

Prediction on Land Market Value Based on the Real Estate Market in USA

Lei Wang

Department the Mathematics, The University of Southern Mississippi, USA

Article history

Received: 20-02-2017

Revised: 21-02-2017

Accepted: 23-05-2017

Email: fiona901587@yahoo.com

Abstract: The Land Market Value, defined as the total value of land price and quantity data are derived from data on housing values, is an important factor in the estimation of structure costs using price indexes for housing and construction costs. In this study, we gather and analyze 34 years' national data on past and present real estate transaction. According to the characteristics of raw data, we try to develop the potential Decomposition, Smoothing, ARIMA and other advanced forecasting models with appropriate transformations. Specifically, we employ an innovation space state underlying certain forecasting model. For regression analysis, we involves GDP, CPI, Construction Cost Index, population, unemployment rate, inflation rate and Purchasing Manage Index in multivariate statistical model. Most importantly, we obtain how to add value to business and apply skills set to real estate in a real world environment. The goal in providing crucial statistical method is to enable government and investors to make informed decisions regarding real estate.

Keywords: Forecasting Model, Land Market Value, Time Series Analysis

Introduction

This paper aims to provide important information of real estate market in USA and potential problems and opportunities or buyer and seller. As housing is a form of wealth, the purchase of a home represents an important investment and it is normally a hot topic for the scholars and investors. Because of scarcity, the fluctuation of land market value will have a great influence of the net worth of business and household. In this regard, Davis and Heathcote (2007) estimate that wings in residential land prices accounted for most of the variation in house prices over 1975-2006 for the United States as a whole. Davis and Palumbo (2008) reach the same conclusion for a large set of metropolitan areas over a somewhat shorter sample period, as do Bostic *et al.* (2007) in their detailed analysis of home price changes within a single metropolitan area (Wichita, Kansas). In addition, the land is an important component of wealth. Also, it is a source of variation in real estate prices and as collateral for loans, only a handful of studies have calculated land price indexes for the nation as a whole or for a broad set of cities. Davis and Heathcote (2007) and Davis and Palumbo (2008) estimate price indexes for residential land, while

Davis (2009) estimates indexes for both residential and commercial land. Also, Sirmans and Slade (2009) use transaction prices to calculate national land prices indexes. The data were collected on the basis of past and present real estate transactions and develop processes which guide future investment by demonstrating the true future value of the investment. To provide students with sufficient understanding and ability to model, analyze and develop forecasts for engineering and business decisions. The emphasis will be on quantitative methods.

Background

After food and medical care, housing is the largest consumer expenditure in the United states. In 1994, personal consumption expenditures on housing were about \$2600 per capita, or 14.9% of household budgets. Further, the bulk of expenditures in one of the next highest categories, household operations, are linked to housing. From the investment side, housing is the largest single form of fixed capital investment in the United States, comprising more than \$9 trillion, or roughly half of this nation's gross fixed private capital. Other than human capital, housing and land are more widely held than any other form of capital.

Table 1. How housing market works

Inputs		Production		Demand
Land	P	Developers	P	Renters ⇒ Homeowners (Income and population)
Finance	R	Builders	R	
Infrastructure ⇒	I	⇒ Landlords ⇒	I	
Labor	C	Homeowners	C	
Materials	E		E	

In United States, as in most countries, the market for housing services can be approximated as a competitive market. Housing production activities have few barriers to entry or large economies of scale. Few landlords or developers are large enough to exert significant market power. The Table 1 showed the mechanism of housing market.

Housing is the largest asset of most American households, so the housing market profoundly affects the distribution of wealth; housing's location and tenure could well affect the behavior of its occupants. Hence, we work on how housing market works? In a good economic situation when house prices are consistently rising, most consumers can afford what is perceived to be full market value for a given property, because the inherent assumption is that value will continue to rise.

Regression Analysis and Dynamic Regression Models

For regression method, as shown in Fig. 1, we mainly measures how the land market value, in terms of GDP, CPI, construction cost index and unemployment rate, inflation rate, population and purchasing manage index, enables a relationship with the real estate market and investment.

Besides, we explore the quantitative and qualitative relationship among these economic variables at risk scenarios. Firstly, we developed two regression models with raw data and log transformation. After checking all the significance of all explanatory variables and residuals of autocorrelation, we obtain Equation 1 and 2. From the Equation 1, we may simply conclude that the land market value is highly related to IR, UR, CCI and PMI. When UR increased 1 unit and other variables keep unchanged, the land market value will decrease 218.85 million. For Equation 2, it involves more explanatory variables than Equation 1:

$$LMV = -8400.287 + 574.933IR - 218.85UR + 154.035CCI + 30.048PMI \quad (1)$$

Where:

$LMV = f(GDP, CPI, CCI, UR, IR, PP, PMI)$

LMV = Land market value (Aggregate market value of residential land)

GDP = Gross Domestic Product (the total value of goods and services within a nation over a period)

CPI = Consumer Price Index (a measure of the weighted average of prices of consumer goods and services)

CCI = Construction Cost Index (Expense incurred by a contractor)

UR = Unemployment Rate (a measure of the prevalence of unemployment)

IR = Inflation Rate (the percentage increase in the price of goods and services)

PP = Population (human beings in general or considered collectively)

PMI = Purchasing Manager Index (an indicator of the economic health of the manufacturing sector)

Considering the multicollinearity within the multiple regressions, we only dropped the variables which are highly related to other variables in the regression, such as GDP and PP in Equation 1. Because such variables GDP and PP are capturing the effect of other variables. However, unemployment usually indicate more economic distress and lower production which causes lower demand for economic purchases including land. Hence, it shows the negative relation to the Land Market Value. Because of the big sample size, we keep some insignificant variables in our model, the effects of those variables, PMI and UR, are negligible. Also, we get rid of the effects of multicollinearity for the selected variables in Equation 1:

$$LLMV = 82.146 - 18.029LPP + 4.2LCCI + 4.45LGDP + 0.505LUR = 0.545LPMI - 0.148LIR \quad (2)$$

The log transformation of the regression model can help stabilize the variance. Hence, we try to fit the log transformed model of Equation 2. The transformed model statistical summary shown below.

However, from the Table 2 and 3, we may conclude that there are some evidence of autocorrelation in the residuals of Equation 1 and 2, it indicates there are some information from the data. Hence, we extended the regression method into the general class of dynamic regression models, which simply combined regression models with ARIMA errors. We take into account several formula as the theoretical foundation:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \dots + \beta_k x_{k,t} + n_t \quad (3)$$

$$(1 - \phi_1 B - \phi_2 B^2) n_t = e_t \quad (4)$$

$$(1 - 1.5024B + 0.7229B^2)n_t = e_t \quad (5)$$

where, n_t denotes the errors from the regression models and e_t denotes the errors from the ARIMA model. Only the ARIMA model errors are assumed to be white noise.

$$n_t = 262.8185 + 1.5024n_{t-1} - 0.7229n_{t-2} + e_t \quad (6)$$

Table 2. Regression table for Equation 1

Model	Coefficients	Standard error	T-value	Significance	VIF	Tolerance
Constants	-8400.287	2857.725	-2.94	0.006		
IR	574.933	219.673	2.617	0.014	1.89	0.529
UR	-218.850	136.198	-1.607	0.119	1.098	0.911
CCI	154.035	12.589	12.236	0.00	1.635	0.612
PMI	30.048	37.207	0.808	0.426	1.37	0.726
Model Equation 1	R-squared 0.935	Adj R-squared 0.875	S.E of Est 1220.773	Sample size 34	F-change 50.570	Significance 0.00

Table 3. Regression table for Equation 2

Model	Coefficients	Standard error	T-value	Significance	VIF	Tolerance
Constants	-0.172	1.156	-0.149	0.883		
LIR	0.059	0.053	1.126	0.269	1.338	0.747
LUR	-0.281	0.132	-2.122	0.043	1.12	0.893
LCCI	2.169	0.12	18.088	0.00	1.299	0.770
LPMI	-0.085	0.245	-0.346	0.732	1.153	0.867
Model Equation 2	R-square 0.937	Adj R-squared 0.928	S.E of Est 0.17765	Sample Size 34	F-value 107.374	Significance 0.00

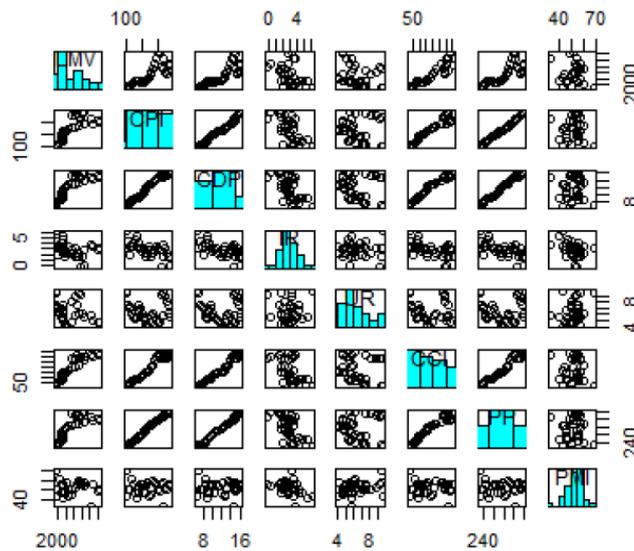


Fig. 1. LMV

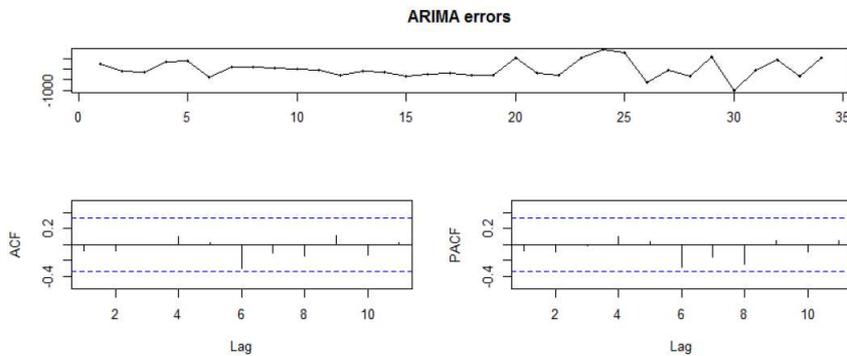


Fig. 2. Dynamic regression

Box.test X-squared df P-value
 Box.Ljung 14.0415 20 0.8284>0.05

From the plots and non-parametric tests of e_t in Fig. 2, we may conclude that the residuals of the AR (2) from the regression model is stationary. Compared to three regression equations, the dynamic method is better than others.

Time Series Decomposition and Smoothing Analysis

For time series decomposition models, we have a couple of options, such as the classical additive decomposition, classical multiplicative decomposition and STL decomposition. The classical decomposition is basic and simple way to forecast the trend. We employ the simple exponential smoothing method, holt’s linear method, exponential smoothing method and additive damped method and multiplicative damped method. Eventually, the exponential smoothing model could have better forecasting on the trend in Fig. 3.

For ETS(M,A,N) model, we take this into innovation by considering multiplicative error equations:

$$y_t = (l_{t-1} + b_{t-1})(1 + \varepsilon_t) \tag{7}$$

$$l_t = (l_{t-1} + b_{t-1})(1 + \alpha \varepsilon_t) \tag{8}$$

$$b_t = b_{t-1} + \beta(l_{t-1} + b_{t-1})\varepsilon_t \tag{9}$$

$$\varepsilon_t = \frac{y_t - (l_{t-1} + b_{t-1})}{l_{t-1} + b_{t-1}} \tag{10}$$

where the $\varepsilon \sim \text{NID}(0, \sigma^2)$, l_t denotes an estimate of the level of the series at time t , b_t denotes an estimate of the trend (slope) of the series at time t , α denotes the smoothing parameter for level. β denotes the smoothing parameter for the trend.

By the method of minimizing the “likelihood”. We estimate the smoothing parameters α , β , b and l . In our model, the estimated parameters are $\alpha = 0.9051$, $\beta = 0.9051$, $l = 836.1576$, $b = 388.3622$. The possible values that the smoothing parameters can take is restricted. Traditionally the parameters have been constrained to lie between 0 and 1 so that the equations can be interpreted as weighted averages. For the state space models, we have set $0 < \alpha < \beta < 1$.

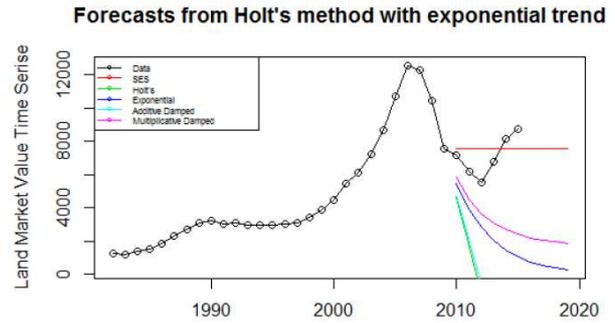


Fig. 3. Holt’s method

Time Series Analysis

In the mature economies LMV illustrate the importance of land as a source of wealth, but in rapidly growing economies land has an even more significant role in determining economic welfare and a host of incentives for the performance of the economy. From the time series graph, we may find that the American land market value shows stable increase from 1982 to 2004, but from 2005 this number increased dramatically and peaked in 2006,12.55. In fact, the economic crisis started in 2006 in USA, the economics crisis led to the increased interest, hence, the LMV rose rapidly. However, LMV decreased from 12.55 to 5.54 which is the relative lower level in 2012. The economy of USA experienced the great recession during this period. Until 2013, the situation recovered and this number rose to 7.594 in 2015.

Conflicting results are very common when performing forecasting competitions between methods. As forecasting tasks can vary by many dimensions (length of forecast horizon, size of test set, forecast error measures, frequency of data, etc.), it is unlikely that one method will be better than all others for all forecasting scenarios. What we require from a forecasting method are consistently sensible forecasts and these should be frequently evaluated against the task at hand.

Obviously, in Fig. 4, Land Market Value is an increasing time series dataset. We tried original data and log transformation data to fit the ARIMA model. The Table 4 showed all the potential models. Eventually, the Box-Ljung test of residual met the assumption of the non-parametric, it indicates that autocorrelation come from the white noise, but the Log transformation reduced the RMSE significantly, which is almost near to 0.1. Hence, we selected the Log ARIMA (2,0,1) for forecasting model.

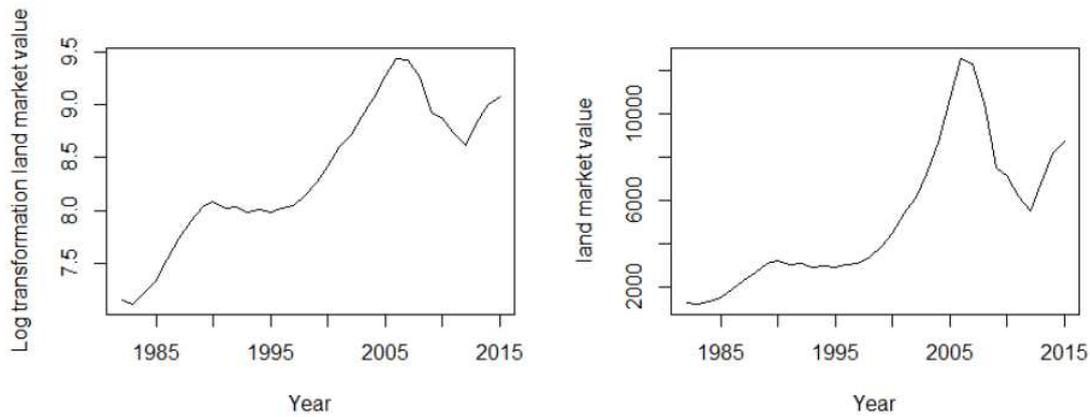


Fig. 4. tslmv

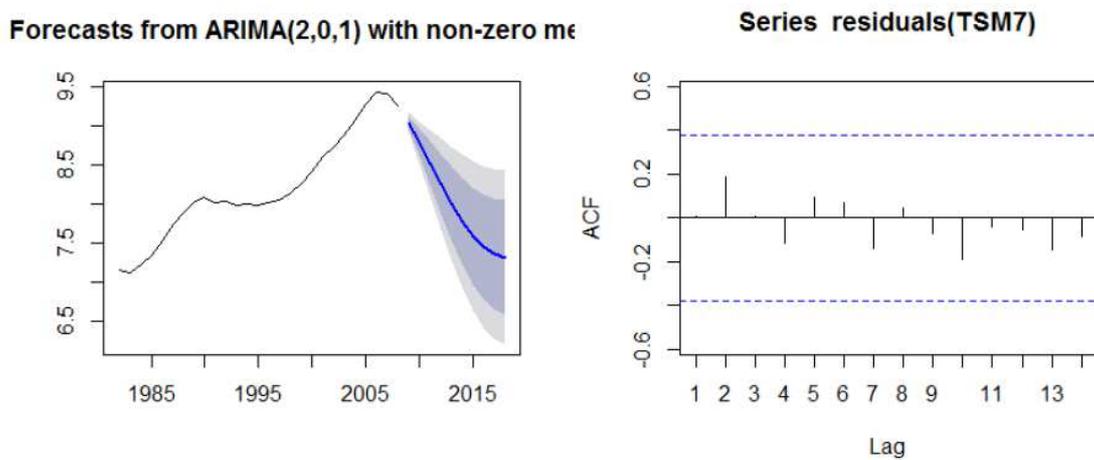


Fig. 5. TSLMV

Table 4. Comparison table

Model	ME	RMSE	MAE	MPE	MAPE	AIC
Regression	-1.103e-13	1127.403	918.1104	0.6624	21.1810	586.37193
Log Regression	-2.7e-17	0.1661	0.1489	-0.0371	1.7665	-12.84073
Regression with AR (2) errors	-5.2326	590.5238	424.728	-1.1753	7.9002	545.45
ETS (M,A,N)	-91.4374	617.2248	339.1251	-0.8840	408.4312	413.6146
ARIMA (1,1,0)	88.7736	713.7366	477.83	2.6121	7.973	532.83
Log ARIMA (2,0,1)	0.001	0.063	0.057	0.006	0.692	N/A

Table 5. Forecasting ARIMA Model

	Actual	Forecast	95% L B	95% U B
2009	8.928	9.041	8.917	9.165
2010	8.878	8.792	8.523	9.061
2011	8.73	8.526	8.1	8.952
2012	8.620	8.259	7.678	8.84
2013	8.821	8.006	7.281	8.731
2014	9.006	7.781	6.931	8.631
2015	9.075	7.594	6.642	8.546
2016	N/A	7.453	6.424	8.481
2017	N/A	7.361	6.28	8.442
2018	N/A	7.321	6.209	8.433

Conclusion and Outlook

As the Fig. 5 indicate that, for the data extending into 2015, Log ARIMA (2,0,1) is the best forecasting model among others. We mainly evaluate forecasting models based on the two performance measures of RMSE and AIC. As was the case with the forecast in Table 5, land market value is projected to continue increase in the following years. It shows the stable increase in the future. This number will significantly rise to 7.3 in 2018.

Land market value, both directly and indirectly, related to the housing market, commercial and residential buildings, construction industry and home price. The forecasting of land market value is more important and necessary for the economy of American, because the tendency of Land Market Value would be helpful for government and investor to examine the problem in housing market, make the appropriate policy and regulate the housing market. Thus, a given forecasting model did a good job of tracking the actual value of land market changes. On the other hand, forecasting techniques are widely used in the area of finance and housing market. As rapidly rising and housing prices are the hot topic in the growing number of metropolitans around the world. Most importantly, forecasting modeling is ever more significant in predicting the direction of future price.

Acknowledgment

We thank Dr. Longhofer for assistance with statistical modeling methodology and Sean Hennessy for comments that greatly improved the manuscript.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

- Bostic, R.W., D.L. Stanley and L.R. Christian, 2007. Land leverage: Decomposing home price dynamics. *Real Estate Econ.*, 35: 183-208.
 DOI: 10.1111/j.1540-6229.2007.00187.x
- Davis, K., 2009. Financial regulation after the global financial crisis. *Australian Econ. Rev.*, 42: 453-456.
 DOI: 10.1111/j.1467-8462.2009.00568.x
- Davis, M.A. and G.P. Michael, 2008. The price of residential land in Large US Cities. *J. Urban Econ.*, 63: 352-84.
 DOI: 10.1016/j.jue.2007.02.003
- Davis, M.A. and H. Jonathan, 2007. The price and quantity of residential land in the United States. *J. Monetary Econ.*, 54: 2595-2620.
 DOI: 10.1016/j.jmoneco.2007.06.023
- Sirmans, C.F. and B.A. Slade, 2009. National transaction-based land price indices. *J. Real Estate Finance Econ.*, 45: 829-845.
 DOI: 10.1007/s11146-011-9306-3

Appendix^a: 1982-2015 Land Market Value Datasets^b [H]

Year	LMV ^c	CPI	GDP ^d	IR	UR	CCI	PP	PMI	Year	LMV	CPI	GDP	IR	UR	CCI	PP	PMI
1982	1274.88	96.5	6.49	6.2	9.7	43.4	231.66	42.8	2000	4509.19	172.2	12.68	3.4	4	75.9	282.16	43.9
1983	1232.25	99.6	7	3.2	9.6	44.70	233.79	69.9	2001	5428.39	177.1	12.71	2.8	4.7	79.7	284.97	45.3
1984	1387.16	103.9	7.4	4.3	7.5	46.7	235.82	50.6	2002	6123.09	179.9	12.96	1.6	5.8	81.7	287.63	51.6
1985	1546.45	107.6	7.71	3.6	7.2	47.9	237.92	50.7	2003	7208.82	184.13	13.53	2.3	6	85.9	290.11	60.1
1986	1879.09	109.6	7.94	1.9	7	50.4	240.13	50.5	2004	8646.18	188.9	13.95	2.7	5.5	93.1	292.81	57.2
1987	2297.13	113.6	8.29	3.6	6.2	52.7	242.29	61	2005	10708.93	195.3	14.37	3.4	5.1	100	295.52	55.1
1988	2678.79	118.3	8.61	4.1	5.5	54.5	244.50	56	2006	12547.31	201.6	14.72	3.2	4.6	106	298.38	51.4
1989	3097.56	124.8	8.85	4.8	5.3	56.4	246.82	47.4	2007	12290.28	207.3	14.99	2.8	4.6	107	301.23	49
1990	3257.63	130.7	8.91	5.4	5.6	58	249.62	40.8	2008	10464.64	215.3	14.58	3.82	5.8	103.3	304.09	33.1
1991	3050.34	136.2	9.02	4.2	6.8	58.2	252.98	46.8	2009	7537.82	214.5	14.54	-0.32	9.3	98.10	306.77	55.3
1992	3089.8	140.3	9.41	3	7.5	58.9	256.51	54.2	2010	7173.83	218.1	14.94	1.64	9.6	96.4	309.35	57.5
1993	2948.23	114.5	9.65	3	6.9	61.8	259.92	55.6	2011	6184.28	224.9	15.19	3.14	8.9	97.4	311.72	53.1
1994	2995.76	148.2	10.05	2.6	6.1	64.6	263.13	56.1	2012	5543.56	229.6	15.43	2.08	8.1	98.4	314.11	50.4
1995	2945.05	152.4	10.28	2.8	5.6	67.3	266.28	46.2	2013	6777.04	233	15.92	1.46	7.4	104.8	316.5	56.5
1996	3033.87	156.9	10.74	3	5.4	68.6	269.39	55.2	2014	8152	237.2	16.29	1.61	6.2	111.8	318.86	55.1
1997	3120.62	160.5	11.21	2.3	4.9	70.6	272.65	54.5	2015	8737.11	242.1	16.3	0.1	5.5	100.37	320.99	53.5
1998	3437.02	163.11	11.77	1.6	4.5	72.5	275.85	46.8									
1999	3886.17	166.6	12.32	2.2	4.2	72.7	279.04	57.8									

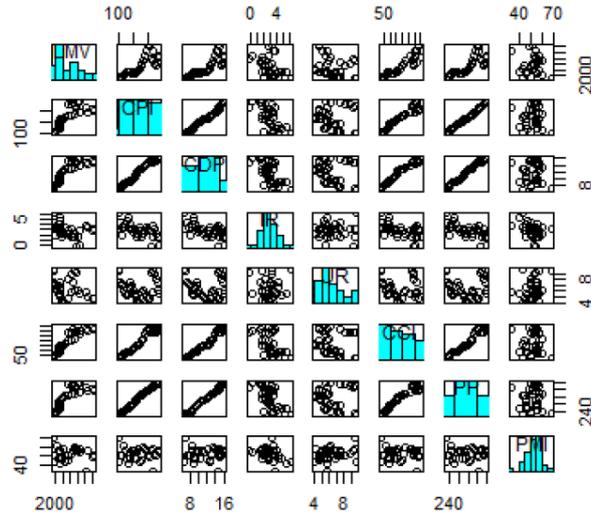
^aThe data was based on the 34 years' national data on past and present real estate transaction from 1982 to 2015.

^b<http://www.statista.com/statistics/188105/annual-gdp-of-the-united-states-since-1990/>

Source: U.S. Bureau of Labor Statistics <https://en.wikipedia.org/wiki/Main-Page>.

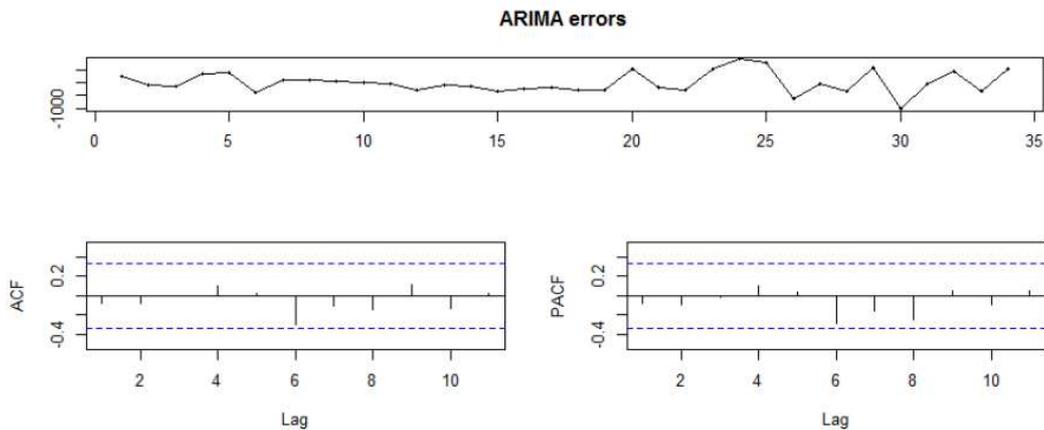
^cThe unit of land market value is million

^dThe unit of GDP is trillion.



```

pane l. his t <- function (x,...){
us r <- par ("us r")
on. exit (par (us r))
par (us r = c (us r [1: 2], 0, 1))
h $<- $ his t (x, plot = FALSE, breaks = "FD")
breaks <- h $ breaks
nB <- length (breaks)
y <- h $ counts
y <- y/max(y)
rect (breaks [-nB], 0, breaks [-1], y, col = "cyan", ...)
pairs (LMV. forecasting2[(2:9)], diag. panel = panel. his t)
fit <- Arima (LMV. forecasting3[1], xreg = LMV. forecasting3[2:8],
order = c (2, 0, 0))
ts display (arima. errors (fit), main = "ARIMA errors")
    
```



```

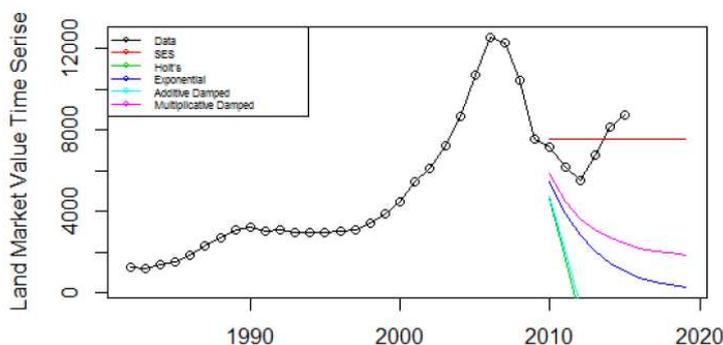
TSLMV1 <- window(TSLMV, start = 1982, end = 2008)
fit1 <- se s (TSLMV1)
fit2 <- holt (TSLMV1)
fit3 <- holt (TSLMV1, exponential = TRUE)
fit4 <- holt (TSLMV1, damped = TRUE)
fit5 <- holt (TSLMV1, exponential = TRUE, damped = TRUE)
plot (fit2$model$state)
plot (fit4$model$state)
    
```

```

plot (fit2$model$state)
  flwd = 1, plot.conf = FALSE)
lines (window(TSLMV, start = 2015), type = "o")
lines (fit1$mean, col = 2)
lines (fit2$mean, col = 3)
  lines (fit4$mean, col = 5)
  lines (fit5$mean, col = 6)
  legend ("topleft", lty = 1, pch = 1, col = 1:6,
    c ("Data", "SES", "Holt's", "Exponential",
      "Additive Damped", "Multiplicative Damped"), cex = 0.75)
fit0 <- ets(TSLMV1)
summary(fit0)
plot(forecast(fit0, h = 8),

```

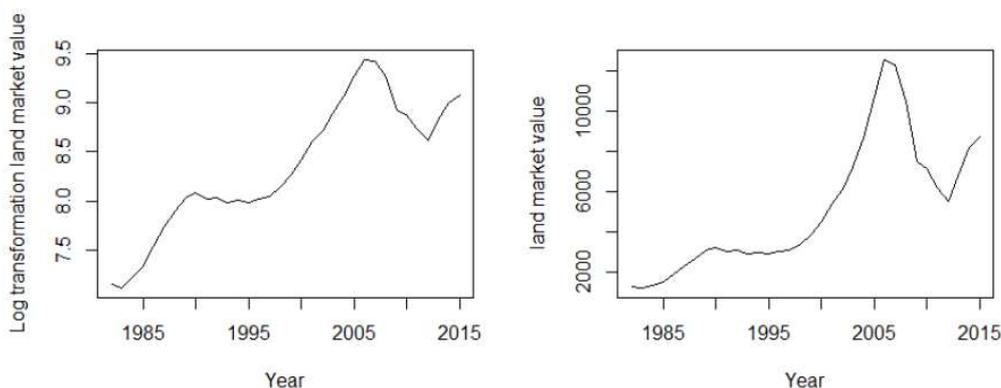
Forecasts from Holt's method with exponential trend



```

  ylab = "Lank Market Value (millions)")
fit1 <- lm(LMV~IR+UR+CCI+PMI, data = LMV. forecasting3)
summary(fit1)
accuracy(fit1)
fit2 <- lm(LLMV ~ LIR+LUR+LCCI+LPMI, data = LMV. forecasting3)
summary(fit2)
accuracy(fit2)
Box.test (residuals (fit), fit df = 5, lag = 10, type = "Ljung")

```

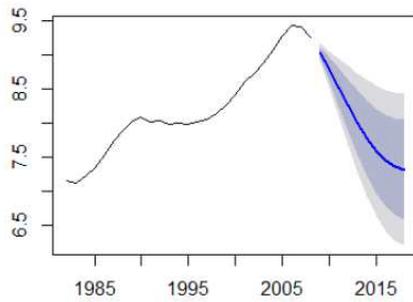


```

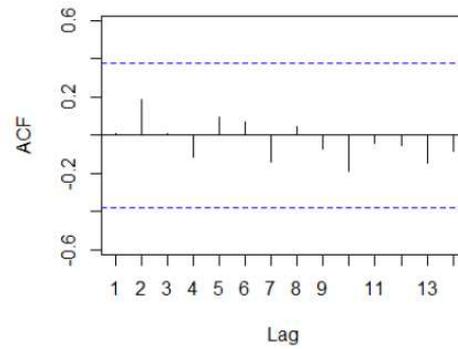
TSLMV <- ts (LMV, start = 1982, frequency = 1)
LTSLMV <- log (TSLMV)
  par (mfrow = c (1, 2))
plot (LTSLMV, ylab = "Log trans formation land market value", xlab = "Year")
plot (TSLMV, ylab = "land market value", xlab = "Year")

```

Forecasts from ARIMA(2,0,1) with non-zero me



Series residuals(TSM7)



```
LTSLMV1 <- window(LTSLMV, start = 1982, end = 2008)
TSM1 <- arima(TSLMV, order = c(1,1,0))
Acf(residuals(TSM1))
summary(TSM1)
accuracy(TSM1)
forecast(TSM1)
plot(forecast(TSM1))
TSM7 <- arima(LTSLMV1, order = c(2,0,1))
Acf(residuals(TSM7))
summary(TSM7)
accuracy(TSM7)
forecast(TSM7)
plot(forecast(TSM7))
```