SESUG Paper 141-2025

Forecasting U.S. Home Vacancy Rate using Arima

Joseph Maina and Abdelmoneam Jornaz

Park University

Abstract

The U.S. home vacancy rate is one of the key indicators of housing market health that shows the unoccupied resident units, which tentatively impacts real estate, economic policies, and related industries. This is among the main contributors to real estate growth. Accurate forecasting of vacancy rate is necessary and crucial for stakeholders in order to anticipate market shifts and have a strategized plan through decision-making.

This study aims to predict the home vacancy rate in the U.S. using a dataset that was gathered from the Federal Reserve Bank of St. Louis over the period of 1986 to 2024 by applying the autoregressive integrated moving average (ARIMA) model.

The ARIMA (2,0,1) model was identified as the best-fitting model, with a forecast of 1.09% for 2025, rising toward 1.66% by 2029.

Keyword: home vacancy rate, ARIMA Model, akaike information criterion (AIC), root mean square error (RMSE).

Introduction:

For many individuals and families, housing vacancy is one of the most important indicators of market dynamics and properties. Vacancy rate not only provide insight into supply and demand but also serve as an attractive metric for investors, especially in international metropolises (Chen, Ong, Zheng, & Hsu, 2017). The development of the housing market can highly influence the economic activities of a country or even the entire world. Hence, the healthy development of a housing market is one of the most important factors strongly associated with the sustainable development of the economy. The housing market is not just an important issue for individuals and businesses. Therefore, vacancy rate prediction has attracted significant attention from various fields, including economy, politics, computer science, etc. (Iacoviello & Minetti, 2008; Park & Bae, 2015; Selim, 2009; Tsai, 2013). However, the main challenge is that the housing vacancy prediction is a nonlinear time series forecasting issue for complex systems, which is influenced by various factors, including economic factors (such as unemployment rate, interest rate, etc.) and crisis events (such as the pandemic).

This paper aims to predict the home vacancy rate, build predictive models using time series techniques (ARIMA), and forecast vacancy rate for the next five years, which will allow decision makers to create informed choices concerning future household decisions.

Data Description

The home vacancy rate data were sourced from the Federal Reserve Economic Data which is a database maintained by the Research Division of the Federal Reserve Bank of St. Louis. The dataset spans from 1986 to 2024 with 39 annual observations.

Analysis and Results

- **Exploratory Data Analysis**

The data analysis was performed using SAS Studio. Exploratory data analysis (EDA) was used to explore and understand the data features.

```
proc sgplot data=HVAC;
    series x=Date y=Vacancy;
    yaxis min=0 max=3 label="Home Vacancy Rate (%)";
    xaxis label="Year";
run;
```

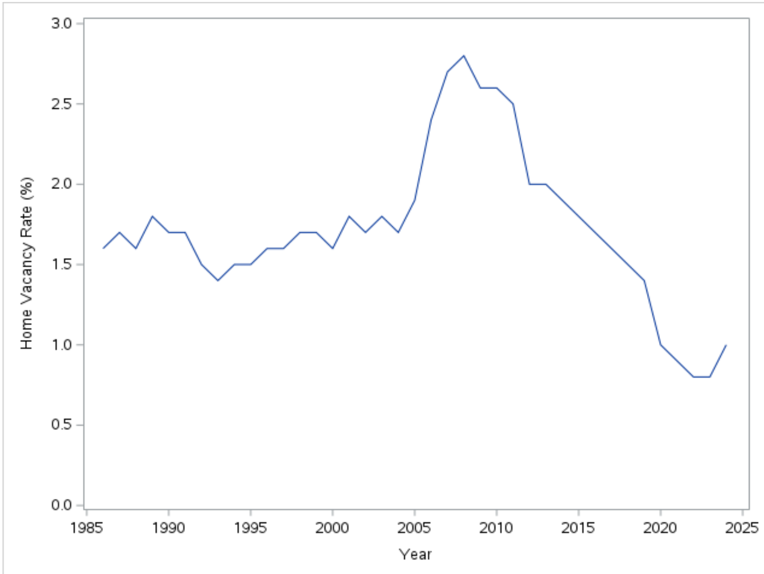

Figure 1: The U.S. Home Vacancy Rate (1986 - 2024)

FigureFigure 1 shows the U.S. home vacancy rate recorded the highest vacancy rate in 2008 (2.8%) during the recession, and the lowest vacancy rate after the pandemic in 2022 (0.8%). Table 1 presents the important statistical characteristics of the home vacancy rate dataset.

```
proc means data=WORK.HVAC n min max mean Median std;
    Var Vacancy;
run;
```

Table 1: Descriptive Statistics of the Home Vacancy Rate

Analysis Variable: Vacancy					
N	**Minimum**	**Maximum**	**Mean**	**Median**	**Std Dev**
39	0.8000000	2.8000000	1.7205128	1.7000000	0.4835066

- **Model Identification**

Based on the results of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) that are presented in figure 2, lags 1 and 2 in the ACF plot were significantly positively associated with the series, and lag 1 in the PACF plot was significantly negatively associated with the series. Several ARIMA models were applied to model the dataset, and the ARIMA (2,0,1) performed the best.

```
proc arima data=HVAC;
    identify var=Vacancy ;
    estimate p=2 q=1;
    forecast lead=5 out=forecastdata;
run;
```

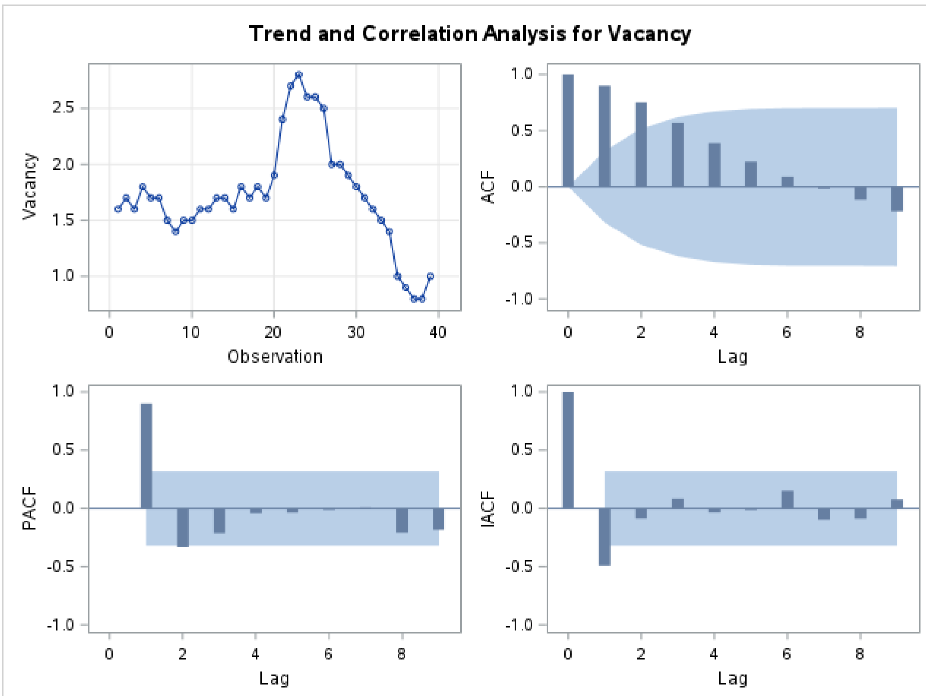

Figure 2: Trend and Correlation Analysis of the Home Vacancy Rate

The parameter estimates using the Conditional Least Squares (CLS) method showed that both the autoregressive and moving average terms were statistically significant.

Table 2: Conditional Least Squares Estimates

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	1.76396	0.08689	20.3	<.0001	0

MA1,1	1	0.1391	7.19	<.0001	1
AR1,1	1.9561	0.05958	32.83	<.0001	1
AR1,2	-0.99567	0.0563	-17.69	<.0001	2

- **Model Evaluation Metrics**

The model evaluation data indicated that ARIMA (2,0,1) had the lowest Akaike Information Criterion (AIC = -29.77), Schwarz Bayesian Criterion (SBC = -23.12), and Root Mean Square Error (RMSE = 0.15) compared to other ARIMA models, suggesting it provided the best balance between model accuracy and complexity.

- **Residual Diagnostics**

Residual analysis confirmed that the model was adequate. The ACF and PACF for the residuals up to 9 lags were not significant (all p-values > 0.05), indicating that no significant autocorrelation remained in the residuals. This means that the model has captured the pattern in the dataset appropriately, and the residuals are considered white noise (Figure 3).

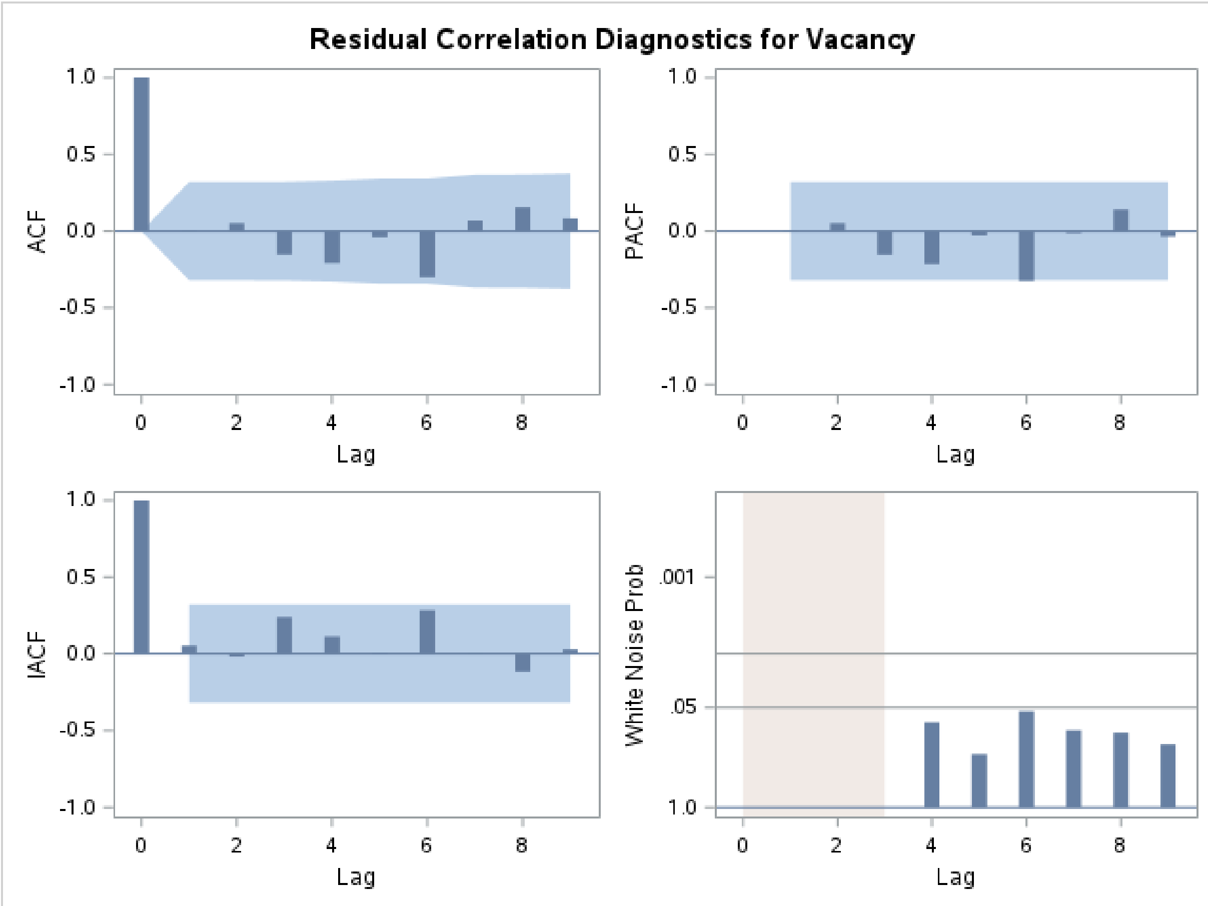

Figure 1: Residual Correlation Diagnostics of the Home Vacancy Rate.

The Q-Q Plot (Figure 4) and residual plot (Figure 5) show that the residuals are approximately normally distributed, independent, centered around 0, and with constant variance, which means that the model was adequate.

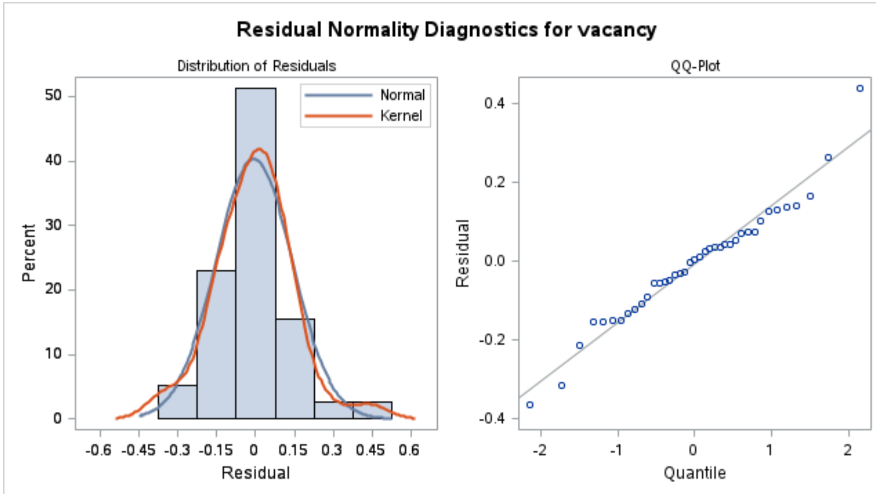

Figure 4: Residual Normality Diagnostics of the Home Vacancy Rate

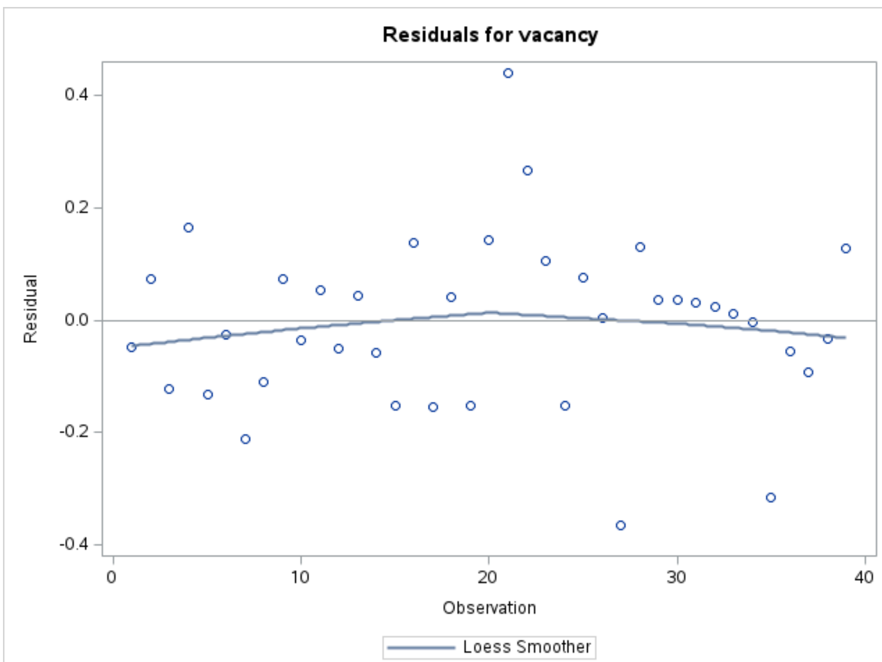

Figure 5: Residuals for Vacancy Scatter Plot

- **Forecasts**

Based on the fitted model, ARIMA (2, 0, 1), the forecasts indicate a consistent rising trend in the series from 1.09 to 1.66 over the next 5 years (Table 3).

Table 3: The Forecasts of Home Vacancy Rate Over the Next Five Years

Year	Forecast	Std Error	95% Confidence Limits	
2025	1.0962	0.1549	0.7926	1.3998
2026	1.2157	0.211	0.8021	1.6293
2027	1.3528	0.2461	0.8706	1.8351
2028	1.5017	0.2677	0.977	2.0265
2029	1.6562	0.2797	1.1079	2.2044

Figure 6 presents the actual values of the home vacancy rate and the model forecasts.

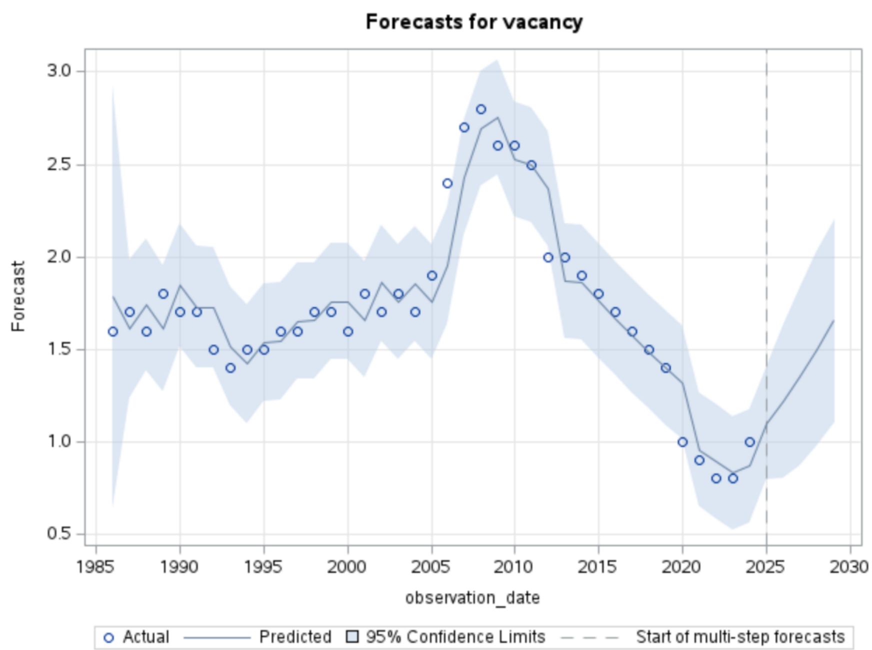

Figure 2: The Forecasts of the Home Vacancy Rate

Conclusion

Predicting the home vacancy rate is important because it helps governments, businesses, and communities make better housing, investment, and policy decisions. The ARIMA (2,0,1) model was identified as the best-fitting model, with the lowest AIC, BIC, and RMSE values compared to other models. The parameter estimates of the autoregressive and moving average terms in the model were statistically significant, and residual diagnostics confirmed that the model captured the underlying pattern of the data adequately.

Based on this model, forecasts indicate a consistent increase in the home vacancy rate over the next 5 years. The forecast for 2025 is 1.09%, rising toward 1.66% by 2029. These results suggest that housing supply may begin to outpace demand, which has important implications for policymakers, developers, and community planners.

Contact Information

Your comments and questions are valued and encouraged. Contact the authors at:

Joseph Maina, Park University, joseph.maina@park.edu

Abdelmonaem Jornaz, Park University, Abdelmonaem.Jornaz@park.edu

References

Abidoye, B. O., & Chan, A. P. C. (2018). Improving property valuation accuracy: A comparison of hedonic pricing model and artificial neural network. Pacific Rim Property Research Journal, 24(3), 219–236. https://doi.org/10.1080/14445921.2018.1436306

Abidoye, B. O., Chan, A. P. C., Abidoye, A. P., & Oshodi, O. S. (2019). Predicting property price index using artificial intelligence techniques. International Journal of Housing Markets and Analysis, 12(6), 1072–1092. https://doi.org/10.1108/IJHMA-11-2018-0095

Chen, Y.-J., Ong, Y.-S., Zheng, J.-H., & Hsu, S.-C. (2017). Forecasting spatial dynamics of the housing market using support vector machine. International Journal of Strategic Property Management, 21(3), 273–283. https://doi.org/10.3846/1648715X.2017.1346391

U.S. Census Bureau, Home Vacancy Rate for the United States [USHVAC], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/USHVAC

Hodson, D., & Quaglia, L. (2009). European perspectives on the global financial crisis: Introduction. Journal of Common Market Studies, 47(5), 939–953. https://doi.org/10.1111/j.1468-5965.2009.02029.x

Maurer, R., Pitzer, M., & Sebastian, S. (2004). Hedonic price indices for the Paris housing market. Allgemeines Statistisches Archiv, 88(3), 359–374. https://doi.org/10.1007/s101820400173

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. Journal of Political Economy, 82(1), 34–55. https://doi.org/10.1086/260169

Temur, A. S., Akgün, M., & Temur, G. (2019). Predicting housing sales in Turkey using ARIMA, LSTM and hybrid models. Journal of Business Economics and Management, 20(5), 920–938. https://doi.org/10.3846/jbem.2019.10447

Yang, Z. (2001). An application of the hedonic price model with uncertain attribute—The case of the People's Republic of China. Property Management, 19(1), 50–63. https://doi.org/10.1108/02637470110366202

Appendix

Table 1: The Actual Values, Forecasted Values, Prediction Interval, and Residual

Year	Vacancy	FORECAST	STD	L95	U95	RESIDUAL
1986	1.6	1.76396	0.15738	1.45551	2.07242	-0.16396
1987	1.7	1.6072	0.15738	1.29874	1.91565	0.0928
1988	1.6	1.7093	0.15738	1.40084	2.01775	-0.1093
1989	1.8	1.61622	0.15738	1.30776	1.92467	0.18378
1990	1.7	1.81392	0.15738	1.50547	2.12238	-0.11392
1991	1.7	1.71689	0.15738	1.40843	2.02534	-0.01689
1992	1.5	1.71942	0.15738	1.41096	2.02787	-0.21942
1993	1.4	1.53073	0.15738	1.22227	1.83918	-0.13073
1994	1.5	1.44556	0.15738	1.13711	1.75402	0.05444
1995	1.5	1.55557	0.15738	1.24712	1.86403	-0.05557
1996	1.6	1.56602	0.15738	1.25756	1.87447	0.03398
1997	1.6	1.67207	0.15738	1.36361	1.98053	-0.07207
1998	1.7	1.67856	0.15738	1.3701	1.98701	0.02144
1999	1.7	1.78065	0.15738	1.4722	2.08911	-0.08065
2000	1.6	1.78319	0.15738	1.47473	2.09164	-0.18319
2001	1.8	1.69011	0.15738	1.38165	1.99856	0.10989
2002	1.7	1.88781	0.15738	1.57936	2.19627	-0.18781
2003	1.8	1.79078	0.15738	1.48232	2.09923	0.00922
2004	1.7	1.88892	0.15738	1.58046	2.19738	-0.18892
2005	1.9	1.79188	0.15738	1.48343	2.10034	0.10812
2006	2.4	1.98563	0.15738	1.67718	2.29409	0.41437
2007	2.7	2.4583	0.15738	2.14985	2.76676	0.2417
2008	2.8	2.71997	0.15738	2.41151	3.02843	0.08003
2009	2.6	2.77854	0.15738	2.47009	3.087	-0.17854
2010	2.6	2.54633	0.15738	2.23788	2.85479	0.05367
2011	2.5	2.51325	0.15738	2.2048	2.82171	-0.01325

Year	Vacancy	FORECAST	STD	L95	U95	RESIDUAL
2012	2	2.38457	0.15738	2.07611	2.69302	-0.38457
2013	2	1.87739	0.15738	1.56894	2.18585	0.12261
2014	1.9	1.86805	0.15738	1.5596	2.17651	0.03195
2015	1.8	1.76311	0.15738	1.45465	2.07156	0.03689
2016	1.7	1.66211	0.15738	1.35366	1.97057	0.03789
2017	1.6	1.56508	0.15738	1.25662	1.87353	0.03492
2018	1.5	1.472	0.15738	1.16354	1.78045	0.028
2019	1.4	1.38288	0.15738	1.07442	1.69133	0.01712
2020	1	1.29771	0.15738	0.98925	1.60617	-0.29771
2021	0.9	0.92967	0.15738	0.62121	1.23813	-0.02967
2022	0.8	0.86428	0.15738	0.55583	1.17274	-0.06428
2023	0.8	0.80286	0.15738	0.4944	1.11131	-0.00286
2024	1	0.841	0.15738	0.53254	1.14945	0.159
2025	.	1.07036	0.15738	0.7619	1.37881	.
2026	.	1.16785	0.21774	0.7411	1.59461	.
2027	.	1.28851	0.25759	0.78364	1.79338	.
2028	.	1.42746	0.28393	0.87097	1.98394	.
2029	.	1.57912	0.29989	0.99134	2.1669	.