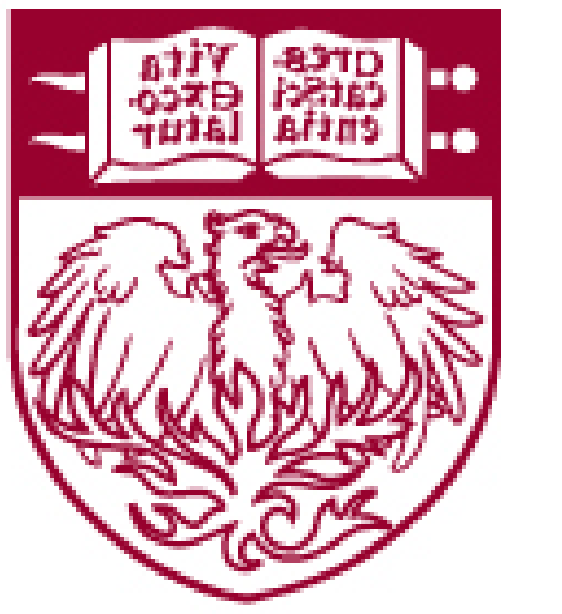# Housing Demand Heat Index: Predicting Chicago Real Estate Market Absorption

James Foster, Susan Parker, Yuanyuan Zha, Vadim Sokolov
Master of Science in Analytics
The University of Chicago Graham School, Chicago, IL

Abstract

Opportunities to support urban economic decision-making with analytical models are extensive in the real estate market. Both buyers and sellers face uncertainty in real estate transactions in large metropolitan areas about when to time a transaction and at what cost. A housing demand index based on microscopic home showings events data can provide decision-making support for buyers and sellers on a very granular time and spatial scale. In the current real estate market, both buyers and sellers make decisions without knowing the present and future state of the large and dynamic real estate market. Consequently, accurate and granular housing market demand forecasts play a valuable role in these decisions. In this paper, we aim to predict housing market demand by developing housing demand indices using high-volume, high-velocity data on home showings, listing events, and historic sales data. By employing a combination of traditional market measures supplemented by the number of home showings, the indices result in timely insight into housing market demand. We demonstrate our analysis using data from seven million individual records sourced from a unique, proprietary dataset that has not previously been explored in application to the real estate market. We then employ a series of predictive models to estimate current and forecast future housing demand. Specifically, we first develop a shorter-term market demand heat index that predicts housing demand for the subsequent week using only past weekly market demand and home showings data. We then develop a longer-term market demand index to assess the ability of home showings and past housing market weekly data to predict the subsequent five weeks of housing market demand. A housing demand index provides insight into the level of demand for a home on the market and to what extent current demand represents future expectation. As a result, these indices provide decision-making support into important questions about when to sell or buy, or the elasticity present in the housing demand market, which impact price negotiations, price-taking and price-setting expectations. This forecast is especially valuable because it helps buyers and sellers to know on a granular and timely basis if they should engage in a home transaction or adjust their home price both in current and future states based on our forecasted housing demand index.

Objectives

- Create a housing market indicator that estimates the demand for homes in the real estate market with ShowingTime's unique real estate showing data
- Through a range of supervised and unsupervised analytical techniques, the team developed a weekly heat index and a longer-term housing demand index that accurately predicts the ratio of homes sold to market availability in Chicago from 2011-2013

Reference
1: Shindler, 2011, "Predictability and Persistence of the Price Movements of the S&P/Case-Shiller House Price Indices."
2: Nagaraja, Brown, Zhao. 2009. "An Autoregressive Approach to House Price Modeling."
3: Meese, Wallace, 1997.
4: Case, Shiller. 2004. "Is there A Bubble in the Housing Market?"
5: Case, Shiller, 1989. "The Efficiency of the Market for Single Family Homes."
6: Genesove, Miller, 2000. "Loss Aversion and Seller Behavior: Evidence from the Housing Market."

Background

Business background:
- ShowingTime is a privately-held, Chicago-based company that launched in 1999 with rapid sales growth
- Products and services are used in 53,000+ real estate offices across more than 200 MLSs by 500,000+ agents from 42 states in US
- Involved in managing more than 2.5 million showings every month
- Main service is scheduling of home showings
- Document all the activities associated with home sold process such as open house, showings, price changes, inspection, in contract, sold, etc.
- Collect agents/buyer feedbacks after the showings

Data structure
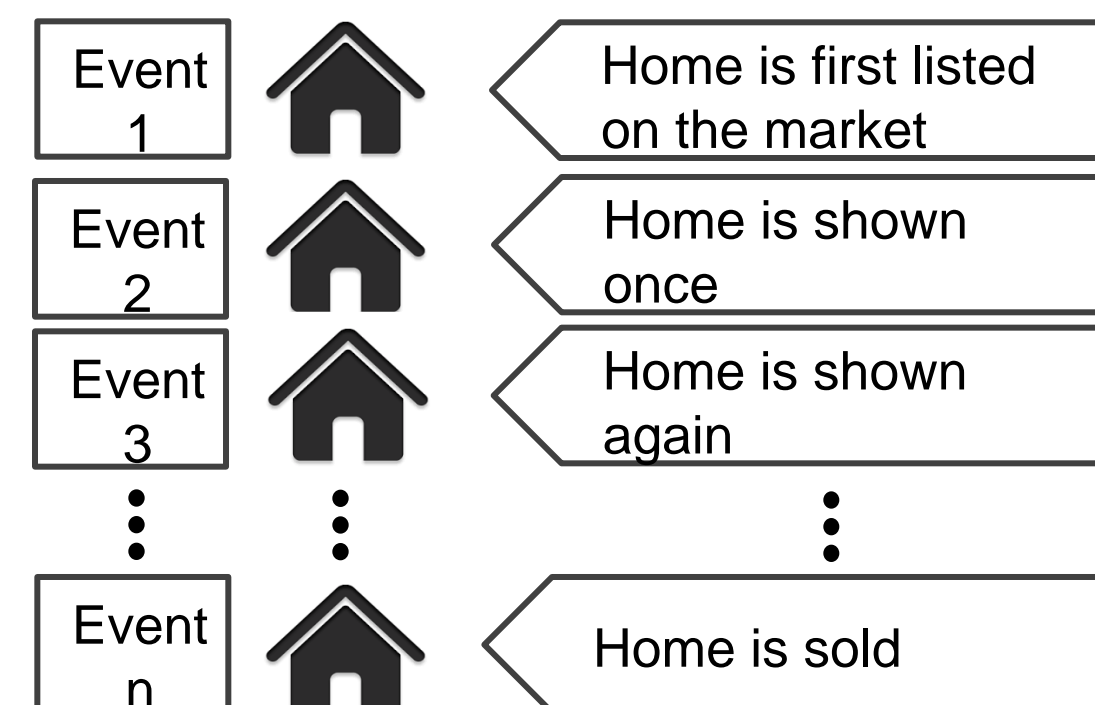
With granular information over time about a single home, ShowingTime's data could provide insights valuable to stakeholders across the real estate market.

Data Preparation

- Join tables to form a single dataset of home showings. The combined dataset contains 6,042,315 property action records including showings, inspections, open house etc. from 2011 to 2013
- Aggregate data and create new variables: weekly number of showings, weekly number of sold properties, weekly total number of properties on the market, mean and median time on the market for the properties sold each week

Table: Excerpt of the prepared weekly data

Year-Week	# of showings	# of sold	# of properties	Mean time	Median time
2011-01	208	2	15402	56.5	56.5
2011-02	11672	58	15850	156.8	122
2011-03	13250	82	16153	135.8	114.5
2011-04	13732	87	16410	159.8	127
2011-05	12978	153	16637	169.1	134

Variable Development I

- Dependent variable:
  - Demand Index = $\dfrac{\text{weekly home sold}}{\text{weekly total porperties on the market}}$
- Independent variables:
  - Time on the market
  - Weekly ratio of # of showings to total # of properties
  - Weekly ratio of # of sold to total # of properties
  - Weekly ratio of # of sold to # of showings

Variable Development II

- Lags up to 10 weeks were introduced for the Independent variables
- All independent variables had at least one week lag time
- First 9 weeks of the data set was incomplete because of introduction of lags and were eliminated from analysis
- All data were coded to the range -1 to +1 for the analysis
- Split data into 80% training set and 20% testing/validation set

Table: Excerpt of the lagged weekly data

Year-Week	Ratio Sold to showings lag 1	Ratio Sold to showings lag 2	Ratio Sold to showings lag 3	Ratio Sold to showings lag 4	Ratio Sold to showings lag 5
2011-01					
2011-02	0.009615				
2011-03	0.004969	0.009615			
2011-04	0.006189	0.004969	0.009615		
2011-05	0.006336	0.006189	0.004969	0.009615	
2011-06	0.011789	0.006336	0.006189	0.004969	0.009615
2011-07	0.013564	0.011789	0.006336	0.006189	0.004969

Baseline Data Characteristics

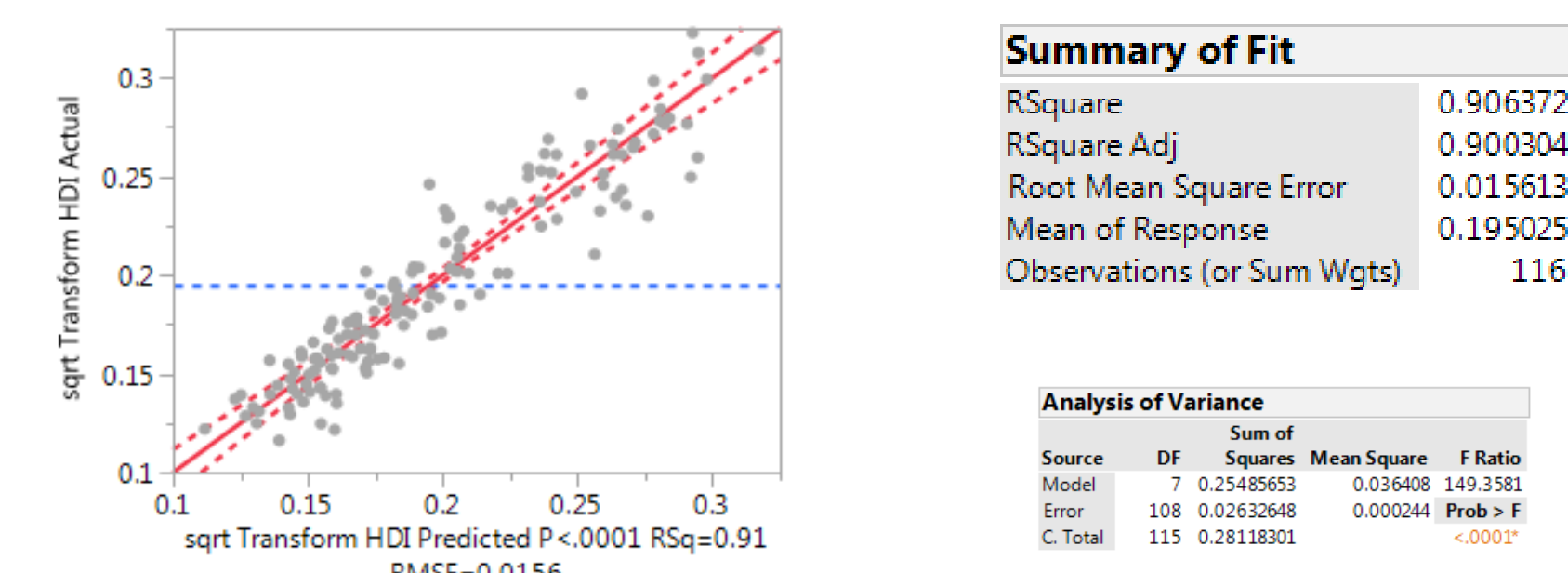As the number of showings increases, the time on the market decreases
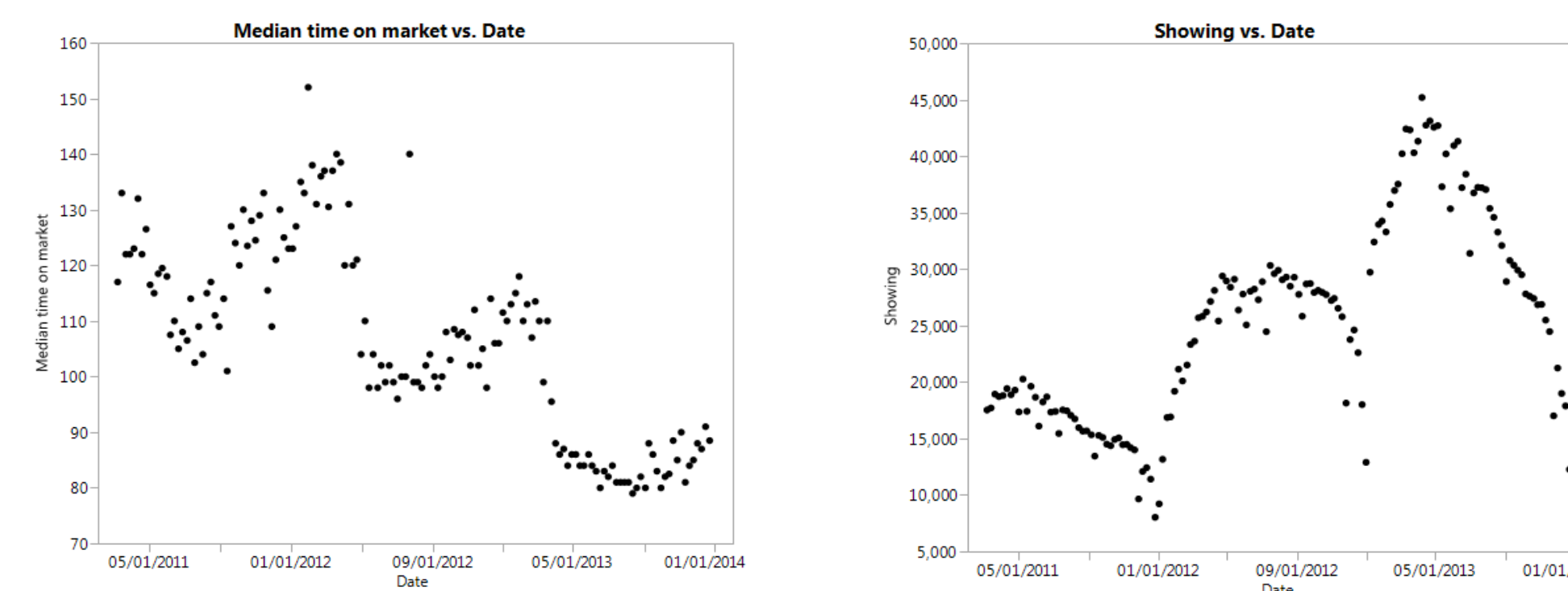
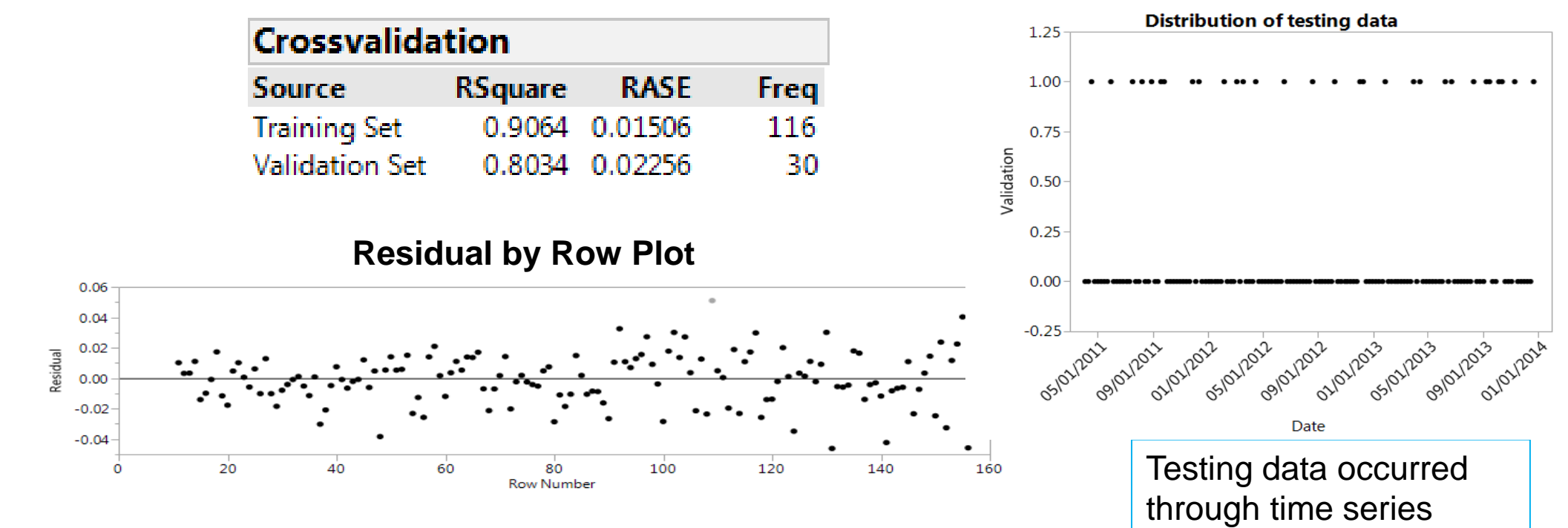The derived housing demand index has a similar trend to the Case Schiller index
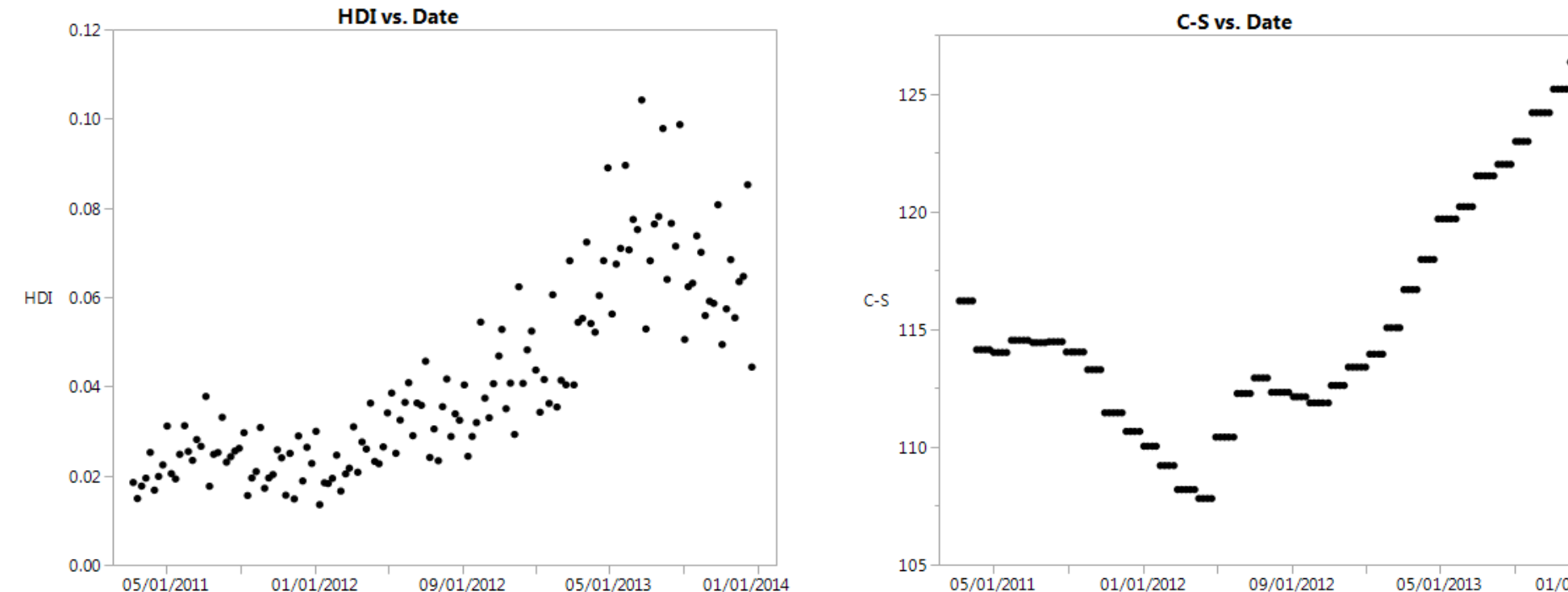
Model Development I – Linear Model

Modelling head index (ratio of sold to property on the market) using stepwise forward regression

Summary of Fit

RSquare	0.906372
RSquare Adj	0.900304
Root Mean Square Error	0.015613
Mean of Response	0.195025
Observations (or Sum Wgts)	116

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	7	0.2548563	0.036408	149.3581
Error	108	0.0263268	0.000244	Prob > F
C. Total	115	0.2811831		<.0001

Model Validation I – Linear Model

Linear model performs well on test data

Crossvalidation

Source	RSquare	RASE	Freq
Training Set	0.9064	0.01506	116
Validation Set	0.8034	0.02256	30

Residual by Row Plot

Distribution of testing data

Testing data occurred through time series

Model Development II – CART tree

Model Validation II – CART tree

Column Contributions

	RSquare	RMSE	Number N of Splits	AICc
Training	0.859	0.0185056	117 5	-586.53
Validation	0.702	0.0277056	30	

Model relatively accurate on test data set

Final Model

0.15 * Linear regression + 0.05 * CART + 0.8*Neural Net

Measures of Fit for sqrt Transform HDI

Predictor	Creator	.2.4.6.8	RSquare	RASE	AAE	Freq
Ensemble 2			0.8710	0.0183	0.0141	30

Residual by Row Plot
Residual by Row Plot for sqrt Transform HDI

Conclusion & Future Directions

- The house demand heat index predicts the next two weeks of housing market demand with 87% accuracy
- Next steps include validating the model using other years of real estate data and different geographic locations
- Further develop the model to improve the prediction accuracy for longer time frame