

A VAR Model for Forecasting Land Market Value in USA

Lei Wang

Department of the Mathematics, The University of Southern Mississippi, United States

Article history

Received: 26-10-2017

Revised: 21-12-2017

Accepted: 29-12-2017

Email: Fiona901587@yahoo.com

Abstract: Forecasts of the future tendency of economic variables such as GDP, inflation rate and unemployment rate, arise many interests from business and government. Also, Modeling the land market at the national level can capture rich dynamic presenting in interdependent economies. In this paper, we studied a Vector Auto-regression (VAR) of Land Market Value and five US macroeconomic variables. We employed the VAR model for forecasting Land Market Value in USA and analyzed annual data on the main macroeconomic variables of interest going back to 1982. Most importantly, we explore the mutual influence between Land market value and selected macroeconomics variables to enable government and investor to make informed decision regarding real estate market.

Keywords: VAR Model, Land Market Value, Multivariate Time Series, Analysis

Introduction

We can define land in many ways and different fields. In the economics sense, we defined land as all naturally occurring resources whose supply is inherently fixed. All types of economic activity require land, either directly or indirectly. The direct use of land is obvious in industries such as farming and construction. But all other forms of commerce ultimately require land as well because workers, equipment and buildings need to be located somewhere. What we studied in this paper is about the land market value. The Land Market Value, defined as the total value of land 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 (Calomiris *et al.*, 2013).

According to Sims (1980), he pioneered Vector auto-regression models, after that this approach have become widely used in applied economic research. In addition, Robertson and Tallman (1999) focused on a VAR model fitted to monthly data and staggered release dates that uses a distributed monthly estimate of quarterly GDP data in the published paper. Accordingly, Crawford and Fratantoni (2003) adopted a Markov-switching (MS) model to house prices in five states and compare forecasting results to linear auto-regressive-moving average and generalized AR conditional heteroscedastic model. In particular, Miles (2008) tried to compare the forecasting performance of linear and Non-linear models of house prices. Kuminoff and Pope (2013) used data on single-family home sales to estimate the land value by hedonic model. Recently, Calomiris *et al.* (2013) applied panel VAR model for

quarterly state-level data indicate that price-foreclosure linkages run both ways. In a range of applications we show that these series can shed light on trends, fluctuations and variation in the market value of land.

VAR Model

There are a variety of methods available for forecasting economic variables. One common type of forecast is Vector auto-regression modeling for multivariate Time Series approach. This type of forecast is predominantly in economics and financial analysis.

A VAR model is an useful and flexible approach to describe the dynamic behavior of economic activity and financial time series dataset; that is, a vector of time series. In this system, we consider one equation for one variable as dependent variable with constant and lags. Each variable is assumed to influence with each other in the system, which makes direct interpretation of the estimated coefficients very difficult (Hyndman and Athanasopoulos, 2014).

We write a multi-dimensional VAR (p) as:

$$Y_t = C + \Phi_1 \begin{bmatrix} LLMV_{t-1} \\ LCPI_{t-1} \\ LUR_{t-1} \\ LPP_{t-1} \\ LCCI_{t-1} \\ LPMI_{t-1} \end{bmatrix} + \Phi_2 \begin{bmatrix} LLMV_{t-2} \\ LCPI_{t-2} \\ LUR_{t-2} \\ LPP_{t-2} \\ LCCI_{t-2} \\ LPMI_{t-2} \end{bmatrix} + \dots + \Phi_{t-p} \begin{bmatrix} LLMV_{t-p} \\ LCPI_{t-p} \\ LUR_{t-p} \\ LPP_{t-p} \\ LCCI_{t-p} \\ LPMI_{t-p} \end{bmatrix} + a_t \quad (1)$$

where, a_t are white noise process. $E(a_t) = 0$ and:

$$E(a_t, a_t') = \begin{cases} 0 & \text{when } t = \tau, \\ \Omega & \text{when } t \neq \tau, \end{cases}$$

In the reduced form, we will include a six variable VAR with one lag in our forecasting model:

$$Y_t = \begin{bmatrix} LLMV_t \\ LCPI_t \\ LUI_t \\ LCCI_t \\ LPP_t \\ LPMI_t \end{bmatrix} \Phi_1 = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} & \phi_{14} & \phi_{15} & \phi_{16} \\ \phi_{21} & \phi_{22} & \phi_{23} & \phi_{24} & \phi_{25} & \phi_{26} \\ \phi_{31} & \phi_{32} & \phi_{33} & \phi_{34} & \phi_{35} & \phi_{36} \\ \phi_{41} & \phi_{42} & \phi_{43} & \phi_{44} & \phi_{45} & \phi_{46} \\ \phi_{51} & \phi_{52} & \phi_{53} & \phi_{54} & \phi_{55} & \phi_{56} \\ \phi_{61} & \phi_{62} & \phi_{63} & \phi_{64} & \phi_{65} & \phi_{66} \end{bmatrix}$$

Coefficient ϕ_{ii} indicates the influence of the i th lag of variable Y_i on itself, while coefficient ϕ_{ij} indicates the influence of the i th lag of variable Y_j on Y_i .

A "VAR in levels" is known as the series modeled are stationary, we forecast them directly by fitting a VAR to the data. A "VAR in differences" is known as the series are non-stationary, we firstly take differences to make them stationary and then we fit a VAR model. In both cases, the models and coefficients are estimated equation by equation using the principle of least squares.

We applied the VAR selection package for forecasting the raw data. The function returns information criteria and final prediction error for sequential increasing the lag order up to a VAR(p)-process which are based on the same sample size. For each equation, the parameters are estimated by minimizing the sum of squared $e_{i,t}$ values.

Before we ran the R software, we take the log transformation of the raw data to stabilize the variance. And then, we set the 80% of the data as training set and the remaining data as the test set (Brockwell and Davis, 2002). The statistical summary of transformed model shown below:

Next, we can check the stability conditions for VAR Systems by the value of Roots of the characteristic polynomial. Because they are all less than 1 in Table 1, we could conclude that the VAR (1) model is stationary.

Firstly, if a VAR (2) model is estimated. The null hypothesis of no autocorrelation is rejected since the p-value of 0.03325 is lower than the significance level of 0.05. Since autocorrelation is an undesirable

feature of the model, we moves on to look for another model that does not have autocorrelation. As shown in Table 1, we estimates a VAR (1) model, tests for autocorrelation and finds that the null of no autocorrelation cannot be rejected because the p-value of 0.3429 is greater than the significance level of 0.05. Since there is not enough evidence of presence of autocorrelation, we satisfied and sticks to the VAR (1) model.

A portmanteau test is used for autocorrelation in errors:

H_0 : There is no evidence show that there are autocorrelation in residuals for some lag p .

H_1 : There are some evidence show that there are autocorrelation in residuals for some lag p :

$$LLMV = 1.198LLMV_{t-1} - 0.341LUR_{t-1} + LCCI_{t-1} + 7.556LPP_{t-1} - 1.496LPMI_{t-1} - 1.886LCPI_{t-1} - 29.122$$

$$LCCI = 1.198LLMV_{t-1} - 0.341LUR_{t-1} + LCCI_{t-1} + 7.556LPP_{t-1} - 1.496LPMI_{t-1} - 1.886LCPI_{t-1} - 29.122$$

When we use a VAR to forecast, we have to make decision on the number of variables (denoted by K) and the number of lags (denoted by p). The number of coefficients to be estimated in a VAR is equal to $K + pK^2$ (or $1 + pK$ per equation). For example, in our VAR model with $K = 6$ variables and $p = 1$ lags, there are 7 coefficients per equation making for a total of 42 coefficients to be estimated. The more coefficients to be estimated the larger the estimation error entering the forecast.

In Table 2 and 3, We list the forecasting model for LLMV and LCCI and ignore the other variables, because we select the variables we interests regarding real estate market. In Table 2, noticed that the p value are extremely small. It indicates that LLMV can be estimated by the $LLMV_{t-1}$, $LCPI_{t-1}$, LPP_{t-1} , $LPMI_{t-1}$, $LCCI_{t-1}$ and LUR_{t-1} . The forecasting model fits very well for LLMV. Our goal is to forecast the tendency of Land market value to provide the information for policy-maker or decision-maker. The construction cost index is highly related to real estate market. In table 2, the forecasting performance for LCCI is influenced by the $LLMV_{t-1}$, LPP_{t-1} , $LPMI_{t-1}$ and $LCCI_{t-1}$. So we mainly studied the performance of the two variables in our forecasting model. Also, the performance of forecasting of other variables are not good enough, so we will not discuss them in this study.

Table 1: VAR estimation results for VAR (1)

Endogenous variables: LLMV, LUR, LPP, LCPI, LPMI, LCCI			
Deterministic variables: Const			
Sample size: 28			
Log Likelihood: 472.079			
Roots of the characteristic polynomial: 0.9685 0.9403 0.9403 0.807 0.5849 0.5849			
Residuals of VAR object VAR	Chi-squared	DF	P-value
Portmanteau test (asymptotic)	333.73	324	0.3429

Table 2: VAR forecasting model for LLMV

Model	Estimate	Standard Error	T-value	Significance
LLMV.11	1.198100	0.140200	8.548	4.13e-08 ***
LCPI.11	-1.886000	0.624300	-3.021	0.006750 **
LPP.11	7.555600	2.253200	3.353	0.003164 **
LPML.11	0.007910	0.001851	4.273	0.000514 ***
LCCI.11	0.031452	0.007257	4.334	0.000450 ***
LUR.11	-0.341100	0.104500	-3.265	0.003874 **
Constant	-29.122300	8.733700	-3.334	0.003304 **

Table 3: VAR forecasting model for LCCI

Model	Estimate	Standard error	T-value	Significance	
LLMV.11	0.077	0.034	2.255	0.035 *	
LCPI.11	-0.15	0.152	-0.989	0.335	
LPP.11	1.418	0.548	2.586	0.018 *	
LPML.11	0.123	0.027	4.475	0.000232 ***	
LCCI.11	0.480	0.177	2.711	0.0135*	
LUR.11	-0.0047	0.025	-0.183	0.857	
Constant	-6.053	2.126	-2.848	0.00995 **	
Mul R-squared	Adj R-squared	Residual.s.e	S. Size	F-stat	P-value
0.9969	0.9951	0.04333 on 17 df	29	544.6on10	< 2.2e-16

Forecasting

VAR model generate the forecasting in a recursive structure. The VAR is a system in which each variable is regressed on a constant and p of its own lags as well as on p lags of each of the other variables in the VAR. To illustrate the process, assume that we have fitted the multi-dimensional VAR (1) described in equations Equation (1) for all observations up to time T .

Then the one-step-ahead forecasts are generated by:

$$\hat{y}_{1,T+1|T} = \hat{c}_1 + \hat{\phi}_{11,1}y_{1,T} + \hat{\phi}_{12,1}y_{2,T} \quad (2)$$

$$\hat{y}_{2,T+1|T} = \hat{c}_1 + \hat{\phi}_{21,1}y_{1,T} + \hat{\phi}_{22,1}y_{2,T} \quad (3)$$

$$\hat{y}_{3,T+1|T} = \hat{c}_1 + \hat{\phi}_{31,1}y_{1,T} + \hat{\phi}_{32,1}y_{2,T} \quad (4)$$

$$\hat{y}_{4,T+1|T} = \hat{c}_1 + \hat{\phi}_{41,1}y_{1,T} + \hat{\phi}_{42,1}y_{4,T} \quad (5)$$

$$\hat{y}_{5,T+1|T} = \hat{c}_1 + \hat{\phi}_{51,1}y_{1,T} + \hat{\phi}_{52,1}y_{5,T} \quad (6)$$

$$\hat{y}_{6,T+1|T} = \hat{c}_1 + \hat{\phi}_{61,1}y_{1,T} + \hat{\phi}_{62,1}y_{6,T} \quad (7)$$

This is the same form as Equation (2) to (7) except that the errors have been set to zero and parameters have been replaced with their estimates.

LMV

From the Fig. 1, 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, 12550 million. In fact, the economic crisis started in 2006 in USA, the economics crisis led to

the increased interest. Hence, the LMV rose rapidly. The economy of USA experienced the great recession during this period. Until 2013, the situation recovered and this number rose to 8737.11 million in 2015. The Table 4 shows that the Land Market Value will increase in the following years. This number will rise to 9.85 in 2020.

CPI

The CPI is used to adjust income. When the CPI increased rapidly, wages have to increase eventually. The Bureau of Labor and Statistic (BLS) uses the CPI to adjust wages, retirement benefits, tax brackets and other important economic indicators. Especially, in 2009, the growth of CPI is -0.4, because of the crises influence, the CPI decreased from 215.3 to 214.5.

UR

In the US, the unemployment rate started to increase in the summer 2007, from a level of 4:6% of the labor force (June 2007). It doubled in less than two years and reached a peak of 9:6% in 2009. Since then, it has been very gradually diminishing and was recorded at 8:9% in 2011. The unemployment rate, which rose from 5.5 percent in 2004, is at its highest level since September 2004.

Population

When population growth is strong and long-lasting this encourages development to occur and new housing is build to cope with the increase in demand. To understand this better, we can find it in the graph that the population has a general rise during this period. In 2000, the Population in USA is approximately 280 million, by 2015 this number has reached approximately 321 million in 2015. During this period, the population has increased 14:6%.

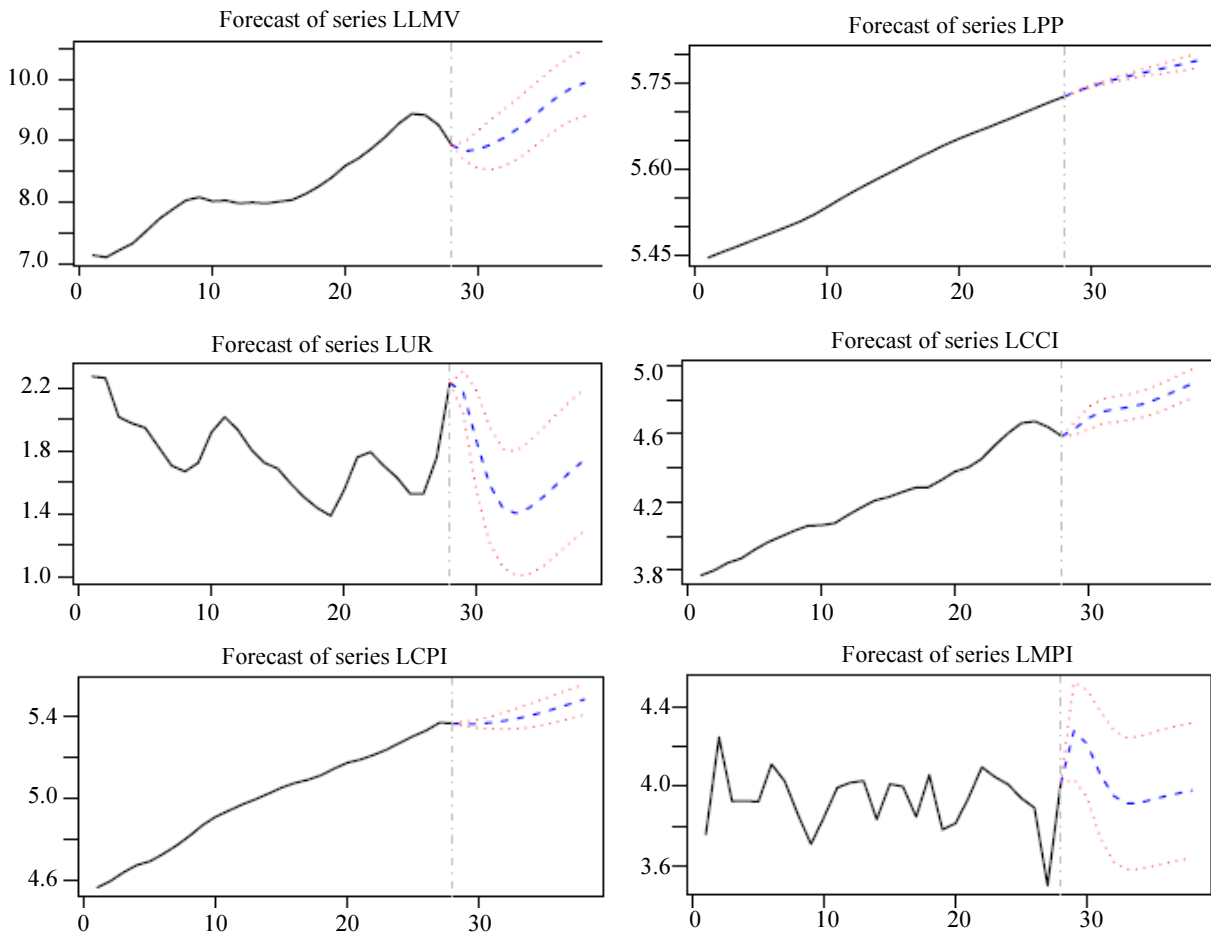


Fig. 1: Forecasting

Table 4: Forecasting VAR Model for LLMV

	Actual	Forecast	95% LB	95% UB	CI
2010	8.878	8.833	8.674	8.992	0.159
2011	8.730	8.863	8.556	9.171	0.308
2012	8.620	8.945	8.538	9.352	0.407
2013	8.821	9.058	8.595	9.522	0.463
2014	9.006	9.203	8.711	9.695	0.492
2015	9.075	9.374	8.871	9.878	0.504
2016	9.085	9.554	9.047	10.062	0.508
2017	N/A	9.72	9.208	10.232	0.512
2018	N/A	9.853	9.330	10.375	0.523
2019	N/A	9.938	9.400	10.476	0.538

As the case of forecasting, the prediction of population fits very well of the actual value. The population of USA is projected to increase to 5.78 in 2018.

CCI

The Fig. 1 shows the national construction cost index increase from 2010 to 2015, from low 96.48 to just 100.37. Our research shows a 5.5% increase in the national average in construction cost from that January 2014 and December 2014.

PMI

The usefulness of the PMI as an early signal of changes in manufacturing output and GDP. Therefore, an index 50.0 means that the variable is unchanged, a number over 50.0 indicates an improvement while anything below 50.0 suggests a decline. The further away from 50.0 the index is, the stronger the change over the month. For example PMI of 53.5 points in 2015 to a stronger increase in a variable than 50.4 in 2012. In

this forecasting model, the trend of PMI will remain around 54 in the next decades.

Conclusion

The VAR forecasting model has been widely used in many area of finance in recent years and it increased the understanding of tendency of land market value. Compared to separate univariate models, the VAR models have the advantage over traditional large-scale macroeconomics variables, but are easily interpreted and available. However, VAR models have also been much criticized, but the criticism usually refers to particular applications and interpretations of empirical results, rather than the methodology itself. For example, if a VAR model deal with a risk as the longer the lags, it will estimate the greater the number of parameters and the fewer the degrees of freedom.

As noted, Land is nature's gift to mankind, which enables life to continue and prosper. Because of its uniqueness of fixed supply and immobility, housing and land are more important for the economy that at any point in recent memory, the better housing forecasts results will be useful and necessary for USA real estate market. The tremendous rise in house prices over the decade has been both a national and global phenomenon. In addition, Land market value, both directly and indirectly, related to the housing market, commercial and residential buildings, commercial banks, construction industry, job-hunting market and home price. Therefore, the improved forecasting promise important benefits for any parties exposed to housing market (Miles, 2008). In other hand, how to measure the forecasting accuracy is also our interest and conclusion from this paper. Measuring the forecasting accuracy is an efficient way to select a better model for the vector time series datasets. Forecasting evaluation is relevant to the decision-maker when choosing on a model specification for subsequent use. The preference or loss function of forecast evaluations the accuracy measures are some forms of average error, typically root mean squared error or mean absolute error, but many other possibilities are available.

The VAR package contains the function VAR select for selection of lags p by four different information criteria: AIC, HQ, SC and FPE. We have met the AIC before, Akaike information criterion (AIC) also created by Akaike. SC stands for Schwarz Criterion after Gideon Schwarz who proposed it. It is also simply another name for Bayesian Information Criterion (BIC). HQ is the Hannan-Quinn criterion and FPE is the "Final Prediction Error" criterion. The criterion of AIC is usually used to choose large numbers of lags. Instead, for VAR models, we prefer to use the AIC and BIC (Wei, 2006).

Based on the same sample size and the scope of different information criterion and the final prediction error are computed:

$$AIC(p) = \ln(|S(p)|) + \frac{2pm^2}{n}$$

$$HQ(p) = \ln(|S(p)|) + \frac{2\ln(\ln(n))pm}{n}$$

$$SC(p) = \ln(|S(p)|) + \frac{2\ln(n)pm}{n}$$

$$FPE(p) = \left(\frac{n+p^*}{n-p^*}\right)^m (|S(p)|)$$

where, $|S(p)|$ is the residual sum of squares and cross-products. P^* is the total number of parameters in each equation. p assigns the lag order.

VAR select (mydatats , lag .max = 8, type = "cons t") \$ selection

AIC(n)	HQ(n)	SC(n)	FPE(n)
4	4	4	3

As you can see, the four criteria are so similar. According to the AIC,HQ,SC the optimal lag number is $p = 6$, whereas the FPE criterion indicates $p = 3$. We estimated for one lag of VAR including a constant and a trend as deterministic regressors and conducted diagnostic tests with respect to the residuals.

Acknowledgement

We thank Dr.cheng for assistance with statistical modeling methodology so that I can improve my paper with some novelty. Also I am grateful for comments and datasets from Sean Hennessy.

Ethics

In this study, I take into account a couple of ethical principles, such as obtain informed consent from potential research participants, minimize the risk of harm to participants, protect their anonymity and confidentiality, avoid using deceptive practices and give participants the right to withdraw from your research. This article meet these five ethical principles and their practical implications when carrying out the research.

References

- Brockwell, P.J. and R.A. Davis, 2002. Introduction to Time Series and Forecasting. 2nd Edn., Springer International Publishing, New York, ISBN-10: 978-3-319-29854-2.
- Calomiris, C.W., S.D. Longhofer and W.R. Miles, 2013. The foreclosure-house price nexus: A panel VAR model for U.S. States, 1981-2009. Real Estate Econom., 41: 709-746. DOI: 10.1111/reec.12011

- Crawford, G.W. and M.C. Fratantoni, 2003. Assessing the forecasting performance of regime-switching, arima and garch models of house prices. *Real Estate Econom.*, 31: 223-243.
 DOI: 10.1111/1540-6229.00064
- Hyndman, R.J. and G. Athanasopoulos, 2014. *Forecasting: Principles and Practice*. 1st Edn., OTexts, ISBN-10: 0987507109, pp: 291.
- Kuminoff, N.V. and J.C. Pope, 2013. The value of residential land and structures during the great housing boom and bust. *Land Econom.*, 89: 1-29.
 DOI: 10.3368/le.89.1.1
- Miles, W., 2008. Boom-bust cycles and the forecasting performance of linear and non-linear models of house prices. *J. Real Estate Finance Econom.*, 36: 249-264. DOI: 10.1007/s11146-007-9067-1
- Robertson, J.C. and E.W. Tallman, 1999. Vector autoregressions: Forecasting and reality. *Federal Reserve Bank Atlanta Econom. Rev.*, 84: 4-18.
- Sims, C.A., 1980. Macroeconomics and reality. *Econometrica*, 48: 1-48. DOI: 10.2307/1912017
- Wei, W., 2006. *Time Series Analysis: Univariate and Multivariate Methods*. 2nd Edn., Pearson Addison Wesley, ISBN-10: 0321322169, pp: 614.

```
> VAR select (mydata7, lag .max = 8, type = "cons t ") $ selection
AIC(n) HQ(n) SC(n) FPE(n)
4      4      4      3
> var <- VAR(mydata7 , p = 2, type = "cons t ")
> serial . test ( var , lags . pt = 10, type = "PT. asymptot ic ")
Portmanteau Test (asymptotic)
data: Residuals of VAR object var
Chi-squared = 3 3 3 . 5 9 , df = 288 , p-value = 0.03325
> var <- VAR(mydata7 , p = 1, type = "cons t ")
> serial . test ( var , lags . pt = 10, type = "PT. asymptot ic")
> pre <- predict ( var )
> plot ( pre )
> print ( pre )
```

Appendix^a: 1982-2015 land market value datasets^b

Year	LMV ^b	CPI	GDP ^c	IR	UR	CCI	PP	PMI	Year	LMV	CPI	GDP ^d	IR	UR	CCI	PP	PMI
1982	1274.88	96.5	6.490	6.2	9.7	43.40	231.66	42.8	2000	4509.190	172.20	12.68	3.4	4.0	75.90	282.16	43.9
1983	1232.25	99.6	7.000	3.2	9.6	44.70	233.79	69.9	2001	5428.300	177.10	12.71	2.8	4.7	79.70	284.97	45.3
1984	1387.16	103.9	7.400	4.3	7.5	46.70	235.82	50.6	2002	6123.000	179.90	12.96	1.6	5.8	81.70	287.63	51.6
1985	1546.45	107.6	7.710	3.6	7.2	47.90	237.92	50.7	2003	7208.820	184.13	13.53	2.3	6.0	85.90	290.11	60.1
1986	1879.09	109.6	7.940	1.9	7.0	50.40	240.13	50.5	2004	8646.180	188.90	13.95	2.7	5.5	93.10	292.81	57.2
1987	2297.13	113.6	8.290	3.6	6.2	52.70	242.29	61.0	2005	10708.93	195.30	14.37	3.4	5.1	100.00	295.52	55.1
1988	2678.79	118.3	8.610	4.1	5.5	54.50	244.50	56.0	2006	12547.31	201.60	14.72	3.2	4.6	106.00	298.38	51.4
1989	3097.56	124.8	8.850	4.8	5.3	56.40	246.82	47.4	2007	12290.28	207.30	14.99	2.8	4.6	107.00	301.23	49.0
1990	3257.63	130.7	8.910	5.4	5.6	58.00	249.62	40.8	2008	10464.64	215.30	14.58	3.82	5.8	103.30	304.09	33.1
1991	3050.34	136.2	9.020	4.2	6.8	58.20	252.98	46.8	2009	7537.820	214.50	14.54	-0.32	9.3	98.10	306.77	55.3
1992	3089.80	140.3	9.410	3.0	7.5	58.90	256.51	54.2	2010	7173.830	218.10	14.94	1.64	9.6	96.40	309.35	57.5
1993	2948.23	114.5	9.650	3.0	6.9	61.80	259.92	55.6	2011	6184.280	224.90	15.19	3.14	8.9	97.40	311.72	53.1
1994	2995.76	148.2	10.05	2.6	6.1	64.60	263.13	56.1	2012	5543.560	229.60	15.43	2.08	8.1	98.40	314.11	50.4
1995	2945.05	152.4	10.28	2.8	5.6	67.30	266.28	46.2	2013	6777.040	233.00	15.92	1.46	7.4	104.80	316.50	56.5
1996	3033.87	156.9	10.74	3.0	5.4	68.60	269.39	55.2	2014	8152.000	237.20	16.29	1.61	6.2	111.80	318.86	55.1
1997	3120.62	160.5	11.21	2.3	4.9	70.60	272.65	54.5	2015	8737.110	242.10	16.30	0.10	5.5	100.37	320.99	53.5
1998	3437.02	163.11	11.77	1.6	4.5	72.50	275.85	46.8									
1999	3886.17	166.6	12.32	2.2	4.2	72.70	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

^bThe unit of land market value is million

^cThe unit of GDP is trillion

^d<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>