average
- 2) Revision – Level of change in initial index values as index series grown over time
- 3) Accuracy – How well can we predict second transaction in a repeat sales pair?
  - In Sample – How well does it fit data?
  - Out of Sample – How well does it prediction in a K-Fold situation?
  - Future Prediction – How well does it predict at t+1?
    - `forecast` package, simple exponential smoothing

Comparison to Existing Methods

- Compute index and series for full data set (~43k transactions, ~5k repeat)
- Full 2010 through 2016 time period (84 months)
- Robust RS and Hedonic Models

Method	Volatility	Revision	In Sample Accuracy	K-Fold Accuracy	Prediction Accuracy
Repeat Sales	.020	.398	.094	.096	.117
Hedonic	.021	**.010**	.095	.097	**.095**
RF – PDP	**.001**	.062	**.093**	**.093**	.100

Zillow

Small Geographic Area Comparison

- 25 King County Assessment Zones
- ~ 2000 sales in each over the 7 year period
- ~ 250 repeat transactions in each

Method	Volatility	Revision	In Sample Accuracy	K-Fold Accuracy	Prediction Accuracy
Repeat Sales	.119	.086	.123	.201	.217
Hedonic	.061	**.013**	.101	.105	**.099**
RF – PDP	**.002**	.077	**.093**	**.095**	.110

Zillow

Takeaways

- Random Forest with Interpretability
  - Smoothest
  - Most accurate In and Out of sample
  - Intuitive explanation (value property X as if sold repeatedly)
  - Computationally expensive
- Repeat Sales
  - High revision + worse accuracy
  - Faster + least data intensive
- Hedonic model
  - Least Revision, Best Prediction accuracy
- ML interpretability models offer some competitive performance
  - At the expense of compute time
  - Struggle at the ‘leading end’ of the index
  - Only tried one model type (RF) and one interpretability method (PDP)
- ‘hpiR’ offer quick way to test a model/method
  - Dataset, benchmark and a framework for comparison