Time Series Analysis of All-Transactions House Price Index for Houston-The Woodlands-Sugar Land

Course number	SYDE 631
Course title	Time series modelling
Course Instructor	Professor Keith W. Hipel

Author	Shuo Tian
Author's department	System Design Engineering
Author's degree	Master of Engineering

UNIVERSITY OF WATERLOO

Contents

Abstract

The purposes of this research are to find a model best fitting the All-Transactions House Price Index for Houston-The Woodlands-Sugar Land (ATNHPI) data and to forecast the ATNHPI in the next three years. The method of maximum likelihood and Minimum Means Square Error were used to estimate the parameters and to forecast the number of ATNHPI in the next three years. The ATNHPI data from 1976-01-01 to 2018-04-01, which have been documented quarterly by the U.S. Federal Housing Finance Agency, reveal that the SARIMA(1,1,1)× $(1,0,1)_{4,}$ model may fit most adequately and forecast that the number of ATNHPI in the next three years will reach approximately 311.34, which exhibits 1.685 times the number of ATNHPI in 2010, under the assumption that current political, social, and economic conditions will be persistent in the future. Policy decision makers may be advised to prepare measures and programs for the increasing ATNHPI so that the new immigrants and locals can be smoothly adjusted to their life in Texas, the U.S..

Introduction

Over the past decade, Texas economy growing fast and has become the top one state by GDP growth of 30.4% in the US. The Texas economy is the second biggest in the U.S., behind only California. Texas ranks first for its growth prospects, thanks to strong employment and income growth forecasts over the next five years. Texas' job growth continued in March, 2018 for the 21st consecutive month with 32,000 jobs added. Many people are attracted from all over the world because of the growing job opportunities. Therefore, the house price has been concerned by many people already in the Texas or prepare to receive a job in Texas, also the government need to make appropriate decisions on the real-estimate market based on the analyzed data. The All-Transactions House Price Index for Houston-The Woodlands-Sugar Land was selected to analyze and forecast.

Nowadays, data science and information technology are adopted to assist various organizations to make a decision or to get the highest benefit from the information. Planning by forecasting trends in the future is a feasible option to apply statistical methods to analyzing data in the past that are related to the current activity. Time Series is an ordered sequence of values of a variable at equally spaced time intervals. Four components were consisted of the time series data: trend, seasonal, effect cyclical, and irregular effect. The time series analysis adopts forecasting methods through the past data. With the supposition that the information will resemble itself in the future, hence the forecast future events from the occurred data can be forecasted.

The present study used All-Transactions House Price Index for Houston-The Woodlands-Sugar Land (ATNHPI), TX (MSA) [ATNHPIUS26420Q] data released by U.S. Federal Housing Finance Agency (U.S. Federal Housing Finance Agency, 2018). The Federal Housing Finance Agency House Price Index (HPI) is a broad measure of the movement of single-family house prices. The HPI is a weighted, repeat-sales index, meaning that it measures average price changes in repeat sales or refinancings on the same properties. This information is obtained by reviewing repeat mortgage transactions on single-family properties whose mortgages have been purchased or securitized by Fannie Mae or Freddie Mac since January 1975. The ATNHPI data in this present study consist of 170 observations, that is to say, four observations from each year from 1976-01-01 to 2018-04-01. Following sections include 1) model specification, 2) model fitting and diagnostics, 3) forecasting (Shumway & Stoffer, 2011; Nenadic & Zucchini, 2012). All statistical analyses will be performed in SPSS and EViews.

Background

Various statistical forecasting methods were developed in the past decade years, such as regression analysis, classical decomposition method, Box and Jenkins and smoothing method. Different accuracy of forecasting models was supplied based on the minimum error of the forecast. Several factors are related with the optimal prediction methods, which are prediction interval, prediction period, characteristic of time series, and size of time series (Makridakis and Wheelwright, 1998).

In this project, we are focused on the time series analysis with the most popular method the Box-Jenkins method (Box andJenkins, 1976). In the 1960s, American Scholar Box and British statistician Jenkins presented a set of time series analysis and predictions methods, which broadly known as the Box - Jenkins modeling method. Box-Jenkins models contain Autoregressive (AR) and Moving Average (MA) models. Combining these two, the Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA). For seasonal time series forecasting, a variation of ARIMA, the Seasonal Autoregressive Integrated Moving Average (SARIMA) model is used. All these models are detailed as follow.

Autoregressive (AR) Models

A common approach for modeling univariate time series is the autoregressive (AR) model:

$$Z_t-\mu=\phi_1(Z_{t-1}-\mu)+\phi_2(Z_{t-2}-\mu)+\cdots+\phi_p(Z_{t-p}-\mu)+a_t$$

where ϕ_i (i = 1,2,...,p) is the ith nonseasonal AR parameter, a_t is white noise, and with μ denoting the process mean. A white noise process, i.e. a sequence of independent and identically distributed (i.i.d) random variables with zero mean and a constant variance σ^2. Generally, the white noise is assumed to follow the typical normal distribution. By introducing the B operator can equivalently be written as

$$(1-\phi_1B^1-\phi_2B^2-\cdots-\phi_pB^p)(Z_t-\mu)= a_t$$

$$\text{or } \phi(B)(Z_t-\mu)= a_t$$

where $\phi(B)= 1-\phi_1B^1-\phi_2B^2-\cdots-\phi_pB^p$ is the nonseasonal AR operator of order p.

An autoregressive model is simply a linear regression of the current value of the series against one or more prior values of the series. The value of p is the order of the AR model.

Moving Average (MA) Model

MA is another common approach for modeling univariate time series models, an MA(q) model uses past errors as the explanatory variables.:

$$Z_t-\mu=a_t-\theta_1a_{t-1}-\theta_2a_{t-2}-\cdots-\theta_qa_{t-q}$$

where θ_j (j = 1,2,...,q) is the jth nonseasonal MA parameter. Z_t is the time series, μ is the mean of the series, a_{t-i} are white noise. The value of q is the order of the MA model. By employing the B operator, the MA(q) process can be presented as

$$Z_t-\mu=a_t-\theta_1Ba_t -\theta_2B^2a_t -\cdots-\theta_qB^qa_t$$

$$=(1-\theta_1B^1-\theta_2B^2-\cdots-\theta_qB^q)a_t$$

$$=\theta(B)a_t$$

where $\theta(B) =1-\theta_1B^1-\theta_2B^2-\cdots-\theta_qB^q$

The Autoregressive Moving Average (ARMA) Models

An ARMA(p, q) model is a combination of AR(p) and MA(q) models and is suitable for univariate time series modeling.Mathematically an ARMA(p, q) model is represented as

$$(Z_t-\mu)-\phi_1(Z_{t-1}-\mu)-\phi_2(Z_{t-2}-\mu)-\cdots-\phi_p(Z_{t-p}-\mu)= a_t-\theta_1a_{t-1}-\theta_2a_{t-2}-\cdots-\theta_qa_{t-q}$$

By implementing the B operator, ARMA(p, q) can be written more conveniently as

$$(1-\phi_1 B^1-\phi_2 B^2-\cdots-\phi_p B^p)(Z_t-\mu)=(1-\theta_1 B^1-\theta_2 B^2-\cdots-\theta_q B^q)a_t$$

or

$$\phi(B) =\theta(B)a_t$$

where $\phi(B)= 1-\phi_1 B^1-\phi_2 B^2-\cdots-\phi_p B^p$ and $\theta(B) =1-\theta_1 B^1-\theta_2 B^2-\cdots-\theta_q B^q$. the model orders p,q refer to p autoregressive and q moving average terms.

Autoregressive Integrated Moving Average (ARIMA) Models

Time series, which contain trend, seasonal, effect cyclical patterns, are also non-stationary in nature. Thus, the ARIMA model is proposed, which is a generalization of an ARMA model to include the case of non-stationarity as well, whereas the ARMA models are inadequate to properly describe non-stationary time series. In ARIMA models a non-stationary time series is made stationary by applying finite differencing of the data points. The formulation of the ARIMA(p,d,q) model is given below:

$$\phi(B)(1-B)^d Z_t =\theta(B)a_t$$

where $\phi(B)= 1-\phi_1 B^1-\phi_2 B^2-\cdots-\phi_p B^p$ and $\theta(B) =1-\theta_1 B^1-\theta_2 B^2-\cdots-\theta_q B^q$, d is the order of differencing.

Seasonal Autoregressive Integrated Moving Average (SARIMA) Models

The ARIMA model is for non-seasonal non-stationary data. However, this model can not properly describe seasonal time series, which are frequently encountered in practice. For this reason, seasonal ARIMA (SARIMA) model was generalized by Box and Jenkins. In this model seasonal differencing of appropriate order is used to remove non-stationarity from the series. For monthly time series s = 12 and for quarterly time series s = 4. This model is generally known as the SARIMA$(p,d,q)\times (P,D,Q)_s$ model.The formulation of a SARIMA$(p,d,q)\times (P,D,Q)_s$ model in terms of lag polynomials is given as follow:

$$\phi(B)\, \boldsymbol{\phi}(B^s)\, \nabla^d \nabla_s^D Z_t =\theta(B)\, \boldsymbol{\theta}(B^s)\, a_t$$

where

$\phi(B) = 1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p$ and $\theta(B) = 1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q$

$\boldsymbol{\phi}(\mathbf{B}) = 1-\boldsymbol{\phi}_1 B^s-\boldsymbol{\phi}_2 B^{2s}-\cdots-\boldsymbol{\phi}_p B^{Ps}$ and $\boldsymbol{\theta}(B) = 1-\boldsymbol{\theta}_1 B^s-\boldsymbol{\theta}_2 B^{2s}-\cdots-\boldsymbol{\theta}_q B^{Qs}$

(p,d,q)= the nonseasonal component,

p= order of the nonseasonal AR operator,

d= order of the nonseasonal differencing operator,

q= order of the nonseasonal MA operator,

$(P,D,Q)_s$= the seasonal component,

s= number of seasons per year,

P= order of the seasonal AR operator,

D= order of the seasonal differencing operator,

Q= order of the seasonal MA operator.

In this paper we focus on the real-world problem in the real estate market in the Texas by analyzing the ATNHPI data from 1976-01-01 to 2018-04-01. The Box-Jenkins methodology was used to construct a best fit model SARIMA(1,1,1)× $(1,0,1)_4$. It uses a three-step iterative approach of model identification, parameter estimation and diagnostic checking to determine the best fit model from a general class of ARIMA models. The Box-Jenkins forecast method is schematically shown as below. The that the number of ATNHPI in the next three years was forecasted. Policy decision makers may be advised to prepare measures and programs for the increasing ATNHPI.

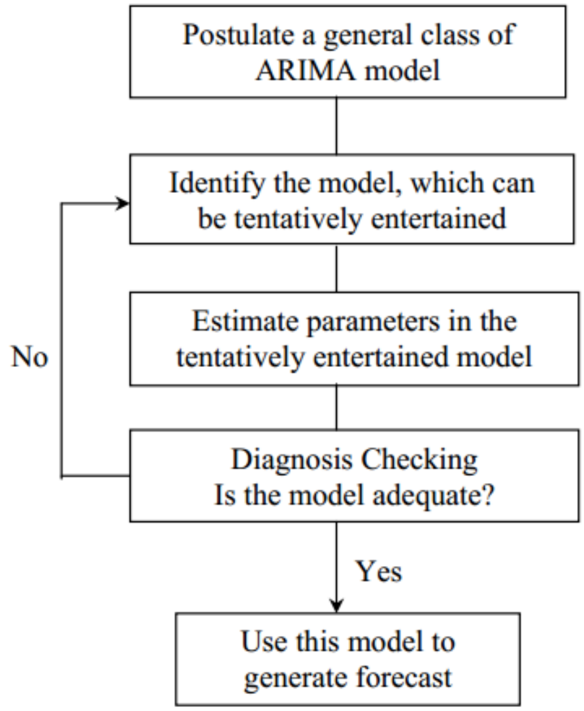


Fig. 1: The Box-Jenkins methodology for optimal model selection (Adhikari and Agrawal,2013)

Exploratory Data Analysis

The plot in Figure 2, which is a scatterplot of Zt = value of the variable Z (ATNHPI) at time t (Quarters), shows the number of ATNHPI has been increasing, behaving a quadratic pattern. More specifically, there was a slow increase in the ATNHPI from 1976-01-01 to 1998-10-01; there has been a steep increase afterwards. Even though there were some level s of drops after 2010-01-01, the increasing trend continued after 2012-01-01. Over all, the ATNHPI data may be represented as a realization of a deterministic trend with a quadratic function.

The sample ACF appear to attenuate shows in Figure 3, which indicates that differencing is needed. The deterministic trend is required to be removed. Moreover, seasonality is also need to be eliminated. Therefore, this study will select a differencing approach, which is known as SARIMA(p,d,q)× $(P,D,Q)_s$ model, for ATNHPI data.

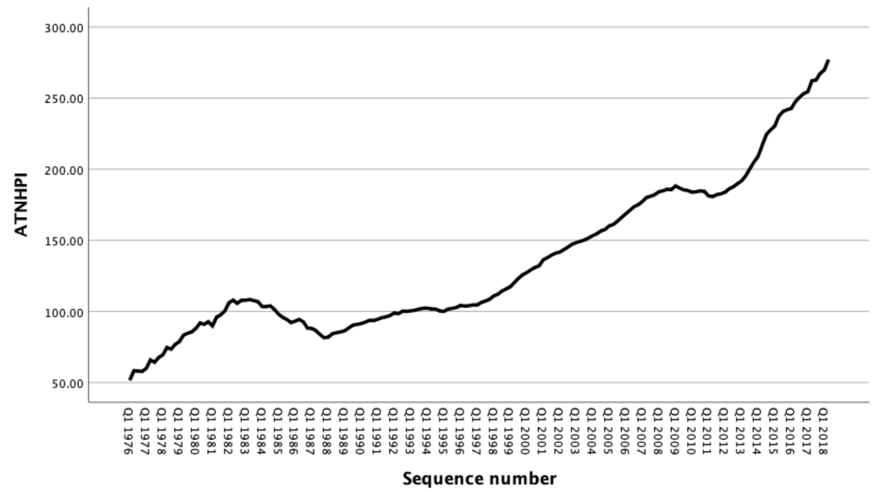


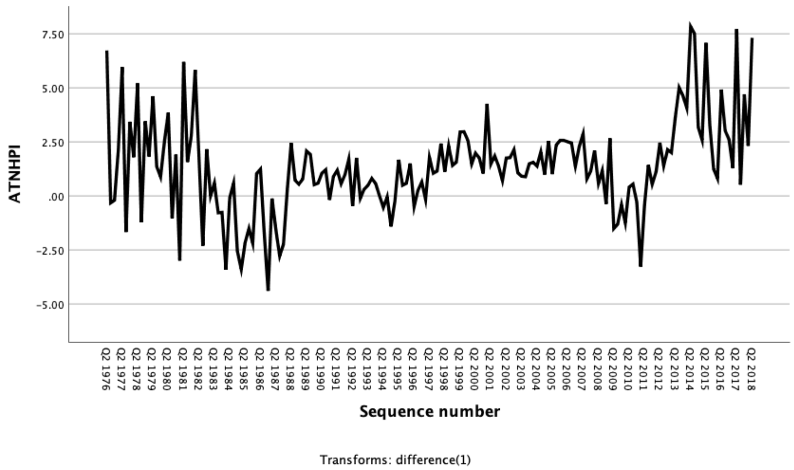
Fig.2: ATNHPI Original Data

Confirmatory Data Analysis

Taking a Difference

The plot of the data in Figure 1 shows an obviously increasing trend, which suggests non-stationarity in the data. A quantitative test, the augmented Dickey-Fuller unit-root test, for stationarity guided by Dickey and Fuller also confirms that the observed time series is not stationary (p value = 0.9849). (See the Output-1 in the appendix). The observed increasing trend and the augmented Dickey-Fuller unit-root test result demonstrate taking a difference of the data for stationarity. Output-2 in the appendix shows that the data did not come to be stationary after the first differencing. However, the plot of the first differencing data in Figure 4 shows that the data is evenly distributed on the upper and lower side of the X-axis, which demonstrate that the first differencing data assume to be stationary. Output-3 in the appendix shows that the second differencing data is stationary (p value = 0.0000). The scatter plot after taking the first and second difference of the data in support the series is now stationary. Because the second difference of the data lose more details than the first one, the first difference of the data was selected.

Fig. 4: First Differencing of the ATNHPI Data



Model Identification

For selecting a SARIMA(p,d,q)× $(P,D,Q)_s$ model for the differenced data, the sample ACF and PACF were plotted. According to Figure 3 in the appendix, the plots show the fluctuations of the sample autocorrelation values within the margin of random error bounds, suggesting a white noise process. Both the sample ACF and PACF appear to die off across the first few lags, which indicates that a nonseasonal AR parameter is needed. Nevertheless, one could interpret the sample PACF as attenuating during the first four lags, this demonstrates that a nonseasonal AR parameter is required. The sample ACF and PACF attenuate at lag 4, 8,12 and 16, means that a seasonal ARMA parameter is need in the model. Based on the results, candidate models for the data may be SARIMA(1,1,1)× $(1,0,1)_{4,}$ SARIMA(4,1,1)× $(1,0,1)_{4,}$ and SARIMA(4,1,11)× $(1,0,1)_4$ model. Hence, the present study will start to fit and diagnose SARIMA(1,1,1)× $(1,0,1)_4$ model, followed by possible candidates including SARIMA(4,1,1)× $(1,0,1)_{s,}$ and SARIMA(4,1,11)× $(1,0,1)_4$ models to select the best model for the data.

Model fitting

A SARIMA(1,1,1)× $(1,0,1)_4$ process can be expressed as

$$(1-B)(1-\phi_1 B)(1-\boldsymbol{\phi}_1 B)\ Z_t=(1-\theta_1 B)(1-\boldsymbol{\theta}_1 B)\ a_t$$

Where

a_t is white noise,

ϕ_1 is the first nonseasonal AR parameter,

θ_1 is the first nonseasonal MA parameter,

$\boldsymbol{\phi}_1$ is the first seasonal AR parameter,

$\boldsymbol{\theta}_1$ is the first seasonal MA parameter,

Table 1 shows that SARIMA(1,1,1)× $(1,0,1)_4$ fit well with the value of R-squared 0.999. To estimate unknown parameters, the method of maximum likelihood was used since this method in fitting time series has some advantages including 1) that parameter estimates are based on the entire observed sample; and 2) maximum likelihood estimators have very nice large-sample distributional properties (Cryer, & Chan, 2008). As table 4 in the appendix, the estimated ϕ_1 =0.878, $\boldsymbol{\phi}_1$=0.932, θ_1=0.622 and $\boldsymbol{\theta}_1$=0.64, t tests indicating that all estimates are statistically different from zero.

Estimated parameters as follows:

$$(1-B)\ (1-0.878B)(1-0.932B^4)\ Z_t = (1-0.622B)(1-0.641B)a_t$$

Model Diagnostics

To check the fit of the SARIMA(1,1,1)× $(1,0,1)_{4,}$ model, the observed residuals $\hat{a}_t$ will be checked since the residuals serve as proxies for the white noise terms $a_{t.}$ If the model is correctly specified and the estimates are reasonably close to the true parameters, the residuals should behave roughly like a sequence of independent, normal random variables with zero mean and constant variance. In examining whether the residuals behave like a white noise process and the SARIMA(1,1,1)× $(1,0,1)_{4,}$ model fit the data adequately, a series of tests will include testing for the normality and independence assumptions and the Ljung-Box test.

Normality Assumption

The histograms and QQ plots of the standardized residuals are used to visually assess the normality assumption. Kolmogorov-Smirnov test is used for the hypothesis tests for normality. As in Figure 5 in the appendix, the histogram and QQ plots exhibit no

gross departures from normality. The Kolmogorov-Smirnov test in Tab. 5 in the appendix confirms the results of the plots: In testing H_0: the standardized residuals are normally distributed,with the observed P-value of 0.983, we failed to reject the null hypothesis at $\alpha = 0.05$ level. In conclusion, there is not sufficient evidence against normality.

Independence Assumption

A time series plot of the residuals is used for visual assessment of the independent assumption and a runs test with the residuals is used for a formal test of the independence assumption. The residual plot displays no discernible patterns and looks to be random in appearance (Fig. 6). The runs test confirms the visual assessment: In testing H_0: the standardized residuals are independent, with the observed P value of 0.817 with median or mean test value, we failed to reject the null hypothesis at $\alpha = 0.05$ level. In conclusion, there is no evidence against independence (Tab. 6).

Ljung-Box Test

The Ljung-Box test is used to check the adequacy of a fitted SARIMA$(1,1,1)\times(1,0,1)_4$ model formally. According to Table 1 in the appendix showing the results of the modified Ljung-Box test, in testing H_0: the SARIMA$(1,1,1)\times(1,0,1)_4$model is appropriate, and the residuals are independent. we do not have sufficient evidence against SARIMA$(1,1,1)\times(1,0,1)_4$ model with p value = 0.297. The plot of the sample autocorrelation function (ACF) of the residuals shows (Fig. 6) that the residuals are approximately uncorrelated, behaving a white noise process also supports the aforementioned hypothesis test.

A crucial step in an appropriate model selection is the determination of optimal modelparameters. In this research, widely used measures for model identification Bayesian Information Criterion (BIC) was selectd, which are defined below:

$BIC(p) = n \ln(\hat{\sigma}_e^2/n)+p+p \ln(n)$

Here n is the number of effective observations, used to fit the model, p is the number of parameters in the model and $\hat{\sigma}_e^2$ is the sum of sample squared residuals. The optimal model order is chosen by the number of model parameters, which minimizes BIC. When comparing the normalized BIC value, the SARIMA(1,1,1)× $(1,0,1)_4$ model has the lowest value of 1.244, whereas the SARIMA(4,1,1)$(1,0,1)_4$ and SARIMA(4,1,11)$(1,0,1)_4$ models has the value of 1.354 and 1.991 respectively (Tab. 2,3). The Ljung-Box Q test also demonstrate that the SARIMA(4,1,11)$(1,0,1)_4$ model reject the H_0: the residuals are independent, which mean that this model is inappropriate. Taking into account all the results from the diagnostics, the SARIMA(1,1,1)× $(1,0,1)_4$ model may be a better model than the SARIMA(4,1,1)$(1,0,1)_4$ and SARIMA(4,1,11)$(1,0,1)_4$ models since the plots favored the SARIMA(1,1,1)× $(1,0,1)_4$ models. The ATNHPI data will be forecasted through the SARIMA(1,1,1)× $(1,0,1)_4$ model in the next section.

Applications and Insights

The minimum mean squared error (MMSE) forecast is adopted as a formal mathematical criterion to calculate model forecasts. From time origin t, let the best forecast at lead time l be written as

$\hat{Z}t\ (l)= \Psi_t a_t+\Psi_{t+1} a_{t-1}+\Psi_{t+2} a_{t-2}+\ldots$

where the weight $\Psi_t, \Psi_{t+1}, \Psi_{t+2}, \ldots$is ith parameter of infinite MA operator and a_t is the innovation sequence distributed as $NID(0,\sigma_a^2)$.

The table 7 consists of the predicted values in the forecast period. The upper confidence limits (UCL) and lower confidence limits (LCL) for the forecasted values (95%) are also included in the table 7 and Fig. 8. The HPI serves as a timely, accurate indicator of house price trends. It also provides housing economists with an improved analytical tool that is useful for estimating changes in the rates of mortgage defaults, prepayments and housing affordability (U.S. Federal Housing Finance Agency, 2018). Home prices increased by 5.9 percent in the Houston metro area from a year earlier, while prices rose nationally by 7 percent during the period. The number of ATNHPI grew briskly in the first quarter of 2018, and in the next three years will reach approximately 311.34, under the assumption that current political, social, and economic conditions will be persistent in the future.

	Q3 2018	Q4 2018	Q1 2019	Q2 2019	Q3 2019	Q4 2019	Q1 2020	Q2 2020	Q3 2020	Q4 2020
Forecast	280.17	283.68	285.99	291.99	294.83	298.06	300.18	305.73	308.35	311.34
UCL	283.46	288.96	293.18	301.05	306.31	311.94	316.43	324.33	329.68	335.38
LCL	276.88	278.40	278.81	282.92	283.35	284.18	283.92	287.13	287.03	287.29

Table 7: Forecast values.

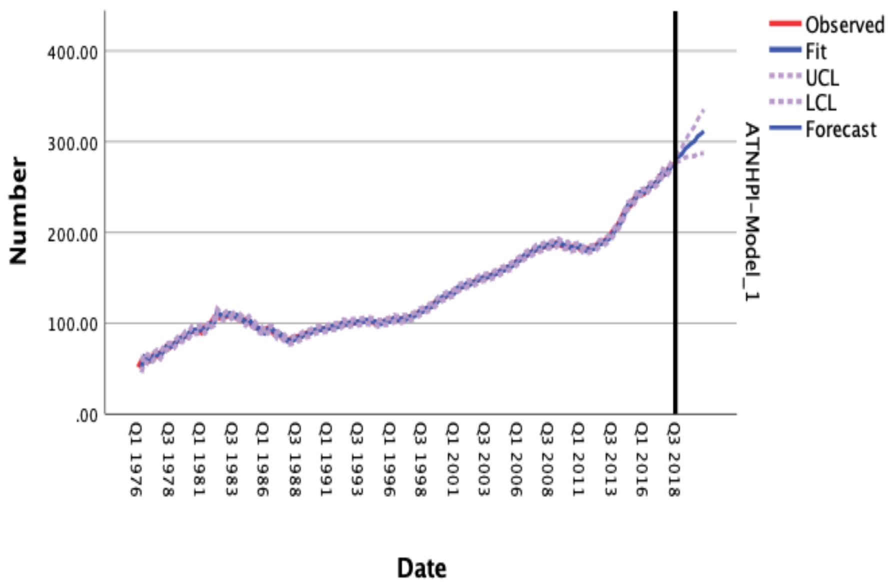


Fig. 8: Forecasting Plot.

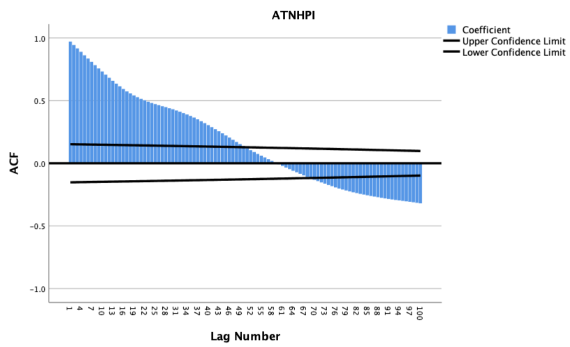
Conclusions

The SARIMA (1,1,1) × $(1,0,1)_{4,}$ model forecast that the number of ATNHPI in the next three years will reach approximately 311.34, which explains 1.685 times the number of ATNHPI in 2010, under the assumption that current political, social, and economic conditions will be persistent in the future. Policy decision makers may be advised to prepare measures and programs for the increasing ATNHPI so that the new immigrants and locals can be smoothly adjusted to their life in Texas, the U.S..

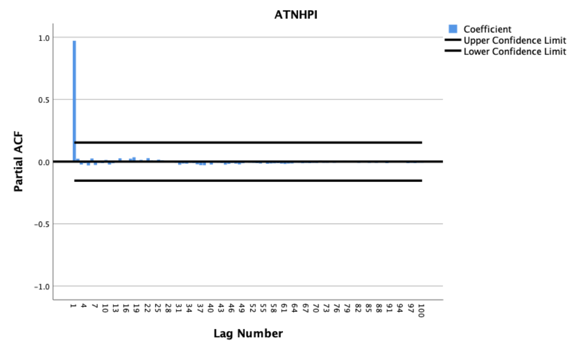
Appendices

Fig. 3: Sample ACF and PACF

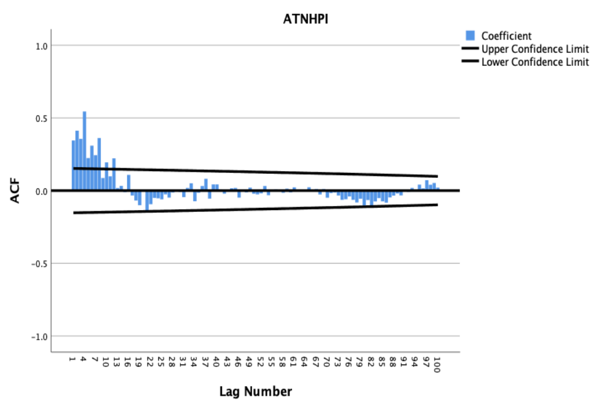
Sample ACF for the original data



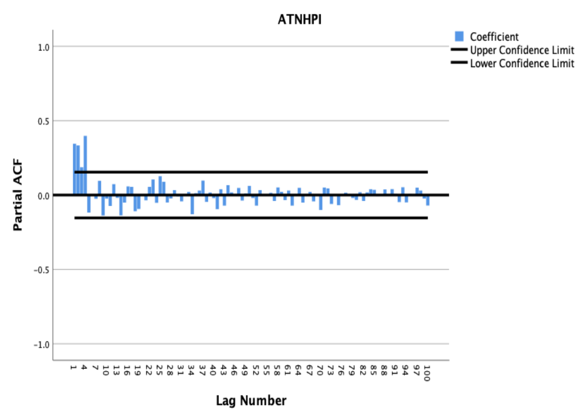
Sample PACF for the original data



Sample ACF for the data after first difference.



Sample PACF for the data after first difference



Output-1: Augmented Dickey-Fuller Unit Root test with the original data

Null Hypothesis: ATNHPI has a unit root
Exogenous: None
Lag Length: 4 (Automatic - based on AIC, maxlag=13)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		1.860922	0.9849
Test critical values:	1% level	-2.579052	
	5% level	-1.942768	
	10% level	-1.615423	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(ATNHPI)
Method: Least Squares
Date: 11/26/18 Time: 15:02
Sample (adjusted): 1977Q2 2018Q2
Included observations: 165 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ATNHPI(-1)	0.002432	0.001307	1.860922	0.0646
D(ATNHPI(-1))	0.083965	0.073823	1.137378	0.2571
D(ATNHPI(-2))	0.197504	0.073085	2.702385	0.0076
D(ATNHPI(-3))	0.099399	0.073843	1.346073	0.1802
D(ATNHPI(-4))	0.405363	0.072006	5.629567	0.0000

R-squared	0.410829	Mean dependent var	1.315939
Adjusted R-squared	0.396099	S.D. dependent var	2.150977
S.E. of regression	1.671547	Akaike info criterion	3.895211
Sum squared resid	447.0512	Schwarz criterion	3.989330
Log likelihood	-316.3549	Hannan-Quinn criter.	3.933417
Durbin-Watson stat	1.858693		

Output-2: Augmented Dickey-Fuller unit root test after taking a first difference of the data

Null Hypothesis: D(ATNHPI) has a unit root
Exogenous: None
Lag Length: 3 (Automatic - based on AIC, maxlag=13)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.300620	0.1781
Test critical values:	1% level	-2.579052	
	5% level	-1.942768	
	10% level	-1.615423	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(ATNHPI,2)
Method: Least Squares
Date: 11/26/18 Time: 14:54
Sample (adjusted): 1977Q2 2018Q2
Included observations: 165 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(ATNHPI(-1))	-0.085185	0.065496	-1.300620	0.1952
D(ATNHPI(-1),2)	-0.799327	0.086778	-9.211169	0.0000
D(ATNHPI(-2),2)	-0.568365	0.090087	-6.309078	0.0000
D(ATNHPI(-3),2)	-0.436846	0.070524	-6.194325	0.0000

R-squared	0.503828	Mean dependent var	0.031333
Adjusted R-squared	0.494583	S.D. dependent var	2.369139
S.E. of regression	1.684285	Akaike info criterion	3.904503
Sum squared resid	456.7271	Schwarz criterion	3.979798
Log likelihood	-318.1215	Hannan-Quinn criter.	3.935068
Durbin-Watson stat	1.868919		

Output-3: Augmented Dickey-Fuller unit root test after taking a second difference of the data

Null Hypothesis: D(ATNHPI,2) has a unit root
Exogenous: None
Lag Length: 7 (Automatic - based on AIC, maxlag=13)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.010376	0.0000
Test critical values:	1% level	-2.579495	
	5% level	-1.942830	
	10% level	-1.615384	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(ATNHPI,3)
Method: Least Squares
Date: 11/26/18 Time: 14:52
Sample (adjusted): 1978Q3 2018Q2
Included observations: 160 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(ATNHPI(-1),2)	-2.661861	0.531270	-5.010376	0.0000
D(ATNHPI(-1),3)	0.904321	0.497458	1.817882	0.0711
D(ATNHPI(-2),3)	0.374178	0.449038	0.833287	0.4060
D(ATNHPI(-3),3)	0.032930	0.390872	0.084247	0.9330
D(ATNHPI(-4),3)	0.009825	0.326451	0.030097	0.9760
D(ATNHPI(-5),3)	-0.076863	0.245888	-0.312593	0.7550
D(ATNHPI(-6),3)	-0.116259	0.160395	-0.724828	0.4697
D(ATNHPI(-7),3)	-0.156932	0.078797	-1.991608	0.0482

R-squared	0.833078	Mean dependent var	0.009813
Adjusted R-squared	0.825391	S.D. dependent var	3.939222
S.E. of regression	1.646053	Akaike info criterion	3.883345
Sum squared resid	411.8427	Schwarz criterion	4.037104
Log likelihood	-302.6676	Hannan-Quinn criter.	3.945781
Durbin-Watson stat	1.954866		

		Model Fit statistics			**Ljung-Box Q(18)**		
Model Type	Number of Predictors	Stationary R-squared	R-squared	Normalized BIC	Statistics	DF	Sig.
SARIMA(1,1,1)(1,0,1)$_4$	0	.361	.999	1.244	16.278	14	.297

Table 1: ARIMA(1,1,1)(1,0,1)$_4$ Model Statistics

		Model Fit statistics			**Ljung-Box Q(18)**		
Model Type	Number of Predictors	Stationary R-squared	R-squared	Normalized BIC	Statistics	DF	Sig.
SARIMA(4,1,1)(1,0,1)$_4$	1	.384	.999	1.354	11.855	11	.375

Table 2: ARIMA(4,1,1)(1,0,1)$_4$ model Statistics

		Model Fit statistics			Ljung-Box Q(18)		
Model Type	Number of Predictors	Stationary R-squared	R-squared	Normalized BIC	Statistics	DF	Sig.
SARIMA(4,1,11)(1,0,1)$_4$	1	.194	.999	1.991	67.571	1	.000

Table 3: ARIMA(4,1,11)(1,0,1)$_4$ model Statistics

		Estimate	SE	t	Sig.
AR	Lag 1	.878	.068	12.921	.000
Difference		1			
MA	Lag 1	.622	.112	5.559	.000
AR, Seasonal	Lag 1	.932	.042	22.324	.000
MA, Seasonal	Lag 1	.641	.100	6.426	.000

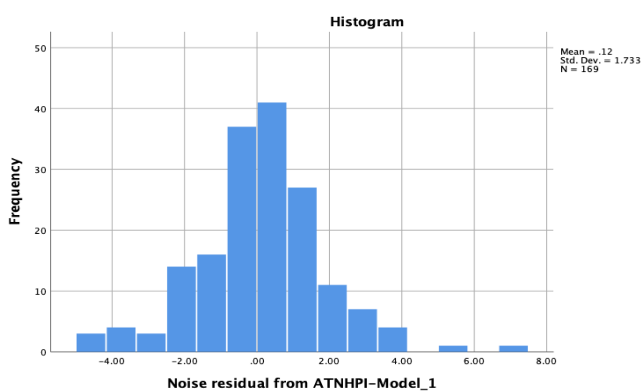
Table 4: SARIMA(1,1,1)(1,0,1)$_4$ model parameters

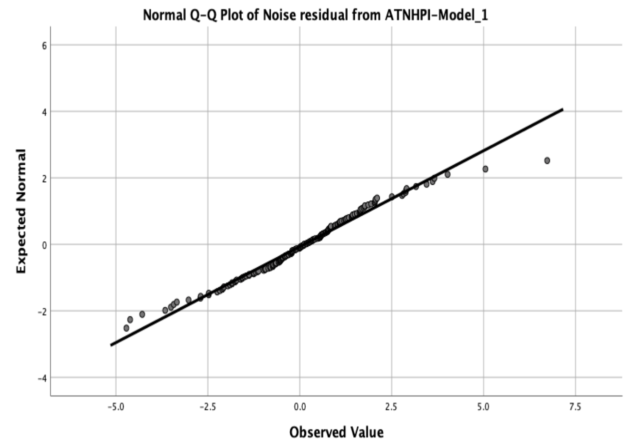
	Statistic	df	Sig.	Statistic
Noise residual from SARIMA(1,1,1)(1,0,1)$_4$	.063	169	.200*	.983

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Table 5: Kolmogorov-Smirnova test (Normality assumption test).





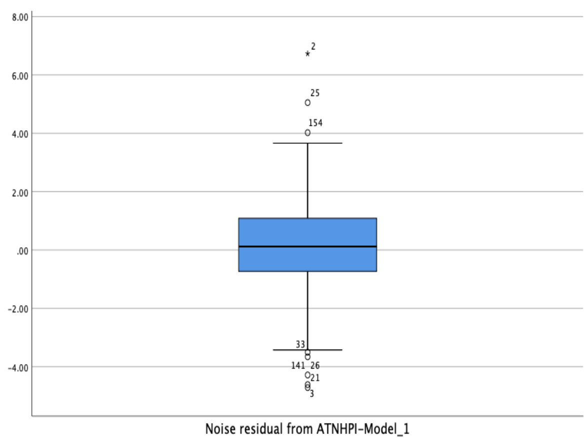


Figure 5: Normality assumption test with residuals from SARIMA(1,1,1)(1,0,1)$_4$ model.

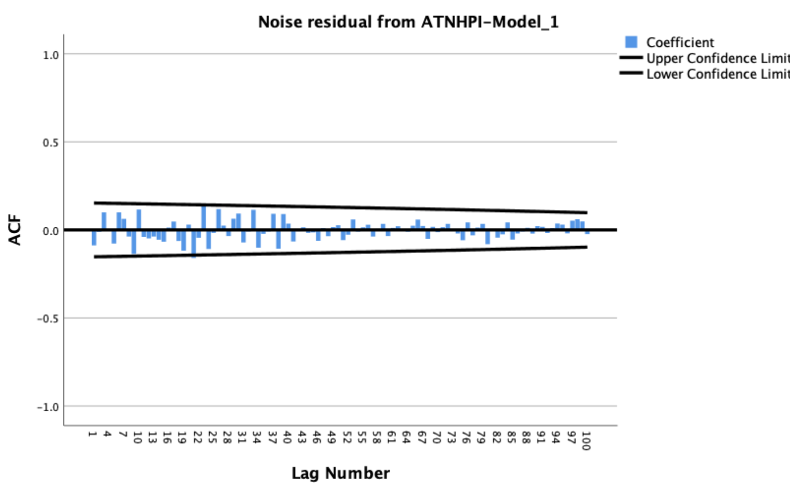
	Noise residual from SARIMA(1,1,1)(1,0,1)$_4$
Test Value[a]	.12
Cases < Test Value	84
Cases >= Test Value	85
Total Cases	169
Number of Runs	87
Z	.232
Asymp. Sig. (2-tailed)	.817

a. Median

	Noise residual from SARIMA(1,1,1)(1,0,1)$_4$
Test Value[a]	.1201
Cases < Test Value	85
Cases >= Test Value	84
Total Cases	169
Number of Runs	87
Z	.232
Asymp. Sig. (2-tailed)	.817

a. Mean

Table 6: Runs test of noise residual from SARIMA(1,1,1)(1,0,1)$_4$.



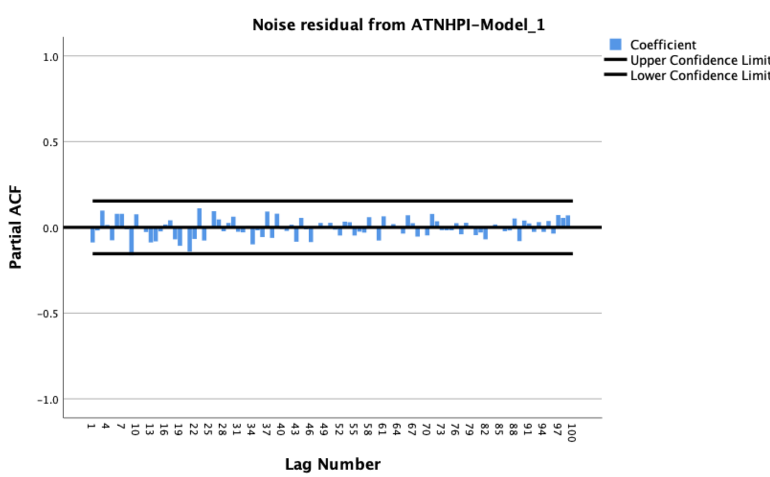


Fig. 6: The ACF and PACF of Noise residual from SARIMA(1,1,1)(1,0,1)$_4$

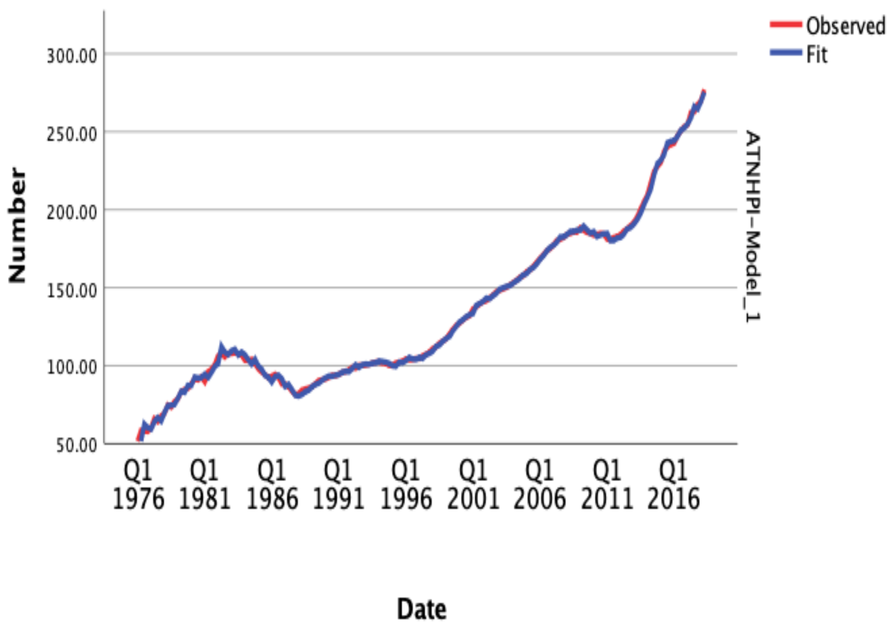


Fig. 7: The fit plot of SARIMA(1,1,1)(1,0,1)$_4$

Data set:

Federal Reserve Economic Data: All-Transactions House Price Index for Houston-The Woodlands-Sugar Land, TX (MSA), Index 1995: Q1=100, Quarterly

Observation_date	ATNHPI	Observation_date	ATNHPI
1976-01-01	51.59	1977-01-01	59.94
1976-04-01	58.32	1977-04-01	65.90
1976-07-01	58.00	1977-07-01	64.23
1976-10-01	57.80	1977-10-01	67.66
1978-01-01	69.45	1979-01-01	78.71
1978-04-01	74.66	1979-04-01	83.32
1978-07-01	73.44	1979-07-01	84.67
1978-10-01	76.89	1979-10-01	85.55
1980-01-01	88.07	1981-01-01	89.81
1980-04-01	91.92	1981-04-01	96.01
1980-07-01	90.89	1981-07-01	97.58
1980-10-01	92.80	1981-10-01	100.37
1982-01-01	106.20	1983-01-01	107.81
1982-04-01	107.95	1983-04-01	108.40
1982-07-01	105.65	1983-07-01	107.62
1982-10-01	107.81	1983-10-01	106.86
1984-01-01	103.45	1985-01-01	97.99
1984-04-01	103.37	1985-04-01	95.83
1984-07-01	103.95	1985-07-01	94.32
1984-10-01	101.38	1985-10-01	92.13
1986-01-01	93.16	1987-01-01	88.10
1986-04-01	94.40	1987-04-01	86.56
1986-07-01	92.61	1987-07-01	83.78
1986-10-01	88.22	1987-10-01	81.55
1988-01-01	81.87	1989-01-01	86.37
1988-04-01	84.32	1989-04-01	88.45
1988-07-01	85.05	1989-07-01	90.37
1988-10-01	85.59	1989-10-01	90.89
1990-01-01	91.48	1991-01-01	94.46
1990-04-01	92.53	1991-04-01	95.65
1990-07-01	93.74	1991-07-01	96.21
1990-10-01	93.56	1991-10-01	97.19

1992-01-01	98.89	1993-01-01	100.38
1992-04-01	98.42	1993-04-01	100.87
1992-07-01	100.18	1993-07-01	101.68
1992-10-01	100.07	1993-10-01	102.25
1994-01-01	102.25	1995-01-01	100.00
1994-04-01	101.71	1995-04-01	101.67
1994-07-01	101.67	1995-07-01	102.16
1994-10-01	100.26	1995-10-01	102.74
1996-01-01	104.24	1997-01-01	104.47
1996-04-01	103.74	1997-04-01	106.25
1996-07-01	103.99	1997-07-01	107.30
1996-10-01	104.63	1997-10-01	108.45
1998-01-01	110.87	1999-01-01	117.28
1998-04-01	111.99	1999-04-01	120.24
1998-07-01	114.31	1999-07-01	123.21
1998-10-01	115.72	1999-10-01	125.78
2000-01-01	127.26	2001-01-01	136.29
2000-04-01	129.24	2001-04-01	137.76
2000-07-01	130.99	2001-07-01	139.61
2000-10-01	132.03	2001-10-01	140.99
2002-01-01	141.70	2003-01-01	148.43
2002-04-01	143.45	2003-04-01	149.34
2002-07-01	145.22	2003-07-01	150.23
2002-10-01	147.36	2003-10-01	151.72
2004-01-01	153.28	2005-01-01	160.22
2004-04-01	154.66	2005-04-01	161.23
2004-07-01	156.69	2005-07-01	163.58
2004-10-01	157.68	2005-10-01	166.15
2006-01-01	168.72	2007-01-01	177.33
2006-04-01	171.22	2007-04-01	180.20
2006-07-01	173.67	2007-07-01	180.99
2006-10-01	175.05	2007-10-01	182.13
2008-01-01	184.22	2009-01-01	188.27
2008-04-01	184.80	2009-04-01	186.76
2008-07-01	185.97	2009-07-01	185.45
2008-10-01	185.60	2009-10-01	185.04
2010-01-01	183.80	2011-01-01	181.18
2010-04-01	184.21	2011-04-01	180.76
2010-07-01	184.76	2011-07-01	182.19

2010-10-01	184.45	2011-10-01	182.73
2012-01-01	183.84	2013-01-01	191.83
2012-04-01	186.30	2013-04-01	195.45
2012-07-01	187.69	2013-07-01	200.46
2012-10-01	189.83	2013-10-01	205.10
2014-01-01	209.17	2015-01-01	230.27
2014-04-01	217.00	2015-04-01	237.35
2014-07-01	224.51	2015-07-01	240.63
2014-10-01	227.68	2015-10-01	241.86
2016-01-01	242.67	2017-01-01	254.50
2016-04-01	247.59	2017-04-01	262.22
2016-07-01	250.61	2017-07-01	262.74
2016-10-01	253.22	2017-10-01	267.44
2018-01-01	269.76		
2018-04-01	277.07		

References

Adhikari R, Agrawal RK, An introductory study on time series modeling and forecasting. arXiv preprint arXiv:1302.6613, 2013 - arxiv.org.

Box, G.E.P. and G. Jenkins, “Times Series Analysis Forecasting and Control,” Holden-Day, San Francisco, CA, 1976.

Best States For Business, Forbes, 2018. https://www.forbes.com/places/tx/

Chuck DeVore, Of The 5 Big States, Texas #1 For Growth, California #1 For Poverty. 2018. https://www.forbes.com/sites/chuckdevore/2018/08/27/of-the-5-big-states-texas-1-for-growth-california-1-for-poverty/#6b9c65b170ef.

Hipel K.W. McLeod A.I, Time Series Modelling of Water Resources and Environmental Systems, Volume 45, 1st Edition, ISBN: 9780080870366,Published Date: 7th April 1994.

Jonathan D. Cryer & Kung-Sik Chan, “Time series analysis: with applications in R,” 2nd ed. New York: Springer, 2008.

Makridakis, S., S.C. Wheelwright and R.J. Hyndman, “Forecasting: Methods and Applications,” 3 ed. Wiley, Inc., New York, 1998.

Oleg Nenadic, Walter Zucchini, “Statistical Analysis with R - a quick start -,” Retrieved November 10, 2012, from http://www.scribd.com/doc/114263559/Statistical-Analysis-With-RA-Quick-Start.

Robert H. Shumway & David S. Stoffer, “Time Series Analysis and Its Applications with R Examples,” 3rd ed. New York: Springer, 2011.

U.S. Federal Housing Finance Agency, All-Transactions House Price Index for Houston-The Woodlands-Sugar Land, TX (MSA) [ATNHPIUS26420Q], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/ATNHPIUS26420Q, November 25, 2018.