Predicting House Price Bubbles in the United States

Zijun Xie

A thesis submitted in partial fulfilment of the requirements for the degree of
MPhil of Philosophy

The University of Sheffield
Faculty of Faculty of Social Sciences
School of Economics

07/11/2024

Abstract

This study investigates new methods to predict house price bubbles in the United States. Using data from 1987 Q1 to 2023 Q2, the methods of Phillips, Shi and Yu (2015) [PSY] and Harvey, Leybourne and Whitehouse (2020) [HLW] are used to identify the timing of bubble episodes and construct a housing bubble indicator for the United States.

Both the PSY and HLW methods identify two episodes of housing market bubbles: one in the 2000s and another in the 2020s. Employing a binary bubble indicator as a dependent variable, this study investigates the relationship between housing bubbles and a number of macroeconomic and financial determinants using a Linear Probability Model as well as Probit and Logit estimation. We find evidence that housing market conditions, monetary policy, and stock market returns provide significant information on the probability of a housing bubble occurring. However, macroeconomic indicators, such as inflation, unemployment, and GDP growth, are not significant explanatory variables for housing bubbles across the full sample period.

In addition, subsample analysis reveals a structural break in the relationship between explanatory variables and the housing bubble indicator before and after the 2008 financial crisis. Moreover, extending the baseline model to include higher order lagged independent variables shows that the housing market reacts quickly to changes in economic conditions, as there is no improvement in in-sample forecast performance from variables lagged by half a year or more.

The study further assesses the out-of-sample forecasting performance of the three regression models using static, rolling, and recursive forecasts. It finds that static and rolling forecasts are inadequate for predicting housing bubbles, and the recursive model does not outperform a naïve random walk model in terms of forecast accuracy. These findings provide valuable insights into forecasting housing market bubbles and the impact of economic conditions.

List of contents

Abstract	page 1
Table of Contents	page 3
List of Tables	page 5
List of Figures	page 6

Please ensure that the following is included in your declaration:

I, the author, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously been presented for an award at this, or any other, university.

Please ensure that any publications arising from the thesis are acknowledged in this section.

MAIN BODY OF THE THESIS/CHAPTERS FOLLOWS

Appendices	page 58
Bibliography and/or References	page 62

Table of Contents

Abstract 1

Section 1 Introduction 7

Section 1.1 Housing market bubbles and financial crisis 7

Section 1.2 Research context 8

Section 1.3 Findings and contributions of the dissertation 8

Section 2 Literature review 10

Section 2.1 Influential factors for house prices 10

2.1.1 Monetary policy and monetary market conditions 10

2.1.2 Alternative investment opportunities 12

2.1.3 Macroeconomic conditions 13

2.1.4 Housing market condition 14

Section 2.2 Time series methods for modelling house prices 14

Section 2.3 Housing price bubbles 15

Section 2.4 Identification of housing bubbles 17

Section 2.5 Knowledge gap and research question 19

Section 3 Methodology 20

Section 3.1 Construction of housing bubble indicator 20

Section 3.2 Regression model 26

Section 3.3 Forecast performance evaluation 28

Section 3.4 Extended analysis 29

Section 3.5 Hypothesis test and expected findings 30

Section 4 Data set and summary analysis 31

Section 4.1 Collection and preparation of data set 31

Section 4.2 Selection and formation of variables 32

Section 4.3 Exploratory analysis 33

Section 5 Empirical findings..37
Section 5.1 Baseline estimation 37
Section 5.2 In-sample estimation evaluation 40
Section 5.3 Extended analysis...43
Section 5.4 Out-of-sample estimation evaluation 49
5.4.1 Static forecast and performance evaluation 50
5.4.2 Dynamic forecast and performance evaluation...51
Section 5.5 Summary of empirical analysis...53
Section 5.6 Limitations and further extension directions 54
Section 6 Discussion and conclusion...56
Appendix 1: List of variables and data sources 58
Appendix 2: Alternative regression model setting..59
Reference 62

List of tables

Table 4.1 Summary statistics by sampling periods..34

Table 4.2 Correlation coefficients..35

Table 5.1 Baseline estimation 37

Table 5.2 Measurements of model performance..41

Table 5.3 Regression with higher order lagged terms..46

Table 5.4a Subsample estimation for pre-2008 period 47

Table 5.4b Subsample estimation for post-2008 period..48

Table 5.5 Training set estimation with static model 50

Table 5.6 Out-of-sample performance evaluation with recursive model....................52

List of figures

Figure 1.1 Median housing price in the United States..7

Figure 3.1 Illustration of housing bubble indicators..25

Figure 5.1 Comparison of three model forecast performance 41

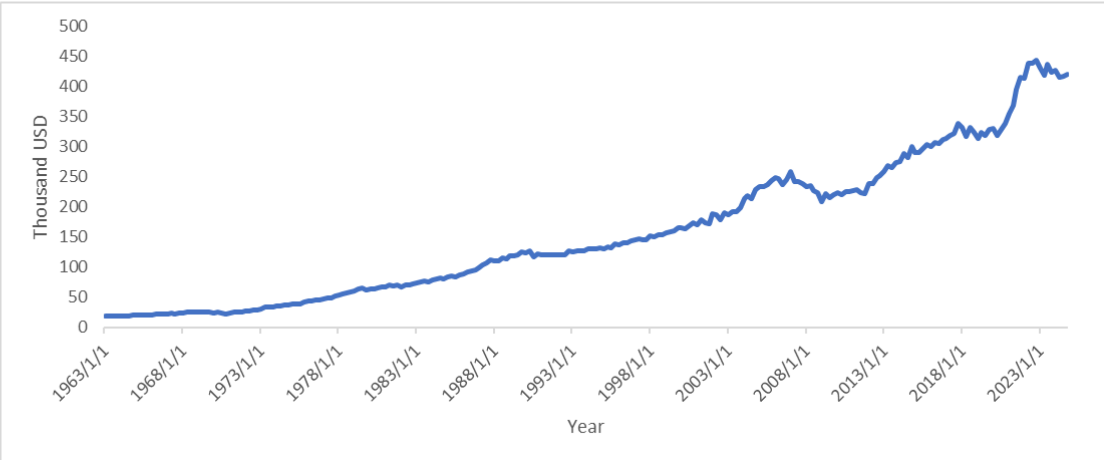
Figure 5.2 Out-of-sample forecasting performance with recursive model..................51

Section 1 Introduction

Section 1.1 Housing market bubbles and financial crisis

The 2008 global financial crisis represents a pivotal moment in economic history. In this period, the failure to predict and detect a housing market bubble had devastating consequences for both the United States [US] economy and the global economy. At the centre of the crisis was a housing market that became inflated by speculative lending, animal spirit and irrational exuberance, as well as financial derivatives related to mortgage-backed securities. As housing prices surged and then collapsed (see figure 1.1), the resulting shock spread across the financial market, triggering a widespread recession with increased unemployment, banking crises, and a sharp contraction in global economic activity.

Figure 1.1 Median housing price in the United States



Source: U.S. Census Bureau & U.S. Department of Housing and Urban Development. Data retrieved from FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/MSPUS

The speed and severity of the collapse demonstrate the consequence of an inability to foresee bubbles in asset markets. The central role of the housing bubble in the financial crisis implies a necessity to better understand, detect, and predict housing market bubbles.

However, correct detection of such bubbles remains a significant challenge. Despite the recent development of econometric bubble detection techniques, econometric forecasting has primarily focused on the prediction of house prices rather than on bubbles themselves.

This dissertation aims to address this gap by providing forecasting models of house price bubbles directly These forecasting models are expected to have crucial policy implications, supporting Central Banks, regulators and financial institutions in achieving financial stability.

Section 1.2 Research context

The housing market is a prominent topic in economics research, with substantial efforts made to understand the determinants of house prices. Over the years, various methods for identifying housing bubbles have been proposed, including the use of price-to-income ratios, the price-to-rent ratio, and more sophisticated statistical and econometric models.

However, the forecasting of house price bubbles remains an underdeveloped area, as many studies focus on predicting house prices rather than bubbles directly. Whilst the development of econometric methods to detect bubbles has improved our ability to identify when a bubble has emerged or collapsed, the prediction of these events is much more complex. Forecasting housing bubbles could provide valuable insights for policymakers, investors, and financial institutions. Such an analysis may ultimately help with the prevention of such financial crises as occurred in 2008.

This dissertation aims to make a novel contribution by predicting house price bubbles directly. By applying recently developed econometric methods of bubble detection to a time series dataset of US house prices, a binary bubble indicator can be constructed. This bubble indicator serves as the dependent variable in regression analysis to identify the determinants of housing bubbles and build an appropriate forecasting model.

Section 1.3 Findings and contributions of the dissertation

The main results of this study are as follows:

This study finds that housing market conditions, monetary policy, and stock market returns are significant predictors of a housing bubble, while traditional macroeconomic variables such as inflation, unemployment, and GDP growth are less important predictors.

When the forecasting models are extended to include higher order lagged independent variables, this study finds that the housing market reacts quickly to changes in economic conditions. This is because historical data beyond six months earlier provides little additional value for bubble forecasting. Furthermore, by conducting a subsample analysis of the pre- and post-2008 periods, the study detects the presence of a structural break in the relationship between explanatory variables and housing bubble indicators. This result suggests that the forces influencing housing market dynamics may evolve since the financial crisis.

Lastly, in terms of the forecast performance, the study finds that static forecasts and rolling forecasts both perform poorly for out-of-sample analysis. Meanwhile, although recursive model correctly predicts up to 40% of observations in out-of-sample evaluation, the predictive power fails to exceed the naïve model of a random walk process. The result indicates the challenge of forecasting house price bubbles directly.

The remainder of the dissertation consists of five chapters, which summarise the related literature, discuss the relevant research methods, introduce the dataset, illustrate the empirical findings, and concludes the discussion respectively.

Section 2 Literature review

Section 2 reviews previous studies to establish the academic background of the research. The content is split into five sections with the following arrangement. Section 2.1 discusses factors affecting house prices, which justifies the selection of predictors in the study. Section 2.2 discusses the time series method for house prices estimation. Section 3 and 4 introduces the formation and identification of housing bubbles respectively, which explain the construction of bubble indicator in the analysis. Section 2.5 sums up the discussion while specifying the gaps of research. The section develops research hypothesis and illustrates the expected contributions of the paper.

Section 2.1 Influential factors for house prices

From a theoretical perspective, the prediction of house prices can be separated into two major categories. On the one hand, microeconomic level analysis pays attention to the differing features of houses and the decision making by individuals. It adopts the hedonic pricing model to explain the price of houses given their features related to consumer utility (Sirmans et al., 2005). On the other hand, macroeconomic level analysis pays attention to the general conditions of aggregate demand and aggregate supply in the housing market (Tsai, 2012). The underlying idea can be traced back to the discounted cash flow framework. That is, when houses are viewed as an investment, the discounted value of the future cash flow (e.g., rental price) determines the fundamental value (Kishore, 1996). From such a point of view, several macroeconomic factors (from either demand side or supply side) could have an influence on house prices either through the change of cash flow or the change of discount rate. This study takes a macroeconomic perspective of housing market analysis and section 2.1 summarizes the main influential factors for house prices.

2.1.1 Monetary policy and monetary market conditions

To begin with, monetary policy, which affects the long term or short-term interest rate, could affect house prices directly as it changes the rate for discounting the future cash flow (Luciani, 2015). In particular, the short-term interest rate (within three-month time horizon) determines the liquidity in the monetary market, while the long-term interest rate (for one year or longer time horizon) determines the cost for long run borrowing as well as guide the formation of market expectation.

For instance, the early study by Harris (1989) points out the negative relationship between real interest rate and house prices due to the shrinkage of market demand with the higher borrowing cost for mortgage loans. Further, McQuinn and O'Reilly (2012) suggest that interest rate also affects house prices through the channel of the borrowing constraint. When interest rate tends to be low, mortgage loans become more affordable. It explains the inverse relationship between house prices and interest rate (Garriga et al., 2019).

Apart from interest rate, another indicator of monetary policy is demonstrated to affect house prices as well: money supply. This is because the supply of money affects the liquidity in the monetary market. It not only boosts the house prices through the direct channel of bank lending, but also boosts the housing market through the support of other business activities. To be specific, when money provision is adequate, from the supply side, we can anticipate more investment in the construction sector. From the demand side, we can anticipate the improved capacity of purchasing by consumers or investors. That is why supply of money is a crucial indicator for house prices (Lastrapes, 2002).

Further research suggests that it is not just monetary policy itself that affects house prices, but that uncertainty about economic policy is also influential (Wang et al., 2020). The underlying mechanism is that the higher uncertainty of monetary policy adds to the volatility and uncertainty of housing market, which translates into the change of house prices when consumers (or investors) are risk averse and reduce investment under a higher volatility.

It is of importance to recognize that the influence of monetary policy on the housing market can be asymmetric. For instance, with a time series analysis of the UK housing market, Tsai (2013) claims the rigidity of housing price for downward movement. That is, while the housing market is responsive to forces that drive up the price level, its adjustment downward is rather slow. Combining behavioural forces like loss aversion, the study argues that there is an overreaction to loose monetary policy and an underreaction for tight monetary policy. Using a state level panel data set in the United States, Bahaman-Oskooee et al. (2024) also indicate that expansionary monetary policy and contractionary monetary policy have asymmetric effects on the housing market. However, contrary to Tsai (2013), they find that the market reaction to contractionary monetary policy is more rapid and of a larger magnitude.

2.1.2 Alternative investment opportunities

Aside from monetary policy, another major macroeconomic force that affects the housing market is the market condition of alternative investment opportunities. For housing market participants who treat real estate as a financial investment, return and risk of other investment choices could have a strong impact on their intention of housing purchase (or sale). Therefore, alternative investment opportunities serve as a critical force that drives the demand and supply in the real estate market (Boone and Girouard, 2003). The relationship between the housing market and other investment markets can be in either direction. On the one hand, through the channel of wealth effect and sentiment spill over effect, the co-movement of different investment markets is expected (Zheng and Osmer, 2020). In detail, when investment from other asset market gains value, wealth of potential house buyers increases. It leads to the possible rise of house price from the demand side force. On the other hand, when investors are confronted with a credit constraint that limits the total scale of investment, an opposite correlation between housing price and market condition of other investment opportunities is anticipated (Bahmani-Oskooee and Ghodsi, 2018). For instance, when the stock market is offering a high return, investors who face a credit constraint may forgo housing investment. The implied correlation between investment return in the alternative asset market and housing market would then be negative.

Empirical analysis confirms the interaction between housing prices and returns in alternative investment markets. For example, Chiang et al. (2020) examine the dynamic interplay between the U.S. stock and housing markets. Using a Markov switching vector autoregressive (MSVAR) approach, the study identifies a long-term co-integration of these two markets specifically during low-volatility periods, indicating a stable, long-run relationship. However, in high-volatility regimes, this relationship reverses, with market spill over effects showing an opposite directional influence between the two markets. This finding demonstrates that while stock and housing markets generally move together over the long term, periods of high volatility alter this connection, likely due to shifts in investor behaviour or capital reallocation. Furthermore, Liow et al. (2019) extend this analysis by showing that it’s not only returns that transmit across markets, but also the risk (volatility) associated with these assets, leading to co-movements in both stock and real estate market volatilities. This highlights that both markets are interconnected not just in terms of returns but also in terms of shared risk, especially during periods of economic uncertainty.

2.1.3 Macroeconomic conditions

In addition to the market condition of other investment products, macroeconomic conditions are another set of critical determinants for housing price (Mohan et al., 2019). This is because the overall performance of the economic system affects the demand side of the housing market through affordability of houses and the anticipation of future market condition (Nneji et al., 2013). Moreover, indicators of macroeconomic conditions could also be associated with housing price from the supply side, as they affect the cost of construction (Adams and Füss, 2010). Specifically, common macroeconomic variables that are documented to have a predictive power for housing price are discussed below.

To start with, the inflation rate is shown to affect housing price in opposite directions in the short run and the long run (Schwab, 1982). In detail, the rising cost of living due to high inflation in the short run is documented to be positively related to housing price. This is because the increase of rental price could be the underlying force of inflation (Kuang and Liu, 2015). However, in the long run, the anticipation for inflation reduces the real value of future cash flows and increases the expectation for a contractionary policy by the central bank, which leads to the inversed relationship between inflation rate and housing price (Baldi, 2014).

Secondly, indicators for the business cycle are shown to have an effect on housing price. For instance, a high unemployment rate increases precautionary savings and pessimistic sentiments among consumers, which reduces housing demand and hence housing prices (Jen-Shi et al., 2012). On the contrary, a high level of GDP (gross domestic product) and a high rate of economic growth imply an improved consumption capacity, which is positively related to housing price (Amidu et al., 2016). Specially, the boom and bust of the housing market are also found to have a reversed influence on macroeconomic conditions (Chi-Wei et al., 2018). A possible underlying mechanism discussed in the paper is the wealth effect and the credit constraint effect. In detail, when housing price goes up, the collateral value of houses rises. It releases the credit constraint for borrowing and in turn stimulates consumption. Meanwhile, the wealth level of households also gets higher with a housing bubble, which according to the idea for lifetime consumption smooth, causes the rise of consumption level. However, when there is a crash of the housing market, the tightening of the credit constraint and the shrinkage of the household wealth would lead to the reduction of consumption and the recession of macroeconomy.

Lastly, residential investment is shown to have an interactive relationship with housing price (Petrini and Teixeira, 2023). As the prosperity of housing market initiates more residential investment, while the increase of residential investment drives up housing market demand as well.

2.1.4 Housing market condition

Apart from the external factors discussed so far, the condition of the housing market itself significantly influences housing prices through supply-side forces (Glaeser and Gyourko, 2018). For instance, research on Chinese cities has shown that the overall quantity of housing units is closely related to housing prices. An increase in housing supply, often through new construction, can stabilize prices by balancing supply and demand, particularly during periods of market overheating (Wu et al. 2016). This is because a sufficient stock of available houses moderates price expectations and helps prevent excessive price increases. In particular, Duca et al. (2021) suggest that macroeconomic policy could interplay with these housing market factors in the determination of housing price.

Section 2.2 Time series methods for modelling house prices

Time series approaches are often adopted for the analysis of housing price at the macro level. Specially, traditional time series methods can be categorized into the following three sets with respect to the application in housing price estimation.

Firstly, ARIMA regression is applied for the univariate analysis of housing market dynamics (Tse, 1997). The approach models the movement of housing price with lagged market shock and autoregressive terms (Crawford and Fratantoni, 2003). It makes possible the identification of serial correlation in housing price (Chirchir et al., 2024). However, ARIMA models may struggle with sudden economic shocks or structural changes. This limitation leads to the combination ARIMA with structural breaks as well as time-varying volatility. These models better capture periods of high uncertainty, such as during economic downturns, by accounting for volatility clustering (Engle, 1982). For example, the univariate framework of housing price estimation can be augmented with the combination of GARCH related models to account for the serial correlation of housing price volatility (Crawford and Fratantoni, 2003).

Secondly, while the univariate estimation captures the autocorrelation of housing price, it fails to recognize the influence of external forces that affect the real estate market. In order to account for the impact of macroeconomic variables on housing price, ARDL and VAR methods are adopted. VAR analysis is particularly useful for examining the effects of policy changes on the housing market. This is especially relevant as central banks increasingly adjust policies to manage inflation, which can directly influence housing affordability and demand. For instance, Vargas-Silva (2008) applies the VAR approach to analyse the effect of monetary policy on housing price. Through the inclusion of a sign restriction, the VAR framework makes possible the identification of asymmetric housing market response to macroeconomic variables. In addition, the VAR framework can be assessing the co-movement of housing price across different regions and countries (Vansteenkiste and Hiebert, 2011).

Lastly, time series analysis can not only account for the short run dynamics of housing price, but also identify the long-term relationship of housing price with macroeconomic variables. Through the applications of VECM (Anoruo and Braha, 2008), both the short run market movement and the long run equilibrium relationship can be identified. By establishing a long-term equilibrium, it works in predicting market corrections when prices deviate from their long-term path. In addition, DSGE framework can be adopted to specify the general equilibrium of an economic system while identifying the response of housing price to a series of exogenous shocks (Iacoviello and Neri, 2010). It provides an integrated framework that accounts for broader economic cycles and potential exogenous shocks. It is proved to be useful in forecasting housing market responses to a wide array of economic disturbances by incorporating endogenous adjustments. For example, the DSGE model can simulate the housing price impact of policy changes, technological advances, or financial shocks, thus offering more nuanced, scenario-based forecasts (Iacoviello and Neri, 2010).

Section 2.3 Housing price bubbles

Housing price bubbles occur when property prices significantly exceed their fundamental values (Smith and Smith, 2006). This rapid increase in housing prices is often considered to be unsustainable over the long term and is typically followed by a sharp decline, or a "burst," when market corrections bring prices back to more realistic levels (Harvey et al., 2020). In essence, a housing bubble is characterized by a departure from the equilibrium between housing prices and fundamental determinants like income levels, rent prices, and housing demand (Cheng, 2014).

The formation of housing bubbles may involve a combination of factors, including speculative behaviour, market sentiment, and favourable economic conditions (Glaeser and Nathanson, 2015). Initially, optimistic expectations drive investors and homebuyers to purchase properties, anticipating that prices will continue to rise. This leads to a self-fulfilling cycle of demand and price increases. High leverage and easy access to credit can amplify this effect, as buyers are more willing to take on debt, expecting future price appreciation to offset their borrowing costs. When financial institutions relax lending standards, as seen in the lead-up to the 2008 global financial crisis, the housing bubble's growth accelerates. As Soo (2018) suggests, swings in market sentiment play a critical role; extreme optimism fosters higher investments and drives up prices, creating the bubble.

Several indicators are commonly used to detect the presence of housing bubbles. One is the price-to-rent ratio, which compares housing prices to rental income. When this ratio is substantially above historical norms, it may signal overvaluation (Bourassa et al., 2019). Another metric is Tobin's Q, which compares property prices to replacement costs. In detail, a value greater than one indicates that asset prices are above replacement costs, which suggests a bubble (Donald and Winkler, 2003). Additionally, speculative trading volume often spikes in bubble periods, as more investors enter the market expecting short-term gains. Shiller (2005) highlights the importance of such volume surges, noting that increased speculative transactions typically precede a price correction or bubble burst. Moreover, market sentiment surveys and transaction volume analysis provide qualitative insights, as increased optimism and speculative buying further signal potential bubble conditions (Marfatia et al., 2022).

Understanding housing bubbles is essential for preventing the economic disruption that typically follows their collapse. When housing bubbles burst, they often lead to significant declines in household wealth, increased foreclosures, and broader economic downturns, as seen in the 2008 crisis. This can affect not only homeowners but also financial institutions, as falling prices undermine the value of mortgage-backed securities and related financial assets (Diewert et al., 2009). Policymakers and economists need to recognize early signs of bubbles to implement regulatory or policy measures that can mitigate their effects, such as tightening lending standards or increasing interest rates. Furthermore, accurate bubble detection methods enable investors and stakeholders to make informed decisions, reducing their risk exposure and enhancing market stability (Arce and López-Salido, 2011).

To sum up, housing market bubbles are the phenomenon of housing price deviation from housing value. Identifying these bubbles early could help with the prevention of financial crises like those seen in 2007-2008. As such, developing robust econometric methods for bubble detection becomes crucial in economic research.

Section 2.4 Identification of housing bubbles

Traditional models of bubble detection, such as the present value model, rely on the rational bubble assumption. This model posits that the price of a housing asset should reflect the discounted future income streams in the absence of bubbles. Any deviation from this, due to speculative behaviour or irrational market exuberance, could signal a bubble (Blanchard and Watson, 1982). Another early approach, variance bounds tests, proposed by Shiller (1981), assessed whether the variance in observed prices exceeds that of the fundamental value. However, these methods are subject to criticism for lacking robustness and sensitivity to certain market conditions.

To solve the limitation, more advanced econometric models evolve. For example, unit root and co-integration tests are introduced by Campbell and Shiller (1987) for bubble detection. These tests aim to detect explosive behaviour in the gap between an asset's price and its fundamental value, where non-stationary prices relative to stationary fundamentals signal a bubble. These methods are applied in various housing markets worldwide, from the UK in the mid-1980s to the US in the mid-2000s. However, their inability to detect periodically collapsing bubbles, where bubbles burst and reform, pose a significant limitation highlighted by Evans (1991). To be specific, Evans (1991) argues that while the method by Campbell and Shiller (1987) helps to signal the presence of a housing bubble, it fails to reveal the exact time of the bubble and cannot be applied for the detection of multiple house price bubbles.

More recently, improvements in econometric techniques address these deficiencies. For instance, the sup Dickey-Fuller (SADF) test by Phillips et al. (2011) introduce a forward recursive right-tailed test to enhance the power of bubble detection, particularly in real-time applications. This method provides consistent estimates of both bubble origination and termination dates. Moreover, to account for multiple bubble episodes, Phillips et al. (2015) extend this approach by varying both the starting and ending points of the sub-sample windows used in the construction of the test statistic, thus improving detection accuracy for multiple bubbles. These tests now form the core of modern bubble detection.

These tests now form the core of modern bubble detection, which serves as the foundation of bubble detection techniques upon which new advancements are made. For instance, Astill et al. (2018) extend the one-shot test framework with the comparison of subsample statistics with the monitoring period versus the training period. With two different approaches for bubble warning based on real-time comparison and threshold value respectively, the study indicates the possibility of reducing false positive rate in bubble detection. In addition, the recent study by Whitehouse et al. (2025) adopts a similar procedure of explosive autoregression for the detection of housing bubbles in OECD countries. However, the study employs an alternative variance standardization which enables the early detection of a housing bubble.

This study takes the approach by Phillips et al. (2015) to construct a bubble indicator for the US housing market. A detailed methodology is provided in section 3.1.

When it comes to the prediction of housing bubbles instead of housing price, what makes the situation more complex is the fact that the outcome variable is a binary indicator instead of a continuous measurement. In such a case, the property of a binary response variable renders the inadequacy of a linear regression model as it suffers from problems including heteroscedastic and non-normal distribution of the error term, while failing to limit the predicted value within feasible range and failing to account for the varying marginal effect of independent variables (Wooldridge, 2015).

To cater for the insufficiency of a linear regression model, maximum likelihood estimation should be adopted. Through the application of a cumulative probability function, methods like Probit and Logit estimation overcome the drawbacks of the linear probability model. Therefore, these methods are widely adopted in the prediction of binary response variables like market recessions and bankruptcies (Maddala, 1983). For instance, Estrella and Mishkin (1988) find that Logit estimation can be applied to effectively forecast market recessions in US. Similarly, these methods are found to be useful in the prediction of bankruptcies in studies like Cole and Gunther (1998) as well as Cole and White (2012).

Section 2.5 Knowledge gap and research question

The review of previous studies in the four sections above summarizes current knowledge regarding the prediction of housing price, the detection of housing bubble, and the relevant econometrics methods.

Most of the existing literature considers modelling and predicting house prices with limited efforts devoted to the prediction of housing bubbles themselves. In fact, there are only some standalone studies, like Dreger and Kholodilin (2012), that notice the importance of a forecasting model for housing bubbles to guide policymakers in the decision of monetary policies and fiscal policies. Nonetheless, while the study by Dreger and Kholodilin (2012) compares the performance of the signalling approach, Probit model, and Logit model in the forecasting of housing price bubbles in 12 OECD countries, the method of bubble construction in the research is rather outdated. It fails to incorporate the most recent idea for the detection of multiple bubbles in the housing market or the idea for the separation of different bubble regimes in the data generation process.

However, considering the triggering role of house price bubbles and crashes in financial crisis, it is of importance to detect housing bubbles in advance for both better investment decision making and economic policy design. To do this, requires econometric methods that are suitable to predict a binary response variable.

Using US time series data, this study will achieve the following:

- Build a forecast model of US house price bubbles.
- Directly examine the relationship between house price bubbles and a range of macroeconomic factors.

Section 3 Methodology

This section introduces the research methodology. The content is split into five sub-sections with the following arrangement. Section 3.1 discusses the construction of the bubble indicator for the housing market, which will be used as the dependent variable for the analysis. Section 3.2 presents the regression model and discusses the choice of estimation methods. Section 3.3 illustrates the method for evaluating the model performance. Section 3.4 discusses the possible extension of the baseline estimation. Section 3.5 presents the expected findings from the empirical analysis.

Section 3.1 Construction of housing bubble indicator

A challenge of this study is that housing bubbles are not directly observable. Instead, this study constructs a binary variable to indicate a housing bubble in the US based on the approach by Phillips et al. (2011, hereafter PWY), Phillips et al. (2015, hereafter PSY), and Harvey et al. (2020, hereafter HLW).

These models are constructed upon the framework of present value model under the assumption for rational bubble. Specifically, as is indicated by Diba and Grossman (1988), asset bubbles manifest as explosive autoregressive processes.

The central idea lies with the non-linear explosive character of the housing price series, which is first established in the detection of an asset bubble with the right tailed DF test.

A simplified framework of the test can be summarized as below.

To start with, an AR (1) model is adopted for the data generation process of the asset price:

$$y_t = \phi y_{t-1} + \varepsilon_t$$

In this model, when $|\phi| < 1$, the time series is covariance stationary. When $\phi = 1$, the asset price follows a random walk process. While when $\phi > 1$, the asset price becomes explosive, which indicates the presence of an asset price bubble (Blanchard and Watson, 1982).

Correspondingly, the formation and the crash of an asset price bubble can be represented as follows:

$$y_t = \begin{cases} y_{t-1} + \varepsilon_t \; t \in [1, \tau_1 T] \\ \phi_1 y_{t-1} + \varepsilon_t \; t \in [\tau_1 T + 1, \tau_2 T] \\ \phi_2 y_{t-1} + \varepsilon_t \; t \in [\tau_2 T + 1, \tau_3 T] \\ y_{t-1} + \varepsilon_t \; t \in [\tau_3 T + 1, T] \end{cases}$$

where $\phi_1 > 1$ captures the formation of the bubble and $\phi_2 < 1$ captures the crash of the bubble, while for the rest of the period, a random walk process is assumed for the asset price.

The associated detection of the asset pricing bubble is based on the right tailed ADF test for a unit root process. That is, the procedure is quite similar as the ADF test in its classical (left-tailed) format, while the critical value for the test is adjusted for the right tailed distribution.

$$y_t = \phi y_{t-1} + \varepsilon_t$$

$$\Delta y_t = \delta y_{t-1} + \varepsilon_t \; (\delta = \phi - 1)$$

$$H_0: \delta = 0 \; H_1: \delta > 0$$

in which the rejection of the null hypothesis would indicate the presence of an asset pricing bubble.

The methods by PWY and PSY extend the ADF test in several aspects. To begin with, their approaches make possible the identification of bubbles with a flexible set of initial observations and allows for the flexible selection of window size. In addition, PWY conducts the calculation for the right tailed ADF test statistics based on the forward recursive regression. That is, the estimation takes a fixed starting point of the data for estimation while extends the ending point of the sample recursively for the calculation of the ADF test statistics. Then, the sup ADF test statistics is applied for the detection of an asset price bubble.

In detail, as is discussed in Evans (1991), the explosion of asset price series can be a temporary phenomenon. As a result, a full sample method for the detection of an asset pricing bubble could be inadequate. For the adjustment, PWY applies the calculation for the right tailed ADF statistics for recursive subsamples of the data set.

That is, for a sample interval of [0,1], the ADF statistics is constructed with the initial window $[0, r_2]$, where $r_1 = 0, \; r_2 \in [r_0, 1]$. Specially, r_2 takes a varying value and the corresponding test statistics are calculated for different subsample periods.

Lastly, the PWY test statistics is found with the supremum for the test statistics calculated above as:

$$SADF = \sup_{r_2 \in [r_0, 1]} ADF_{r_2}$$

The approach also makes possible the identification of the starting point and the ending point for the asset price explosion as:

$$\hat{r}_e = \inf_{\mathrm{r}_2 \in [r_0, 1]} r_2 \colon ADF_{r_2} > CV_{ADF}(r_2)$$

$$\hat{r}_f = \inf_{\mathrm{r}_2 \in [\widehat{r_e}, 1]} r_2 \colon ADF_{r_2} < CV_{ADF}(r_2)$$

Nonetheless, there is a central limitation of the PWY approach as it allows for the identification of a single bubble in the sampling period only. As a result, the method is less appropriate to time stamp the subsequent asset pricing bubbles. Such a drawback gets addressed in the approach of PSY with a more flexible recursive regression that has varying number of observation and varying initial observations for the estimation. In particular, PSY calculates the sup ADF test statistics based on backward expanding samples with a flexible window size of regression. It allows for the exact time stamp of multiple asset bubbles during the sampling period.

The central difference for the construction of ADF test statistics for these two approaches lies with the starting point for the estimation window. In the PWY framework, the starting point of the subsample r_1 is a fixed one. However, for PSY, it is allowed to vary over the range of $[0, r_2 - r_0]$. In this way, multiple bubbles can be detected with the test statistics formed as:

$$GSADF = \sup_{r_1 \in [0, r_2 - r_0], r_2 \in [r_0, 1]} ADF_{r_2}^{r_1}$$

with rejection occurring when GSADF test statistics exceeds the right tailed test critical value.

Similar as the framework for PWY, the method by PSY allows for the specific identification of the date for bubble and crash based on the backward SADF statistics with the formula as:

$$BSADF_{r_2}(r_0) = \sup_{r_1 \in [0, r_2 - r_0]} ADF_{r_2}^{r_1}$$

$$\hat{r}_e = \inf_{r_2 \in [r_0, 1]} r_2 \colon BSADF_{r_2} > CV_{SADF}(r_2)$$

$$\hat{r}_f = \inf_{r_2 \in [\hat{r}_e, 1]} r_2 \colon BSADF_{r_2} < CV_{ADF}(r_2)$$

Apart from the improvement with multiple bubble detection by PSY, the study by HLW further enhances the accuracy of the identification of date for bubble and crash with a two-step procedure.

For the first step of the HLW approach, PSY method is applied. It helps with the identification of bubble intervals to split the sample into different windows. We need the assumption that each date interval contains only a single explosive bubble for the asset.

For the second step of the HLW approach, we determine the most appropriate model for bubble and crash and estimate the change points for the model.

For instance, if a sampling period contains a total of N bubbles, the regime of the asset price will be denoted as:

$$y_t = \begin{cases} y_{t-1} + \varepsilon_t \ t \in [1, \tau_{j1}T] \\ \phi_{j1} y_{t-1} + \varepsilon_t \ t \in [\tau_{j1}T + 1, \tau_{j2}T] \\ \phi_{j2} y_{t-1} + \varepsilon_t \ t \in [\tau_{j2}T + 1, \tau_{j3}T] \\ y_{t-1} + \varepsilon_t \ t \in \left[\tau_3 T + 1, \tau_{(j+1)1}T\right] \\ \cdots \end{cases}$$

where we denote the regime of each bubble again with τ_{j1} and τ_{j2}, while the ending time for the crash regime is denoted with τ_{j3}.

The associated data generation process can be separated into four types:

Type 1: $0 < \tau_{j1} < 1, \phi_{j1} > 1, \tau_{j2} = 1$

It means that the asset price is first a unit root process and then becomes explosive toward the ending period.

Type 2: $0 < \tau_{j1} < \tau_{j2} < 1, \phi_{j1} > 1, \phi_{j2} = 1, \tau_{j3} = 1$

It means that the asset price is first a unit root process and then becomes explosive but transforms again to a unit root process without a crash until the ending period.

Type 3: $0 < \tau_{j1} < \tau_{j2} < 1, \phi_{j1} > 1, \phi_{j2} < 1, \tau_{j3} = 1$

It means that the asset price is first a unit root process. It then becomes explosive and stationary collapse toward the ending period.

Type 4: $0 < \tau_{j1} < \tau_{j2} < \tau_{j3} < 1, \phi_{j1} > 1, \phi_{j2} < 1$

It means that the asset price is first a unit root process. It then becomes explosive and collapse. Finally, the process returns to a unit root toward the ending period.

To apply the HLW method for the precise determination of time for asset pricing bubble, we first utilize the GSADF test statistics discussed above to check whether the sampling period involves asset bubbles. In the case where an asset bubble is detected, we then adopt the BSADF test statistics to identify the time of the bubble regime.

Based on the regime specified with PSY approach, we can split the sampling period as follow:

$$s_j = \begin{cases} 1 \; j = 1 \\ e_{j-1} + 1 \; j > 1 \end{cases}$$

$$e_j = \begin{cases} \left|\widehat{\tau_{j2}}^{PSY} T\right| + \dfrac{\left|\widehat{\tau_{(J+1)1}}^{PSY} T\right| - \left|\widehat{\tau_{j2}}^{PSY} T\right|}{2} \; j < \widehat{N} \\ T \; j = \widehat{N} \end{cases}$$

Then for the next step, we can match each of the case of the $\widehat{N}$ bubbles with any of the four types. With the type of each bubble categorized, change point is found by minimizing the residual sum of squared for any possible dates in the bubble regime time with the following model form:

Type 1 model:

$$\Delta y_t = \hat{\mu}_1 D_t(\tau_{j1}, 1) + \widehat{\delta_{j1}} D_t(\tau_{j1}, 1) y_{t-1} + \hat{v}_{1t}$$

$$\widehat{\tau_{j1}} = argmin\, SSR_{j1}(\tau_{j1})$$

Type 2 model:

$$\Delta y_t = \hat{\mu}_1 D_t(\tau_{j1}, \tau_{j2}) + \widehat{\delta_{j1}} D_t(\tau_{j1}, \tau_{j2}) y_{t-1} + \hat{v}_{2t}$$

$$(\widehat{\tau_{j1}}, \widehat{\tau_{j2}}) = argmin\, SSR_{j2}(\tau_{j1}, \tau_{j2})$$

Type 3 model:

$$\Delta y_t = \hat{\mu}_1 D_t(\tau_{j1}, \tau_{21}) + \hat{\mu}_2 D_t(\tau_{j2}, 1) + \widehat{\delta_{j1}} D_t(\tau_{j1}, \tau_{j2}) y_{t-1} + \widehat{\delta_{j2}} D_t(\tau_{j2}, 1) y_{t-1} + \hat{v}_{3t}$$

$$(\widehat{\tau_{j1}}, \widehat{\tau_{j2}}) = argmin\, SSR_{j3}(\tau_{j1}, \tau_{j2})$$

Type 4 model:

$$\Delta y_t = \hat{\mu}_1 D_t(\tau_{j1}, \tau_{21}) + \hat{\mu}_2 D_t(\tau_{j2}, \tau_{j3}) + \widehat{\delta_{j1}} D_t(\tau_{j1}, \tau_{j2}) y_{t-1} + \widehat{\delta_{j2}} D_t(\tau_{j2}, \tau_{j3}) y_{t-1} + \hat{v}_{4t}$$

$$(\widehat{\tau_{j1}}, \widehat{\tau_{j2}}, \widehat{\tau_{j3}}) = argmin\ SSR_{j3}(\tau_{j1}, \tau_{j2}, \tau_{j3})$$

In particular, the selection of the correct model is based on the modelling performance judged with BIC statistics.

This study constructs two bubble indicators as follows, using bubble date estimation from both PSY and HLW:

$$bubble_t^{PSY} = \begin{cases} 1 & if\ PSY\ indicates\ a\ bubble\ at\ time\ t \\ 0 & otherwise \end{cases}$$

$$bubble_t^{HLW} = \begin{cases} 1 & if\ HLW\ indicates\ a\ bubble\ at\ time\ t \\ 0 & otherwise \end{cases}$$

Figure 3.1 Illustration of housing bubble indicators

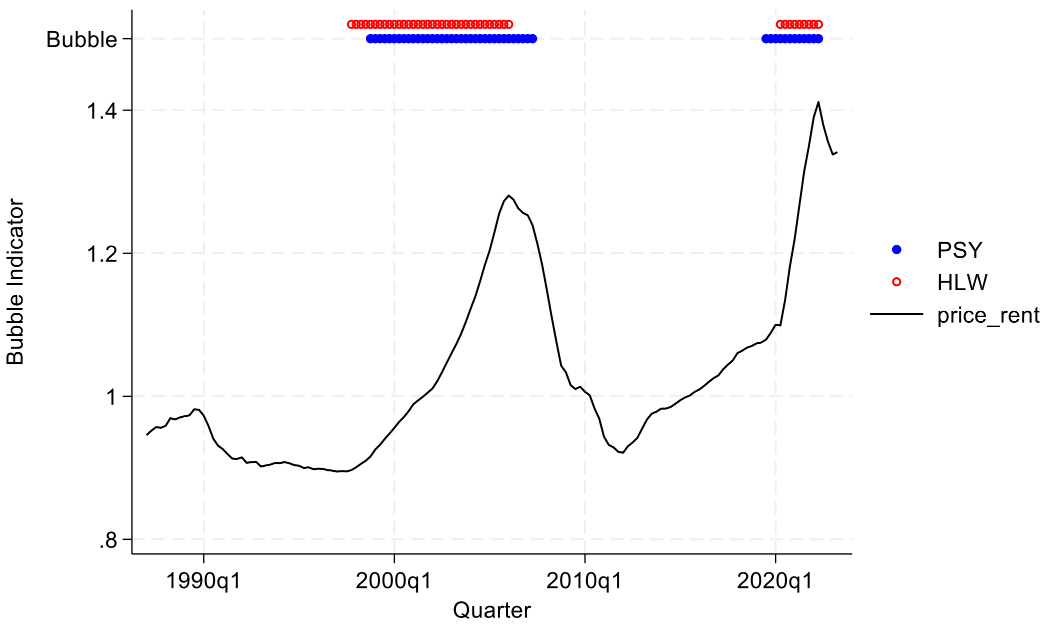


Figure 3.1 displays the house-price rent ratio index (1972 Q4 =1) in the US along with the two bubble indicators. Specially, the detection of housing bubbles is based on the price to rent ratio rather than the raw price series. This is because rent is a fundamental determinant of housing valuation, as it reflects the economic return a homeowner could earn from their property as an income-generating asset. In turn, fair price of house is

determined with the discounted value of future rent series. An unreasonably high ratio of housing price to rent indicates the explosive pattern of housing price, which signals the occurrence of a housing market bubble.

The figure indicates that American house prices experience two rounds of peak and trough during the sampling period. Both two bubble indicators are able to capture the explosive nature of the house prices during the period of a fast house price increases. Meanwhile, their identification of the bubble period is quite close, where the HLW indicator leads the detection for the first bubble and the PSY indicator leads the detection for the second bubble respectively. Since the two bubble indicators are closely correlated and have a mostly overlapping period of bubble or non-bubble time, for the rest of the study, attention is focused with the forecast of PSY bubble indicator.

Section 3.2 Regression model

For a baseline analysis, the study first estimates a linear probability model (OLS). Since the outcome variable is a binary one, the estimation captures the conditional probability for bubble occurrence in the housing market (LPM: linear probability model). The regression function is denoted as:

$$\begin{aligned} bubble_t = \beta_0 &+ \beta_1 PHU_{t-1} + \beta_2 VIP_{t-1} + \beta_3 PIGrowth_{t-1} + \beta_4 PRGrowth_{t-1} \\ &+ \beta_5 Growth_{t-1} + \beta_6 Unemployment_{t-1} + \beta_7 Inflation_{t-1} \\ &+ \beta_8 ShortRate_{t-1} + \beta_9 LongRate_{t-1} + \beta_{10} M1_{t-1} + \beta_{11} Stock_{t-1} \\ &+ \epsilon_t \end{aligned}$$

where variables are defined in section 4.2.

That is, the one period lagged values for the independent variables are applied for the forecast of a housing bubble.

However, LPM has four major drawbacks, which render the model inappropriate for the forecast of a housing bubble (Stock and Watson, 2020). Firstly, since the regression model aims to estimate the conditional probability for a housing bubble, the predicted value should lie within the range from zero to one. However, the fitted value with a linear regression model is not bounded to this feasible range. Secondly, when a linear regression model is adopted, the marginal effect of each independent variable on the occurrence of a housing bubble is assumed to be constant. However, in the case of a binary response variable, the assumption of a constant marginal effect is not likely to be valid. Thirdly, for a binary response variable, the variance is related to the mean value of itself. It implies the correlation between variance of error term and values of the independent variable, which causes the issue of heteroscedasticity. Lastly, while the assumption of a normally distributed error term is assumed for a linear regression to have a valid student-t distribution of test statistics, the assumption of normality cannot hold for the case of a linear response variable.

In order to deal with the limitations of the LPM estimation discussed above, the study also adopts the method of MLE (maximum likelihood estimation) and conducts Probit & Logit regressions instead. The alternative model specification is denoted as follow:

$$
\begin{aligned}
P(bubble_t = 1) \\
&= F(\beta_0 + \beta_1 PHU_{t-1} + \beta_2 VIP_{t-1} + \beta_3 PIGrowth_{t-1} \\
&+ \beta_4 PRGrowth_{t-1} + \beta_5 Growth_{t-1} + \beta_6 Unemployment_{t-1} \\
&+ \beta_7 Inflation_{t-1} + \beta_8 ShortRate_{t-1} + \beta_9 LongRate_{t-1} \\
&+ \beta_{10} M1_{t-1} + \beta_{11} Stock_{t-1})
\end{aligned}
$$

Where the function $F()$ on the right-hand side is the cumulative distribution function for a standard normal distribution (logistic distribution) for the Probit estimation (Logit estimation). With the adoption of a cumulative distribution function, the two alternative estimation methods ensure that the predicted value is restricted to the range of zero to one. Meanwhile, they cater for the problem associated with the error term with MLE for coefficient calculation. Moreover, the marginal effect of variables now becomes:

$$
\begin{aligned}
\frac{\partial P(bubble_t = 1)}{\partial x} \\
&= \beta_x f(\beta_0 + \beta_1 PHU_{t-1} + \beta_2 VIP_{t-1} + \beta_3 PIGrowth_{t-1} \\
&+ \beta_4 PRGrowth_{t-1} + \beta_5 Growth_{t-1} + \beta_6 Unemployment_{t-1} \\
&+ \beta_7 Inflation_{t-1} + \beta_8 ShortRate_{t-1} + \beta_9 LongRate_{t-1} \\
&+ \beta_{10} M1_{t-1} + \beta_{11} Stock_{t-1})
\end{aligned}
$$ [1]

Where the probability density function on the right-hand side implies that the marginal effect is non-constant but dependent on the values of independent variables.

[1] The study reports the marginal effect calculated at the sample mean values of independent variables in the Probit and Logit estimations.

Section 3.3 Forecast performance evaluation

For the evaluation of forecast performance, the study considers two sets of diagnostic tests.

For one thing, the study utilizes the following three metrics for the assessment of absolute model performance of each individual regression (Montgomery et al., 2021; Osius and Rojek, 1992).

- The rate of a type 1 error (failure for not detecting a housing bubble) and type 2 error (incorrect alarm for a housing bubble).
- The deviance test, which compares the log likelihoods between the regression model and the saturated model.
- The Pearson chi squared test, which evaluates the goodness of fit with the comparison of observed and expected response variables from the estimation.

This study will conduct the evaluation of modelling performance both for in-sample prediction and out-of-sample forecast. In particular, for out-of-sample performance evaluation, the study considers the random walk process as the benchmark model to make comparisons. It makes the forecast of housing market bubbles with the one term lagged value as:

$$bubble_t = bubble_{t-1} + e_t$$

As for the forecast construction, the study first considers a static model established with the first 75% of observations (from 1987 Q1 to 2014 Q2). Then, to account for the potential structural change and non-stationarity of the housing bubble in the time series, a rolling model and a recursive model are utilized respectively.

For the rolling model, the coefficients to be used for forecasting are obtained from the rolling window regression with data from $t - k$ to $t - 1$ periods (where we allow for the choice of rolling window length to take varying values for comparison). For the recursive model, the coefficients to be used for forecasting are obtained from the recursive regression with data up to $t - 1$ period. In this way, new information can be updated for the adaptation of the estimation model each time.

$$
\begin{aligned}
\widehat{bubble_t} = I\{G(&\widehat{\beta_0} + \widehat{\beta_1}PHU_{t-1} + \widehat{\beta_2}VIP_{t-1} + \widehat{\beta_3}PIGrowth_{t-1} + \widehat{\beta_4}PRGrowth_{t-1} \\
&+ \widehat{\beta_5}Growth_{t-1} + \widehat{\beta_6}Unemployment_{t-1} + \widehat{\beta_7}Inflation_{t-1} \\
&+ \widehat{\beta_8}ShortRate_{t-1} + \widehat{\beta_9}LongRate_{t-1} + \widehat{\beta_{10}}M1_{t-1} + \widehat{\beta_{11}}Stock_{t-1}) \\
&> 0.5\}
\end{aligned}
$$

where $G()$ refers to the cumulative probability distribution function for Probit (standard normal distribution) and Logit (logistic distribution) respectively.

Section 3.4 Extended analysis

Apart from the baseline estimation model discussed so far, the study further considers two extended empirical analysis.

For one thing, in the baseline estimation, the study only considers a one term lagged value in the forecast of a housing bubble. However, it is possible that the housing market has a lagged response to certain independent variables greater than one period. In this case, we can anticipate the predictive power of higher order lagged terms of the regressors for a housing bubble. In particular, since the study is based on data of a quarterly frequency, there is a possibility for seasonal pattern that is reflected only with higher order lagged terms of the independent variables. In order to capture such a possibility, the study conducts alternative estimations with up to fourth lagged terms of the independent variables[2].

For the other, in the baseline estimation, the response of housing market condition to the set of independent variables are assumed to be constant over time. In other words, the baseline regression does not account for the possibility of a structural break. However, events such as the financial crisis in 2008 call attention to the pricing of houses and the effect of housing market bubbles to national economy. Such a change of social awareness for the issue of housing bubbles could alter the way in which the housing market react to macroeconomic variables. In order to reflect such a possibility, the study conducts a subsample estimation with the data set separated by the financial crisis in 2008.

---

[2] The study only considers models with the same order of lagged terms for all independent variables. The selection of optimal lagged length is based on the information criteria of estimation (AIC and BIC statistics).

Section 3.5 Hypothesis test and expected findings

The study pays attention to the forecast of housing bubbles with four sets of variables introduced in section 4.2. The hypothesis test for the study is based on the student-t statistics of individual estimation coefficient from the estimation. Specially, according to the discussion of previous studies for housing price forecast, the study anticipates the following relationships between a housing bubble and the independent variables:

- The prosperity of housing market is associated with a higher risk of a housing bubble. That is, PHS and VIP are positively related to the chance of a housing bubble.
- The prosperity of the national economy is associated with a lower risk of a housing bubble. That is, a higher growth rate of the national economy reduces the chance of a housing bubble, while the issue of stagnation (a high unemployment rate with a high inflation rate) increases the chance of a housing bubble.
- The expansionary monetary policy increases the risk of a housing bubble. That is, when interest rate is at a low level and money supply is at a high level, the chance of a housing bubble is high.
- The prosperity of the stock market reduces the risk of a housing bubble. This is because the stock market and the real estate market are competitive for market investment.

Section 4 Data set and summary analysis

The fourth section discusses the preparation of data and performs the exploratory analysis of the data set. The content is split into three sections with the following arrangement. Section 4.1 discusses the collection and preparation of the data set, which includes the detailed steps for cleaning up and combining the raw data. Section 4.2 discusses the choice of variables and the process of variable formation. Section 4.3 presents the summary statistics of the data set.

Section 4.1 Collection and preparation of data set

The study combines data from four sources to construct the time series data set for the forecast of housing bubbles in the US. To start with, for the set of variables indicating the housing market conditions, data is collected from the US census bureau (Gupta, 2013). Next, for the set of variables indicating the macroeconomic conditions, data is collected from IMF (international monetary fund) or OECD (organization for economic cooperation and development) respectively (Jerven, 2016; Spencer and Liu, 2010). Thirdly, for the set of variables indicating the condition of financial market, data is collected from WRDS (Warton research data service) platform (Wachowicz, 2010). Lastly, for the set of variables indicating the monetary policy, data is collected from FRED (Federal reserve economic data) platform (McCracken and Ng, 2016).

The raw data series are collected in mixed frequency from weekly to quarterly, with which the study constructs a final data set in quarterly frequency. Specially, when merging data from a higher frequency to a lower frequency, the study considers two approaches (Hamilton, 2020). For one thing, the average value of the high-frequency variables is applied, which reflects the arithmetic mean value of the variable during a quarter horizon. For the other, the end of period value for the high-frequency variables is applied, which reflects the latest information from that variable at the end of a quarter. These two approaches are found to yield comparable empirical findings. The discussion of estimation results in the paper is based on the first approach as the average value over a time horizon is less affected by irregular short-term spikes or drops, which reflects the stable trend of the series.

For the sampling period of study, the maximum overlapping period for these variables is adopted to construct the most comprehensive data for analysis. The corresponding sample for analysis spans from 1987 Q1 to 2023 Q2, with a total of 146 quarterly observations.

Section 4.2 Selection and formation of variables

In accordance with the discussion of major predictors for housing price in section 2, the study considers four sets of independent variables for the forecast of housing bubbles. The detailed list of variables is attached as follow.

Firstly, in order to reflect the condition of housing market from the supply side, the study considers four variables:

- Private housing starts (PHS): It represents the number of new private residential construction projects started.
- Construction value put in place (VIP): It captures the dollar value of completed construction projects, including both residential and non-residential properties.
- Growth rate of price to rent ratio (PR_growth): It reflects the increase of housing price to rent ratio on an annual basis.
- Growth rate of price to income ratio (PI_growth): It reflects the increase of housing price to income level on an annual basis.

Secondly, in order to reflect the condition of the macroeconomy, the study considers three variables:

- Economic growth rate (Growth): It is constructed as the annual log growth rate of real GDP.
- Unemployment rate (Unemployment): It is the fraction of unemployed but active job searchers over the whole size of working aged population.
- Inflation rate (Inflation): It is constructed as the annual log growth rate of CPI (consumer price index).

Thirdly, in order to reflect the implementation of monetary policy, the study considers three variables:

- Short term interest rate (ShortRate): It is the three-month treasury bill rate, which captures the liquidity in the monetary market and the cost for borrowing at periods less than a year.
- Long term interest rate (LongRate): It is the ten-year treasury bill rate, which captures the long-term guidance of monetary policy by the central bank with the interest rate for financial assets with a maturity longer than one year.

- Narrow money supply (M1): It is the total of all physical currency in circulation and demand deposits in commercial banks, which reflects the supply of market liquidity by the central bank.

Lastly, in order to reflect the condition of overall financial market and the alternative investment opportunities, the study considers the stock market indicator:

- Stock return (Stock): It is the quarterly investment return of S&P 500 index (Robertson, 2023).

Specially, for the formation of variables, we adopt the concept of annual log growth rate for two reasons. On the one hand, compared with the variable in raw terms, the application of growth rate caters for the potential of non-stationarity in the time series variable. It helps to avoid the problem for a spurious regression and to ensure that the regression model carries a causal interpretation (Granger and Newbold, 1974). On the other hand, for the calculation of growth rate, instead of finding the quarter-by-quarter change of the macroeconomics variable, the annual growth rate considers the year by year change rate that compares the value to that of the same quarter in the previous year. It caters for the possible seasonal pattern in the time series variable over a year time and hence provides a more credible measurements for the growth rate of variables like price level and output level. The associated formula is:

$$growth\ rate_t = 100 * (\ln(y_t) - \ln(y_{t-4}))$$

Section 4.3 Exploratory analysis

To understand the general pattern of the data set, table 4.1 reports the summary statistics of variables for the full sampling period as well as for the two subsampling periods separated by the presence of a housing bubble or not.

The summary statistics indicates a high quality of the data set. This is because the data set does not contain any missing value and the distribution of variables shows no obvious outlying values[3].

For the full sampling period, table 4.1 suggests that the average level of housing market

[3] The rule of 1.5IQR (interquartile range) is applied for the identification of outlying values. That is, when an observation is 1.5IQR below the lower quartile or 1.5IQR above the upper quartile, the value is detected as an outlier. By comparing the minimum and maximum values of each variable with the interquartile range, we confirm that the data set does not suffer from the issue of outlying values.

features 333.9 PHS and 229.0 VIP, reflecting a relatively active housing market over the entire period. The macroeconomic conditions during this period feature an average annual economic growth rate of 4.79%, an unemployment rate of 5.79%, and an inflation rate of 2.73%, suggesting a generally stable macroeconomic environment.

Table 4.1 Summary statistics by sampling periods

Full sample	Mean	SD	Min	Median	Max	Skewness	Kurtosis
phs	333.86	100.569	114.4	338	575.3	-0.118	2.637
vip	229.033	96.763	89.446	218.213	508.497	0.714	3.019
pr_growth	1.044	4.557	-12.734	1.421	14.717	-0.087	4.664
pi_growth	0.363	5.177	-13.818	0.200	24.863	0.645	6.788
growth	4.792	2.573	-7.056	4.877	15.559	-0.637	8.58
unemployment	5.788	1.686	3.34	5.418	12.887	1.161	4.583
inflation	2.732	1.583	-1.637	2.611	8.283	0.805	4.888
shortrate	3.341	2.658	0.1	3.137	9.6	0.372	1.992
longrate	4.606	2.267	0.65	4.292	9.207	0.332	2.074
m1	105.81	165.69	24.212	45.299	682.861	2.798	9.273
stock	0.768	1.657	-6.367	0.939	5.157	-1.35	7.462
Non-bubble	Mean	SD	Min	Median	Max	Skewness	Kurtosis
phs	292.196	86.26	114.4	297.3	480.2	-0.154	2.439
vip	207.349	94.449	89.446	183.624	508.497	0.959	3.601
pr_growth	-0.762	3.651	-12.734	0.084	4.849	-1.315	4.493
pi_growth	-1.180	4.331	-13.818	-1.017	9.265	-0.499	4.171
growth	4.551	2.131	-3.771	4.753	9.021	-1.391	6.835
unemployment	6.098	1.668	3.5	5.753	10.4	0.617	2.638
inflation	2.669	1.601	-1.637	2.607	8.001	0.414	4.051
shortrate	3.525	2.86	0.113	3.23	9.6	0.307	1.862
longrate	4.907	2.468	1.563	3.887	9.207	0.274	1.578
m1	82.103	120.943	24.212	45.748	676.817	4.174	19.887
stock	0.779	1.646	-6.336	1.047	3.767	-1.444	7.005
Bubble	Mean	SD	Min	Median	Max	Skewness	Kurtosis
phs	421.621	66.514	298.8	419.1	575.3	0.39	2.529
vip	274.706	85.824	153.177	252.003	490.1	0.756	2.555
pr_growth	4.848	3.907	-2.830	3.445	14.717	0.774	3.603
pi_growth	3.614	5.349	-6.630	2.799	24.863	1.624	7.404
growth	5.278	3.263	-7.056	5.764	15.559	-0.427	7.597
unemployment	5.137	1.548	3.34	4.813	12.887	2.947	14.921
inflation	2.86	1.555	0.364	2.616	8.283	1.71	6.439
shortrate	2.953	2.15	0.1	2.103	6.627	0.253	1.604
longrate	3.972	1.62	0.65	4.303	6.48	-0.738	2.398
m1	155.745	227.196	36.14	44.764	682.861	1.613	3.746
stock	0.745	1.7	-6.367	0.733	5.157	-1.164	8.325

Then, for the comparison of the summary statistics during the period with and without a housing bubble, table 4.1 provides a preliminary result regarding the relationship between housing bubbles and the economic condition. To be specific, the summary statistics suggests that the period with a housing bubble features a higher intensity of house construction (a higher PHS and VIP). In addition, a housing bubble is more likely to occur during economic prosperity characterized by a higher GDP growth rate, a lower unemployment rate, and a higher inflation rate. Moreover, the occurrence of a housing bubble is related to a period with loose monetary policies (a lower interest rate and a larger narrow money supply). Besides, compared with the period having no housing bubble, the period of a housing bubble has a lower stock market return on average. Most importantly, the annual growth rate for price to rent ratio and price to income ratio are found to be two critical indicators for a housing bubble. To be specific, for the period without a housing bubble, there is a general negative growth rate for the two variables at -0.76% and -1.18% respectively. However, for the period with a housing bubble, these two variables are shown to growth at a rate of 4.85% and 3.61% separately.

Table 4.2 Correlation coefficients

		1	2	3	4	5	6	7	8	9	10	11	12
1	bubble	1	0.656	0.33	0.658	0.497	0.147	-0.306	0.038	-0.053	-0.127	0.095	-0.046
2	phs	0.629	1	0.137	0.592	0.485	0.515	-0.5	0.249	0.329	0.269	-0.196	0.204
3	vip	0.311	0.172	1	0.393	0.384	-0.154	-0.424	-0.207	-0.554	-0.792	0.877	-0.138
4	pr_growth	0.591	0.614	0.353	1	0.857	0.283	-0.349	-0.073	-0.248	-0.29	0.262	0.168
5	pi_growth	0.452	0.502	0.37	0.869	1	0.192	-0.382	-0.107	-0.272	-0.302	0.299	0.086
6	growth	0.133	0.472	0.021	0.504	0.477	1	-0.393	0.584	0.462	0.438	-0.301	0.256
7	unemployment	-0.265	-0.588	-0.306	-0.345	-0.408	-0.523	1	-0.217	-0.352	-0.002	-0.15	0.033
8	inflation	0.057	0.247	0.09	0.153	0.221	0.625	-0.335	1	0.528	0.488	-0.323	-0.12
9	shortrate	-0.078	0.309	-0.499	-0.184	-0.2	0.343	-0.38	0.443	1	0.87	-0.765	-0.022
10	longrate	-0.169	0.231	-0.745	-0.222	-0.222	0.309	-0.123	0.365	0.881	1	-0.92	0.048
11	m1	0.201	0.1	0.748	0.426	0.411	0.217	-0.133	0.383	-0.304	-0.461	1	-0.088
12	stock	-0.011	0.273	-0.092	0.319	0.22	0.393	-0.147	-0.06	-0.018	0.035	-0.022	1

Table 4.2 provides the correlation coefficients between variables. Specially, for the top right part, table 4.2 illustrates the Spearman rank correlation coefficients. Meanwhile, for the bottom left part, table 4.2 illustrates the Pearson linear correlation coefficients. There are two main pieces of conclusions to be drawn from table 4.2.

For one thing, in terms of the relationship between each independent variable with the presence of a housing bubble, table 4.2 offers a similar conclusion as discussed above with the summary statistic. That is, a housing bubble is more likely to occur in a time period with a higher PHS, VIP, price to rent growth rate, price to income growth rate, economic growth rate, inflation rate, and money supply. On the contrary, a lower unemployment rate, a lower interest rate, and the unsatisfying performance of the stock market are associated with a housing bubble. In particular, from the aspect of economic magnitude, PHS and the growth rate of price to rent are found to be the two most important indicators for a housing bubble.

For the other, the correlation coefficients in table 4.2 suggests that independent variables are not of a high level of linear correlation. It guarantees that the Gauss-Markov assumption for no perfect correlation is valid and ensures that the regression model will not suffer from the issue of high collinearity.

Section 5 Empirical findings

The fifth section demonstrates and interprets the research findings. The content is split into three sections with the following arrangement. Section 5.1 illustrates the estimation from baseline model specification. Section 5.2 conducts model performance evaluation. Section 5.3 depicts the findings from the extension analysis.

Section 5.1 Baseline estimation

Table 5.1 Baseline estimation

	(1)	(2)	(3)	(4)	(5)
	LPM	Probit		Logit	
		Estimation	ME	Estimation	ME
phs	0.0034***	0.028***	0.00014	0.048***	0.00018
	(0.00053)	(0.0080)	(0.00033)	(0.015)	(0.00029)
vip	-0.0019**	-0.0022	-0.000011	-0.00025	-0.00000094
	(0.00092)	(0.012)	(0.000075)	(0.023)	(0.000087)
pr_growth	0.053***	0.98***	0.0051	1.91**	0.0071
	(0.017)	(0.34)	(0.011)	(0.74)	(0.010)
pi_growth	0.021*	0.067	0.00035	0.13	0.00048
	(0.012)	(0.14)	(0.00087)	(0.24)	(0.00097)
growth	-0.033	-0.48	-0.0025	-0.89	-0.0033
	(0.020)	(0.33)	(0.0056)	(0.62)	(0.0051)
unemployment	0.012	0.15	0.00079	0.23	0.00084
	(0.029)	(0.45)	(0.0030)	(0.84)	(0.0035)
inflation	0.036	-0.78	-0.0040	-1.58	-0.0059
	(0.031)	(0.60)	(0.0087)	(1.13)	(0.0085)
shortrate	0.049	1.79***	0.0092	3.23***	0.012
	(0.030)	(0.61)	(0.020)	(1.11)	(0.018)
longrate	-0.17***	-1.86**	-0.0096	-3.11**	-0.012
	(0.048)	(0.77)	(0.022)	(1.39)	(0.020)
m1	-0.00017	-0.0068	-0.000035	-0.013	-0.000049
	(0.00034)	(0.0052)	(0.000079)	(0.0099)	(0.000076)
stock	-0.061***	-0.86***	-0.0045	-1.52**	-0.0057
	(0.020)	(0.33)	(0.010)	(0.61)	(0.0090)
_cons	0.25	-5.08		-9.89	
	(0.38)	(5.97)		(11.2)	
N	141	141	141	141	141
R^2	0.617				
adj. R^2	0.585				
pseudo R^2		0.818		0.818	

Note: Table 5.1 reports the baseline estimation results. The first column corresponds to the LPM regression; the second and the third columns correspond to the raw estimation and marginal effect from the Probit model; the fourth and the fifth columns correspond to the raw estimation and marginal effect from the Logit model. Robust standard errors are calculated with adjustment to possible heteroscedasticity and serial correlation. The notations for the statistical significance follow the rule as: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5.1[4] illustrates the regression from the baseline model specification. It reports the estimation obtained from LPM, Probit, and Logit models separately. Specially, for the Probit and Logit estimation parts, the marginal effect calculated at the sample mean values of independent variables are derived. The estimation obtained from three models are in general comparable. In detail, except for the statistical significance of PI_growth and short run interest rate, for the other variables, both the sign of the estimation and the statistical significance are the same across all three regression models. For simplicity, the following discussion is based on LPM, whose coefficient estimation can be directly interpreted as the marginal effect of variables for the chance of a housing bubble. The associated marginal effect obtained from Probit and Logit estimation is reported in the parentheses.

To begin with, the housing market condition is a strong indicator for the occurrence of a housing bubble. Ceteris paribus, a one thousand housing unit increase of PHS leads to a higher risk of a housing bubble by 0.34% (Probit 0.014% and Logit 0.018%) while a one million of dollars increase of VIP leads to a lower risk of a housing bubble by 0.19% (Probit 0.0011% and Logit 0.0001%). The effect of these two variables is documented to be statistically significant at the level of 1%. In addition, since the overall risk of a housing bubble in the data set is just 32.19%, the impact of these two variables is found to be of a considerable economic magnitude. In detail, for a one standard deviation increase of PHS (100.57 units of change), the chance of a housing bubble is found to be higher by 34.19%. Meanwhile, for a one standard deviation increase of VIP (96.76 units of change), the chance of a housing bubble is found to be lower by 18.38%. The estimation result is no exactly aligned with the anticipation. While a higher PHS indeed forecast a higher risk of a housing bubble, the opposite

[4] Table 5.1 considers PHS, VIP, and M1 in level form. The choice is made majorly due to the stationarity check of these variables. In detail, growth rate is adopted for variables that are non-stationary in their level format as it parallels the first difference of log transformed series that help to solve the issue of unit root. Meanwhile, for stationary variables (for this case, the three variables are found to be trend stationary), level terms are adopted in the regression analysis. In addition, from the perspective of model fitness and the forecasting performance of the regression, the function with these three variables included in level for yields a better overall model fitness in contrast with the function with these three variables included in growth rate form.

However, thanks to the examiners, an alternative specification based on the growth rate of these variables are considered. The corresponding estimation results are reported in appendix 2 for reference. The major research findings remain robust when the alternative regression model is adopted.

holds for VIP. A potential explanation lies in the fact that when the house prices is overstated, new construction is of a high intensity. On the other hand, a higher value of completed construction indicates an abundant market supply of houses, which reduces the pressure of excessive demand in the market and hence has the effect to cool down the housing market.

Apart from the speed of construction, the change of price in the housing market is indicative to the presence of a housing bubble as well. In detail, for a one percent higher growth rate of price to rent ratio, the risk of a housing bubble is expected to increase by 5.3% (Probit 0.21% and Logit 0.71%) ceteris paribus. Meanwhile, for a one percent faster rise of price to income ratio, the risk of a housing bubble increases by 2.1% (Probit 0.035% and Logit 0.048%) on average. Both of the two variables are demonstrated to be statistically significant predictors for a housing bubble at the level of 5%. Moreover, from the economic perspective, for a one standard deviation change of the two variables (which is 4.56 for PR_growth and 5.18 for PI_growth), the increased risk of a housing bubble is 24.15% and 10.87% separately. It reinforces the crucial role of housing market conditions for the prediction of a housing bubble.

Next, for all of the three macroeconomic variables, we do not observe a statistically significant effect for any of them on the occurrence of a housing bubble. In detail, table 5.1 indicates that for a one percent rise of economic growth rate, unemployment rate, and inflation rate, holding all other variables constant, there is an expected change of housing bubble probability by -3.3% (Probit -0.25% and Logit -0.33%), 1.2% (Probit 0.079% and Logit 0.084%), and 3.6% (Probit -0.4% and Logit -0.59%)respectively. However, at the level of 10%, none of these three variables is shown to have a statistically distinguishable influence on housing bubble occurrence.

Thirdly, however, monetary policy is shown to have a considerable influence on the occurrence of a housing bubble. In detail, for a 1% rise of short-term interest rate, ceteris paribus, the opportunity of a housing market bubble is expected to be higher by 4.9% (Probit 0.92% and Logit 1.2%). Meanwhile, for a 1% increase of long-term interest rate, the expected probability of a housing bubble is lower by 17% (Probit 0.96% and Logit 1.2%) holding all else equal. Specially, the marginal effect of the long-term interest rate is shown to be statistically significant by 5% level. Moreover, in terms of economic significance, the change of short-term interest rate by one standard deviation is related to a higher rate of housing bubbles by 13.02%. Meanwhile, the marginal effect from a one standard deviation changes of the long-term interest rate is expected to cause a lower rate of housing bubbles by 38.54%. However, the supply of money is not detected to have a considerable effect for the occurrence of a housing bubble either from the statistical perspective or from the economic perspective.

Lastly, the negative effect of the stock market performance on the probability of a housing market bubble is confirmed. All other variables held constant; a 1% rise of stock market return is associated with a lower rate of housing bubble by 6.1% (Probit 0.45% and Logit 0.57%). The result is statistically significant by 1% level. In addition, the economic magnitude of the variable is comparable with that from the short-term interest rate.

Putting together, the baseline estimation is able to confirm the effect of housing market condition, monetary policy, and stock market performance on the occurrence of a housing bubble. Nonetheless, the nexus between key macroeconomic variables and the occurrence of a housing bubble is not detected. In addition, contrary to the expected findings illustrated in Section 3.5, the analysis of American data reveals that VIP has a negative effect on housing bubble occurrence, while short term interest rate is not shown to mitigate the risk of the housing market. Overall, the model performance is satisfying. This is because that the R squared from the estimation suggests that nearly two thirds of the variation in the occurrence of a housing bubble.

Section 5.2 In-sample estimation evaluation

Except for the assessment of model performance based on the R squared reported from the regression, Section 5.2 further evaluates the goodness of fit from the three estimation approaches. Table 5.2 reports the forecast performance of the three models.

In terms of the overall rate of correct classification, LPM performs a little bit worse than the other two methods. In fact, Logit model gives the highest rate of correct specification for 95.04%, which is followed by 94.33% from the Probit model and 92.91% from the LPM estimation.

Table 5.2 Measurements of model performance

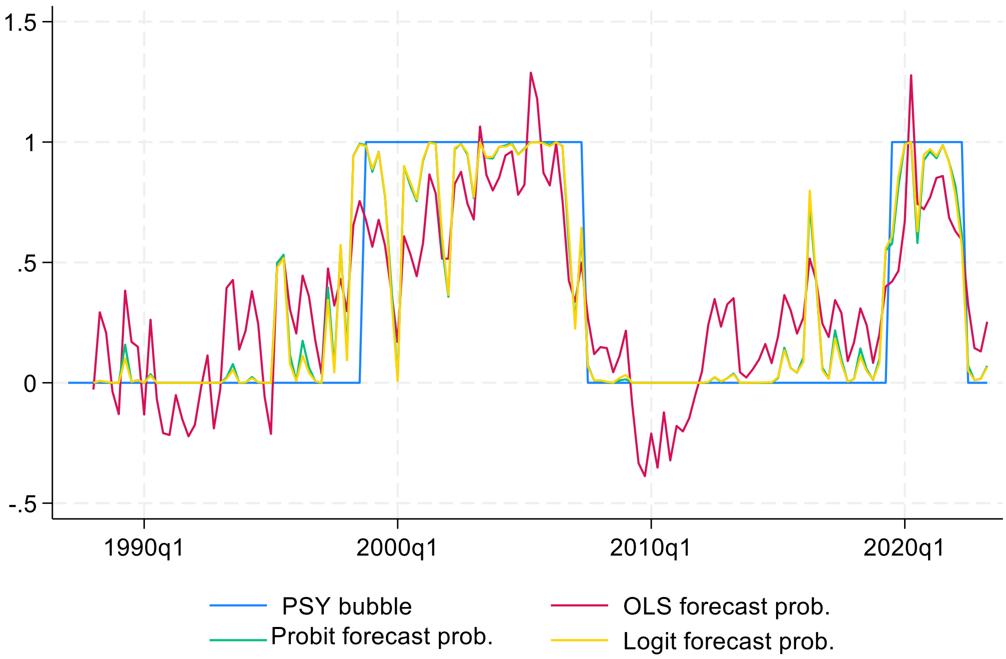
Measurements	LPM	Probit	Logit
Correct bubble prediction	39	43	44
Correct non-bubble prediction	92	90	90
False bubble prediction	2	4	4
False non-bubble prediction	8	4	3
Sensitivity	82.98%	91.49%	93.62%
Specificity	97.87%	95.74%	95.74%
Positive predictive value	95.12%	91.49%	91.67%
Negative predictive value	92.00%	95.74%	96.77%
False bubble rate	2.13%	4.26%	4.26%
False non-bubble rate	17.02%	8.51%	6.38%
False bubble rate among classified bubbles	4.88%	8.51%	8.33%
False non-bubble rate among classified non-bubbles	8.00%	4.26%	3.23%
Correct classified rate	92.91%	94.33%	95.04%

However, that is not to say that the LPM forecast is entirely worse than the other two methods for the forecast of a housing bubble. In fact, for those aiming to reduce the rate of a falsely predicted bubble, LPM yields a better performance (2.13% against 4.26%). Nonetheless, for those aiming to reduce the rate of falsely predicted non-bubbles, Probit and Logit perform much better than LPM (8.51% and 6.38% against 17.02%).

In addition, when evaluating the goodness of fit for the two MLE approaches, we notice that the deviance test statistics for Logit estimation is 32.61 and that for Probit estimation is 32.65. Both of them are associated with a p value larger than 5%. It provides evidence that for the in-sample performance, both Probit and Logit methods are able to cover sufficient information for the series of housing bubble indicator. Meanwhile, the Pearson chi squared test statistics for the Probit model is 48.85 and that for the Logit estimation is 61.62. At the p value of 5%, both of them pass the Pearson chi squared test and can be taken as a satisfactory model to forecast housing bubbles.

Moreover, in terms of the predicted probability of a housing bubble, the comparison of MSE (mean squared error) from Probit (MSE=0.0462) and Logit (MSE=0.0458) shows no detectable difference at 10% level (p value=0.665). In terms of the predicted binary variable of a housing bubble, Probit and Logit yields exactly the same forecast and hence have the same goodness of fit.

Figure 5.1 Comparison of three model forecast performance



The graphical illustration of the three-model forecast performance is attached in figure 5.1. It shows that Probit and Logit models yield quite close forecasted probability of a housing bubble, while all three models are able to capture the general pattern for the increase and decrease for the chance of a housing bubble. Specially, an advantage of the three-bubble forecasting model lies with the fact that they provide a way to track the process of bubble formation. In detail, even before the occurrence of first bubble, the probability forecast obtained with the three methods show a trend of increase. It means that the three models are able to reflect the rising risk for a housing market bubble even before its actual occurrence. Therefore, we can not only apply the three models for the forecast of a bubble itself, but also apply the three models to monitor the risk of a housing market bubble before its true occurrence.

Aside from the forecasting performance evaluated as above, figure 5.1 also indicates the changed behaviour of the growth of a housing market bubble for the two periods. In detail, for the housing market bubble in 2000s, there is a prolonged period with the risk of a housing market bubble increases steadily. Meanwhile, its crash during the financial crisis in 2008 is rather drastic, which causes a remarkable shock and a long period of chaos to the national economy. On the other hand, for the housing market bubble during 2020s, we fail to observe a long period with a rising risk for its occurrence. In detail, although there is a sharp rise of bubble risk during late 2016, the market soon cools down. Nonetheless, in late 2019, the risk for a housing market bubble soon increases. Then, compared with the first housing market bubble, the second one is of a rather shorter length. A potential issue is the confounded effect of Covid-19, which leads to the recession for both demand side and supply (Barlow and Vodenska, 2021; Gamber et al., 2023). In turn, the collapse of the housing market bubble gets accelerated.

Section 5.3 Extended analysis

Established upon the baseline estimation model, Section 5.3 digs into the possible extension of the modelling framework to improve the forecasting performance.

For table 5.3, we report the LPM estimation with up to the fourth order lagged terms[5]. For table 5.4a and table 5.4b, we report the estimation with three methods separately for the sampling period before 2008 financial crisis and after that.

When higher order lagged terms are taken into consideration, table 5.3 indicates that the inclusion of a second order lagged term is sufficient to capture all historical information. This is because the AIC and BIC statistics are both smallest for the model with second order lagged values of the independent variables, which indicates the best balance between model complexity and model performance. The estimation results partially confirm the discussion in table 5.1, which show evidence for the lagged response of housing market to the set of regressors.

To start with, for the housing market activities, table 5.3 confirms the importance of PHS and VIP for the forecast of a housing bubble. Ceteris paribus, a one shot increase of PHS by one unit leads to a higher risk of a housing bubble by 0.29% in the next quarter, while a permanent increase of PHS by one unit is associated with a higher risk of a housing bubble by 0.40%. On the other hand, the short-term influence of a one unit rise of VIP on housing bubble probability is -0.23%. However, the effect reverses direction in the half year horizon to a positive impact by 0.20%. Overall, the permanent effect of a one-unit change in VIP on the probability of a housing bubble is -0.03% holding all else equal. In particular, for both the first and the second term lags, the effect

[5] For the analysis of higher ordered terms, for the complexity of calculating the cumulative marginal effect in the non-linear estimation models, the analysis is concentrated with the LPM regression only.

of PHS and VIP on the chance of a housing bubble is shown to be statistically significant by 10% level.

From the aspect of house prices change, table 5.3 reinforces its strong effect for the prediction of a housing bubble. In detail, for short run impulse response, a one percent higher growth rate for price to rent leads to a higher risk of housing bubble by 3.6% in the next quarter and a higher risk of housing bubble by 4.7% in a half year horizon. In addition, that for the growth rate of price to income ratio is 0.93% and 3.8% respectively. However, statistical significance is only detected for the second lagged term of price to income ratio growth.

Next, the findings for the three macroeconomic variables are different from the baseline estimation in table 5.1. In fact, when higher order lagged terms are considered, for the one quarter horizon, a higher GDP growth rate predicts a lower housing bubble risk while a higher unemployment rate predicts the opposite. The statistical significance is detected at 10% level. Moreover, in terms of the economic magnitude, we should notice that the marginal effect of these three variables is in fact stronger than the effect from housing market activities. Besides, the estimation suggests that while the effect of economic growth on housing market bubbles gets strengthened with time, there is a reversed effect of unemployment rate and inflation rate with time.

Thirdly, from the aspect of monetary policy, the estimation suggests that the positive relationship between short term interest rate and housing bubbles lies mainly with the half year horizon. Meanwhile, the negative impact of the long-term interest rate on housing bubble occurrence lies mainly with the one quarter horizon. In addition, we now detect a statistically significant negative effect of money supply on housing bubble occurrence for the one quarter horizon while the effect reverses in direction for the half year horizon. A potential explanation lies in the fact that quantitative easing, which increases money supply, is often applied during a recession to boost the market recovery when overall housing bubble risk is low.

Lastly, from the aspect of stock market performance, while table 5.3 confirms that a rise of stock price cools down the real estate market, but no statistical significance is detected for the first or second order lagged terms. Nonetheless, in terms of economic magnitude, the estimation is quite comparable with table 5.1. In detail, table 5.3 shows that for a 1% rise of stock index return, ceteris paribus, for the next quarter, the risk of a housing bubble will be lower by 2%. Moreover, the effect continues to accumulate with time. When the stock market return keeps being high, on the half year horizon, the cumulative effect on housing market bubble occurrence is -4.9%. It depicts the strong substitution effect of stock market investment against real estate investment.

As for the overall model performance, table 5.3 indicates the improved model fitness with the inclusion of higher order lagged terms. In detail, the R squared suggests that more than 72.8% of the variation in housing bubble occurrence is captured by the model with two lagged terms for forecast. It suggests a quite satisfying performance of LPM with higher order lagged terms in the forecast of housing bubbles. However, when looking into the in-sample performance of the model with a higher ordered length, the frequency of a falsely predicted non-bubble increases to 6 and the frequency of a falsely predicted bubble increases to 9. It suggests the possible problem of overfitting, since the error of forecast (both type 1 error and type 2 error) gets even larger than the previous model with only the first order lagged terms. The implication of the result is that information about a housing market bubble is short lived. Or put it another way, the housing market response to macroeconomic variables is quite speedy. In this case, the inclusion of more lagged terms into the regression model fails to yield a considerable performance improvement.

Table 5.3 Regression with higher order lagged terms

	(1)		(2)		(3)	
	bubble		bubble		bubble	
L.phs	0.0029***	(0.00076)	0.0019*	(0.00096)	0.0012	(0.0013)
L2.phs	0.0011	(0.00072)	0.0012	(0.0012)	0.00066	(0.0014)
L3.phs			0.0010	(0.00089)	0.00090	(0.0013)
L4.phs					0.0013	(0.0012)
L.vip	-0.0023*	(0.0012)	-0.0018	(0.0017)	0.00059	(0.0036)
L2.vip	0.0020	(0.0014)	0.0014	(0.0021)	-0.0018	(0.0036)
L3.vip			-0.00037	(0.0016)	0.0028	(0.0036)
L4.vip					-0.0038	(0.0037)
L.pr_growth	0.036	(0.030)	0.058	(0.044)	0.081	(0.051)
L2.pr_growth	0.047	(0.035)	-0.015	(0.067)	-0.013	(0.072)
L3.pr_growth			0.071*	(0.039)	0.055	(0.068)
L4.pr_growth					0.047	(0.042)
L.pi_growth	0.0093	(0.012)	0.016	(0.015)	0.032	(0.021)
L2.pi_growth	0.038***	(0.014)	0.032**	(0.015)	0.042**	(0.017)
L3.pi_growth			-0.018	(0.017)	-0.012	(0.018)
L4.pi_growth					-0.025	(0.018)
L.growth	-0.047*	(0.028)	-0.036	(0.033)	-0.046	(0.044)
L2.growth	-0.0041	(0.022)	-0.0077	(0.033)	-0.0047	(0.037)
L3.growth			-0.019	(0.025)	-0.051	(0.035)
L4.growth					0.0079	(0.031)
L.unemployment	0.099**	(0.048)	0.12*	(0.069)	0.051	(0.12)
L2.unemployment	-0.036	(0.040)	0.032	(0.067)	0.037	(0.15)
L3.unemployment			-0.075	(0.058)	0.11	(0.14)
L4.unemployment					-0.13**	(0.063)
L.inflation	0.065	(0.042)	0.058	(0.046)	0.036	(0.050)
L2.inflation	-0.036	(0.042)	-0.031	(0.061)	-0.0013	(0.063)
L3.inflation			0.000094	(0.046)	-0.0073	(0.063)
L4.inflation					0.0019	(0.048)
L.shortrate	-0.10	(0.076)	-0.0067	(0.095)	0.060	(0.11)
L2.shortrate	0.15**	(0.072)	-0.092	(0.15)	-0.17	(0.16)
L3.shortrate			0.18**	(0.088)	0.12	(0.17)
L4.shortrate					0.092	(0.094)
L.longrate	-0.090	(0.087)	-0.19*	(0.098)	-0.19*	(0.10)
L2.longrate	-0.039	(0.082)	0.12	(0.13)	0.060	(0.14)
L3.longrate			-0.098	(0.088)	0.037	(0.14)
L4.longrate					-0.13	(0.093)
L.m1	-0.0038**	(0.0017)	-0.0039	(0.0027)	-0.0021	(0.0036)
L2.m1	0.0036**	(0.0018)	0.00073	(0.0043)	-0.0015	(0.0072)
L3.m1			0.0031	(0.0023)	-0.0028	(0.0077)
L4.m1					0.0071	(0.0045)
L.stock	-0.020	(0.023)	-0.00032	(0.025)	-0.0026	(0.028)
L2.stock	-0.029	(0.021)	-0.0080	(0.028)	0.0016	(0.031)
L3.stock			-0.027	(0.025)	0.0013	(0.031)
L4.stock					-0.043	(0.028)
_cons	-0.72	(0.51)	-0.59	(0.72)	0.035	(0.88)

N	140	139	138
R^2	0.728	0.767	0.800
adj. R^2	0.677	0.694	0.706
AIC	50.8	51.9	53.2
BIC	118.4	151.7	184.9

Note: Table 5.3 reports the extended estimation results with up to fourth order lagged independent variables in the LPM regression. Robust standard errors are calculated with adjustment to possible heteroscedasticity and serial correlation. The notations for the statistical significance follow the rule as: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Apart from the extension with higher ordered lagged terms, we perform the subsample estimation for the period before and after 2008 financial crisis to check for the potential of a structural break. As is discussed with the series of housing market bubble and the predicted risk for the housing market bubble in figure 5.1, the pattern of the housing market bubble is quite distinct for the rounds in 2000s and 2020s separately. The estimation in table 5.4a and table 5.4b indeed demonstrates the changed responsiveness of housing market to predictors in the two sampling periods. In particular, for the subsample analysis, both Probit and Logit methods are able to correctly predict all occurrence of housing market bubbles in the estimation.

Parallel the baseline estimation, the subsample estimation is conducted with LPM, Probit, and Logit respectively. Then, as the estimation results are comparable over the three different estimation methods, the discussion is again based on LPM estimation for simplicity.

Table 5.4a Subsample estimation for pre-2008 period

	(1) LPM	(2) Probit	(3) Logit
phs	-0.0012	0.037	0.21
vip	0.0025^{*}	0.056	-0.38
pr_growth	0.13^{***}	11.9	28.0
pi_growth	-0.022	2.31	-0.039
growth	-0.097^{***}	6.20	3.43
unemployment	-0.15^{**}	6.61	10.1
inflation	0.011	9.73	24.0
shortrate	-0.040	12.1	26.6
longrate	-0.035	-49.9	-83.7
m1	-0.014	-1.33	3.95
stock	-0.031	-6.89	-17.3
_cons	2.56^{***}	160.6	59.7
N	80	80	80
R^2	0.830		
adj. R^2	0.802		
pseudo R^2		1.000	1.000

Note: Table 5.4a reports the estimation results for the period before 2008 financial crisis. The first column corresponds to the LPM regression; the second column corresponds to the raw from the Probit model; the third column corresponds to the raw estimation from the Logit model. The notations for the statistical significance follow the rule as: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

For the period before 2008 financial crisis, we obtain the same result as with the baseline estimation that PR_growth has a positive influence on housing bubbles. However, the influence of VIP reverses in direction and the effect from PHS and PI_growth becomes statistically insignificant.

Then, in contrast with the baseline estimation, the influence of the macroeconomic variables on housing bubbles can be detected in the first subsampling period. In detail, for a 1% rise of unemployment rate, the rate of a housing bubble is expected to be lower by 15% holding other things equal. Meanwhile, the rise of GDP growth rate by 1% is associated with a lower probability of a housing bubble by 9.7% ceteris paribus. The influence of these two variables is of statistical significance as well as a considerable economic magnitude.

Table 5.4b Subsample estimation for post-2008 period

	(1) LPM	(2) Probit	(3) Logit
phs	0.0021	0.34	2.23
vip	-0.00027	-0.25	-1.74
pr_growth	0.0029	0.74	10.1
pi_growth	-0.010	0.92	-1.53
growth	0.018	6.39	28.1
unemployment	0.051	6.23	32.6
inflation	-0.021	-7.65	-2.72
shortrate	0.064	25.2	152.9
longrate	-0.25^{***}	-32.3	-225.8
m1	0.00052	0.013	-0.066
stock	-0.019	-2.31	-11.4
_cons	-0.16	-38.9	-154.5
N	61	61	61
R^2	0.614		
adj. R^2	0.530		
pseudo R^2		1.000	1.000

Note: Table 5.4b reports the estimation results for the period before 2008 financial crisis. The first column corresponds to the LPM regression; the second column corresponds to the raw from the Probit model; the third column corresponds to the raw estimation from the Logit model. The notations for the statistical significance follow the rule as: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Meanwhile, from the aspect of monetary policy, the influence estimated in table 5.4a is shown to differ from the full sample estimation as well. In the subsample estimation, table 5.4a suggests that none of the three monetary policy indicators carry a statistically significant impact on housing bubble occurrence.

Lastly, table 5.4a reinforces the substitution effect between stock investment and real estate investment. It shows that when stock market return is higher by 1%, ceteris paribus, the risk of a housing bubble is lower by 3.1%. However, the effect is not statistically significant by 5% level.

Overall, for the first subsample periods, the predictors are found to carry a considerable power to make a forecast of the housing bubble. Together, 83% of the variation in housing bubble occurrence is explained by the model.

In contrast, for the second subsample period, the overall model fitness reduces to 61.4% for the LPM estimation. In addition, only long-term interest rate remains to have a considerable effect to make forecast, while all the rest of regressors become statistically insignificant. It implies the structural break of the housing market since 2008 financial crisis.

Section 5.4 Out-of-sample estimation evaluation

While the previous assessment of the modelling performance is conducted with the in-sample estimation, in order to reflect the actual forecasting performance of the model in practice, this study conducts an out-of-sample evaluation.

Specifically, the study takes the naïve random walk model as the benchmark model to compare the forecasting performance, in which the one quarter lagged indicator for the presence of a housing market bubble is adopted as the reference model. In this way, we can actually achieve a high precision of forecast as only the starting period of a bubble and the ending period of a bubble are not correctly forecasted by the naïve model.

As for the construction of the out-of-sample forecasts, the study considers three possibilities for model estimation. First, we consider a static regression model that is estimated over a fixed period of data, without updating, and uses the estimated coefficients from this fixed period regression to produce forecasts over the rest of the sample. Second, we consider a rolling regression where the fixed-length window over which the model is estimated is moved forward by one observation each time to produce a series of one step ahead forecasts. Finally, we consider estimating the model recursively such that all historical information is used for model construction, with the estimation window growing by one observation each time, again producing a series of one step ahead forecasts. Particularly, for all three sets of estimation approaches, the study considers the same set of regressors in table 5.1, while compares the performance with an augmented regression model that further incorporates the one period lagged indicator of a housing market bubble.

5.4.1 Static forecast and performance evaluation

The study takes the first 75% of the observations as the initial estimation period for model construction (110 observations from 1987 Q1 to 2014 Q2) and the last 25% of the observations as the out of sample period (36 observations from 2014 Q3 to 2023 Q2). With such a separation, the in sample estimation period and the out of sample period will contain the first and the second detected housing market bubbles respectively.

The estimation result for the in sample period is presented in table 5.5. It again indicates that both Probit model and Logit model correctly predict all housing bubbles, while LPM also has a satisfactory performance that explains more than four fifth of the variation in housing bubble occurrence. However, such a satisfactory in sample performance fails to transfer into a useful out-of-sample forecast.

Table 5.5 Training set estimation with static model

	(1) LPM	(2) Probit	(3) Logit
phs	-0.00075	-4.26	-1.35
vip	0.0021^{**}	4.62	1.37
pr_growth	0.13^{***}	156.2	59.4
pi_growth	-0.041^{**}	-0.041	-0.24
growth	-0.074^{***}	-141.2	-63.1
unemployment	-0.12^{***}	-275.6	-88.2
inflation	0.040	78.0	16.1
shortrate	-0.046	-60.5	2.11
longrate	-0.10^{**}	-130.5	-89.6
m1	-0.022^{***}	-38.2	-14.5
stock	-0.021	6.99	-3.52
_cons	2.78^{***}	5020.8	1977.6
N	106	106	106
R^2	0.811		
adj. R^2	0.789		
pseudo R^2		1.000	1.000

Note: Table 5.5 reports the estimation results for training set. The first column corresponds to the LPM regression; the second column corresponds to the raw from the Probit model; the third column corresponds to the raw estimation from the Logit model. The notations for the statistical significance follow the rule as: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

In fact, as we move on to the assessment of the forecasting performance in the testing set with these three models, none of the three models are able to forecast the second round of housing market bubble. That is, throughout the out of sample period, the regression models yield an out-of-sample forecast of no bubble. It indicates that while these models are able to have a quite remarkable in-sample performance, when a static model is trained for the forecast of housing market bubbles in the future, the model performance is poor. The underlying mechanism is overfitting, which causes a precise in-sample prediction but a poor out-of-sample forecast. In fact, these models will be even worse than the simple random walk model.

Nonetheless, such a result is not surprising for two reasons. On the one hand, as is discussed for the subsample estimation in the previous Section, the determination of a housing market bubble contains a structural break over time. It indicates that if a static model is constructed for its prediction, the out-of-sample performance could be poor after a structural break.

5.4.2 Dynamic forecast and performance evaluation

In order to improve the out-of-sample performance of the model, we consider the alternative modelling approaches for the construction of forecasting model in a dynamic style.

On the one hand, the study sets out to make the forecast with a rolling model. However, the sparse nature of bubbles and the concentration of bubble period and non-bubble period means that a long enough window is required to capture both bubble and non-bubble dynamics. However, when such a long rolling window (for example, an estimation horizon of 20 years with 80 observations) is used, it reduces the benefit from a rolling estimation to capture only the latest information. As a result, the out-of-sample forecast falls into a similar problem as the static approach and fails to forecast most of the bubble periods.

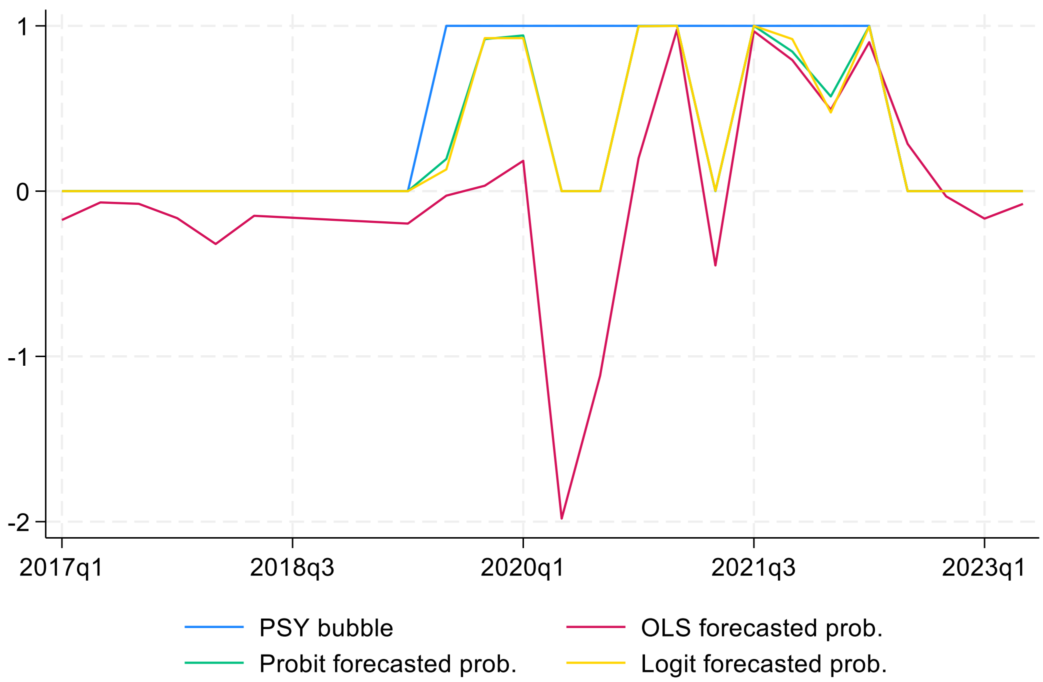
On the other hand, we consider the recursive model that always take the estimation during $[0, t]$ to make a forecast in $t + 1$. The associated out of sample forecasting performance is depicted in table 5.6. The corresponding MSE for the Probit and Logit model are quite close: 0.264 versus 0.233 in terms of forecasted bubble probability and 0.63 versus 0.63 in terms of the forecasted bubble occurrence. The two models are found to have a similar performance in the forecast of bubbles out-of-sample.

Although we are able to achieve a much better forecasting performance this time, it is worth noting that we fail to beat the benchmark model using the random walk property of the series of housing market bubble. While such a finding is quite frustrating, it is able to reveal the truth about housing market bubble forecast: it is hard to be captured by a certain regression model due to the issue of structural break and non-stationarity.

Table 5.6 Out-of-sample performance evaluation with recursive model

Measurements	LPM	Probit	Logit
Correct bubble prediction	3	7	6
Correct non-bubble prediction	10	9	9
False bubble prediction	9	5	6
False non-bubble prediction	13	14	14
Sensitivity	18.75%	33.33%	30.00%
Specificity	52.63%	64.29%	60.00%
Positive predictive value	25.00%	58.33%	50.00%
Negative predictive value	43.48%	39.13%	39.13%
False bubble rate	47.37%	35.71%	40.00%
False non-bubble rate	81.25%	66.67%	70.00%
False bubble rate among classified bubbles	75.00%	41.67%	50.00%
False non-bubble rate among classified non-bubbles	56.52%	60.87%	60.87%
Correct classified rate	37.14%	45.71%	42.86%

Figure 5.2 Out-of-sample forecasting performance with recursive model



In fact, if we recognize the non-stationarity property of the occurrence of housing market bubble and add the one term lagged value of the bubble occurrence into the estimation model, all three forecasting model construction approaches are able to achieve the same performance as the naïve model. That is, only the first period for bubble occurrence and the last period before bubble disappearance are not correctly predicted. Specially, in these cases, the one term lagged bubble indicator is shown to carry the largest predictive power in all three models.

Section 5.5 Summary of empirical analysis

To sum up, the estimation in this chapter indicates that housing market activities, macroeconomic indicators, monetary policy settings, and stock market condition are all influential factors for housing market bubbles before 2008 financial crisis, while only long-term interest rate is shown to be an effective predictor for housing bubbles after 2008 financial crisis. The estimation indicates the presence of a structural break for the models of housing bubble prediction after the financial crisis in 2008. Meanwhile, it suggests a quite satisfying performance of the Logit and Probit models for the forecast of housing bubbles.

As for the linkage of the study with previous ones, since research about the forecast of housing bubbles is rather limited, the research findings can only be compared with past studies concerning the prediction of house prices. Specially, from the viewpoint that an abnormally high rate of house prices to rental price is an indicator of the housing market bubble, we should anticipate the direction of the influence from independent variables in the house prices equation to be the same as that in the housing bubble forecast equation.

Indeed, the discussion above can be reconciled with previous studies for the determination of house prices, especially for the first subsample period. For instance, Ghent and Owyang (2015) indicate the pro-cyclical behaviour of house prices, which is consistent with our research findings that the risk of a housing bubble moves in the opposite direction with the unemployment rate while moves in the same direction with the inflation rate. Besides, Favara and Imbs (2015) argue that the supply of credit is able to boost house prices as it makes mortgage loans more accessible and cheaper. The result is consistent with the study since the drop of interest rate is associated with a higher risk for housing bubbles. Meanwhile, Rünstler and Vlekke (2018) claim the co-movement of housing construction, business cycle, and house prices, which implies that housing market activities (such as PHS) can be predictive for the risk of housing bubbles. Therefore, it is safe to reach the conclusion that the study reinforces some previous finding for the influential factors of house prices while extends the framework for the forecast of a housing bubble.

Section 5.6 Limitations and further extension directions

As for the directions of a further analysis, there are five aspects worthy of a discussion.

Firstly, the search of additional macroeconomic factors for an improved model to forecast housing market bubbles is of interest. In particular, while our study is constrained to domestic indicators of economic and financial conditions, the possible spillover of information and market sentiment across countries should not be neglected. For a further extension of the study, it is of importance to take into consideration the possible influence from the integration of the global financial market and to include macroeconomic indicators (or even housing market bubble indicators) from other major economic players into the regression model.

Secondly, while the current study is restricted to the toolbox of traditional econometrics, the potential of new statistical methods like machine learning and neural network could be employed for the forecast of housing market bubbles in the future. In particular, given the complex nature of the formation and collapse of a housing market bubble, methods such as automated learning and reinforcement learning are promising to enhance the forecast performance.

Thirdly, the study focuses primarily on low frequency macroeconomic indicators for the forecast model construction, while there are several high frequency series that could provide additional information about the housing market condition. In particular, the usage of data of a higher frequency could be combined with the technique of machine learning. For instance, through natural language programming (NLP), we would be able to derive information about the sentiment of market participants from online discussions and social media posts. Considering the important role of investor sentiment for the occurrence of a housing market bubble, the inclusion of such an indicator of market sentiment has a considerable potential to improve the model performance.

Fourthly, while the current study is based primarily on PSY bubble indicator for model construction, it is of importance to check whether the research findings are robust with the adoption of HLW bubble indicator instead. More importantly, since these two methods yield slightly different time points for the starting period and ending period of a housing market bubble, lagged information from alternative bubble indicators could be a strong explanatory variable in a regression model for housing bubble forecast. How to use such information from the lead lag relationship of housing market bubbles identified from these two models should attract more efforts in the future.

Finally, while the current study pays attention to the timing of a market bubble, it does not distinguish between different scenarios at the end of a bubble, i.e. whether the bubble collapses or instead prices revert back to a unit root regime. Forecasting a housing market bubble collapse is clearly of great policy relevance and an area for future research.

Section 6 Discussion and conclusion

Using a time series data set of US house prices and macroeconomic conditions during 1987 Q1 to 2023 Q2, this study constructs a binary housing market bubble indicator and a forecasting regression model to predict house price bubbles.

The study shows that the two methods of housing bubble detection used, PSY and HLW, yield similar timings of house price bubbles in the US. They both indicate that there are two episodes of housing market bubbles during the sampling period, in the2000s and 2020s respectively. Then, based on LPM, Probit, and Logit estimation, the study finds that housing market conditions, monetary policy, and stock market returns carry useful information about the occurrence of housing bubbles. Meanwhile, macroeconomic indicators, such as inflation rate, unemployment rate, and GDP growth rate, fail to be important explanatory variables. The study shows that the in-sample performance for the three regression models is all satisfactory.

Aside from the baseline results, the study further conducts subsample analysis for the period before and after 2008 financial crisis. The estimation indicates the presence of a structural break in the relationship between various explanatory variables and the housing bubble indicator. In addition, with the extension of the baseline regression model to include more lagged terms of the independent variables, the study finds that housing market reaction to change in economic conditions is quite instant, as historical information from half year or longer periods ago fail to improve the forecast performance.

Lastly, the study evaluates the out-of-sample performance of the three regression models using static forecast, rolling forecast, and recursive forecast respectively. Based on the out-of-sample performance, the study shows that static forecast and rolling forecast are not appropriate for the forecast of housing bubbles. In addition, compared with the naïve model of random walk, the recursive model fails to yield a better forecast performance either.

The study shows the difficulty in constructing a regression model to forecast house price bubbles as the out-of-sample performance of the three estimation methods is poor relative to a naïve benchmark. However, that is not to say that the constructed model in the study is of no use. In fact, when it comes to the prediction of a housing market bubble, the two most important time points to be forecasted are the starting point and the ending point for a bubble series. This is because these two points are of the most value both to policy makers that aim to stabilize the housing market and to speculative traders that aim to make a financial return on housing. The naïve model provides no information about these two data points. In contrast, while the regression models constructed in the study does not make a precise forecast for the occurrence of a housing market bubble out-of-sample, they do provide valuable information for the rise and fall of the housing bubble risk. Such information could be useful to policy makers as it could guide the implementation of certain stabilising policies before the actual appearance of a housing market bubble (or its collapse). This study therefore provides a useful starting point for the prediction of house price bubbles that should provide value to policymakers and institutional investors alike.

Appendix 1: List of variables and data sources

Categories	Variables	Source	Unit
Housing market conditions	Private Housing Starts (PHS)	U.S. Census Bureau	Thousands of units
	Construction Value Put in Place (VIP)	U.S. Census Bureau	Billions of USD
	Growth rate of price to income (PI_growth)	OECD database	%
	Growth rate of price to rent (PR_growth)	OECD database	%
Macroeconomic conditions	Unemployment Rate (Unemployment)	IMF	%
	Inflation rate (Inflation)	IMF	%
	GDP growth rate (growth)	OECD database	%
Monetary policy	Short-term Interest Rate (ShortRate)	OECD database	%
	Long-term Interest Rate (LongRate)	OECD database	%
	Narrow money (M1)	OECD database	USD
Stock market condition	Stock return (Stock)	OECD database	%

Appendix 2: Alternative regression model setting

Table A1 Additional estimation

	(1)	(2)	(3)	(4)	(5)
	LPM	Probit		Logit	
		Estimation	ME	Estimation	ME
phs_growth	-0.0042	-0.065**	-0.00043	-0.11*	-0.00069
	(0.0032)	(0.032)	(0.00073)	(0.059)	(0.00077)
vip_growth	-0.0078	-0.011	-0.000069	-0.027	-0.00018
	(0.0073)	(0.066)	(0.00047)	(0.12)	(0.00082)
pr_growth	0.14***	1.50***	0.0099	2.66***	0.018
	(0.020)	(0.34)	(0.017)	(0.65)	(0.018)
pi_growth	0.046***	0.099	0.00065	0.21	0.0014
	(0.015)	(0.13)	(0.0015)	(0.23)	(0.0022)
growth	-0.064**	-0.55**	-0.0036	-0.95**	-0.0062
	(0.026)	(0.22)	(0.0065)	(0.42)	(0.0072)
unemployment	-0.044	-0.58	-0.0038	-1.00	-0.0066
	(0.032)	(0.36)	(0.0066)	(0.64)	(0.0073)
inflation	0.0030	-1.40***	-0.0093	-2.42***	-0.016
	(0.031)	(0.41)	(0.015)	(0.74)	(0.016)
shortrate	0.021	0.98***	0.0064	1.69***	0.011
	(0.035)	(0.36)	(0.011)	(0.64)	(0.012)
longrate	-0.021	-0.25	-0.0017	-0.46	-0.0030
	(0.038)	(0.38)	(0.0038)	(0.67)	(0.0053)
m1_growth	-0.0029	-0.00053	-0.0000035	-0.0034	-0.000022
	(0.0022)	(0.027)	(0.00018)	(0.050)	(0.00033)
stock	-0.010	-0.37*	-0.0024	-0.65	-0.0042
	(0.026)	(0.21)	(0.0044)	(0.39)	(0.0049)
_cons	0.85***	3.70**		6.48**	
	(0.21)	(1.55)		(2.93)	
N	141	141	141	141	141
R^2	0.488				
adj. R^2	0.444				
pseudo R^2		0.709		0.708	

Note: Table A1 reports the alternative model specification to check for the robustness of baseline estimation results. The first column corresponds to the LPM regression; the second and the third columns correspond to the raw estimation and marginal effect from the Probit model; the fourth and the fifth columns correspond to the raw estimation and marginal effect from the Logit model. Robust standard errors are calculated with adjustment to possible heteroscedasticity and serial correlation. The notations for the statistical significance follow the rule as: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A1 reports the analysis with PHS, VIP, and M1 included in growth rate form instead of the level form. In contrast with the baseline estimation reported in table 5.1, such a specification is able to capture the influence of housing construction speed and money creation on housing price bubble. Compared with the benchmark estimation depicted in table 5.1, we notice that the overall model fitness reduces, with the R squared of LPM gets lower from 0.617 to 0.488 and the pseudo-R squared of Probit

(Logit) regression reducing from 0.818 (0.818) to 0.709 (0.708), respectively.

Nonetheless, in terms of the main findings of the estimation, we are able to obtain several similar results as with table 5.1. Specially, paralleling the discussion with the benchmark estimation, the analysis below is mainly based on LPM estimation for convenience, while the marginal effect constructed with Probit and Logit regression is reported in parentheses.

To begin with, table A1 indicates that the sped-up construction of new houses is able to cool down the housing market and reduce the likelihood of a housing market bubble. In detail, for a one percent higher growth rate of PHS and VIP, the chance of a housing market bubble is found to be lower by 0.42% and 0.78%, respectively (0.043% and 0.0069% in Probit model; 0.069% and 0.0018% in Logit model). Although statistical significance is not detected at 5% level, the estimation indicates the importance of supply-side forces in the determination of the probability for a housing market bubble.

Next, similar as the benchmark regression, the rise of housing price to income level and the housing price to rent price level is found to be associated with a higher probability of a housing market bubble. In detail, with a one percent higher growth rate of housing price to income, the chance of a housing market bubble is expected to be higher by 4.6% (0.065% in Probit and 0.14% in Logit). The effect is documented to be statistically significant by 5% level. Meanwhile, when the growth rate for the price-to-rent ratio gets higher by one percent, the chance of a housing market bubble is higher by 14% (0.99% in Probit and 1.8% in Logit). The effect is again statistically significant by 5% level. More importantly, from the perspective of economic magnitude, both two indicators have a strong impact on the risk of a housing market bubble, which is consistent with the discussion of benchmark regression in table 5.1.

Then, from the aspect of macroeconomic indicators, table A1 indicates some different findings as with table 5.1. To be specific, while the economic growth rate remains to have a negative association with the risk of a housing market bubble, the statistical significance and the economic magnitude of the influence both increases. Meanwhile, similar to table 5.1, the unemployment rate is not shown to have a statistically detectable relationship with the risk of a housing market bubble. However, it is worth noting that the direction of the relationship reverses in table A1. Finally, when it comes to the influence of the inflation rate, similar to table 5.1, table A1 indicates an inconsistent estimation result based on LPM versus Probit and Logit model. While the inflation rate is shown to have a positive and statistically insignificant effect on the risk of a housing market bubble in the LPM estimation, it is found to have a positively significant role in Probit and Logit estimations. Nonetheless, its marginal effect at the sample mean values is not statistically significant in either of Probit or Logit regression.

Moreover, table A1 again indicates the relationship between monetary policy proxy variables and the risk of a housing market bubble. While the direction of the estimation is comparable with table 5.1, there is a reduction of statistical significance for the estimation of long run interest rate. In detail, table A1 suggests again that short-term interest rate is of a positive relationship with the risk of a housing market bubble. It implies that short-term interest rate adjustment is applied as the reaction by the central bank to stabilize housing market prosperity. Meanwhile, the negative relationship between long-term interest rate and the risk of a housing market bubble suggests that the guidance of central bank for the long run monetary policy (and associated inflation rate) plays a critical role for the dynamic movement of the housing market.

Lastly, in terms of the relationship between housing market and stock market, the substitution effect discussion in table 5.1 is again detected. In detail, table A1 shows that when stock market return is higher by one percent, the risk of a housing market bubble reduces by 1% (0.24% in Probit and 0.42% in Logit). Although the statistical significance level of the estimation reduces, the economic importance of the variable remains. It indicates the competition for capital investment between the housing market and the stock market, as the prosperity of the stock market attracts funds away from the housing market and reduces the risk of a housing market bubble.

To sum up, table A1 is able to demonstrates similar research findings as with table 5.1 in terms of the role for supply side forces, macroeconomic variables, monetary policy, and asset market conditions. It suggests that the forecast of a housing market bubble is in general not sensitive to the choice of the functional form of the estimation. Besides, table A1 confirms that Probit and Logit estimation is in general, consistent with the LPM estimation from the aspect of coefficient signs. However, it shows the difference in terms of economic magnitude and statistical significance, which indicates that different choices of regression methods may have an unneglectable impact on the performance of the forecast for a housing market bubble.

Reference

Adams, Z. and Füss, R., 2010. Macroeconomic determinants of international housing markets. *Journal of Housing Economics*, *19*(1), pp.38-50.

Amidu, A.R., Agboola, A.O. and Musa, M., 2016. Causal relationship between private housing investment and economic growth: An empirical study. *International Journal of Housing Markets and Analysis*, *9*(2), pp.272-286.

Anoruo, E. and Braha, H., 2008. Housing and stock market returns: An application of GARCH enhanced VECM. *The IUP Journal of Financial Economics*, *32*(2), pp.30-40.

Arce, Ó. and López-Salido, D., 2011. Housing bubbles. *American Economic Journal: Macroeconomics*, *3*(1), pp.212-241.

Astill, S., Harvey, D.I., Leybourne, S.J., Sollis, R. and Robert Taylor, A.M., 2018. Real-time monitoring for explosive financial bubbles. *Journal of Time Series Analysis*, *39*(6), pp.863-891.

Bahaman-Oskooee, M., Ghodsi, H., Hadzic, M. and Marfatia, H., 2024. Do expansionary and contractionary monetary policy have a symmetric impact on housing permits across the USA?. *International Journal of Housing Markets and Analysis*, *17*(4), pp.1034-1049.

Bahmani-Oskooee, M. and Ghodsi, S.H., 2018. Asymmetric causality between the US housing market and its stock market: Evidence from state level data. *The Journal of Economic Asymmetries*, *18*, p.e00095.

Baldi, G., 2014. The economic effects of a central bank reacting to house price inflation. *Journal of Housing Economics*, *26*, pp.119-125.

Barlow, J. and Vodenska, I., 2021. Socio-Economic Impact of the COVID-19 Pandemic in the US. *Entropy*, *23*(6), p.673.

Blanchard, O. and Watson, M., 1982. Bubbles, Rational Expectations and Financial Markets. Crises in the Economic and Financial Structure. P. Wachtel. Lexington.

Boone, L. and Girouard, N., 2003. The stock market, the housing market and consumer behaviour. *OECD economic studies*, *2002*(2), pp.175-200.

Bourassa, S.C., Hoesli, M. and Oikarinen, E., 2019. Measuring house price bubbles. *Real Estate Economics*, *47*(2), pp.534-563.

Campbell, J.Y. and Shiller, R.J., 1987. Cointegration and tests of present value models. *Journal of political economy*, *95*(5), pp.1062-1088.

Cheng, I.H., Raina, S. and Xiong, W., 2014. Wall Street and the housing bubble. *American Economic Review*, *104*(9), pp.2797-2829.

Chiang, M.C., Sing, T.F. and Wang, L., 2020. Interactions between housing market and stock market in the United States: A Markov switching approach. *Journal of Real Estate Research*, *42*(4), pp.552-571.

Chirchir, D., Mwangi, M. and Iraya, C., 2024. Modeling Nairobi Residential Real Estate Prices using ARIMA. *European Journal of Business and Management Research*, *9*(4), pp.30-36.

Chi-Wei, S., Yin, X.C., Tao, R. and Zhou, H., 2018. Are house pricess improving GDP or vice versa? A cross-regional study of China. *Applied Economics*, *50*(29), pp.3171-3184.

Cole, R.A. and Gunther, J.W., 1998. Predicting bank failures: A comparison of on-and off-site monitoring systems. *Journal of Financial Services Research*, *13*(2), pp.103-117.

Cole, R.A. and White, L.J., 2012. Déjà vu all over again: The causes of US commercial bank failures this time around. *Journal of financial services Research*, *42*, pp.5-29.

Crawford, G.W. and Fratantoni, M.C., 2003. Assessing the forecasting performance of regime-switching, ARIMA and GARCH models of house prices. *Real Estate Economics*, *31*(2), pp.223-243.

Diba, B.T. and Grossman, H.I., 1988. Explosive rational bubbles in stock prices?. *The American Economic Review*, *78*(3), pp.520-530.

Diewert, W.E., Nakamura, A.O. and Nakamura, L.I., 2009. The housing bubble and a new approach to accounting for housing in a CPI. *Journal of Housing Economics*, *18*(3), pp.156-171.

Diebold, F.X. and Mariano, R.S., 2002. Comparing predictive accuracy. *Journal of Business & economic statistics*, *20*(1), pp.134-144.

Donald Jud, G. and Winkler, D.T., 2003. The Q theory of housing investment. *The*

Journal of Real Estate Finance and Economics, *27*, pp.379-392.

Dreger, C. and Kholodilin, K.A., 2012. *An early warning system to predict the speculative house price bubbles* (No. 2012-44). Economics Discussion Papers.

Duca, J.V., Muellbauer, J. and Murphy, A., 2021. What drives house price cycles? International experience and policy issues. *Journal of Economic Literature*, *59*(3), pp.773-864.

Estrella, A. and Mishkin, F.S., 1998. Predicting US recessions: Financial variables as leading indicators. *Review of Economics and Statistics*, *80*(1), pp.45-61.

Evans, G.W., 1991. Pitfalls in testing for explosive bubbles in asset prices. *The American Economic Review*, *81*(4), pp.922-930.

Favara, G. and Imbs, J., 2015. Credit supply and the price of housing. *American economic review*, *105*(3), pp.958-992.

Gamber, W., Graham, J. and Yadav, A., 2023. Stuck at home: Housing demand during the COVID-19 pandemic. *Journal of Housing Economics*, *59*, p.101908.

Garriga, C., Manuelli, R. and Peralta-Alva, A., 2019. A macroeconomic model of price swings in the housing market. *American Economic Review*, *109*(6), pp.2036-2072.

Ghent, A.C. and Owyang, M.T., 2010. Is housing the business cycle? Evidence from US cities. *Journal of Urban Economics*, *67*(3), pp.336-351.

Glaeser, E.L. and Nathanson, C.G., 2015. Housing bubbles. In *Handbook of regional and urban economics* (Vol. 5, pp. 701-751). Elsevier.

Glaeser, E. and Gyourko, J., 2018. The economic implications of housing supply. *Journal of economic perspectives*, *32*(1), pp.3-30.

Glaeser, E.L., Gyourko, J. and Saiz, A., 2008. Housing supply and housing bubbles. *Journal of urban Economics*, *64*(2), pp.198-217.

Granger, C.W. and Newbold, P., 1974. Spurious regressions in econometrics. *Journal of econometrics*, *2*(2), pp.111-120.

Gupta, R., 2013. Forecasting house prices for the four census regions and the aggregate US economy in a data-rich environment. *Applied Economics*, *45*(33), pp.4677-4697.

Hamilton, J.D., 2020. *Time series analysis*. Princeton university press.

Harris, J.C., 1989. The effect of real rates of interest on house pricess. *The Journal of Real Estate Finance and Economics*, *2*, pp.47-60.

Harvey, D.I., Leybourne, S.J. and Whitehouse, E.J., 2020. Date-stamping multiple bubble regimes. *Journal of Empirical Finance*, *58*, pp.226-246.

Iacoviello, M. and Neri, S., 2010. Housing market spillovers: evidence from an estimated DSGE model. *American economic journal: macroeconomics*, *2*(2), pp.125-164.

Jen-Shi, N., Shuen-Shi, H. and Yu, W., 2012. Interest rate, Unemployment rate, and Housing Market in US. *Journal of Modern Accounting and Auditing*, *8*(6), p.837.

Jerven, M., 2016. Data and statistics at the IMF: Quality assurances for low-income countries. *Background Paper, Independent Evaluation Office of the International Monetary Fund, Washington DC, Feb*, *25*.

Jin, Y. and Zeng, Z., 2004. Residential investment and house prices in a multi-sector monetary business cycle model. *Journal of Housing Economics*, *13*(4), pp.268-286.

Kishore, R., 1996. Discounted cash flow analysis in property investment valuations. *Journal of Property Valuation and Investment*, *14*(3), pp.63-70.

Kuang, W. and Liu, P., 2015. Inflation and House Prices: Theory and Evidence from 35 Major Cities in China. *International Real Estate Review*, *18*(2).

Lastrapes, W.D., 2002. The real price of housing and money supply shocks: time series evidence and theoretical simulations. *Journal of Housing Economics*, *11*(1), pp.40-74.

Liow, K.H., Huang, Y. and Song, J., 2019. Relationship between the United States housing and stock markets: some evidence from wavelet analysis. *The North American Journal of Economics and Finance*, *50*, p.101033.

Luciani, M., 2015. Monetary policy and the housing market: A structural factor analysis. *Journal of applied econometrics*, *30*(2), pp.199-218.

Maddala, G.S., 1983. *Limited-dependent and qualitative variables in econometrics* (Vol. 149). Cambridge University Press.

Marfatia, H.A., André, C. and Gupta, R., 2022. Predicting housing market sentiment: The role of financial, macroeconomic and real estate uncertainties. *Journal of*

Behavioral Finance, *23*(2), pp.189-209.

McCracken, M.W. and Ng, S., 2016. FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, *34*(4), pp.574-589.

McQuinn, K. and O'Reilly, G., 2008. Assessing the role of income and interest rates in determining house prices. *Economic modelling*, *25*(3), pp.377-390.

Mohan, S., Hutson, A., MacDonald, I. and Lin, C.C., 2019. Impact of macroeconomic indicators on house pricess. *International Journal of Housing Markets and Analysis*, *12*(6), pp.1055-1071.

Montgomery, D.C., Peck, E.A. and Vining, G.G., 2021. *Introduction to linear regression analysis*. John Wiley & Sons.

Nneji, O., Brooks, C. and Ward, C.W., 2013. House price dynamics and their reaction to macroeconomic changes. *Economic Modelling*, *32*, pp.172-178.

Osius, G. and Rojek, D., 1992. Normal goodness-of-fit tests for multinomial models with large degrees of freedom. *Journal of the American Statistical Association*, *87*(420), pp.1145-1152.

Petrini, G. and Teixeira, L., 2023. Determinants of residential investment growth rate in the US economy (1992–2019). *Review of Political Economy*, *35*(3), pp.702-719.

Phillips, P.C., Wu, Y. and Yu, J., 2011. Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values?. *International economic review*, *52*(1), pp.201-226.

Phillips, P.C., Shi, S. and Yu, J., 2015. Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International economic review*, *56*(4), pp.1043-1078.

Robertson, A.Z., 2023. The (mis) uses of the S&P 500. *The University of Chicago Business Law Review*, *2*(1), p.3.

Rünstler, G. and Vlekke, M., 2018. Business, housing, and credit cycles. *Journal of Applied Econometrics*, *33*(2), pp.212-226.

Schwab, R.M., 1982. Inflation expectations and the demand for housing. *The American*

Economic Review, *72*(1), pp.143-153.

Shiller, R., 1981. Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review, 71*(4), pp.421–436

Sirmans, S., Macpherson, D. and Zietz, E., 2005. The composition of hedonic pricing models. *Journal of real estate literature*, *13*(1), pp.1-44.

Smith, M.H. and Smith, G., 2006. Bubble, bubble, where's the housing bubble?. *Brookings Papers on Economic Activity*, *2006*(1), pp.1-67.

Soo, C.K., 2018. Quantifying sentiment with news media across local housing markets. *The Review of Financial Studies*, *31*(10), pp.3689-3719.

Spencer, P. and Liu, Z., 2010. An open-economy macro-finance model of international interdependence: The OECD, US and the UK. *Journal of banking & finance*, *34*(3), pp.667-680.

Stock, J.H. and Watson, M.W., 2020. *Introduction to econometrics*. Pearson.

Tsai, I.C., 2012. Housing supply, demand and price: construction cost, rental price and house price indices. *Asian Economic Journal*, *26*(4), pp.381-396.

Tsai, I.C., 2013. The asymmetric impacts of monetary policy on house pricess: A viewpoint of house prices rigidity. *Economic Modelling*, *31*, pp.405-413.

Tse, R.Y., 1997. An application of the ARIMA model to real-estate prices in Hong Kong. *Journal of Property Finance*, *8*(2), pp.152-163.

Vansteenkiste, I. and Hiebert, P., 2011. Do house price developments spillover across euro area countries? Evidence from a global VAR. *Journal of Housing Economics*, *20*(4), pp.299-314.

Vargas-Silva, C., 2008. Monetary policy and the US housing market: A VAR analysis imposing sign restrictions. *Journal of Macroeconomics*, *30*(3), pp.977-990.

Wachowicz, E., 2020. Wharton research data services (WRDS). *Journal of Business & Finance Librarianship*, *25*(3-4), pp.184-187.

Wang, S., Zeng, Y., Yao, J. and Zhang, H., 2020. Economic policy uncertainty, monetary policy, and house prices in China. *Journal of Applied Economics*, *23*(1), pp.235-252.

Whitehouse, E.J., Harvey, D.I. and Leybourne, S.J., 2025. Real-time monitoring procedures for early detection of bubbles. *International Journal of Forecasting*.

Wooldridge, J.M., 2015. Introductory Econometrics: A Modern Approach. Cengage AU.

Wu, J., Gyourko, J., & Deng, Y., 2016. Evaluating the risk of Chinese housing markets: What we know and what we need to know. *China Economic Review*, 39, 91-114.

Zheng, Y. and Osmer, E., 2021. House prices dynamics: The impact of stock market sentiment and the spillover effect. *The Quarterly Review of Economics and Finance*, *80*, pp.854-867.