INTERNATIONAL REAL ESTATE REVIEW

2020 Vol. 23 No. 4: pp. 505 – 536

Forecasting Housing Markets from Number of Visits to Actual Price Registration System

Tsoyu Calvin Lin*
Department of Land Economics, National ChengChi University, Address: No.64, Sec. 2, ZhiNan Rd., Wenshan District, Taipei City 11605, Taiwan. Email: tsoyulin@nccu.edu.tw.

Shih-Hsun Hsu
Department of Economics, National ChengChi University, Address: No. 64, Sec. 2, ZhiNan Rd., Wenshan District, Taipei City 11605, Taiwan. Email: shhsu@nccu.edu.tw.

Taiwan launched the actual price registration system for real estate transactions in 2012. Real estate–related information, for e.g., prices, area and location, can be obtained through a search on this platform. Most market participants, including potential buyers and sellers, obtain property information before making their transaction decision. If the search behavior can be transferred into supply or demand action, then the number of visits to a website can be used as a leading indicator of price changes or transaction volume. This study has collected the number of visits to the actual price registration system in New Taipei City in Taiwan and other macro-economic variables from 2014 to 2019 and applied a model with vector auto-regression with exogenous variables (VARX) for empirical analysis. We find two important results in our analysis: 1. the transaction volume significantly leads house prices and the number of visits to this system in most districts, and 2. the number of visits leads transaction volume only in the district with a very good transportation system and infrastructures, and leads the house prices only in districts that have affordable house prices or deemed to be a "good value". This is the first empirical study done after Taiwan launched the actual price registration system. Governments in other countries can launch similar systems and market participants can apply the findings of this study to their future policy and investment decision making process.

* Corresponding author.

Keywords

Actual price registration system, Hit rate, Search behavior, Big data, Vector auto-regression with exogenous variables

1. Introduction

As internet use becomes more prevalent and convenient, people now have the tendency to search and gather information online as a preparatory step before making a decision (Askitas and Zimmerman, 2009). The search and transaction records which contain private information of consumers, can accumulate continuously and rapidly. Demchenko et al. (2013) use the 5Vs to explain the difference between big data and traditional data. Big data implies the “volume”, “velocity”, “variety”, ”veracity” and “value” of the data. It also transcends the limits of traditional data which are screened subjectively. If we can foresee the needs of internet users though their search behavior, then business owners, manufacturers, marketing strategists or government personnel can react, or even pro-act, more efficiently for future business strategies and policy making.

The real estate industry is usually considered to have “information asymmetry”. Sellers obtain more information than buyers, for e.g., the quality of the construction, living environment and records of previous selling prices. To address the disparity in information, Taiwan launched the actual price registration system (hereinafter the System) in 2012, which requires sellers to register the transaction information, including transaction price, month of the transaction, floor area, age of the dwelling, location, etc., on this official platform. Potential buyers or sellers may search the transaction-related information on the internet before making the decision to purchase or sell, respectively. According to the statistics submitted by the real estate brokers, over 80% of the house buyers in Taiwan use this platform for information. The System has accumulated a vast amount of transaction data and attracted over 140 million visits from 2012 to 2019. If the search behaviors become real transaction activities, it is expected that the number of visits to the System can serve as a valid leading indicator of the housing market. We have thus collected the number of visits to the System to conduct an empirical analysis to explore the relationships among the number of visits to the System, transaction volume and prices.

2. Literature Review

As the search and transaction records on websites continue to accumulate, studies that apply “big data” (large volumes of complex data that can be mined for information) for marketing or predicting prices or transaction volume have become prevailing. Choi and Varian (2012) include the keywords on “Google

Trends” to predict car sales volume. They conclude that keyword searches can serve as leading signals of the markets or economic conditions. Vosen and Schmidt (2011) also indicate that the predictive validity of Google Trends outperformed the Consumers Sentiment (or Confidence) Index, thus implying that the search information on Google Trends has higher predictive power than surveying.

“Big data” has also been commonly used to monitor epidemic diseases and election activities. Althouse et al. (2011) find that “Google Flu Trends” which estimates influenza activity, can effectively predict the peak of dengue fever in Asia. Polykalas et al. (2013) conclude that the search preference of voters in German elections is positively related to the results. Jun et al. (2018) analyze the trends in research changes and conclude that the focus of research has shifted to predicting changes, from surveillance and monitoring to the application of big data sources (e.g., Google Trends).

In the application of big data to the real estate industry, Wu and Brynjolfsson (2015) find that house buyers tend to browse websites to search for market prices and transaction volume. They also indicate that the search records on Google Trends can effectively predict house prices and transaction volume. Kulkarni et al. (2009) also indicate that the search index on Google Trend has significant prediction power of the housing markets in 20 cities in the U.S. Van Dijk and Francke (2018) find that potential house buyers start their search for a house by browsing the internet and claim that their search behavior on the websites could provide useful indications of current or future demand after examining internet search data (click data) and the number of listed houses from *Funan.nl* (the largest housing website in the Netherlands) as well as the transaction data (house price index and rate of sale) from the Dutch Brokerage Association to examine housing market dynamics. In the tourism industry, Sun et al. (2019) apply machine learning and internet search indexes to predict tourist arrivals to popular destinations. They find causality and a co-integration relationship between the internet search indexes generated by Google and Baidu and tourist arrivals to Beijing.

Maclennan and O'Sullivan (2012) discuss the spatial process as part of the search activity before a housing purchase in their econometric model of housing market search. The results show that their expanded model can provide more information on the behavior of consumers to plan housing policies. Wu and Deng (2015) constructed a measure of information flow in Chinese housing markets based on search records from Google Trend, including the name of the city and house price (such as “Beijing” + “house price”) among others. They find evidence of the diffusion of information on house price from larger to smaller cities. Lin (2019) includes the keywords “buying house” and “real estate broker” on Google Trend Taiwan with other macro-economic variables in an econometric model to explore the housing market. He finds that the search numbers on Google Trend have significant leading effects on the prices and transaction volume in the housing market in Taiwan.

From the studies reviewed above, it is evident that online search records can serve as valid leading indicators or forecasting tools in many industries. Taiwan launched the System for real estate transactions in 2012. The System has accumulated a large volume of transactions and search records (over 140 million visits as of November 2019) for potential buyers and sellers. We intend to collect the number of visits to the System to explore how search behaviors affect house prices and transaction volume as well as their interrelation. Besides these three main variables, we are also interested in including some of the frequently-used macro-economic variables in the housing markets, such as construction cost index, money supply (M2) and consumer price index (CPI) in the proposed framework to control the effects of the changes of the underlying economic situation while analyzing the relationships among house prices, number of visits and transaction volume. For the empirical analysis, we establish a model with vector auto-regression with exogenous variables (VARX), that is, three endogenous variables (the number of visits, house prices and transaction volume) and three exogenous variables (construction cost index, CPI, and M2) in this study. If the number of visits is found to have significant effects on the changes in house prices or transaction volume, then governments, the industry and potential buyers and sellers will have another efficient, timely and valid indicator for policy and investment decision making.

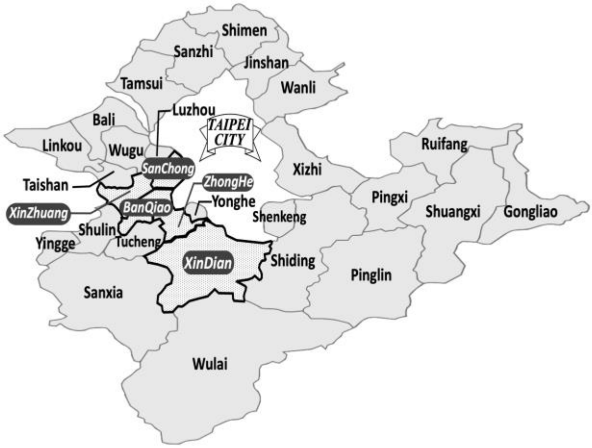
3. Variables and Research Method

3.1 Sample and Variables

To investigate the relationships among house price, transaction volume and the number of visits to the System, we focus on data from five popular residence districts that are highly and densely populated; that is over 300,000 people and more than 20,000 people/sq.km in New Taipei City (i.e., previously Taipei County), including BanQiao (BQ), XinZuang (XZ), ZongHe (ZH), SanChung (SC) and XinDien (XD).[1] Figure 1 is a map of these districts in New Taipei City.

[1] The reason for choosing New Taipei City as the sample area is because it has the largest population in Taiwan. Surrounding Taipei City, New Taipei City had over 4 million residents in 2019, as compared to Taipei City of 2.65 million. Moreover, the house price-to-income (PTI) ratio in New Taipei City and Taipei City is 12.03 and 14.15, respectively, according to the statistics of the Ministry of the Interior in Taiwan in 2019. Even though potential home buyers search for the housing information in Taipei City, many of them resort to buying houses in other cities due to lack of affordability. In Taipei City The high housing prices have also caused the population in Taipei City to drop from 2.68 million in 2017 to 2.67 million in 2018, a 0.6% decline within a period of one year. To enhance the validity of using search behavior to predict purchase action, New Taipei City is an appropriate city for empirical analysis in Taiwan. Furthermore, the number of visits at the city level on the System is too large and diversified to determine the validity of the search purpose. Therefore, we select districts in New Taipei City as the sample instead of the cities in Taiwan.

Figure 1 Map of Districts in New Taipei City, Taiwan



These five districts in the sample surround Taipei City or are in its proximity. BQ is adjacent to Taipei City as well as the capital of New Taipei City with three rails (Taiwan Rail, Taiwan High Speed Rail and the Taipei Mass Rapid Transit (MRT)) that serve the area. With municipal offices and commercial demand for real estate space, BQ therefore has the highest house prices in New Taipei City[2]. The second highest is ZH district due to its proximity to Taipei City and limited house supply. The next two districts that have the more expensive house prices are XD and SC, which are also both adjacent to Taipei City. XD is the largest of these five districts with natural wonders (mountains and rivers) and a convenient bus system to Taipei 101 (tourist attraction; third tallest building in the world). SC is right next to the Taipei Main Station (railway and metro station). Finally, the district with the least expensive housing is XZ, since it is the farthest from Taipei City.

The three considered endogenous variables are housing price index (*HPI*), transaction volume (*TRANSACTION*), and number of visits (*HIT*). Monthly data of housing prices are collected from "HousePlus.com.tw",[3] and transaction volume and the number of visits from the System.

---

[2] The average housing prices in these five districts are NTD (New Taiwan Dollar) 394K (US$13.8K)/3.305 sqm in BQ, 372K (US$13.1K) in ZH, 358K in XD, 357K (US$12.5K) in SC, and 316K (US$11.1K) in XZ. These prices are estimated from over 2500 properties from January 2019 to October 2020 based on the System (http://lvr.land.moi.gov.tw/homePage.action).

[3] There are several house price indices in Taiwan. The XinYi House Price Index is an index of the existing house prices, and Cathay House Price Index of new and pre-sale house prices. Both are quarterly based. The former is a monthly index extracted from the actual price registration system through the repeat-sales method. Due to the volume of the data, we use the prices of the latter (monthly basis) for the empirical analysis.

In addition, three exogenous variables which are common to all districts are also taken into account in the model. They are construction cost index (*CONSTRUCTION_NT*), consumer price index of New Taipei City (*CPI_NT*), and the money supply *(M2)* in Taiwan. The information on M2 is collected from the Central Bank of Taiwan, and that of the other two variables are obtained from statistics provided by the New Taipei City Government. In this study, we focus on investigating the relationships among month-over-month (MoM) rates of change in these endogenous and exogenous variables. The sample ranges from March 2014 to January 2019 since there did not appear to be a consistent number of visits until 2014. The dynamics of all of these variables during the investigation period are discussed as follows.

3.2 Endogenous Variables

3.2.1 House Price Index

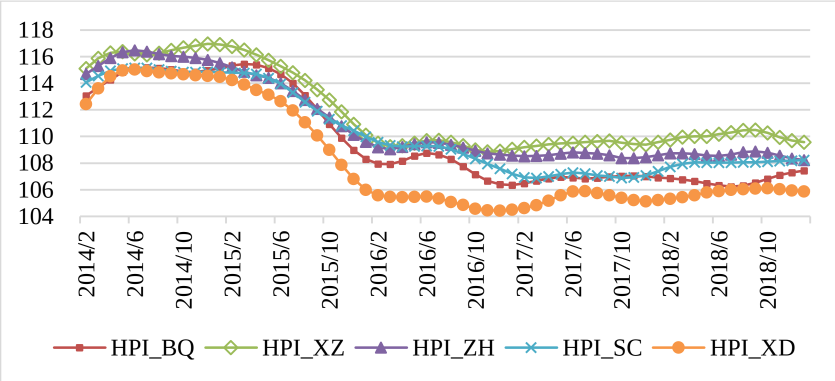
Figure 2 shows the dynamics of the levels of the HPI in the five concerned districts of New Taipei City. We can observe that these levels of the HPI declined from late 2014 and became relatively stable after 2016. Besides, the HPI are calculated by using current prices that are compared with the base year of the district, instead of a mutual comparison. These five districts are all linked to the MRT system of Taipei City. Among them, BQ has the most convenient transportation system since it offers access to 3 rail systems. It is also the capital of New Taipei City with sufficient public infrastructures. XD offers convenient entry/exit to highways aside from the MRT system to Taipei City. Natural wonders (e.g., mountains and rivers) also attract residents to XD. SC is the closest in proximity to the Taipei Main Station; however, the bridge to the Taipei Main Station causes a traffic bottleneck.

The Taiwan Central Bank launched a quantitative easing (QE) policy to bail out the financial markets after the global financial crisis in 2008. The M2 increased 8 folds in 2009, which caused real estate prices to escalate by around 20% annually until 2014 in Taiwan. The surge of house prices also deteriorated housing affordability for most households, which increased the house price to income (PTI) ratio to 13 in New Taipei City in 2013-2014.[4] To curb the real estate speculation fervor, Taiwan introduced the “luxury tax” in 2011, which requires real estate sellers to pay a 15% transaction tax if the deal is transacted within 1 year of purchase, and 10% tax within 1-2 years. To avoid the “luxury lax”, most investors wait for over two years and then look for opportunities to resell their property. This seems to be a wise decision, however, the supply of houses for sale contracted and prices rose slightly during those two years. After a two-year waiting period to avoid the luxury tax, a large number of houses started to flow into the market from investors. Besides, in mid-2014, the Taiwan Central Bank also launched a policy to limit the loan to value (LTV) ratio with a ceiling of 50% for the third purchased house. This policy effectively reduced

[4] The PTI ratio in Taipei City was 16 during the same period of time.

the purchasing power of potential buyers. With the effect of the luxury tax and deterioration of the purchasing power of houses, an obvious decline in the HPIs that started in late 2014 can be observed, and most of them peaked before the first quarter of 2015; see Figure 2.

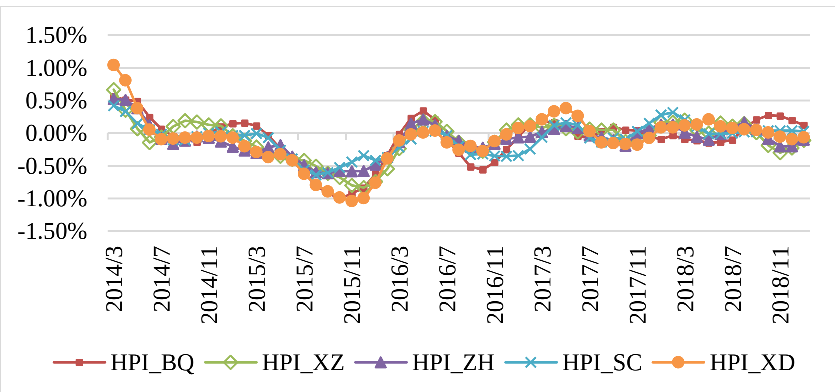
Figure 2 Trends of HPI in Five Districts of New Taipei City



Note: The price indices of these districts are compared with their own base year, not other districts. The ranking of house price for these districts is shown in Footnote 2.

On the other hand, Figure 3 shows the MoM rates of change in the HPI in these five districts of New Taipei City. Roughly speaking, except for a large decline in the growth rates that started in late 2014, most of them peaked (some of the rates of change were reduced to 1% in particular) before the end of 2015. These five rates of change in the HPI are stable, varying from -0.5% to 0.5% in most periods.

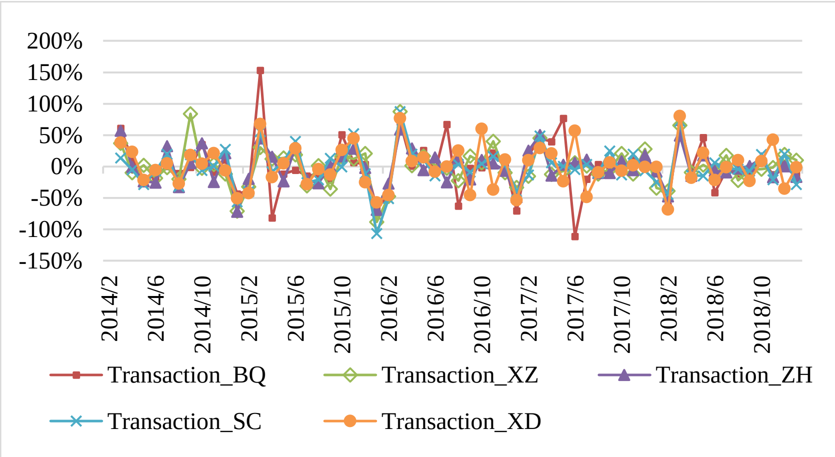
Figure 3 MoM Rates of Change in HPI of Five Districts in New Taipei City



3.2.2 Transaction Volume

Figure 4 is the MoM rates of change in transaction volume in the housing markets of the five districts. In general, the dynamics of these rates of change are slightly different among the five districts. The only exception is the period from the end of 2015 to the beginning of 2016, where there was a large reduction in transaction volume in all five districts. The significant decline in transaction volume is attributed to the amendment of the Income Tax Act with the introduction of a capital gains tax in 2016, which levies heavy taxes for short trading in real estate.[5]

Figure 4 MoM Rates of Change in Transaction Volume of Housing Markets in Five Districts of New Taipei City

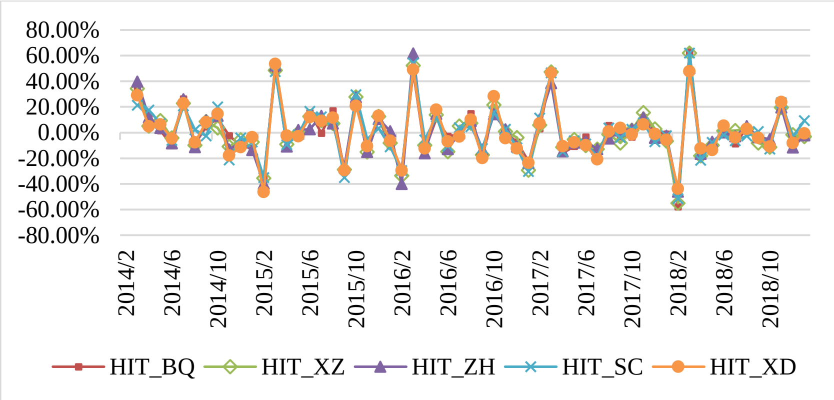


3.2.3 Number of Visits

The number of visits (*HIT*) to the System is the focus in this study. According to the studies discussed in the literature review, search behavior can serve as the leading signal in many aspects. The most important goal of this study is to therefore determine the inter-relation among the rates of change in the *HIT* to the System and those of house prices and transaction volume. Figure 5 shows the MoM rates of change in the *HIT* to the System in the five districts. Obviously, the dynamics of the rates of change in the *HIT* are quite similar. Four major peaks are observed in March from 2015 to 2018. This shows a seasonal pattern since there are more calendar days in March than February so that more visits are made by potential home buyer and/or sellers.

[5] To curb short trading in real estate, Taiwan amended the Income Tax Act with a capital gains tax in 2016, which levies 45% of the capital gains for real estate transactions within a 1-year holding; 35% tax for 1-2 years; 20% for 2-10 years; and 15% for over 10 years.

Figure 5 MoM Rates of Change in *HIT* of Five Districts in New Taipei City



3.3 Exogenous Variables

Three important exogenous variables are considered in the VARX model to approximate (or control) the effects of the changes in the underlying economic situation while analyzing the relationships among the three endogenous variables (construction cost index, CPI, and M2).

The construction cost index is the development cost of constructing houses. The price of the final product is usually equal to the total cost plus the profit margin. Therefore the construction cost is highly and positively related to house prices. However, as the construction costs surge, the willingness to develop and the purchasing power may both decline, and lead to a lower transaction volume.

The CPI is the most commonly-used indicator of inflation. Real estate is usually considered as a hedge tool for inflation. However, as inflation causes house prices to surge, housing unaffordability or other living costs may lead to the decline of the transaction volume.

M2 is the source of capital for investment. As M2 significantly increases, asset prices usually follow. The lax monetary policy in Taiwan in 2009 after the subprime mortgage crisis led to the escalation of housing prices, which is the best example to illustrate the effect of "helicopter money" on the surge of house prices. We introduce these three important macro variables to approximate the underlying economic situation and their dynamics in the MoM rates of change (see Appendix A).

3.4 VARX Models

For every district in New Taipei City, we establish a corresponding VARX model with the endogenous variables of *HPI*, *TRANSACTION*, and *HIT*, coupled with three exogenous variables that are common to all districts, including the construction cost index, CPI, and M2. All of the endogenous and exogenous variables selected in the following VARX models are their MoM rates of change instead of the raw series.[6] Furthermore, all these rates of change are stationary since all reject the null for the unit roots in the augmented Dickey-Fuller (ADF) tests. The test results are summarized in Appendix B in detail.[7]

Let Y_t=$[HPI_t, TRANSACTION_t, HIT_t]$' denote the vector of the endogenous variables in a given district and X_t=$[CONSTRUCTION\_NT_t, CPI\_NT_t, M2_t]$' the vector of exogenous variables common to all districts in New Taipei City at time *t*, then the corresponding VARX(p, q) model is:

$$Y_t = C + \sum_{i=1}^{p} \emptyset_i Y_{t-i} + \sum_{j=0}^{q} \beta_j X_{t-j} + \varepsilon_t, \qquad (1)$$

where C is a 3×1 vector of the intercepts, $\emptyset_i$ $(i = 1, \ldots, p)$ and β_j $(j = 0, \ldots, q)$ are 3×3 matrices of the coefficients, and ε_t is a 3×1 vector of the disturbances (corresponding to the endogenous variables Y_t at time *t*) with the mean and covariance matrix as:

$$\mathrm{E}[\varepsilon_t] = 0, \qquad E[\varepsilon_t \varepsilon_\tau] = \begin{cases} \Omega, t = \tau \\ 0, \text{otherwise} \end{cases}$$

Given this framework, we intend to investigate the Granger causality and the generalized impulse response functions (GIRFs) among the endogenous variables after estimating the model.

---

[6] By using ADF tests on raw data for the three endogenous variables, except for HPI, the other two endogenous variables reject the null. The vector error correction model (VECM) is not a candidate to model these three endogenous variables. Since all three exogenous variables did not reject the null in their ADF tests, we thus construct a VARX model for the rates of change in all of the variables in this study. Besides, we establish the VARX model for the realized rates of change in all of the endogenous and exogenous variables instead of the "level" ones, the unobserved fixed effect (which is local to each district) on the "level" of the variables can be eliminated if each variable in the model is differenced over time and the resulting estimator is known as the first-differenced estimator in the panel data framework. More details can be obtained in Chapter 13 of Wooldridge (2019).

[7] In this study, all the tests and estimations are conducted by using the corresponding packages in EViews 9.5.

3.4.1 Granger Causality

For the variable of interest, Y_1, assume that the Granger causality hinges on additional prediction ability from the previous realized values of another variable, Y_2 given all the other realized values of the endogenous and exogenous variables in the constructed VARX model. More specifically, if Y_2 Granger causes Y_1, this means that the past information of Y_2 can further help to predict the current Y_1 even given all the other information (including the past realized values of all the other endogenous and exogenous variables). Otherwise, Y_2 does not Granger cause Y_1. This causality detection can easily be completed by testing the corresponding entries in the coefficients $\emptyset_i$ $(i = 1, \ldots, p)$ of Equation (1), i.e., we can test the significance of excluding other endogenous variables for the estimated model of any endogenous variable in this VARX model (Equation 1).

3.4.2 Generalized Impulse Response Function

Given the unexpected disturbance $\varepsilon_{jt} = \delta_j$ for the jth endogenous variable at time t in the VARX model (Equation 1), the GIRF measures how the expected h-step ahead values of all of the endogenous variables change in the following:

$$\begin{aligned} GIRF\left(Y_{t+h}, \delta_j, I_{t-1}\right) = E\left[Y_{t+h} \middle| \varepsilon_{jt} = \delta_j, I_{t-1}\right] \\ -E\left[Y_{t+h} \middle| I_{t-1}\right], \end{aligned} \tag{2}$$

where $I_{t-1} = \{Y_{t-1}, Y_{t-2}, \ldots\}$ denotes the information at time $t - 1$. Note that the GIRF is a function of h and thus reveals the dynamics of the expected responses of the VARX model varying h while facing unexpected impulses. Besides, the GIRF is also free of pre-specifying the particular ordering of endogenous variables in the model since the orthogonal impulse response function typically relies on the Cholesky ordering of the endogenous variables. Moreover, δ_j is usually set as the estimated standard deviation of ε_{jt}, and the simulated confidence interval for GIRF can also be provided in practice. For more discussion on GIRF, please see Koop et al. (1996) and Pesaran and Shin (1998).

4. Empirical Results

Given the realized rates of change in the endogenous and exogenous variables above for the five districts in New Taipei City from March 2014 to January 2019, we intend to examine the inter-relations among the endogenous variables through the corresponding VARX(p, 1) model for each district, where the optimal lag length p is determined by the Schwarz information criterion (SIC). As such, the VARX(2,1) model is selected for the BQ, SC and XD districts, and VARX(1,1) for the XZ and ZH districts. Overall, for *HPI*, all the R squares are higher than 0.85 while higher than 0.54 for the two other endogenous variables, *TRANSACTION* and *HIT*. The detailed estimation results for the five districts are reported in Tables C.1 to C.5 in Appendix C. In the following, we focus on investigating the relationships among the endogenous variables via Granger causality tests and GIRFs.

4.1 Granger Causality

Based on the estimated VARX model with the three endogenous variables for each district, we can investigate the Granger causality between any two of the endogenous variables. More precisely, whether Y_1 Granger causes Y_2 can be inferred by jointly testing the significance of excluding all lags of Y_1 in the equation of the dependent variable Y_2 in the VARX model; the null that Y_1 does not Granger cause Y_2 is equivalent to that of all the lags of Y_1 excluded from the model of the dependent variable Y_2. These joint tests are conducted by using chi-square (chi-sq) test statistics. All the test results for the five districts are summarized in Panels (a) to (e) in Table 1.

Table 1 Granger Causality Tests

		Dependent variable					
		HPI		*TRANSACTION*		*HIT*	
	Excluded variable	*TRANS-ACTION*	*HIT*	*HPI*	*HIT*	*HPI*	*TRANS-ACTION*
(a) BQ	Chi-sq	5.503*	0.613	0.517	9.115**	2.342	7.580**
	p-value	0.072	0.736	0.772	0.011	0.310	0.023
(b) SC	Chi-sq	10.920***	1.551	0.390	0.735	5.717*	9.382***
	p-value	0.004	0.461	0.823	0.693	0.057	0.009
(c) XD	Chi-sq	4.710*	12.815***	0.234	0.272	5.209**	0.259
	p-value	0.095	0.002	0.889	0.873	0.074	0.879
(d) XZ	Chi-sq	1.274	0.260	0.083	0.199	0.405	4.152**
	p-value	0.259	0.610	0.773	0.656	0.524	0.042
(e) ZH	Chi-sq	1.283	0.389	0.246	2.211	0.550	2.751*
	p-value	0.257	0.533	0.620	0.137	0.458	0.097

Notes: 1. The degrees of freedom for all chi-square test statistics are 2 in Panels (a) (b) and (c) while they are equal to 1 in Panels (d) and (e).
2. "*", "**" and "***" denotes rejection of the null of the zero coefficient at 10%, 5% and 1% significance levels, respectively.

First, when the dependent variable is *HPI*, the results of the chi-sq in the third column (for the excluded variable *TRANSACTION*) show that the test significantly rejects the null which excludes all lags of *TRANSACTION* in BQ

(Panel (a)), SC (Panel (b)) and XD (Panel (c)) with p-values of 0.072, 0.004, and 0.095, respectively. This indicates that the past information on *TRANSACTION* has additional predictability on the *HPI* in these three districts. This result also conforms with the principle of the transaction volume leading the prices. Combining the results of the VARX analysis in Appendix C (Tables C1-C3), *TRANSACTION* has the negative leading effect on *HPI* in these three districts. The reason is that the house market was in a downward trend during 2014 (as shown in Figure 3). Therefore, the increase in transaction volume may imply the supply increase from the sellers side and then a decline in price.

In addition to *TRANSACTION, HIT* in XD (Panel (c)) also significantly Granger causes the *HPI*. The VARX result in Table C.3 (Appendix C) also shows that the *HIT* positively and significantly leads the *HPI* in XD. Both results show that the transaction volume and search behavior have a leading effect on the house prices in XD.

Secondly, when the dependent variable is *TRANSACTION*, it is significantly Granger caused by the *HIT* only in BQ (Panel (a)) at the 5% level, but there is no significant effect caused by the *HPI* in all of the districts. The VARX result in Table C.1 (Appendix C) also shows that the *HIT* positively and significantly leads *TRANSACTION* in BQ. This implies that previous realized data of *HPI* cannot help to predict *TRANSACTION* since the past data of *TRANSACTION* and all exogenous variables are given for all districts. This may also be explained by the deterioration of housing affordability so that the escalated house prices did not lead to an effective transaction volume.

Thirdly, when the dependent variable is *HIT*, the final column in Table 1 indicates that *TRANSACTION* significantly Granger causes *HIT* in 4 districts at the 10% level with the exception of XD (Panel (c)), since the chi-sq tests significantly reject the null that all the past realizations of *TRANSACTION* can be excluded from the model with the dependent variable *HIT*. This implies that the transaction volume may have a leading effect on the search behavior of home buyers in most districts (i.e., BQ, XZ, ZH, and SC). The VARX results in Appendix C show that *TRANSACTION* positively and significantly leads the *HIT* in SC, XZ and ZH, but the effect in BQ is negative. It is reasonable that the transaction positively leads the number of visits to the System since potential home buyers are encouraged by the market transaction activity to search for houses. The reason that BQ shows a negative sign, however, may be attributed to its high prices (the most expensive in New Taipei City). Since many residents in New Taipei City commute to Taipei City for work, the substitution effect may lead potential house buyers to search in other districts if their job is not in BQ.

Finally, *HIT* only Granger causes *TRANSACTION* in BQ (Panel (a)) and *HPI* in XD (Panel (c)) among all of the cases. The VARX results in Tables C.1 and C.3 (Appendix C) also show that *HIT* has positive leading effects on *TRANSACTION* in BQ and *HPI* in XD. Recall that XD has a living environment with natural features and relatively affordable house prices (in comparison to

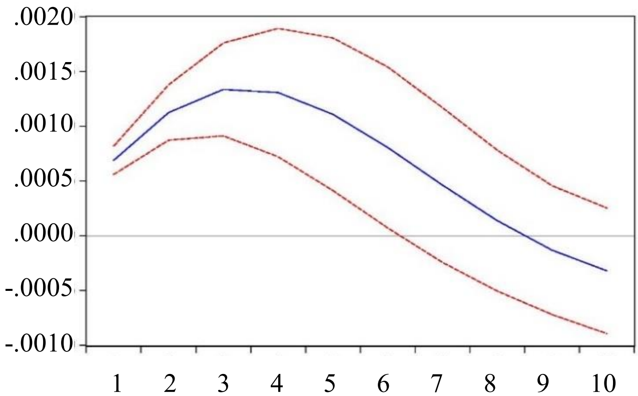
BQ). BQ is the capital of New Taipei City with the most convenient transportation system. Potential house buyers (including investors) and sellers may actually trade houses after searching for houses in these two districts, but the effect on increasing the price in BQ is not significant due to affordability and demand. We can therefore conclude that the past information of *HIT* can help to predict current *HPI* and *TRANSACTION,* even given all the other information (including the past realized values of all the other endogenous variables and all exogenous variables) from the perspective of Granger causality.

4.2 Generalized Impulse Response Function

In Appendix D, Figures D.1 to D.5 summarize the GIRFs (solid curve) and their corresponding 95% confidence intervals (dotted curve) for all of the endogenous variables in the five districts. In each figure, the horizontal axis indicates the specified value of h in the GIRF (see Equation (2) in Section 3.4.2) while the vertical axis represents the corresponding magnitude of the response. There are 9 inset figures in every figure, where we plot the generalized responses of the three endogenous variables in the VARX model while facing one standard deviation (S.D.) of the unexpected innovations from the three endogenous variables.

First, while facing one S.D. of unexpected shock from the variable itself, the GIRFs are shown in the three insets in the diagonal of Figures D.1 to D.5. For *HPI* (the top left inset figure), all of the GIRFs are in general positively significant away from zero in the few beginning periods and then becoming insignificant in the following periods. See the response of *HPI* to *HPI* in BQ in Figure 6 for example. The persistent dynamics of price is not surprising and commonly observed in time series data.

Figure 6 Response of *HPI* to *HPI* in BanQiao



On the other hand, almost all of the GIRFs for *TRANSACTION* and *HIT* (the center and bottom right insets) while facing their own shocks are significantly positive when $h = 1$, but immediately turns to negative when $h = 2$, and then insignificantly away from zero. See Figures 7 and 8 for instance. This phenomenon for the transaction volume and search behavior is interesting.

Figure 7 Response of *TRANSACTION* to *TRANSACTION* in BanQiao

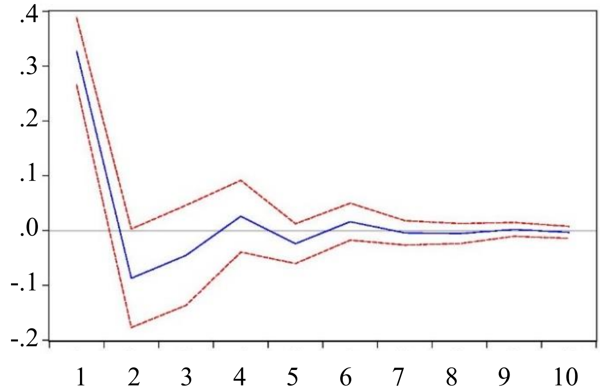
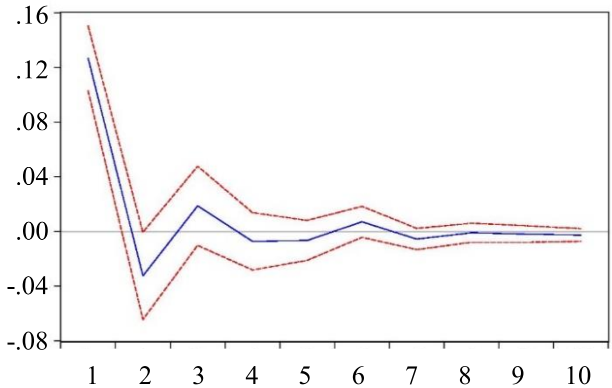


Figure 8 Response of *HIT* to *HIT* in BanQiao



Secondly, the insets in the non-diagonal represent the GIRFs of an endogenous variable while facing unexpected innovations from the other endogenous variables, and most of them show that the GIRFs are not statistically significant from zero at the 5% level. However, the responses of *TRANSACTION* to *HIT* and those of *HIT* to *TRANSACTION* in XZ, XD and ZH are significantly positive when $h = 1$; for example, Figures 9 and 10 show the GIRFs between *TRANSACTION* and *HIT* in XZ. This suggests that the previous positive shock on *HIT* (resp. *TRANSACTION*) inspires the following *TRANSACTION* (resp.

HIT), the realized and potential trades in the housing markets may increase accordingly. These results show that as the transaction volume increases, home buyers may be encouraged to search for houses for residence or investment, which accords with the behavior of investors. In addition, these search behaviors may also trigger future transactions.

Figure 9 Response of *TRANSACTION* to *HIT* in XinZuang

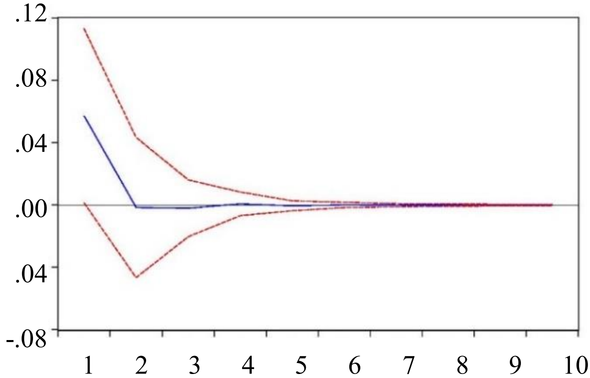
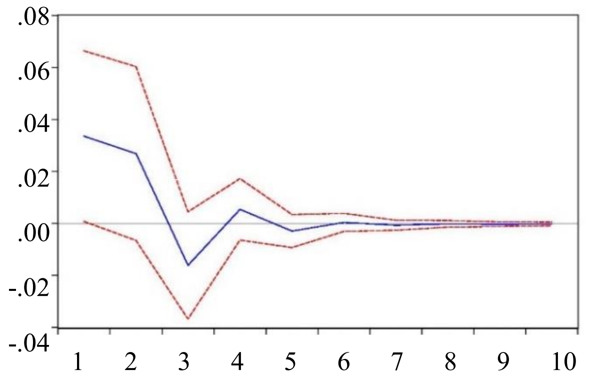


Figure 10 Response of *HIT* to *TRANSACTION* in XinZuang



5. Conclusion and Implications

5.1 Conclusion

The empirical results in this study show that the number of visits to the System significantly leads the transaction volume in BQ, which is the capital of New Taipei City with a very convenient transportation system and high demand for real estate space. The advantage of BQ may attract buyers with residence or investment needs to make a purchase decision after visiting the System. Furthermore, the number of visits leads the house prices only in XD, which is more affordable (than BQ) with natural features. This result suggests that the visits to the System by potential house buyers may only result in purchase action in districts with affordable prices or good value with natural features. Both results indicate that the number of visits to the System would eventually transfer into a "buying" decision and drive up the demand in the housing market, and significantly lead the transaction volume and house prices in districts of better value, including those with affordable prices, natural features and a convenient transportation system.

The effect of the transaction volume positively leads house prices in three districts (BQ, SC and XD) This result is consistent with the traditional convention that transaction volume usually leads prices. The transaction volume positively leads the number of visits in most districts, thus implying that there is a "chasing trend" in the housing market, i.e., the previous transaction volume may attract the interest of buyers in this market.

For the other variables, construction costs have positive effects on the appreciation of housing prices. This result conforms with the pricing theory, i.e., the final price of products include various costs and profit margin in normal situations. The increase of material costs usually drives up the prices of the final products of normal goods.

Surprisingly, the CPI has negative effects on the change in house prices in some districts and in almost all of the transaction volume and number of visits to the System. These results seem to contradict the traditional conception that real estate is a perfect hedge of inflation, which is worth discussing. After the 2009 QE policy, the house prices in Taiwan soared until 2014 and then started to decline due to affordability issues and tax reforms that deterred real estate speculation (i.e., the "luxury tax" in 2011 and the capital gains tax in 2016). During the same period, the CPI stagnated. The sample period in this study starts from 2014, right after the peak of the housing market. This is the reason why the empirical results show that the CPIs are negatively related to the house prices in some of the districts. The opposite relation between house prices and CPI in those districts also indicates that house buyers or investors should take into account the purchasing power for housing aside from the inflation (or deflation) trend in their decision making.

As for the effects of increase in M2, they have a significantly negative effect on transaction volume and number of visits in some of the districts. This unusual result indicates that the slight increase in M2 after the QE policy in 2009 depressed trades in the housing markets, and discouraged house searching behaviors due to affordability. Further insignificant expansion of the monetary supply may result in a "burnout effect" and the policy effect on the housing market and search interest of house buyers (i.e., the number of visits to the System) in most districts may be reduced. Central banks should bear in mind that a lax monetary policy may drive up house prices in the beginning, but then also lose its effectiveness as affordability deteriorates. The "lost decades" in

Japan (1990s economic stagnation) is a vigilant lesson that show the loss of housing affordability.

5.2 Implications

As the information from searching or transaction activities on websites continuously accumulate and is recorded, market movement may be predicted by the number of visits to search engines or industry information websites like the System. If the number of visits can be efficiently transferred into purchase action regardless of affordability, a higher predictive validity may be attained. This paper is the first empirical study to explore the relation between search behavior on the System in Taiwan and housing. The results show that the number of visits to the System leads the demand for housing in terms of transaction volume or prices in districts of high housing value, e.g., those with natural wonders, affordable prices, convenient transportation systems or sufficient infrastructures and public services. These results imply that the number of visits or search behavior can serve as a leading indicator for the housing market that offer good value or meet buyer needs, and that transaction volume may elicit the interest of more potential buyers to search for a house. Furthermore, a monetary policy may lose its effectiveness in bolstering house prices when affordability is deteriorated. Governments, home buyers and investors can refer to the results of this study and make use of the information on a similar platform in the future for a timely, efficient and precise policy or investment decision making.

Acknowledgement

The authors are grateful for the support from the Ministry of Science and Technology in Taiwan (MOST 107-2410-H-004-179).

References

Althouse, B., Ng, Y. and Cummings, D., (2011). Prediction of Dengue Incidence Using Search Query Surveillance. *PLOS Neglected Tropical Diseases*, August 2011, 5(8), e1258.

Askitas, N. and Zimmerman, K., (2009). Google Econometrics and Unemployment Forecasting, *Applied Economics Quarterly*, 55(2), 107-120.

Choi, H. and Varian, H., (2012). Predicting the Present with Google Trends, *Economic Record*, 88, 2–9.

Demchenko, Y., Grosso, P., de Laat, C. and Membrey, P., (2013). Addressing Big Data Issues in Scientific Data Infrastructure, IEEE International Conference on Collaboration Technologies and Systems (CTS), 48-55

Jun, S.P., Yoo, H.S. and Choi, S., (2018). Ten Years of Research Change Using Google Trends: From the Perspective of Big Data Utilizations and Applications, *Technological Forecasting and Social Change*, 130, 69-87.

Koop, G., Pesaran, M.H. and Potter, S.M., (1996). Impulse Response Analysis in Nonlinear Multivariate Models, *Journal of Econometrics*, 74, 119-47.

Kulkarni, R., Haynes, K.E., Stough, R.R. and Paelinck, J.H. (2009). Forecasting Housing Prices with Google Econometrics, GMU School of Public Policy Research Paper (2009-10).

Lin, T.C., (2019). Predicting Housing Markets through the Searching Behavior on Internet, *Taiwan Journal of Applied Economics*, 105, 219-254 (in Chinese).

Maclennan, D. and O'Sullivan, A., (2012). Housing Markets, Signals and Search, *Journal of Property Research*, 29(4), 324–340.

Ministry of the Interior (2019). Housing Information Platform, https://pip.moi.gov.tw/Eng/Default.aspx?pg=sitemap&type=C

Pesaran, M.H. and Shin, Y., (1998). Generalized Impulse Response Analysis in Linear Multivariate Models, *Economics Letters*. 58 (1), 17-29.

Polykalas, S., Prezerakos, G. and Konidaris, A., (2013). An Algorithm Based on Google Trends' Data for Future Prediction. Case Study: German Elections, In proceedings of *IEEE International Symposium on Signal Processing and Information Technology*, 69-73.

Sun, S.L., Wei, Y.J., Tsui, K.L. and Wang, S.Y., (2019). Forecasting Tourist Arrivals with Machine Learning and Internet Search Index, *Tourism Management*, 70, 1-10.

Van Dijk, D. and Francke, M., (2018). Internet Search Behavior, Liquidity and Prices in the Housing Market, *Real Estate Economics*, 46(2), 368-403.

Vosen, S. and Schmidt, T., (2011). Forecasting Private Consumption: Survey-based Indicators vs. Google Trends, *Journal of Forecasting*, 30, 541-564.

Wooldridge, J.M., (2019). Introductory Econometrics: A Modern Approach (7th edition). Boston, MA: Cengage.

Wu, L. and Brynjolfsson, E., (2015). How Google Searches Foreshadow Housing Prices and Sales. In Goldfarb, A., Greenstein, S., and Tucker, C. (Eds.), *Economic Analysis of the Digital Economy*, University of Chicago Press. pp. 89–118.

Wu, J. and Deng, Y., (2015). Intercity Information Diffusion and Price Discovery in Housing Markets: Evidence from Google Searches. *The Journal of Real Estate Finance and Economics*, 50(3), 289-306.

Appendices

Appendix A Dynamics of Three Exogenous Variables

Figure A.1 MoM Rates of Change in Construction Cost Index in New Taipei City

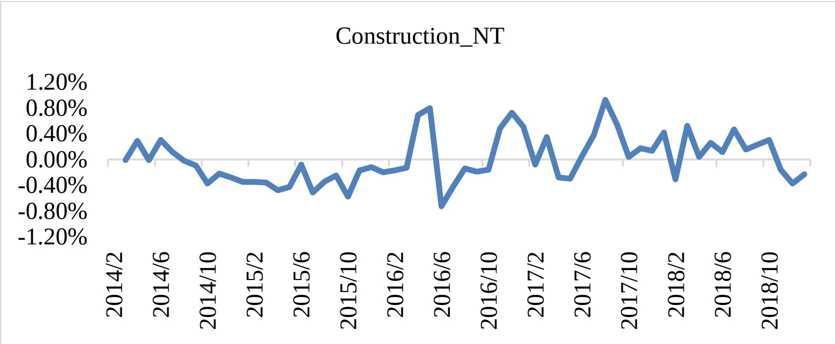


Figure A.2 MoM Rates of Change in CPI in New Taipei City

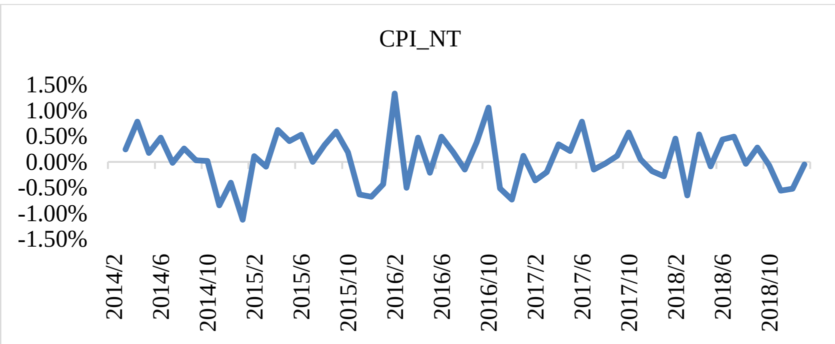
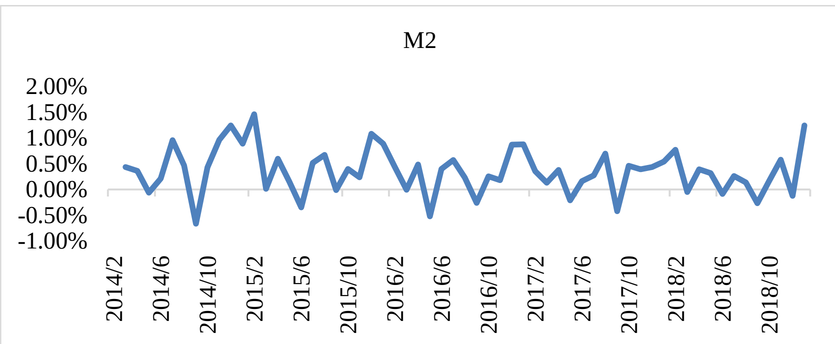


Figure A.3 MoM Rates of Change in M2 in Taiwan



Appendix B Augmented Dickey-Fuller Tests for Unit Roots

Table B.1 ADF Unit Root Tests on MoM Rates of Change

Deterministic term	Intercept		Intercept and linear trends		None	
Variable	*p*-value	Lag	*p*-value	Lag	*p*-value	Lag
(a) Endogenous Variable						
HIT_BQ	0.000	1	0.000	0	0.000	1
HIT_SC	0.000	0	0.000	0	0.000	0
HIT_XD	0.000	0	0.000	0	0.000	0
HIT_XZ	0.000	1	0.000	0	0.000	1
HIT_ZH	0.000	0	0.000	0	0.000	0
HPI_BQ	0.000	1	0.001	0	0.000	1
HPI_SC	0.007	1	0.013	0	0.001	1
HPI_XD	0.007	1	0.014	0	0.001	1
HPI_XZ	0.015	1	0.051	0	0.002	1
HPI_ZH	0.001	1	0.001	0	0.000	1
TRANSACTION_BQ	0.000	0	0.000	0	0.000	0
TRANSACTION_SC	0.002	9	0.005	0	0.000	9
TRANSACTION_XD	0.000	0	0.000	0	0.000	0
TRANSACTION_XZ	0.000	1	0.000	0	0.000	1
TRANSACTION_ZH	0.000	1	0.000	0	0.000	1
(b) Exogenous Variable						
CONSTRUCTION_NT	0.000	0	0.001	0	0.000	0
CPI_NT	0.000	0	0.000	0	0.000	0
M2	0.000	0	0.000	0	0.004	1

Notes: Three ADF tests (where the deterministic terms are with intercept, with intercept and linear trends and none) are used to test all rates of change. The optimal lag (0 to 10) in the ADF model for every series is determined by using a Schwarz information criterion (SIC) and the p-values of the test statistics are reported in Table B.1. Except for the testing results on *HPI_XZ* (where two *p*-values are slightly larger than 0.05), the ADF tests suggest that all of the other variables are not supposed to have unit roots; they are stationary.

Appendix C Estimation Results of VARX Models for Five Districts

Table C.1 VARX Models for BanQiao

	HPI_BQ		*TRANSACTION_BQ*		*HIT_BQ*	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
(a) Endogenous variable						
HPI_B(-1)	1.637***	17.746	29.180	0.667	-20.737	-1.224
HPI_B(-2)	-0.772***	-8.949	-22.426	-0.548	22.989	1.451
TRANSACTION_BQ(-1)	-0.000	-0.719	-0.291**	-2.127	0.077	1.464
TRANSACTION_BQ(-2)	-0.001**	-2.281	-0.296**	-2.101	-0.093*	-1.714
HIT_B(-1)	0.001	0.734	0.565	1.750	-0.318**	-2.540
HIT_B(-2)	0.000	0.157	0.921***	3.013	0.070	0.588
C	-0.000**	-2.304	0.027	0.300	0.023	0.674
(b) Exogenous variable						
CONSTRUCTION_NT	0.054*	1.823	-13.481	-0.968	-0.924	-0.171
CPI_NT	-0.021	-0.929	-22.647**	-2.122	-9.803**	-2.373
M2	0.023	0.844	-15.171	-1.182	-4.598	-0.926
CONSTRUCTION_NT(-1)	0.027	0.819	-2.841	-0.181	0.036	0.006
CPI_NT(-1)	0.045*	1.646	7.333	0.562	4.561	0.902
M2(-1)	0.073**	2.763	-12.386	-0.994	-9.877**	-2.048
MARCH	-0.001*	-1.820	0.825***	3.249	0.445***	4.529
R-squared		0.968		0.548		0.715
S.E. equation		0.001		0.327		0.127
F-statistic		101.454		4.002		8.282
Log likelihood		342.007		-9.210		44.896
Akaike information criterion		-11.509		0.814		-1.084
Schwarz information criterion		-11.007		1.316		-0.582

Note: "*", "**" and "***", respectively, denote the rejection of the null of zero coefficient at the 10 %, 5% and 1% significance levels based on t-statistics.

Table C.2 VARX Models for SanChung

	HPI_SC		*TRANSACTION_SC*		*HIT_SC*	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
(a) Endogenous variable						
HPI_SC(-1)	1.565***	15.465	2.343	0.064	-48.642**	-2.368
HPI_SC(-2)	-0.723***	-7.463	-9.940	-0.283	41.097**	2.089
TRANSACTION_SC(-1)	-0.001***	-2.699	-0.045	-0.312	0.244***	3.040
TRANSACTION_SC(-2)	-0.001**	-2.489	-0.315**	-2.065	0.030	0.352
HIT_SC(-1)	0.000	0.822	0.095	0.455	-0.354***	-3.028
HIT_SC(-2)	0.001	1.193	0.171	0.850	-0.022	-0.192
C	0.000	0.188	0.092	1.494	0.008	0.218
(b) Exogenous variable						
CONSTRUCTION_NT	0.039	1.520	13.871	1.506	-4.368	-0.846
CPI_NT	-0.011	-0.532	-14.717	-1.888	-7.802***	-1.787
M2	-0.027	-1.100	-27.854***	-3.143	-5.562	-1.120
CONSTRUCTION_NT(-1)	0.038	1.520	-16.285*	-1.703	-3.157	-0.589
CPI_NT(-1)	-0.041*	-0.532	5.755	0.696	6.072	1.311
M2(-1)	-0.023	-0.921	-12.180	-1.333	-8.587*	-1.677
MARCH	-0.000	-0.639	0.381**	2.113	0.555***	5.494
R-squared		0.943		0.635		0.739
S.E. equation		0.001		0.218		0.122
F-statistic		56.080		5.754		9.346
Log likelihood		349.968		14.088		47.105
Akaike information criterion		-11.788		-0.003		-1.162
Schwarz information criterion		-11.287		0.499		-0.660

Note: "*", "**" and "***", respectively, denote the rejection of the null of zero coefficient at the 10 %, 5% and 1% significance levels based on t-statistics.

Table C.3 VARX Models for XinDien

	HPI_XD		*TRANSACTION_XD*		*HIT_XD*	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
(a) Endogenous variable						
HPI_XD(-1)	1.468***	16.370	14.310	0.477	-34.361**	-2.220
HPI_XD(-2)	-0.606***	-7.561	-12.683	-0.473	25.784*	1.863
TRANSACTION_XD(-1)	-0.001**	-2.146	-0.427**	-2.449	0.044	0.486
TRANSACTION_XD(-2)	-0.001	-1.319	-0.150	-0.792	0.036	0.368
HIT_XD(-1)	0.003***	3.470	0.096	0.336	-0.254*	-1.732
HIT_XD(-2)	0.002**	2.101	0.141	0.495	-0.026	-0.175
C	-0.000	-1.301	0.133*	1.781	0.018	0.467
(b) Exogenous variable						
CONSTRUCTION_NT	0.048	1.456	4.339	0.396	-4.319	-0.764
CPI_NT	-0.041	-1.425	-15.977*	-1.668	-6.399	-1.294
M2	-0.013	-0.410	-15.410	-1.505	-6.546	-1.239
CONSTRUCTION_NT(-1)	0.059*	1.692	-16.505	-1.406	5.155	0.851
CPI_NT(-1)	-0.024	-0.812	5.899	0.594	4.380	0.854
M2(-1)	0.028	0.912	-32.088***	-3.175	-12.960**	-2.484
MARCH	0.001	1.120	0.444**	2.259	0.515***	5.072
R-squared		0.958		0.546		0.664
S.E. equation		0.001		0.257		0.133
F-statistic		76.160		3.983		6.531
Log likelihood		335.881		4.575		42.269
Akaike information criterion		-11.294		0.331		-0.992
Schwarz information criterion		-10.792		0.833		-0.490

Note: "*", "**" and "***", respectively, denote the rejection of the null of zero coefficient at the 10 %, 5% and 1% significance levels based on t-statistics.

Table C.4 VARX Models for XinZuang

	HPI_XZ		*TRANSACTION _XZ*		*HIT_XZ*	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
(a) Endogenous variable						
HPI_XZ(-1)	0.772***	13.455	3.113	0.289	-4.035	-0.637
TRANSACTION_XZ(-1)	0.001	1.129	-0.218	-1.530	0.171**	2.038
HIT_XZ(-1)	-0.001	-0.510	0.089	0.446	-0.264**	-2.246
C	-0.001*	-1.941	0.167***	2.813	-0.005	-0.140
(b) Exogenous variable						
CONSTRUCTION_NT	0.086*	1.776	6.928	0.761	-0.566	-0.106
CPI_NT	0.006	0.147	-14.525**	-2.042	-6.025	-1.441
M2	0.022	0.500	-22.900***	-2.754	-2.760	-0.565
CONSTRUCTION_NT(-1)	0.096*	1.858	-14.982	-1.544	-2.202	-0.386
CPI_NT(-1)	0.000	0.003	3.468	0.438	10.115**	2.176
M2(-1)	0.043	0.922	-35.010***	-3.955	-5.296	-1.018
MARCH	0.001	1.224	0.525***	3.975	0.481***	6.195
R-squared		0.857		0.584		0.684
S.E. equation		0.001		0.217		0.127
F-statistic		28.102		6.598		10.191
Log likelihood		316.219		12.530		43.357
Akaike information criterion		-10.525		-0.053		-1.116
Schwarz information criterion		-10.134		0.338		-0.725

Note: "*", "**" and "***", respectively, denote the rejection of the null of zero coefficient at the 10 %, 5% and 1% significance levels based on t-statistics.

Table C.5 VARX Models for ZongHe

	HPI_ZH		*TRANSACTION_ZH*		*HIT_ZH*	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
(a) Endogenous variable						
HPI_ZH(-1)	0.786***	13.105	-6.120	-0.496	-5.652	-0.741
TRANSACTION_ZH(-1)	0.001	1.133	-0.322**	-2.455	0.134*	1.659
HIT_ZH(-1)	0.001	0.623	0.289	1.487	-0.248**	-2.070
C	-0.000*	-1.889	0.069	1.282	-0.003	-0.090
(b) Exogenous variable						
CONSTRUCTION_NT	0.070*	1.710	8.228	0.981	-2.821	-0.545
CPI_NT	0.027	0.860	-8.774	-1.334	-6.091	-1.499
M2	0.007	0.194	-8.251	-1.076	-2.123	-0.448
CONSTRUCTION_NT(-1)	0.036	0.822	-7.289	-0.816	3.011	0.546
CPI_NT(-1)	0.005	0.152	6.490	0.874	12.010***	2.619
M2(-1)	0.019	0.531	-29.706***	-4.029	-7.725*	-1.697
MARCH	0.001	1.361	0.603***	4.983	0.454***	6.075
R-squared		0.858		0.552		0.695
S.E. equation		0.001		0.199		0.123
F-statistic		28.507		5.785		10.719
Log likelihood		326.455		17.492		45.453
Akaike information criterion		-10.878		-0.224		-1.188
Schwarz information criterion		-10.487		0.167		-0.797

Note: "*", "**" and "***", respectively, denote the rejection of the null of zero coefficient at the 10 %, 5% and 1% significance levels based on t-statistics.

Appendix D Generalized Impulse Response Functions

Figure D.1 GIRFs with 95% Confidence Interval in BanQiao

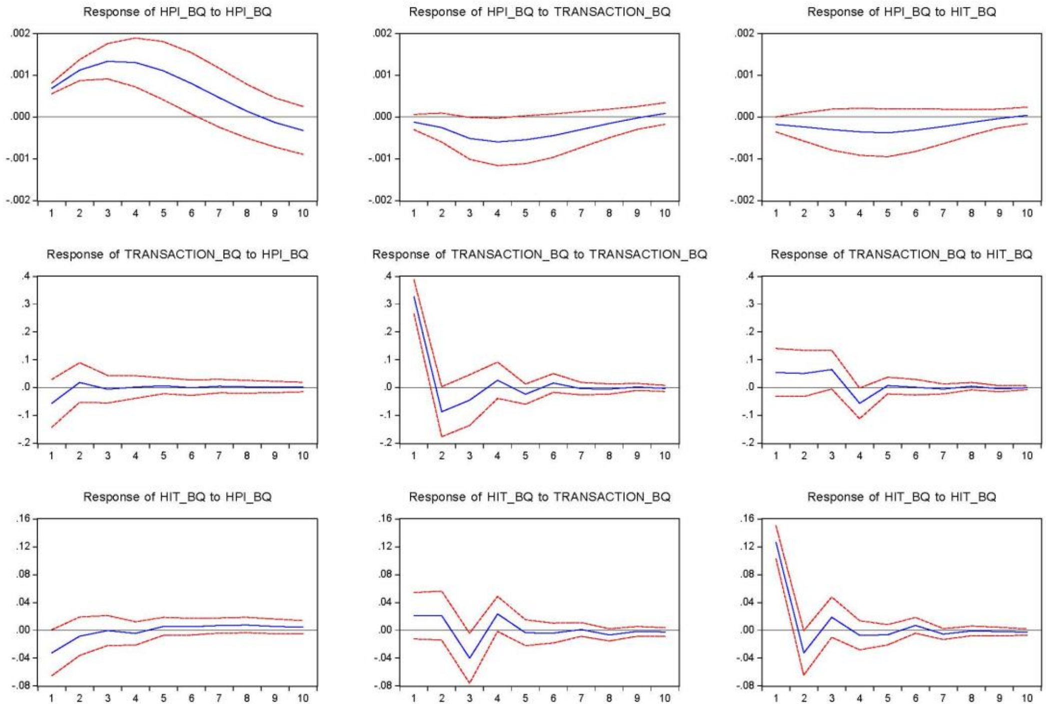


Figure D.2 GIRFs with 95% Confidence Interval in SanChung

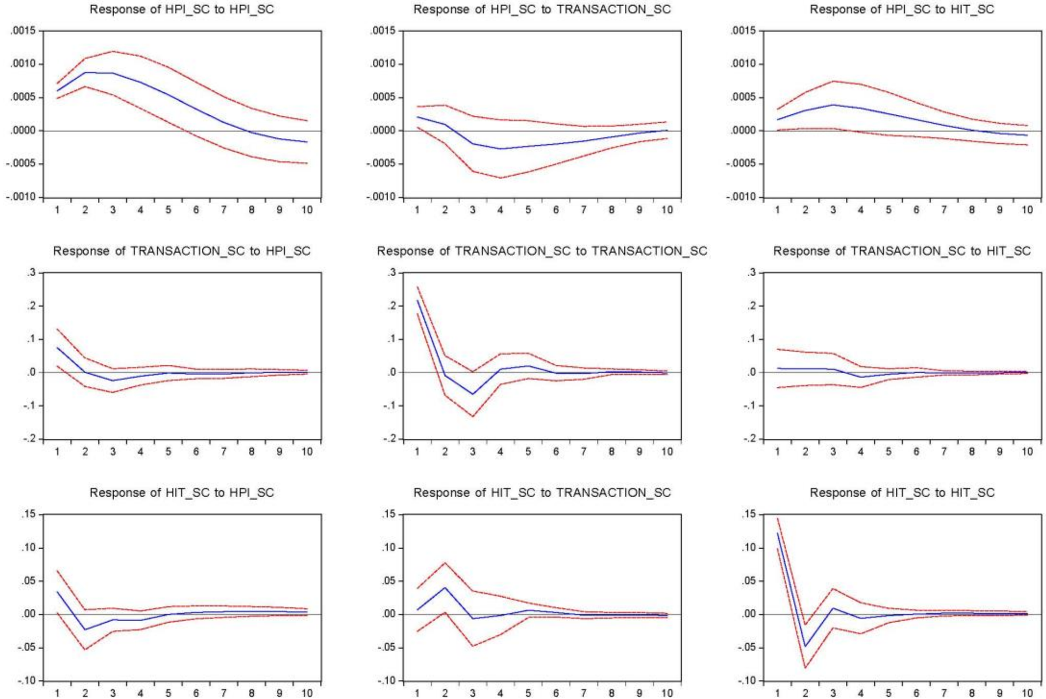


Figure D.3 GIRFs with 95% Confidence Interval in XinDien

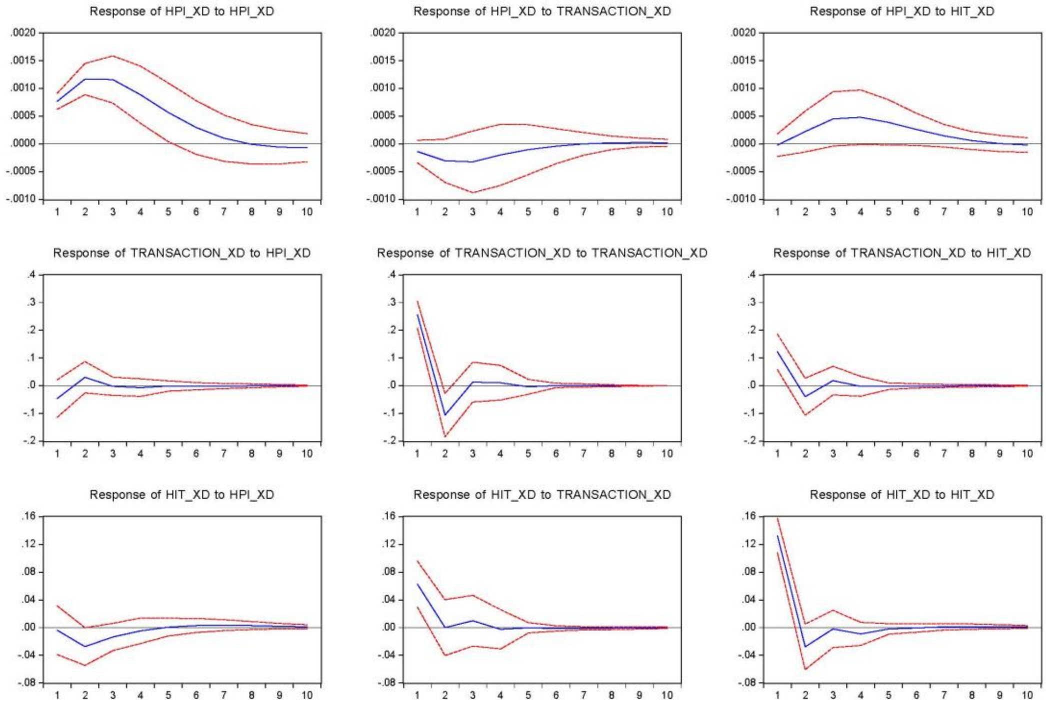


Figure D.4 GIRFs with 95% Confidence Interval in XinZuang

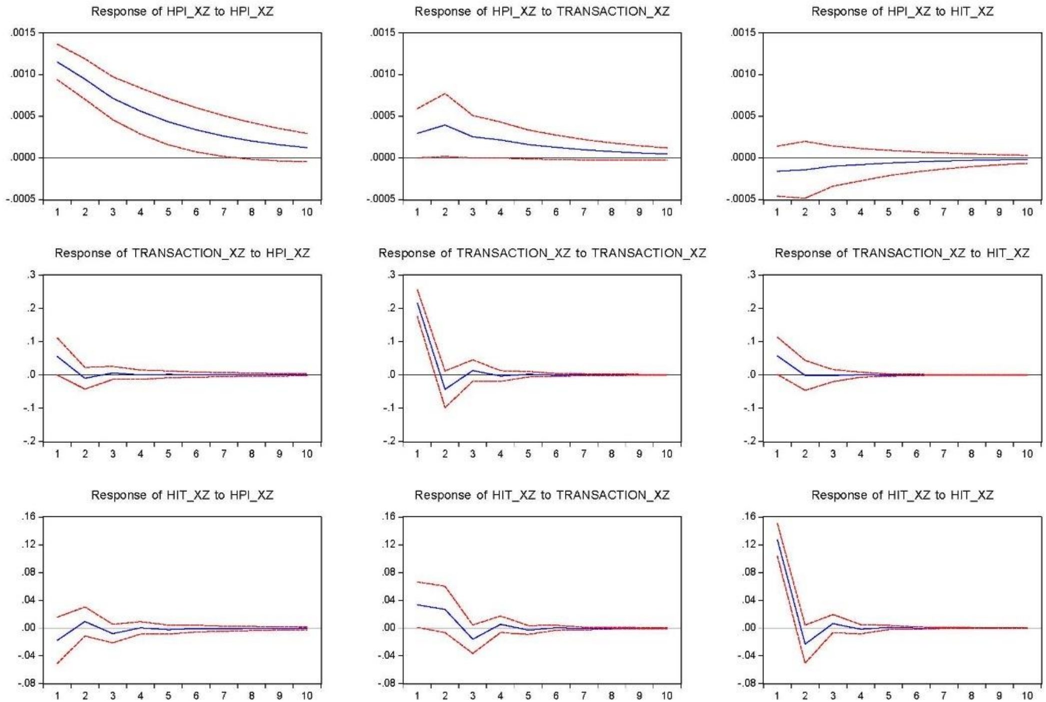


Figure D.5 GIRFs with 95% Confidence Interval in ZongHe