

Analysis of Tianjin Residuals Price's Influential Factors

Peng Xue

College of Management, Tianjin Polytechnic University, Tianjin, China

Email: autumnwords_xp@126.com

Abstract: Market is dynamic, which can be divided into three stages: the past, present and future. Predictions for the future are based on the past and present and appropriate mathematical methods. This paper uses linear regression analysis method to establish the forecasting model and gets the result as follows: Tianjin residual price is decided by investment completed of residential builds in current year (INV) and floor space of residential builds under construction (CON), Therefore the real estate market of Tianjin is a non-perfect market rather than a perfect competitive market like other commercial market.

Keywords: Tianjin residuals price; Linear regression analysis; Stepwise regression analysis; Non-perfect market

1. Introduction

Market is dynamic, which can be divided into three stages: the past, present and future. Predictions for the future are based on the past and present and choose appropriate mathematical model to predict the future trends. Here we use linear regression analysis method to establish the forecasting model.

2. Model constructed

The first step to build a linear regression model is to choose the dependent and independent variable. Obviously, as the main research object of Tianjin commercial residential building price is independent variable. But it

is diversity to choose dependent variable. In this paper the dependent variables chosen are Tianjin per capita disposable income (DI), Tianjin per capita GDP (PGDP), Tianjin floor space of residential builds under construction (CON), Tianjin floor space of residential builds completed (COM), Tianjin investment completed of residential builds in current year (INV), because those variables includes the information of supply and demand which reflects Tianjin housing prices. In order to get the five variables and the dependent variable, this paper got the data in the table 1.

Let $PGDP = GDP/P$, $RAP = S/SA$, we get table 2 which contain the five independent variable and the dependent variable.

Table 1. Basic data from 2000 to 2006

Year	P	GDP	PDI	COM	CON	INV	SA	S
2000	1001.14	1701.88	8140.55	532.63	1582.15	900198	378.34	860428
2001	1004.06	1919.09	8958.70	626.639	1590.62	963985	514.59	1187443
2002	1007.18	2150.76	9337.56	673	1746.1319	1031788	538.26	1299236
2003	1011.3	2578.03	10312.90	750.67	1953.50	1509252	720.64	1723256
2004	1023.67	3110.97	11467.16	1014.46	2352.9828	1752377	997.51	3482318
2005	1043	3697.62	12638.60	1270.96	2827.87	2349240	1264.38	5041379
2006	1075	4344.27	14283.60	1308.95	3396.45	4309000	1332.49	6195083

Table 2. Calculated data from 2000 to 2006

Year	PGDP	PDI	COM	CON	INV	RAP
2000	1.70	8140.55	532.63	1582.15	900198	378.34
2001	1.91	8958.70	626.639	1590.62	963985	514.59
2002	2.14	9337.56	673	1746.13	1031788	538.26
2003	2.55	10312.90	750.67	1953.50	1509252	720.64
2004	3.04	11467.16	1014.46	2352.98	1752377	997.51
2005	3.55	12638.60	1270.96	2827.87	2349240	1264.38
2006	4.04	14283.60	1308.95	3396.45	4309000	1332.49

Multiple regression analysis in SPSS is applied to process the data and got the results in table 3, $R^2 = 0.996$, $F = 27.205$, $P = 0.144$ and $F = F_{0.144} > F_{0.05}$. Therefore, model constructed through multiple linear regression can satisfied test of goodness-of-fit but can't satisfied significant test.

Because the regression equation constructed by multiple linear regression method is not significant, a new model must be constructed by other mathematic method. Therefore, we constructed another model through stepwise regression method. Through stepwise regression, Tianjin CON, INV enter into the model but PGDP, PDI, COM, and detail calculated result are shown in table 4 and table 5. This result is not only different from results that other scholars have gotten but also disaccord with supply and demand principle.

The coefficients of regression equations can be obtained in table 6 and the regression equation of the model is $p = 1573.153 + 3.381CON - 0.02INV$. The following is a model for the inspection. But the model will be proved as true must be meet various related hypothesis testing.

Table 3. ANOV

Model	Sum of Square	Df	Mean Square	F	Sig
1 Regression	5501029.033	5	1100205.807	27.205	.144
Residual	40442.042	1	40442042		
Total	5541471.074	6			

Table 4. Variables into/remove

Model	Variables Entered	Variables Removed	Mthod

1	CON	Stepwise(Criteria: Probability-of-F-to Enter<=.050, Probability-of-F-to Remove<=.100)
2	INV	Stepwise(Criteria: Probability-of-F-to Enter<=.050, Probability-of-F-to Remove<=.100)

Table 5. Non-included variable data table

Model	Beta In	t	Sig.	Partial Correlation	Tolerance	VIF	Minimum Tolerance
1 INV	-1.456	-2.803	.049	-.814	.010	98.732	.010
COM	.329	.960	.392	.433	.056	17.876	.056
GDDP	-.143	-.267	.803	-.132	.028	36.317	.028
DI	-.265	-.464	.667	-.226	.024	42.356	.024
2 COM	.084	.305	.780	.173	.046	21.685	.006
GDDP	-.079	-.218	.841	-.125	.027	36.515	.008
DI	-.030	-.074	.945	-.043	.022	45.126	.009

Table 6. Stepwise regression coefficient table

Model	Unstandardized Coefficients B	Unstandardized Coefficients Std.Error	Standardized Coefficients Beta	t	Sig.	Tolerance	VIF
1(Constant)	52.504	257.211		.204	.846	1.000	1.000
CON	1.367	.122	.984	12.215	.000		
2(Constant)	-1573.153	603.633		-2.606	.047		
CON	3.381	.722	2.433	4.681	.009	.010	98.732
INV	-0.002	.001	-1.456	-2.803	.049	.010	98.732

2. Model test

Through the figure 1 and figure 2, residual of the model follow normal distribution, thus residual need to obey the normal distribution hypothesis conditions are met. This paper considers whether the model exists heteroskedasticity or serial correlation (autocorrelation).

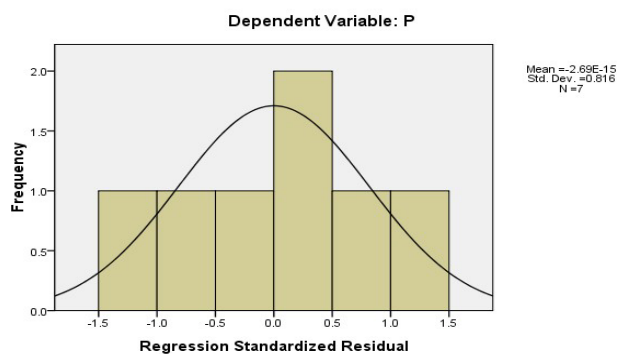


Figure 1. Standard residual normal test

Scatterplot of Regression Standardized Residual to Regression Standardized Predicted Value can be applied to make sure whether the model exist Heteroskedasticity. Because Standardized Residual didn't increase with Standardized Residual Predicted Value (Figure 3).

The model doesn't exist Heteroskedasticity. It is usu-

ally to do autocorrelation test by Durbin - Watson test which can be process through SPSS. After Durbin - Watson test, Durbin - Watson's value equal to 2.451 (from table 7). Because there are 7 years data collected and only 2 variables enter into model, $n = 7$ $k = 2$ is selected in Durbin - Watson significant test table and $DU = 0.294$, $DL = 1.676$. $2 < 2.451 < 4 - DU$, Therefore Heteroskedasticity can't be sure through Durbin - Watson test. Therefore, we combined figure 4 to judge whether model existing serial correlation. There doesn't exist obvious trend in figure 4, therefore the model does not exist serial correlation.

Normal P-P Plot of Regression Standardized Residual

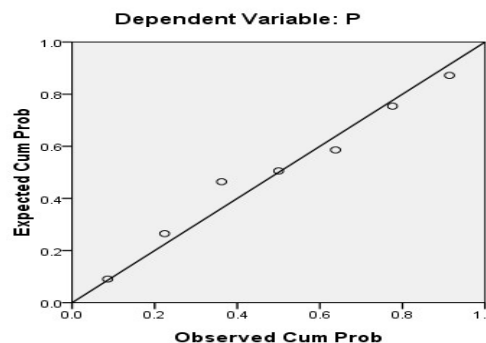


Figure 2. P-P Plot of Regression Standardized Residual

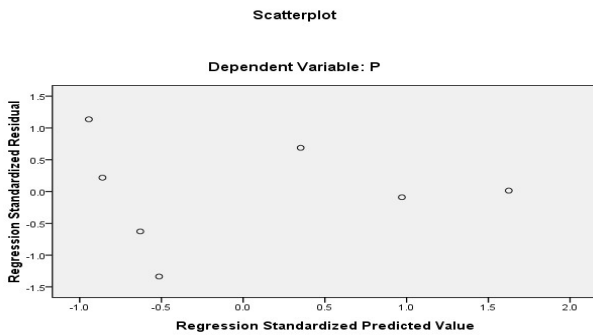


Figure 3. Standards for standard residual prediction of a scatterplot chart

Table 7. Stepwise regression model

Model	R	R Square	AdjustedR Square	Std. Error of the Estimate	Durbin-Watson
1	.984	.968	.961	189.56692	
2	.995	.989	.984	123.11415	2.251

In table 7, $R = 0.995$, $R^2 = 0.989$, Adjusted $R^2 = 0.984$, Therefore the model meet test of goodness-of-fit. In table 8, $F = 180.801$, $P = 0.000$, Therefore the model meet significant test. In table 6, $t_0 = -1573.153$, $t_1 = 3.381$, $t_2 = -0.002$, $p_2 = 0.049$, Therefore the model meet parametric test.

Table 8. ANOVA of Stepwise Regression

Model	Sum of Square	Df	Mean Square	F	Sig
1 Regression	5361792.987	1	5361792.987	149.205	.000
Residual	179678.0872	5	35935.617		
Total	5541471.074	6			
2 Regression	5480842.702	2	2740421.351	180.801	.000
Residual	60628.37219	4	15157.093		
Total	5541471.074	6			

In table 6, tolerance of CON and INV are 0.010 and 0.010 respectively. Therefore the model exist multicollinearity. There are several kinds to process multicollinearity such as PCA, Ridge Regression, PLS. Because SPSS program can't process Ridge Regression and PLS directly, many scholars adopt PCA method. But the in this model, only two variables are entered, thus in this paper we action through inaction. So far, the model passes all hypothesis testing. Therefore it is suitable to use this model as Tianjin residual price trend.

3. Model analysis

The model equation constructed through Stepwise Regression is $P = 1.573.153 + 3.381CON - 0.02INV$. The CON and INV is respectively 407.487 million square meters and ¥332818.7 million in 2007, therefore the Tianjin residual price predicted is ¥5548 which is

just a litter different from the price that is ¥5575 in Tianjin statistical year book. According to the data comparison of predicted and actual value, the model is proved to be right more.

The commodity price is up to supply and demand based on the economics principle view. In this model, the INV is supply variable, however the CON can't be not viewed as demand variable. So the model mismatch with economics principle. The reasons for this may exist in the following three aspects. First, Tianjin real estate market started late, which is an imperfectly competitive market in the process of growing now. Secondly, the condition set is too strict to allow variable enter into model. Thirdly, it is lack of some necessary variable. Actually, the second and third reason can be ignored. INV reflects the current supply and CON reflects future supply. Therefore Tianjin residual price is mainly affected by current supply and future supply and Tianjin real estate market is a supplying market which the commodity price are decided by supplier. In fact, real estate industry is the pillar industry in china, which has great influence to national economy, has established higher entry barrier. All enterprises which entered into this real estate industry in china have market power and can affected residual price partly.

Huge capital is necessary to real estate development including cost of land-use and construction cost which are unable to recover in the short term. In order to reduce the funding gap, the real estate enterprise must raise funding by several ways such as advance-sale, bank loan and raising housing price and so on. Because the real estate enterprise has market power, it can cost less to obtain more profit by raising housing price. So the phenomenon we saw is that the hosing price goes up and up rapidly. But the development mode has its fault that the economy crisis would occur when the capital chain is break like the US sub-prime crisis. Because the real estate market of china is non-perfect market, the government has to step in. In Tianjin real estate industry, the government must govern the related enterprise and increase the supply of housing.

References

- [1] Yang Guizhong, Deng Xuefen. Regression Analysis and After – Simulation of House Price in Chengdu City.[J];Value Engineering, 2007
- [2] Zhou Dan. Chinese real estate industry development factors of stepwise regression analysis. [J]Market Modernization.2009
- [3] Qin Yingxia, Xi Jinping, Wu Runheng. Multiple Linear Regression Model of the Real Estate Market in China [J];Mathematics in practice and theory.2009
- [4] Gao Huixuan. Applied Multivariate Statistical Analysis[M]. Beijing University Press.2005
- [5] Wang Xuemin. Applied Multivariate Statistical Analysis [M]. Shanghai university press.2004.