 doi:10.20944/preprints202210.0434.v1

HOW DO MASS RAPID TRANSIT AND CONVENIENCE STORES AFFECT HOUSING PRICES?

Qingshan Qian

ABE International Business College. Email: qian.qingshan.econ@gmail.com

ABSTRACT

Real estate has the dual characteristics of consumption and investment. Under the inherent concept of "where there is land, there is wealth", real estate has become the investment target of the main asset. In addition to the supply and demand of the market will affect the price of real estate, many general economic factors will also have a certain impact on the price of real estate. This paper pays a special focus on the effects of MRT on housing prices, finding the closer a house is to an MRT station, the higher its price. In addition, convenience stores are very popularized, and this research report is also interested in whether the number of convenience stores within walking distance will affect the house price. In a word, these two factors do affect housing prices significantly.

KEYWORDS

Housing Prices; MRT; Prediction; Machine Learning

1. INTRODUCTION

Intuitively speaking, a property commands a higher price ranging from 10% to 50% when it is near to an MRT station. The closer a house is to an MRT station, the higher its price. The MRT system can improve urban traffic congestion and improve the accessibility of space, so the construction of the MRT system usually has a positive impact on its surrounding real estate. In the past, most of the relevant literature also confirmed the positive capitalization effect of the MRT system on real estate prices. In addition to whether the MRT system has a capitalization effect on real estate prices and the direction of its impact, the location difference of the MRT stations' impact on the surrounding real estate prices is also the focus of the study. Because the MRT station is the gateway to the MRT system, the convenience of the real estate closer to the MRT station is relatively high, and the price consumers are willing to pay for this positive utility is also higher.

 doi:10.20944/preprints202210.0434.v1

The speed and convenience brought by the mass rapid transit system have made it a means of transportation that people in many metropolitan areas rely on. The MRT system improves the transportation accessibility of nearby real estate, and the accompanying convenience will be reflected in its price through the capitalization effect, which will then drive the price of surrounding real estate to rise. Empirical results also support that MRT has a positive impact on housing prices, and the effect decreases with increasing distance from MRT stations (McMillen and McDonald, 2004). Bajic (1983) examined changes in housing prices in the Toronto area before the opening of the subway in 1971 and after it opened in 1978, and found that commuting time savings due to improved subway access increased housing prices by an average of $2,237. Voith (1991) found that increased accessibility to MRT system service increases home prices by an average of $5,714 across the Philadelphia area. Coffman and Gregson (1998), Craig et al. (1998) found that transportation improvements have a significant positive effect on land prices.

Dziauddin (2022) shows a typical condo/serviced residence unit located within 0.4 km of the nearest MRT station (the treatment area) and traded after the system operates on the northwest side of the city can achieve a premium of approximately 9.5%. Ng (2019) shows accessibility to the downtown core via MRT lines does have an impact on housing prices. Cao et al (2019) identify the distance to the nearest mass rapid transit system may significantly affect home resale prices. Lee et al. (2020) discuss the impact of the mass rapid transit (MRT) system's construction and operation on neighbourhood housing prices. Wang (2022) show that the distance to the MRT has positive effects on housing prices. Torzewski (2020) show that metro has a positive impact on property prices if it is located at most 1.5 km from the property. The greatest increase in price of 8-13 per cent is observed when metro is located 400-800 meters from the property. Singhal and Tyagi (2021) indicate that a station node shows a negative trend during the planning and construction period. However, the operation period has produced a significant price premium associated with commercial properties, connected with improved accessibility. Lin et al. (2022) explore the relationship between metro systems and urban development, with particular focus on the comprehensive impacts of metro development on the economic, environmental and social development of cities. Xin (2019) reveals that new MRT lines do bring about a positive premium for homes within close proximity to the new station.

 doi:10.20944/preprints202210.0434.v1

2. DATA

Data are from the UCI Machine Learning Repository[1]. Including various statistical data, the market historical data set of real estate valuation comes from Xindian District, New Taipei City, Taiwan.

The original data has 414 observations in 6 variables as the following:

The dependent variable is:

Y - house price per unit area (10,000 New Taiwan dollars / ping, ping is the local unit, 1 ping = 3.3 square meters)

The chosen explanatory variables are as follows:

X1 - transaction date (e.g. 2013.250=March 2013, 2013.500=June 2013, etc.)

X2 - House Age (unit: Years)

X3 - the distance to the nearest MRT station (unit: meters)

X4 - The number of convenience stores in the pedestrian life circle (unit: number)

X5 - geographic coordinates, latitude. (unit: degree)

X6 - geographic coordinates, longitude. (unit: degree)

From the figures 1 and 2, it can be judged first that X3 (the distance from the MRT station) has a negative relationship with the house price, while X4 (the number of convenience stores within walking distance) has a positive relationship with the house price.

[1] https://archive.ics.uci.edu/ml/datasets/Forest+Fires

Preprints (www.preprints.org) | NOT PEER-REVIEWED | "Posted: 27 October 2022" doi:10.20944/preprints202210.0434.v1

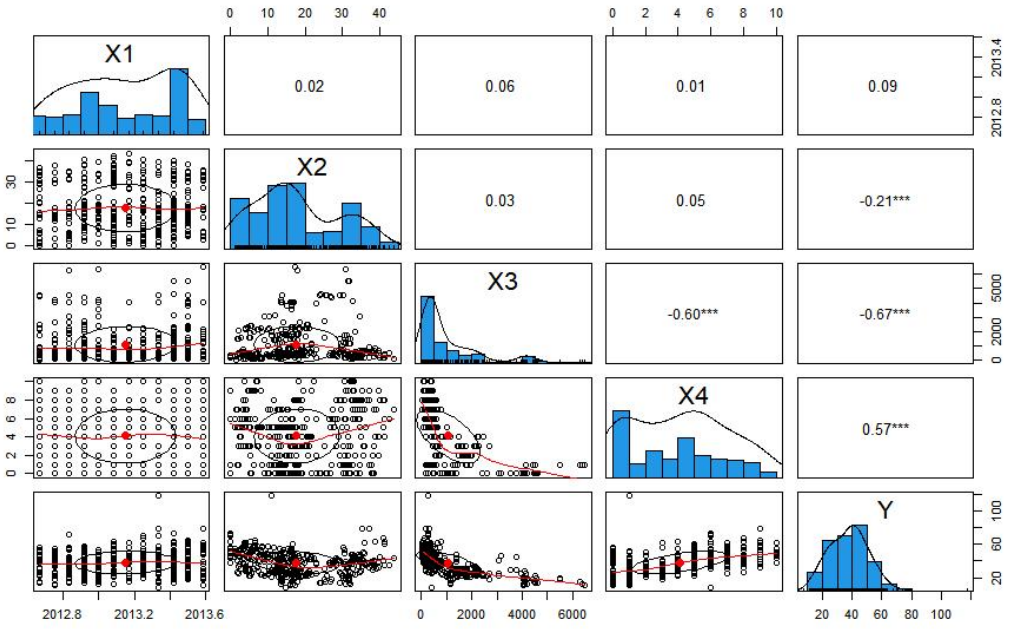

Figure 1.

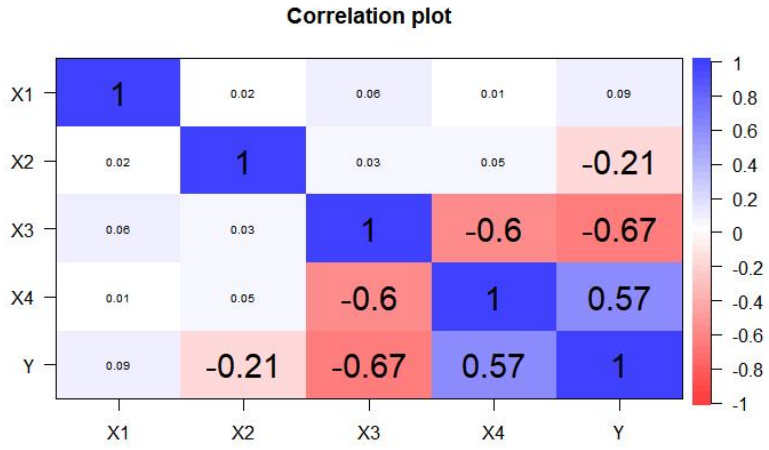

Figure 2.

3. Methodology

First, use random sampling to divide the data according to the ratio of two to eight, and divide it into 331 pieces of training data (training data) and 83 pieces of test data (test data). Residuals of 331 pieces of data are analyzed to confirm whether the residuals conform to the assumptions of the regression. After the correction, do a residual analysis to see if the residuals have met the assumptions of the regression. If so, the variables are screened. After

Preprints (www.preprints.org) | NOT PEER-REVIEWED | "Posted: 27 October 2022" doi:10.20944/preprints202210.0434.v1

screening, the outliers are determined, and the problematic data are deleted and reviewed. If this data affects the overall data information too much, this data will be deleted, and finally the 83 data of the test data set will be brought into the model established by the training data set for prediction.

3.1 Check the major assumptions of linear regression

3.1.1 Normality assumption (using Q-Q diagram)

The Q-Q map (quantile map) drawn by the 331 samples in the training set does not approximate a straight line, as in figure 3, so we can infer that this sample may not come from the mother of normal distribution, so we remove outliers and then re-step one. After removing outliers, it is closer to a straight line (figure 4), so it conforms to the assumption of normality.

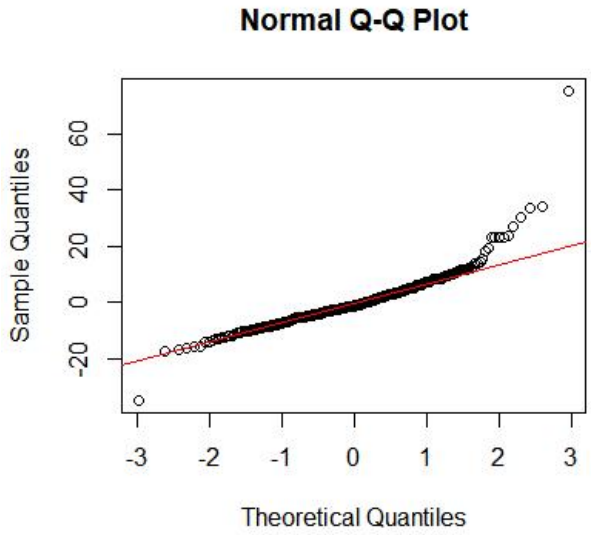

Figure 3.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | "Posted: 27 October 2022" doi:10.20944/preprints202210.0434.v1

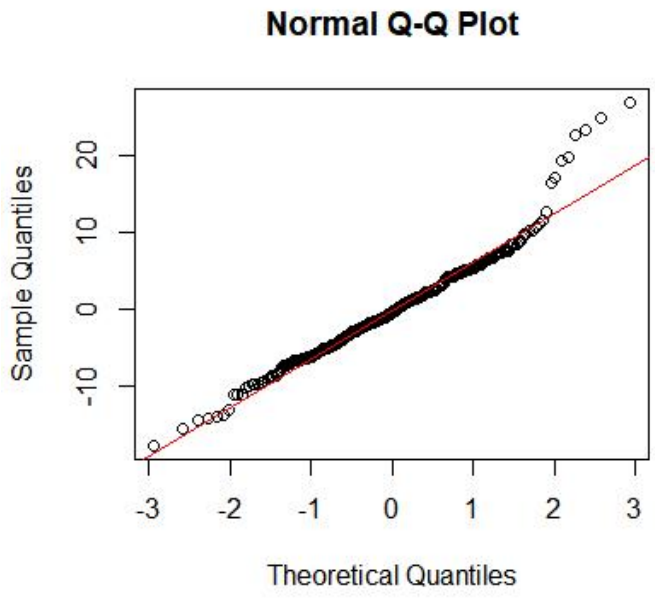

Figure 4.

3.1.2 Independence assumption (using the Durbin-Watson self-correlation test)

H0: There is no correlation between residuals V.S H1: There is correlation between residuals

Its test statistic is 1.944293 and the corresponding p-value is 0.642. Because the p-value is greater than 0.05, we have no evidence to reject the null hypothesis, so we conclude that the residuals of this regression model are not self-correlated.

3.1.3 The residuals have a stable amount of variation at each level of x

- The residual points are evenly centered around 0 (center point) and the beating is symmetrically averaged → No problem with function form → model is correct

- Residual points do not have a symmetrical average beating around 0 (center point) → there is a problem with the function form, and the scattered points exist regularly → the model is incorrect

Preprints (www.preprints.org) | NOT PEER-REVIEWED | "Posted: 27 October 2022" doi:10.20944/preprints202210.0434.v1

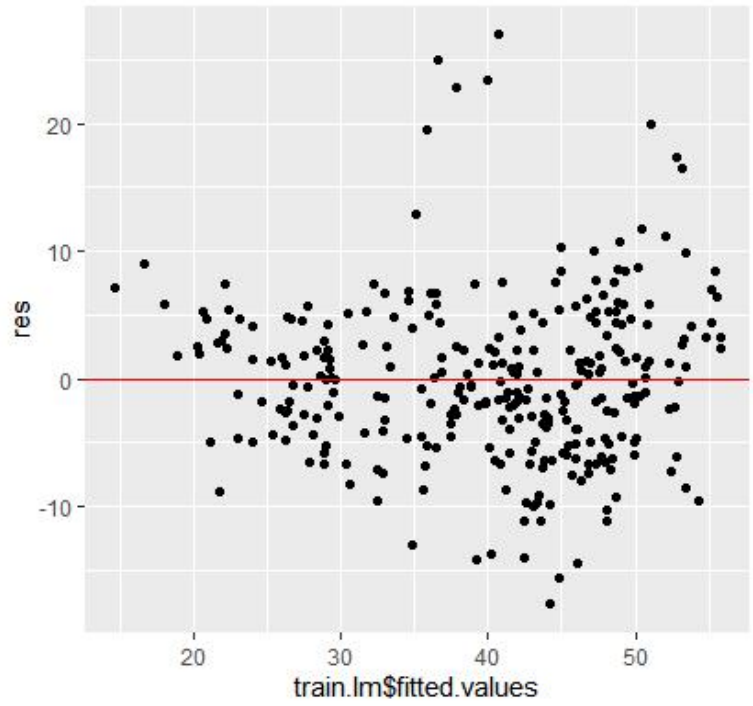

Figure 5.

From figure 5, it can be seen from the residual plot that the residuals do not become more dispersed as the fit value becomes larger.

3.1.4 Homogeneity assumption:

- Uniform distribution of residual points → uniform variance

- Residual points have a trumpet-like spread → inconsistent variance

It can be seen from the above residual analysis diagram (figure 5) that the residual points are uniformly distributed, so it conforms to the homogeneity assumption. Here, bptest checks whether the residual variances are consistent. The assumptions of its test: H0: the variance is consistent V.S H1: the variance is inconsistent. We obtain p-value = 0.1573 > 0.05, H0 is not rejected, so the variance uniformity assumption is met.

4. Results

First, we remove outliers by using Cooks distance to remove outliers.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | "Posted: 27 October 2022" doi:10.20944/preprints202210.0434.v1

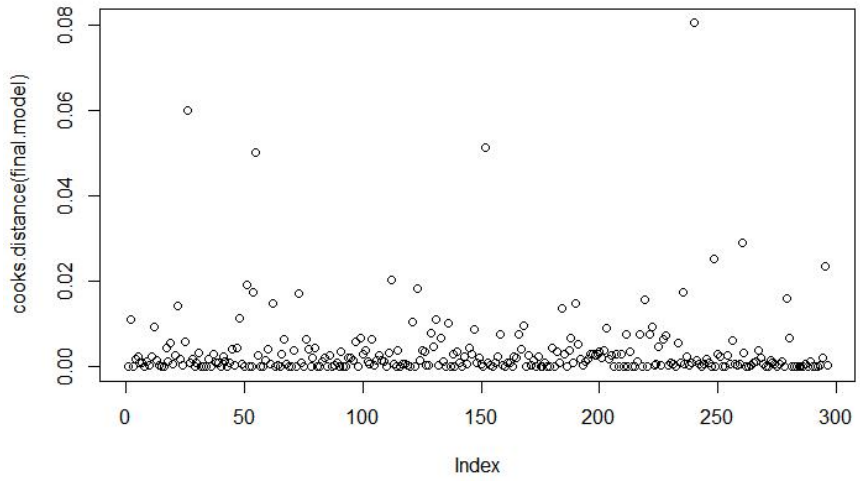

Figure 6.

It can be observed that there are four outliers in figure 6, data points 48, 129, 362 and 390, which were removed and remodeled then check again (shown in figure 7).

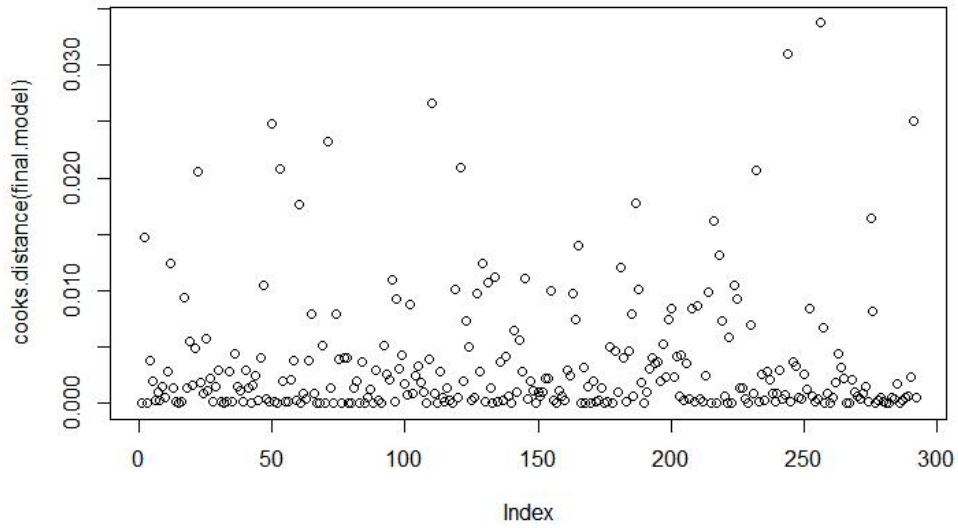

Figure 7.

We use the backward stepwise regression method to build the model and select the variables. First, we will use all the variables to build the model, and then eliminate the variables that are not statistically significant. The revised R value of the initial model was 0.6672 and the BIC was 1993.943, in figure 8. We can find that X6 is the least significant variable, and we remove it first.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | "Posted: 27 October 2022" doi:10.20944/preprints202210.0434.v1

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.637e+04  6.228e+03  -2.628  0.00904 **
X1           4.113e+00  1.394e+00   2.951  0.00343 **
X2          -3.139e-01  3.319e-02  -9.457  < 2e-16 ***
X3          -8.998e-03  8.561e-04 -10.511  < 2e-16 ***
X4           8.836e-01  1.644e-01   5.373 1.59e-07 ***
X5           2.521e+02  4.427e+01   5.694 3.04e-08 ***
X6           1.516e+01  4.453e+01   0.340  0.73375
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.582 on 289 degrees of freedom
Multiple R-squared:  0.674,     Adjusted R-squared:  0.6672
F-statistic: 99.57 on 6 and 289 DF,  p-value: < 2.2e-16
```

Figure 8.

The corrected R value of the eliminated model is increased to 0.6682, and the BIC is 1988.371, seem from figure 9. It can be found that X6 has no way to help explain the model. All variables in the final model are now statistically significant and the BIC does get progressively smaller. For “X3”, it is significant but with very small effect on housing prices. We cannot conclude that the MRT affect a lot in housing prices. However, for “X4”, it is significant and with large effect. This tells us that the number of convenience stores within walking distance will affect the house price.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.451e+04  2.973e+03  -4.879 1.76e-06 ***
X1           4.139e+00  1.390e+00   2.979  0.00314 **
X2          -3.140e-01  3.314e-02  -9.474  < 2e-16 ***
X3          -9.162e-03  7.062e-04 -12.975  < 2e-16 ***
X4           8.833e-01  1.642e-01   5.380 1.54e-07 ***
X5           2.492e+02  4.337e+01   5.745 2.32e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.572 on 290 degrees of freedom
Multiple R-squared:  0.6738,    Adjusted R-squared:  0.6682
F-statistic: 119.8 on 5 and 290 DF,  p-value: < 2.2e-16
```

Figure 9.

Bring testing data into the model to verify that its predictive ability is accurate. We calculate three metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean

Preprints (www.preprints.org) | NOT PEER-REVIEWED | "Posted: 27 October 2022" doi:10.20944/preprints202210.0434.v1

Absolute Error (MAE); their values are 74.47334, 8.629794 and 5.904799 respectively. It seems that the predicted results are not bad, see figure 10.

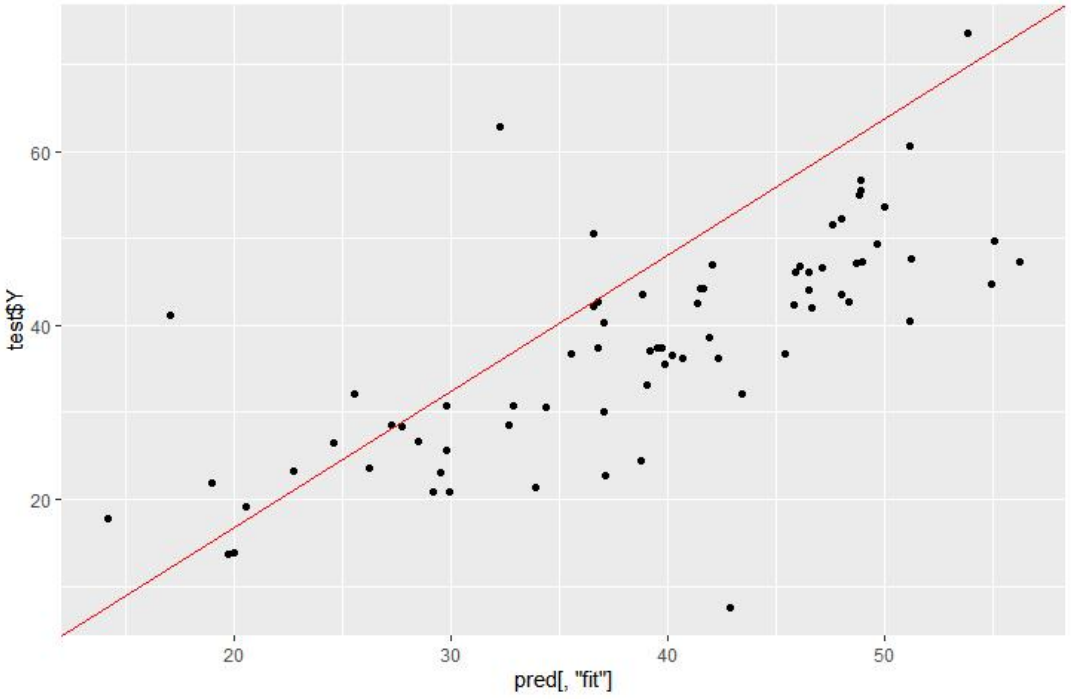

Figure 10.

After we add the prediction interval and observe, almost all data points are within the predicted range, indicating that the overall model is well established.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | "Posted: 27 October 2022" doi:10.20944/preprints202210.0434.v1

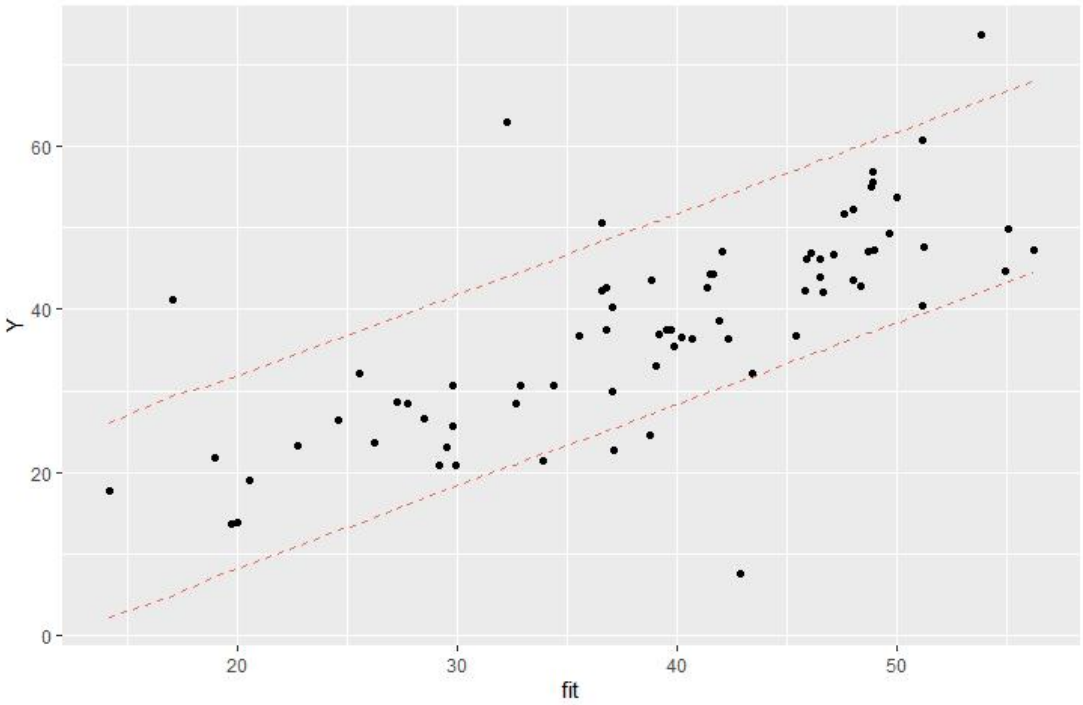

Figure 11.

5. CONCLUSION

Real estate has the dual characteristics of consumption and investment. Under the inherent concept of "where there is land, there is wealth", real estate has become the investment target of the main asset. In addition to the supply and demand of the market will affect the price of real estate, many general economic factors will also have a certain impact on the price of real estate. In addition, convenience stores are very popular in Taiwan, and this research report is interested in whether the number of convenience stores within walking distance will affect the house price. Indeed, it does affect significanly.

REFERENCES

Bajic, V., "The Effect of a New Subway Line on Housing Prices in Metropolitan Toronto", *Urban Studies*, Vol. 20, 1983, pp.147-158, doi: 10.1080/00420988320080291.

Cao, K., Diao, M. and Wu, B., "A big data – based geographically weighted regression model for public housing prices: A case study in Singapore", Annals of the American Association of Geographers, Vol. 109, No. 1, 2019, pp.173-186, doi: 10.1080/24694452.2018.1470925.

Coffman, C. and Gregson, M. E., "Railroad Development and Land Value", *Journal of Real Estate Finance and Economics*, Vol. 16, No. 2, 1998, pp.191-204, doi:

10.1023/A:1007707801970.

Craig, L. A., Palmquist, R. B., and Weiss, T., "Transportation Improvements and Land Values in the Antebellum United States: A Hedonic Approach", *Journal of Real Estate Finance and Economics*, Vol. 16, No. 2, 1998, pp.173-89, doi: 10.1023/A:1007755717900.

Dziauddin, M. F., "Estimating the Impact of Mass Rapid Transit (MRT) on Residential Property Prices in Greater Kuala Lumpur, Malaysia", *Journal of Asian Geography*, Vol. 1, No. 1, 2022, pp. 40-47.

Lee, C., Liang, C., and Hong, H., "The impact of a mass rapid transit system on neighborhood housing prices: an application of difference-in-difference and spatial econometrics", *Real Estate Management and Valuation*, Vol. 28, No. 1, 2020, pp. 28-40, doi: 10.2478/remav-2020-0003.

Lin, D., Broere, B. and Cui, J. "Metro systems and urban development: Impacts and implications", *Tunnelling and Underground Space Technology*, Vol. 125, 2022, doi: 10.1016/j.tust.2022.104509.

McMillen, D. P. and McDonald, J., "Reaction of Prices to a New Rapid Transit Line: Chicago's Midway Line, 1983-1999", *Real Estate Economics*, Vol. 32, No. 3, 2004, pp. 463- 486, doi: 10.1111/j.1080-8620.2004.00099.x.

Ng, L. T. L., "Does Accessibility to the Downtown Core Affect Housing Prices?", 2019.

Singhal, S. and Tyagi, Y. "Analyzing the Influence of Metro Stations on Commercial Property Values in Delhi: A Hedonic Approach", *Real Estate Management and Valuation*, Vol.29, No.4, 2021, pp.10-22. doi: 10.2478/remav-2021-0026.

Torzewski, M., "Public Transport Accessibility and the Prices of Nearby Properties: The Case of the First Metro Line in Warsaw", *European Journal of Transport and Infrastructure Research*, Vol. 20, No. 2, 2020, pp. 41-59, doi:10.18757/ejtir.2020.20.2.3979.

Voith, R., "Transportation, Sorting and House Values", *Real Estate Economics*, Vol. 19, No. 2, 1991, pp.117-137, doi: 10.1111/1540-6229.00545.

Wang, G. Y., "The Effect of Environment on Housing Prices: Evidence from the Google Street View", *Journal of Forecasting*, 2022, doi: 10.1002/for.2907.

Xin, T. Y. "Exploring The Validity Of The Golden Housing Wealth Rule of Residential Capital Appreciation Along Transport Nodes: Evidence From The Downtown Line", 2019.