SAS® GLOBAL FORUM 2016

IMAGINE. CREATE. INNOVATE.

Predicting Current Market Value of a Housing Unit across the Four Census Regions of the United States

#SASGF

Predicting Current Market Value of a Housing Unit across the Four Census Regions of the United States

Mostakim Tanjil and Goutam Chakraborty

Oklahoma State University

INTRODUCTION

- Housing bubble, started in early 2006, affected over half of the American states.
- Overvaluation of housing units resulted in increased foreclosures and credit crisis, leading to high and prolonged unemployment rate.
- Housing market directly affected mortgage market, home builders, real estates, investment banks, home supply retail outlets etc.
- The aftermath of the 2007-2008 recession has led to a large reservoir of potential housing demand.
- Housing price index has increased substantially over the years, reported by Federal Housing Finance agency.

PURPOSE

- To predict current market value (MV) of a housing unit across the five metropolitan statistical areas (MSA) in the four census regions of the United States.
- To determine factors that affect current MV of a housing unit.

METHODS

Data collection and preparation

- Data was collected from The Housing Affordability Data System of the US Department of Housing and Urban Development.
- Main data source – American Housing survey conducted in 2013.
- 36,675 observations and 20 variables.
- Input variables – metro3, region, built, zadeq (adequacy, 1 – 4), structure type, lmed (median income), l80 (low income limit), ipov (poverty income), aplmed (median income adjusted for number of person), fmr (fair market rent), bdrms (number of bedrooms), nunits (number of units), rooms (total rooms), utility (monthly utility cost), other cost, burden, zinc2 (total household income), and target variable – value
- Four variables had missing values – imputed by tree method
- Average of MV increases with number of bedrooms, except for efficiency and one-bed room units(Fig 1).
- Insurance and other costs are higher for efficiency and one-bed room units

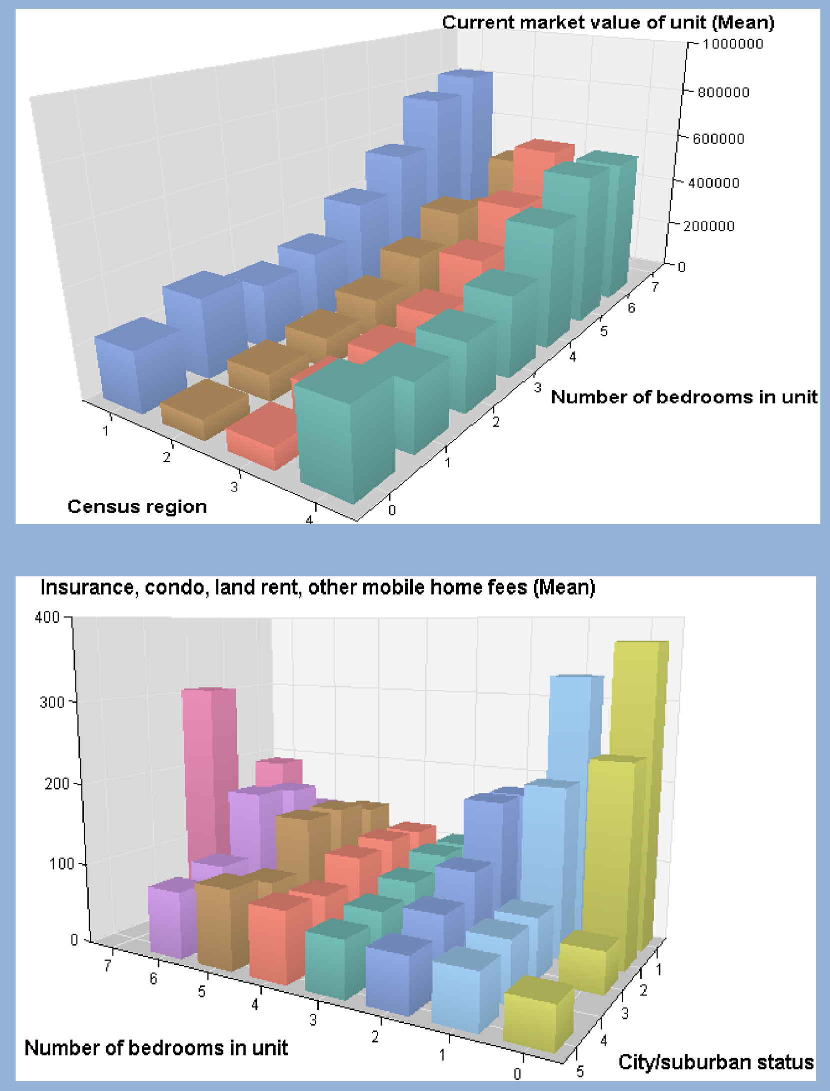

Fig 1. (top) average current MV, and (bottom) other costs.

Model building

- Data were split into – 40% training, 30% validation and 30% test.
- Six variables were transformed to obtain symmetric distribution.
- Multivariate outlier detection was conducted – one observation was considered outlier, left unchanged.
- Categorical variable built was consolidated from 29 to 5 levels – decision tree method.
- Reduction of input variables
  - LARS, LASSO, Adaptive LASSO, Variable Selection, Stepwise Regression, Clustering, PCA with numeric variables and PCA with all variables
- Average Squared Error (ASE) was primary selection criteria. It is computed using the following formula:

$$\text{ASE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2$$

Where $\hat{Y}$ is the prediction and Y_i is the observed value.

- Following models were built
  - Decision tree – different number of branches and depth
  - Neural network – Multilayer perceptron, ordinary radial, normalized radial and generalized linear
  - Polynomial regression – Two factor interaction with polynomial degree 3
  - Partial Least Square regression – NIPALS, SVD, Eigenvalue and RLGW algorithm
  - Gradient boosting – Square error and Huber M-regression loss function
  - Memory based reasoning – with Principal component analysis

RESULTS

- Neural network passed through adaptive LASSO turned out to be the best model.
- Selected input variables for the best model:
  - zadeq (adequacy, 1 – 4), structure type, utility, ipov (poverty income), rooms, built, region, zinc2 (total household income), metro3, other cost, burden, l80 (low income limit), aplmed (median income adjusted for number of person), fmr, bdrms.

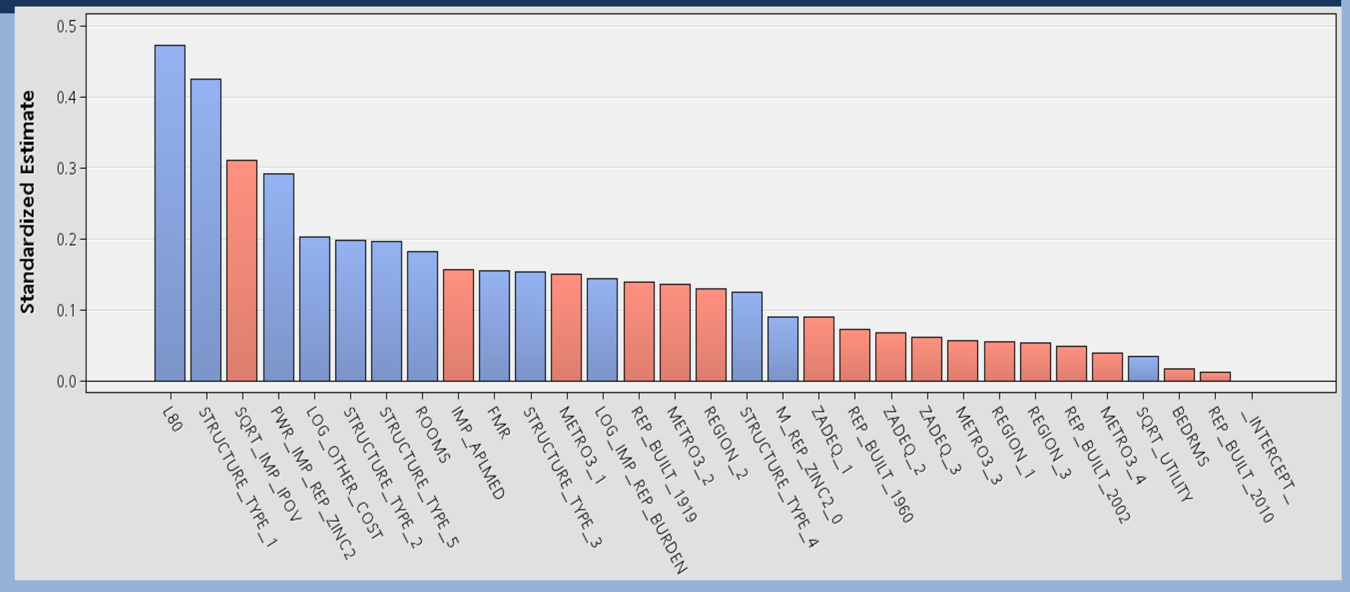

Fig 2. Parameter estimates (absolute values) of adaptive LASSO

Predicting Current Market Value of a Housing Unit across the Four Census Regions of the United States

Mostakim Tanjil and Goutam Chakraborty

Oklahoma State University

Model	Validation ASE	Test ASE
Neural Network (Stepwise Regression)	0.311388	0.322367
Neural Network (Adaptive LASSO)	0.311388	0.322367
Polynomial Regression	0.348389	0.351355
Partial Least Square (PLS)	0.354991	0.360689
Decision Tree	0.385548	0.394083
MBR (all variables)	0.401272	0.410471
MBR (numeric variables)	0.423757	0.44369
Gradient Boosting	0.423372	0.445096

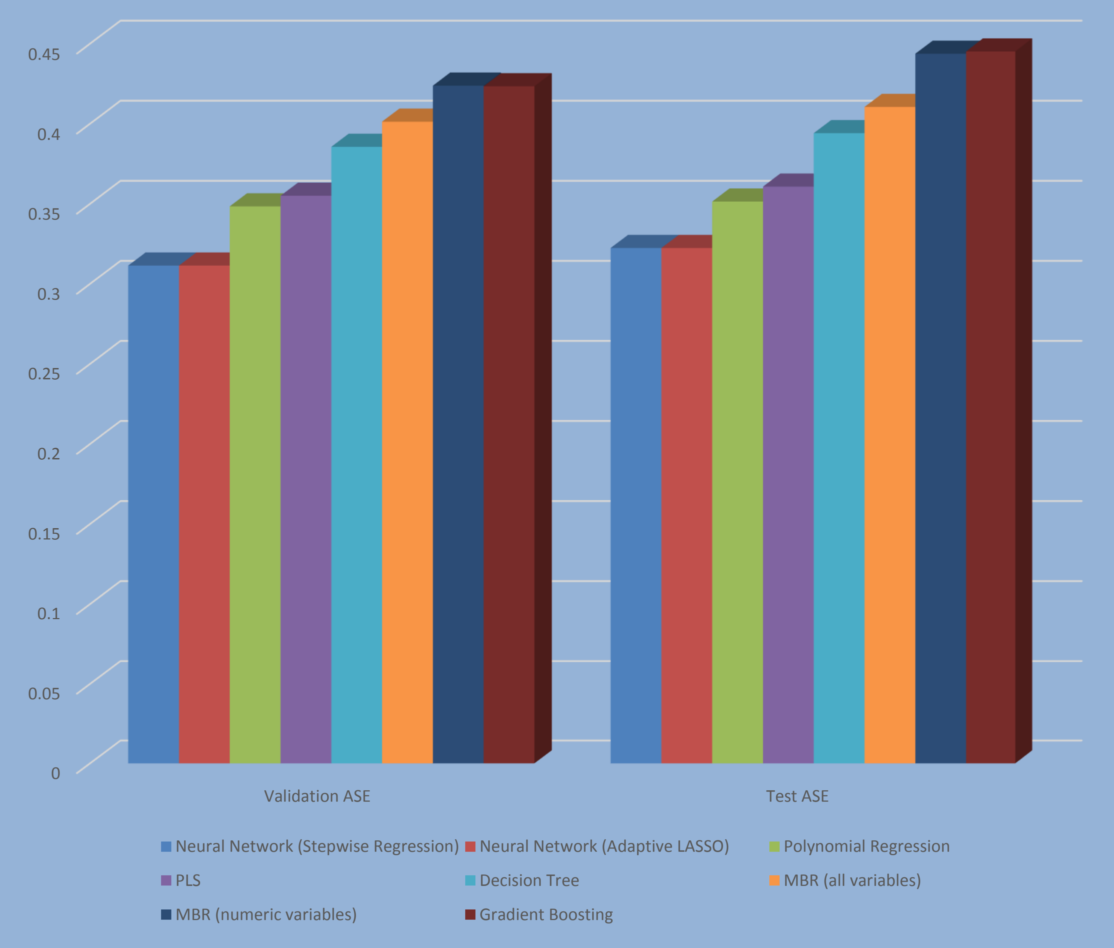

Fig 3. Comparison of top eight models

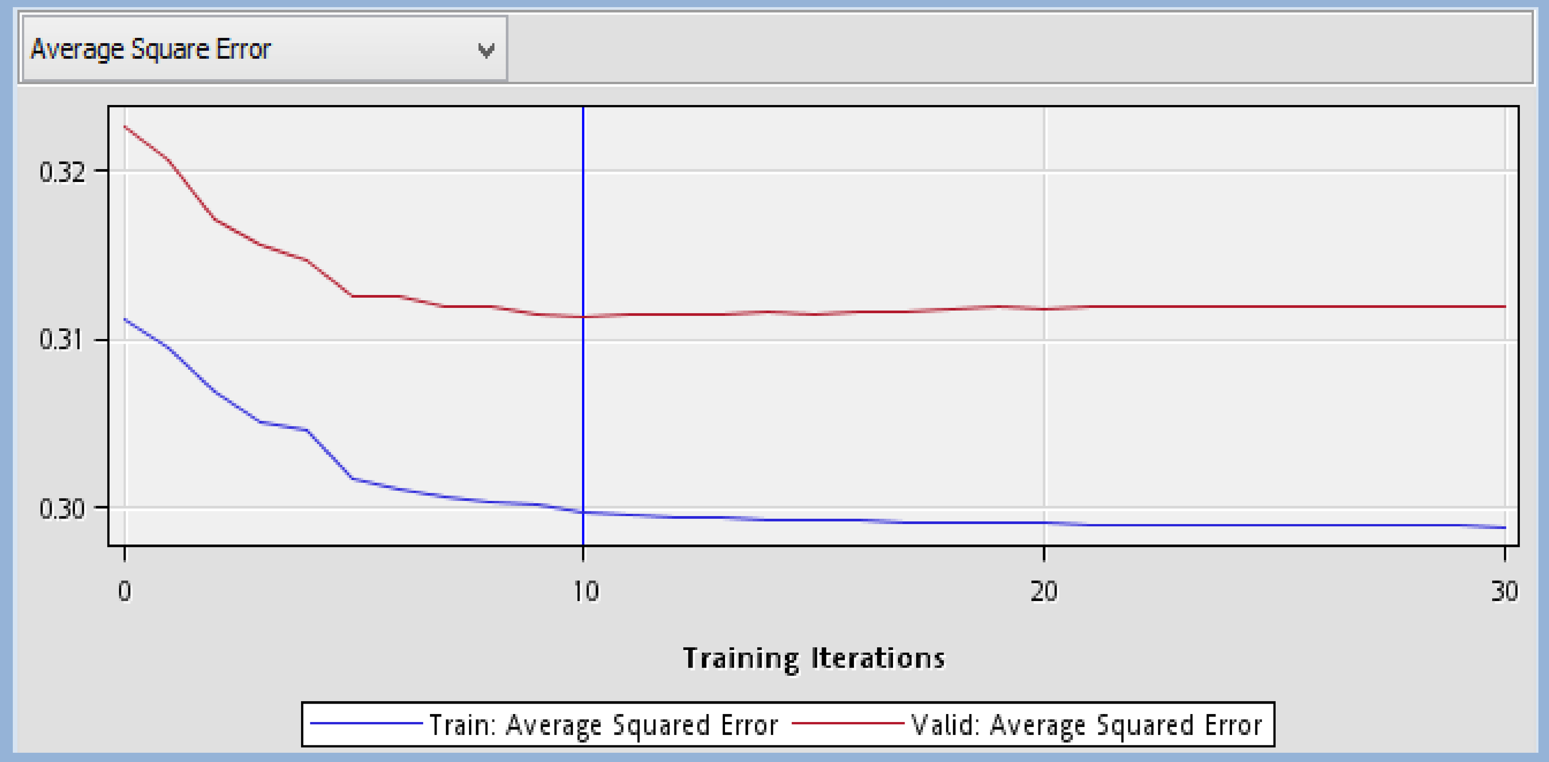

Fig 4. Iteration plot for neural network

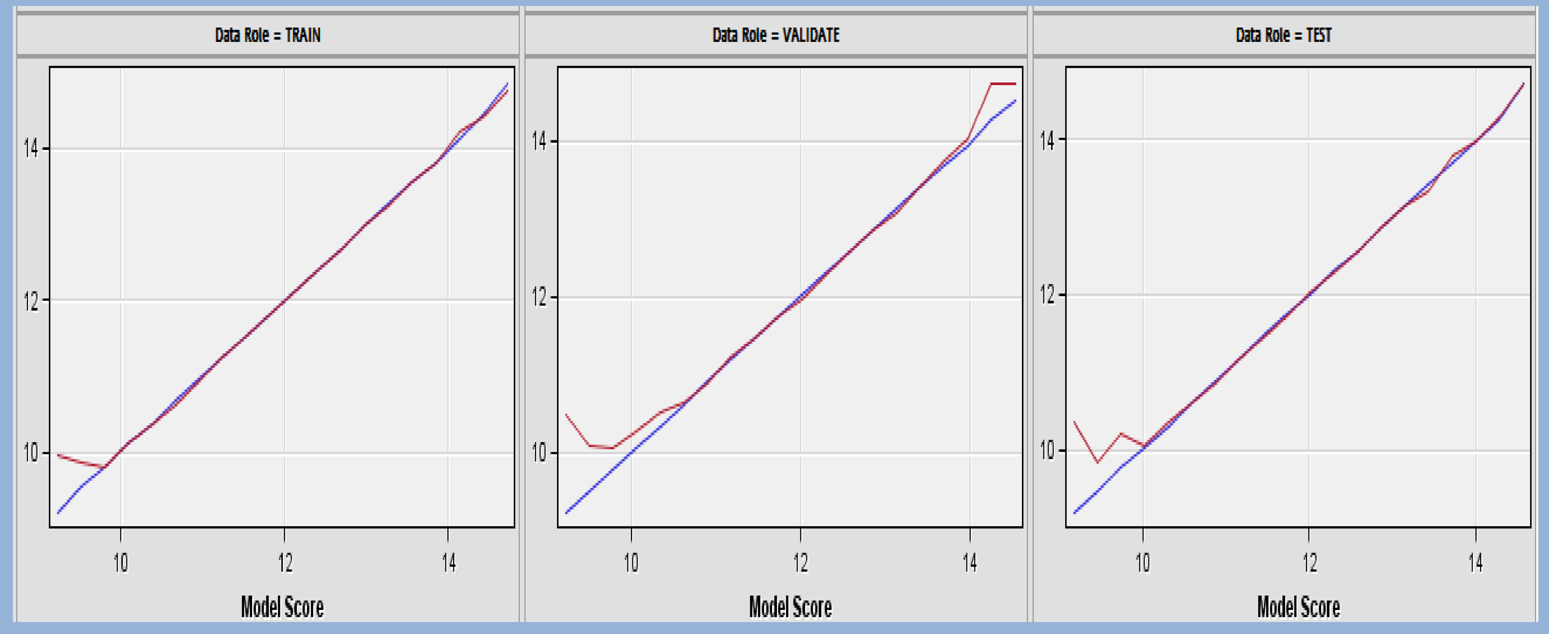

Fig 5. Mean predicted vs mean target

- Convergence criteria was satisfied for 10 iterations (Fig 4).
- Predicted values lie very close to actual values (Fig 5).
- To explain the architecture of the neural network, it was passed through a surrogate tree .
- Some of the rules of the surrogate tree are:
  - fmr > $1,892.5 and other cost < $54 then MV = $261,000
  - Structure type = mobile homes and $1,231.5 < fmr < $1,793.5, then MV = $48, 400
  - Structure type = mobile homes and $1,231.5 < fmr < $1,892.5 and $28 < other cost < $54, then MV = $48, 400

- Single houses and apartment complex in the Northeast and West will have similar MV if fmr and other costs are similar. Similarly, single houses and apartment complex in Midwest and South region will have comparable MV.
- Higher household and median income are associated with higher market value of housing a unit when all other features are controlled.
- Number of units in the building does not affect current market value of a housing unit.
- fmr and other costs are two most important factors that are used several times to determine MV.

CONCLUSIONS

- All business are interconnected. Contraction in one business sector directly affects all other sectors. Hence, any downfall of the US housing market will directly effect other sectors.
- To avoid onset of housing bubble, proper valuation of housing market cannot be overlooked.
- Lending companies, banks, individual home owners etc. all can be benefitted by proper valuation of housing units.

REFERENCES

Leamer, E. 2015. "Housing Really is the Business Cycle: What Survives the Lessons of 2008-09?". *Journal of Money, Credit and Banking,* 47(S1): 43-50.

Reuters. 2015. "Strong U.S. Groundbreaking, Building Permits Boost Housing Outlook". Accessed July 17, 2015. http://www.reuters.com/article/us-usa-eonomy-housing-idUSKCN0PR1CM20150717.

Federal Reserve Bank of St. Louis. 2015. "All-Transactions Housing Price Index for the United States". Accessed, July 1, 2015. https://research.stlouisfed.org/fred2/series/USSTHPI.

Federal Housing Finance Agency. 2015. "Housing Price Index". Accessed July 17, 2015. http://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index.aspx.

U.S. Department of Housing and Urban Development. 2016. "American Housing Survey: Housing Affordability Data System". Accessed May 10, 2015. https://www.huduser.gov/portal/datasets/hads/hads.html.

Rahman, M.G. and Islam, M. Z. 2013. "Missing Value Imputation Using Decision Trees and Decision Forests by Splitting and Merging Records: Two Novel Techniques". *Knowledge Based System.* 53: 51-65.

Rousseeuw, P. J. and Van Zomeren, B. C. 1990. "Unmasking Multivariate Outliers and Leverage Points". *Journal of the American Statistical Association*, 85(411): 633-639.

SAS. 2008. "Sample 30662: Mahalanobis Distance: From Each Observation to the Mean, from Each Observation to a Specific Observation, between all Possible Pairs". Accessed July 15, 2015. http://support.sas.com/kb/30/662.html.

Freund, R. J. and Wilson, W. J. 2003. *Statistical Methods*. 2nd ed. San Diego, CA: Academic Press

Zou, H. 2006. "The Adaptive LASSO and its Oracle Properties". *Journal of the American Statistical Association*, 101(476): 1418-1429.

SAS® GLOBAL FORUM 2016

IMAGINE. CREATE. INNOVATE.

LAS VEGAS | APRIL 18-21

#SASGF