House Prices in New York

CC901E2[a]

[a]The University of Sydney

November 3, 2022

Abstract here

keywords1 | keywords2 | keywords3 | keywords4 | keywords5

1. Introduction

Introduction here

To cite, edit the ref.bib document; and use Eddelbuettel and Balamuta (2017) here.

Once you cited, the references will automatically show up at the end of the article.

1.1. two "#" will make a 'subtitle'. Introduction subsection if needed

2. Data set description

Data description here

3. Analysis

4. Model Stability

5. Model Evaluaion and Results

In Analysis and Model Stability sections, three valid models are found, namely, backward stepwise regression model (backward model), forward stepwise regression model (forward model), and stable model.

In this section, we further compare their in-sample performances and out-of-sample performances.

5.1. In-sample performaces. In-sample perfomances are evaluated by calculating r^2, and adjusted r^2 values of the three models within the dataset (Table 1). We found that backward and forward models have the same r^2 (0.655) and adjusted r^2 values (0.652). Both values are slightly higher the stable model's ones is slightly lower but very similar (0.633 and 0.632 respectively).

Models	r^2	Adjusted r^2
Backward	0.655	0.652
Forward	0.655	0.652
Stable	0.633	0.632

Table 1. r^2 and adjusted r^2

5.2. Out-of-sample performances. A 10-fold cross validation shows that backward and forward models have slightly lower average MAE (Mean Absolute Error) and RMSE (Root Mean Square Error) than the stable model (Table 2).

Figures (Figure 1, Figure 2) of RMSE and MAE of the three models further confirms it, though the range of RMSE and MAE of the three models have overlaps.

Models	RMSE	MAE
Backward	58190.066	41531.593
Forward	58106.615	41467.353
Stable	59832.246	42709.202

Table 2. RMSE and MAE

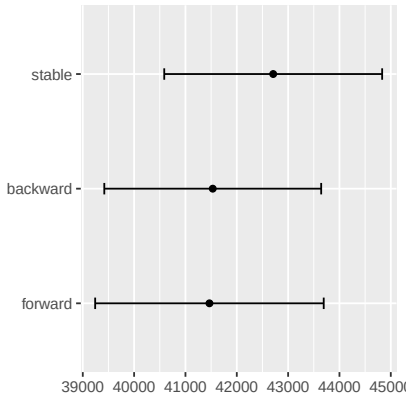

Fig. 1. RMSE

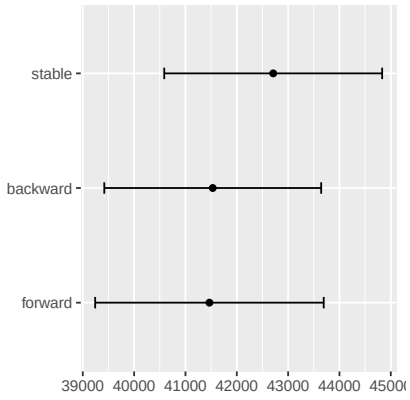

Fig. 2. MAE

6. Discussion, Conclusion, and Limitation

6.1. Discussion. In-sample performance evaluation indicates that the backward and forward models explain the data slightly more than the stable model due to their higher r^2 and adjusted r^2 values.

Out-of-sample performance evaluation further shows that the backward and forward models may be slightly more accurate than the stable model due to their lower RMSE and MAE values.

However, as mentioned in the Model Stability section, the backward and forward models with 14 parameters are not stable. In comparison, the stable model is selected in 78% of bootstrap resamples.

Meanwhile, though the stable model has higher r^2 and adjusted r^2 values than the backward and forward models, it still manages to explain approximately 63.27% (according to its r^2 value) or 63.18% (according to its adjusted r^2 value) of the total variation in the house price data. This is reasonable. Furthermore, the stable model does not have much higher error rates than the backward and forward models.

6.2. Conclution. Balancing stability, in-sample performances and out-of-sample performances, we conclude that the stable model may be the best among the four models.

6.3. Limitation. As mentioned, when compared to the backward and forward models, the stable model is more stable. However, it has relatively slightly lower r^2 and adjusted r^2 values, as well as relatively higher error rates (RMSE and MAE). To some extend, it is a compromise between stability and accuracy.

If more information about the dataset is provided, a domain knowledge expert may make better judgement on which model to choose.

References

Eddelbuettel D, Balamuta JJ (2017). "Extending *R* with *C++*: A Brief Introduction to *Rcpp*." *PeerJ Preprints*, **5**, e3188v1. ISSN 2167-9843. doi:10.7287/peerj.preprints.3188v1. URL https://doi.org/10.7287/peerj.preprints.3188v1.