Housing price forecasting in selected US cities during the COVID-19 pandemic

Moritz Jäger, Adam Sahoun, Philipp Voit and Lars Wrede

Introduction to Data Analytics in Business, Masters in Applied Data Science, Frankfurt School of Finance & Management - (Dated: November 26, 2021)

The COVID-19 pandemic has affected the entire world since its onset in the first quarter of 2020 with an economic shock on the whole economy including the financial and real estate markets. Ongoingly, the pandemic is still having an impact on a wide range of working and living conditions dependent on the epidemic prevalence and magnitude in specific regions and accompanying measures in place. Strong impacts on social activities especially interpersonal relationships as well as governmentally enforced rules such as social distancing, home working, and conditions for leisure activities have focused individuals' awareness for their accommodation.
This one pager examines the housing price dynamics within a univariate time series forecast for three cities in the United States (US) during the COVID-19 pandemic. The seasonal autoregressive integrated moving average model (SARIMA) was used as it explains the housing market based on its own past values, in detail its own lags as well as the lagged forecast errors. Subsequently, the resulting excess trend was regressed against total infections for each of the three cities.
Prices on the US real estate market, represented by selected local housing markets, are growing continuously despite the COVID-19 pandemic with an even accelerated trend for the observed characteristics.

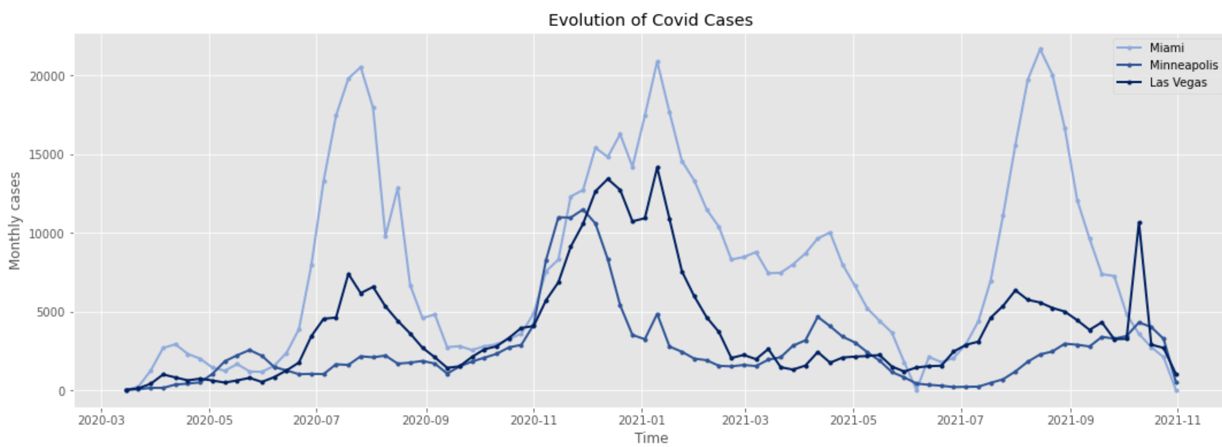
Having inspected already conducted studies on the impact of COVID-19 on the real estate market from Belej [1] for polish cities, finding a continuous growing trend, and the area of Naples from De Toro et al. [2], the latter indicating findings with regards to structural changes in the real estate market with regards to searching for accommodation in the post COVID era.
Following this and focusing on the United States, housing price dynamics in the United States are examined both before and during the COVID-19 pandemic and analyzed in connection with actual COVID-19 positive cases.

Data, Methodology and Prerequisites
The research was carried out at city level for the three US cities Miami, Minneapolis and Las Vegas which are suitable for the examination with regards to their comparability in terms of size and inhabitants as shown hereafter. As data source for our analysis, we took monthly median sales prices [3] covering monthly values from 2008-02-02 until 2021-08-31 from Zillow research, the research department of Zillow Inc, an online real estate marketplace.

Monthly values are used since the characteristics of real estate markets can be described as heterogeneous, delayed affected from macro-economic events due to their inertia and highly variable in terms of volume of transactions and features of the assets in each period [1]. Consequently, monthly median prices are the result of a variable number of transactions in a specific period. A brief overview of the basic statistical descriptors of monthly median prices time series in selected towns as well as additional descriptive characteristics with regards to the real estate market [4]:

	Mean	Minimum	Maximum	Population (in m)	Average Income (in USD)	Homeowner rate (in %)
Miami	235.445	140.000	367.750	2,7	92.800	36,6
Minneapolis	226.783	153.700	350.000	1,8	59.000	47,2
Las Vegas	224.289	128.000	375.000	2,2	44600	52,7

COVID-19 data for the three cities subject to our analysis has been taken from the NY Times [5]. Details as displayed below are similar to what has been experienced all over the United States with high number of cases towards the winter and slightly reduced cases in warmer month.



The time series within our data subject to our analysis can be decomposed into components trend, a consistent directional movement in a time series, seasonal variation, often related to holidays or seasons in business related cycles and a remainder not captured by one of the aforementioned [6]. Following this the time series was made stationary corroborated with Augmented Dickey Fuller Test as Unit Root Test.

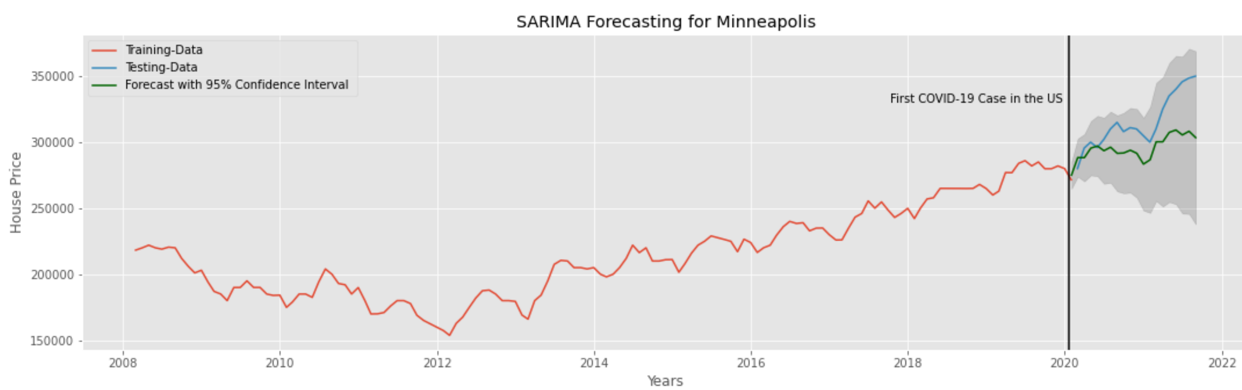
Next, autocorrelation, the similarity between observations as a function of the time lag between them, and partial autocorrelation, the correlation that results after removing the effect of any correlations due to the terms at shorter lags [7]. An autoregressive process expresses a dependent variable *Yt* as a function of past values of the dependent variable [8]. If we have *p* different lagged values of *Y*, the equation is often referred to as a "*p*th-order" autoregressive process and if the first differences do not produce a stationary series, then first differences of this first-differenced series can be taken [8].

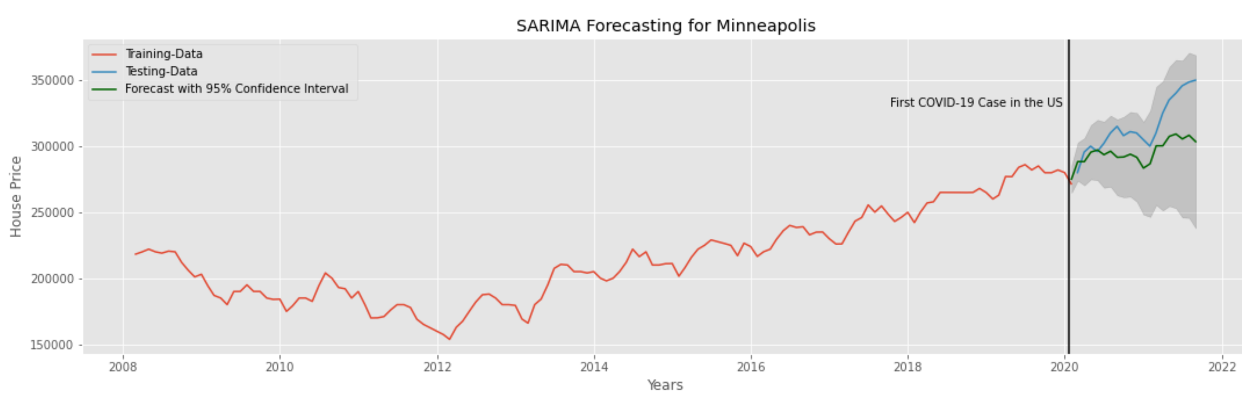
Combining stationary data with an autoregressive model and a moving average model, we obtain an (S)ARIMA model. This model is targeted at modeling the seasonal and non-seasonal component embedded in the time series. In more detail, AR*(p)* represents autoregression, with the maximum lag in the model is referred to as *p*. I*(d)* represents order of integration or number of differences needed to make the series stationary and MA*(q)* for moving average, meaning current error depends

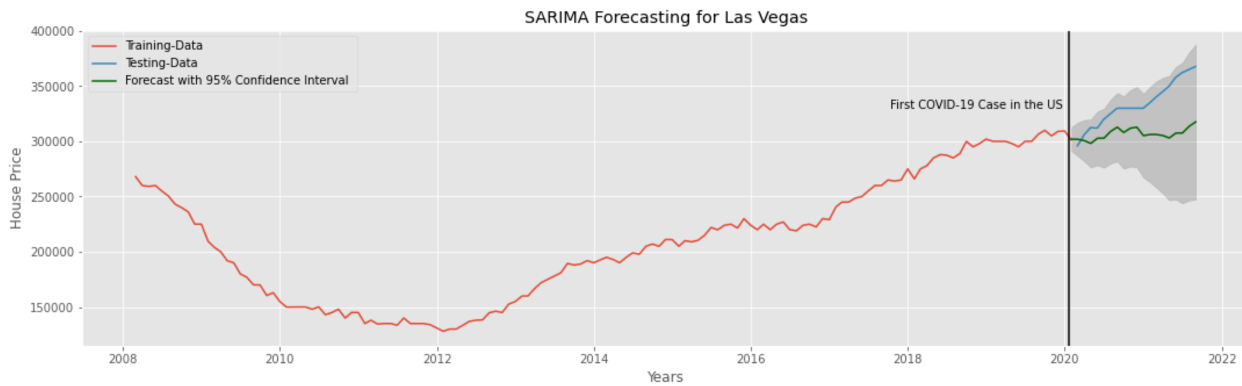
on the previous with some lag, which is *q* [9]. Similarly, parameters in the SARIMA model are *P, D, Q* for the seasonal component with *m* as parameter for recurrent periodicity, such as 12 for monthly as in our model.
Overall application of traditional time series model ARIMA is deemed reasonable due to complete dataset and constant time frames without missing values and applied through 'pm.auto_arima' function in Python. It returns the best SARIMA model according to AICc (Akaike information criterion corrected), BIC (Schwarz Bayesian criterion or SIC for Schwarz information criterion), and HQC value (Hannan–Quinn criterion) [10]. The final parameter values have been chosen when the AICc reach the lowest values since information criterion has been widely used in time series analysis to determine the appropriate order of a model [11].

Results

Taking the data part from beginning of February 2008 until the onset of COVID-19 within the US as training set and predicting the subsequent periods resulted in a forecast of real estate prices which has never experienced any COVID-19 related impact.
Next, calculating the residuals between the forecasted prices (displayed as green line) and the actual prices within the initial data (displayed as blue line) for the period until 2021-08-31 with a total of 18 periods and applying an ordinary least square regression with number of confirmed COVID-19 cases per US county as independent variable and the calculated residuals as dependent variable.







Findings show a moderate positive correlation between COVID-19 cases in two cities Miami and Las Vegas while Minneapolis shows a stronger correlation.

	coefficient	std error	t	P>|t|	[0.025	0.975]
MSP	14,921	2,668	5,593	0,000	9,293	20,550
MIA	0,234	0,056	4,209	0,001	0,117	0,352
LAS	0,862	0,296	2,913	0,010	0,238	1,486

Conclusion

Although figures related to COVID-19 bear ambiguity we used positive tested cases which can be seen as primary benchmarks with limited restrictions due to the focus on the US only. Impact of number of positive cases on the analysis have to be considered carefully with respect to the different outcomes within the three examined cities due to individual characteristics in each city. Additionally, expanding the analysis towards other US cities could enhance the assertions and detail findings. Furthermore, taking factual deaths into the analysis would lead to more differentiated results.

Overall, we created a supervised-regression and time series modeling framework that allows us to perform real estate price forecasts using historical data. This framework generates results to analyze risk and forecasted impact before taking decisions.

Further research to be conducted could inspect reasons of our findings and similar inspection from different cities in the US [12].

Bibliography

[1] M. Bełej1, „ Housing Price Forecasting in Selected Polish Cities During the COVID-19 Pandemic,“ *GEOMATICS AND ENVIRONMENTAL ENGINEERING • 2021,* Bd. 15, Nr. 4, p. 59–80, 2021.

[2] F. N. a. F. B. P. De Toro, „“Real Estate Market Responses to the COVID-19 Crisis: Which Prospects for the Metropolitan Area of Naples (Italy)?,”,“ *Urban Science,* Bd. 5, Nr. 1, p. 23, Feb. 2021..

[3] Zillow Inc., „Metr_median_Sale_price_uc_sfrcondo_week,“ 17 07 2021. [Online]. Available: https://www.zillow.com/research/data/. [Zugriff am 25 10 2021].

[4] U.S. Department of Commerce, „United States Census Bureau,“ U.S. Department of Commerce, 10 October 2021. [Online]. Available: https://www.census.gov/quickfacts/fact/table/miamibeachcityflorida,minneapoliscityminnesota,lasvegascitynevada/NES010218. [Zugriff am 26 10 2021].

[5] NY Times, „github.com/nytimes,“ 1 November 2021. [Online]. Available: https://github.com/nytimes/covid-19-data. [Zugriff am 25 Oktober 2021].

[6] H. e. a. Tatsat, „Machine Learning and Data Science Blueprints for Finance, „ 2020.,“ O'Reilly Media, Incorporated, https://ebookcentral.proquest.com/lib/franksfm/detail.action?docID=6362988., 2020.

[7] A. V. M. S. Cowpertwait, Introductory Time Series with R, New York, NY : Springer-Verlag, 2009.

[8] A. H. Studenmund, Using Econometrics: a Practical Guide, EBook, Global Edition., Harlow: Pearson Education, Limited., 2016.

[9] T. G. e. a. Smith, „ http://www.alkaline-ml.com/pmdarima,“ 03 11 2021. [Online]. Available: http://alkaline-ml.com/pmdarima/1.8.3/tips_and_tricks.html. [Zugriff am 25 10 2021].

[10] E. J. a. B. G. Q. Hannan, „“The Determination of the Order of an Autoregression.”,“ *Journal of the Royal Statistical Society.,* Bde. %1 von %2Series B (Methodological) 41, (1979): , Nr. no. 2 , p. 190–95. , 1979 .

[11] A. M. Shittu. O.I., „Comparison of Criteria for Estimating the Order of Autoregressive Process: A Monte Carlo Approach,“ *European Journal of Scientific Research,* Bde. %1 von %2ISSN 1450-216X , Nr. Vol.30 No.3 (2009), pp. pp.409-416, 2009.

[12] federalreserve, „federalreserve.gov,“ federalreserve.gov, 08 07 2021. [Online]. Available: https://www.federalreserve.gov/econres/notes/feds-notes/housing-market-tightness-during-covid-19-increased-demand-or-reduced-supply-20210708.htm. [Zugriff am 31 10 2021].