

Machine Learning Models to Predict the Home Price Index in 2020 [2021?] In New York: What Are the Effects of The COVID-19 Pandemic?

Sophia Zhou

Worcester Polytechnic Institute

Abstract:

The COVID-19 pandemic significantly affected people's daily lives as well as global and national economic activities. While the unemployment rate surged in the United States (US) after March 2020, mortgage rates remained low. And yet, the S&P/Case-Shiller national home price index (HPI) and the S&P/Case-Shiller home price indexes of many different areas in the United States increased sharply. Prior predictive models for housing markets have not encountered a similar situation, and it is desirable to build predictive models for the 2020 housing market index with high accuracy. In this work, long short-term memory (LSTM) and linear regression models were built, using macroeconomics data from March 2003 to October 2020, to predict the S&P/ Case-Shiller HPI in New York in 2020. The data was collected from the Federal Reserve Bank and Freddie Mac. Different combinations of input variables and time lags were studied by experiments. It was found that the selection of input variables and time lags is important in building an accurate predictive model. Also, inclusion of unemployment rate data always resulted in models with low accuracy. In previous studies of housing market index models for other time periods, the unemployment rate did not show such an impact. This study also suggests that some macroeconomic variants commonly used in creating housing price index models have interdependence, although interdependence in input variables should be avoided. The study obtained LSTM models with high accuracy in predicting the abrupt changes in the HPI from March 2020 to December 2020. The linear regression models were not as accurate as the LSTM models.

Key words: home price index, real estate market, predictive model, machine learning, linear regression.