HEC MONTRÉAL

Predicting Housing Prices in Canada Using News Sentiments, Bubble Indicators, and Google Trends

by

Mohammad Ebrahimi

Simon van Norden

HEC Montréal

Research Director

Vincent Grégoire

HEC Montréal

Research Codirector

Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Finance (M.Sc.)

Mémoire présenté en vue de l'obtention du grade de maîtrise ès sciences en finance (M. Sc.)

August 2024

Résumé

L'objectif principal est d'évaluer si le sentiment des nouvelles, les indicateurs de bulle et Google Trends ont un pouvoir prédictif pour les prix de l'immobilier et de comparer la performance de ces prédicteurs entre eux. L'analyse utilise des modèles de langage de grande envergure (LLMs) pour extraire le sentiment des nouvelles et le convertir en un indice quantifiable. De plus, l'étude intègre les indicateurs de bulle SADF et BSADF issus du test PSY, ainsi que les indicateurs A et S du test WHL. Un nouvel indice de recherche immobilière, dérivé de Google Trends, est également introduit. Les prévisions sont réalisées à l'aide d'un modèle autorégressif avec entrée exogène (ARX), où chaque prédicteur est évalué séparément pour sa capacité à améliorer la précision des prévisions et sa signification statistique. Les résultats révèlent que, dans certains cas, l'incorporation du sentiment des nouvelles et des indicateurs de bulle améliore la précision des prévisions. Notamment, les indices dérivés de Google Trends surpassent les autres prédicteurs sur un horizon d'un mois.

Mots clés: Prix de l'immobilier, Sentiment des nouvelles, Indicateurs de bulle, Google Trends, Modélisation prédictive, Apprentissage automatique, Modèle autorégressif, et Prévision.

Méthodes de recherche: Analyse économétrique, Apprentissage automatique, Analyse de sentiment, Analyse des séries temporelles, Analyse en composantes principales, Tests de détection de bulle

Abstract

The primary objective is to evaluate whether news sentiment, bubble indicators, and Google Trends have predictive power for real estate prices and to compare the performance of these predictors against each other. The analysis employs large language models (LLMs) to extract sentiment from news and convert it into a quantifiable index. Additionally, the study incorporates SADF and BSADF bubble indicators from the PSY test, as well as A and S indicators from the WHL test. A novel Housing Search Index, derived from Google Trends, is also introduced. Forecasting is conducted using an Autoregressive model with exogenous input (ARX), where each predictor is evaluated separately for its ability to improve forecasting accuracy and statistical significance. The findings reveal that, in certain cases, incorporating news sentiment and bubble indicators improves forecast accuracy. Notably, indices derived from Google Trends outperform other predictors in the one-month horizon.

Keywords: Real Estate Pricing, News Sentiment, Bubble Indicators, Google Trends, Predictive Modeling, Machine Learning, Autoregressive Model, and Forecasting.

Research Methods: Econometric Analysis, Machine Learning, Sentiment Analysis, Time Series Analysis, Principal Component Analysis, Bubble Detection Tests.

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

ADF: Augmented Dickey-Fuller

AI: Artificial Intelligence

ANN: Artificial Neural Networks

BIC: Bayesian Information Criterion

BSADF: Backward Supremum Augmented Dickey-Fuller

CV: Cross Validation

FRB: Federal Reserve Bank

GT: Google Trend (series)

HPI: House Price Index

HSI: Housing Search Index

iid: Independent and identically distributed

KDE: Kernel Density Estimate

LLM: Large Language Model

LTS: Long-Term Sentiment

MAFE: Mean Absolute Forecast Error

MDM: Modified Diebold Mariano

MSFE: Mean Squared Forecast Error

OLS: Ordinary least squares

PCA: Principal Component Analysis

PPI: Property Price Index

SADF: Supremum Augmented Dickey-Fuller

STS: Short-Term Sentiment

SVM: Support Vector Machine

UHSI: Ultimate Housing Search Index

Acknowledgments

I express my sincere gratitude to my supervisors, Professors Simon van Norden and Vincent Grégoire. Their steadfast guidance, expertise, and encouragement have been pivotal in shaping both this research and my academic development. Their mentorship and support have been truly invaluable, and I am privileged to have had the opportunity to work under their supervision.

I am also deeply grateful to my wife, Sepideh, for her unwavering support and inspiration throughout the journey of completing my master’s thesis. Her love and understanding have been my greatest strengths.

1. Introduction

1.1 Introduction to Real Estate Price Prediction

The real estate market is a crucial component of any nation's economy, reflecting broader economic trends and individual consumer behavior. Predicting real estate prices has been a focal point of academic and industry research due to its implications for investment, policy-making, and economic stability. This thesis explores the predictive capabilities of unconventional data sources—specifically, news sentiment, Google Trends, and bubble indicators—in forecasting real estate prices in Canada. By employing news sentiment, this research aims to develop a forecasting model that captures market sentiment. The inclusion of Google Trends provides insights into public interest and potential demand, while bubble indicators help identify periods of overvaluation or speculative activity.

The academic discourse surrounding real estate price prediction has evolved significantly over the years. Traditional models often relied on economic fundamentals such as interest rates, income levels, and employment rates (Tabales et al., 2013). However, these models frequently fell short of capturing the full complexity of market dynamics, particularly the psychological and behavioral factors influencing buyer and seller decisions. As a result, researchers have increasingly turned to alternative data sources that can serve as proxies for market sentiment and demand.

One innovative approach involves analyzing news sentiment as a predictor of real estate price movements. This method leverages textual analysis of news articles, aiming to quantify public sentiment and its potential impact on market behavior. Previous studies have demonstrated that shifts in public sentiment, as captured by media narratives, can significantly influence investor decisions and market trends (e.g., Baker et al., 2016). Such sentiment indicators provide a real-time, dynamic measure of market sentiment, offering a more immediate reflection of market mood than traditional economic indicators.

Another promising area of research involves using Google Trends data as a proxy for consumer interest and demand. By analyzing search queries related to real estate, researchers can gain insights into public interest and potential future market activities. This approach has been validated in various contexts, demonstrating that increased search volume often correlates with subsequent market movements (e.g., Choi and Varian, 2012). The use of Google Trends provides a unique advantage in capturing the public's immediate concerns and interests, which traditional economic indicators might overlook.

Lastly, the consideration of bubble indicators acknowledges the presence of irrational behavior and speculative bubbles in real estate markets. These indicators, often derived from metrics like price-to-rent ratios and deviation from long-term trends, help identify periods of overvaluation driven by investor exuberance rather than fundamental value. The study of bubbles has gained traction, especially in the aftermath of the global financial crisis, highlighting the need to account for irrational market behavior in predictive models (e.g., Shiller, 2000).

This thesis aims to investigate these unconventional data sources—news sentiment, Google Trends, and bubble indicators—for predicting real estate prices in Canada. By utilizing these diverse indicators, the research seeks to develop a more holistic and accurate forecasting framework that captures the multifaceted nature of real estate markets. This approach not only contributes to the academic literature but also provides practical insights for investors, policymakers, and other stakeholders in the Canadian real estate market.

While promising, the use of these diverse predictors faces several challenges. These include the availability and reliability of sentiment data, the identification of accurate bubble indicators, and the interpretation of Google Trends data in the context of real estate. Moreover, the dynamic nature of real estate markets and economic conditions make real estate forecasting more complex (Cristescu et al., 2022)

1.2 Artificial Intelligence in Price Prediction

The rapid advancements in Artificial Intelligence (AI) have revolutionized various fields, including finance and economics, where AI techniques are increasingly being employed to enhance predictive accuracy. In the context of real estate markets, AI offers powerful tools for analyzing complex datasets and identifying patterns that traditional methods might overlook. Many studies explored the role of AI in price prediction, highlighting its ability to integrate diverse sources of information to provide more accurate and timely forecasts.

Ardila et al. (2016) introduced a hybrid model combining the Log Periodic Power Law Singular (LPPLS) model with a diffusion index to forecast real estate bubbles. This approach integrates AI with econophysics to identify and predict bubble dynamics, showcasing how AI can be employed to understand complex market phenomena.

Zhang et al. (2018) proposed a novel stock price trend prediction system that can predict both the movement and the growth (or decline) rate of stock prices. While focused on the stock market, the methodology and principles, including unsupervised learning and classification, have relevance for real estate price forecasting, illustrating the adaptability and potential of AI techniques in analyzing financial data. Additionally, "Predicting Property Price Index Using Artificial Intelligence Techniques" by Abidoye et al. (2019) explored the utility of AI for predicting the property price index (PPI), demonstrating the superior performance of artificial neural networks (ANNs) over support vector machine (SVM) and ARIMA models in real estate price prediction.

1.3 News Sentiment Analysis in Real Estate

Academic literature has increasingly recognized the importance of psychological and behavioral factors in financial markets. In the context of real estate, news sentiment analysis has emerged as a powerful tool for capturing market sentiment. News articles, opinion pieces, and media reports often reflect or shape public perception and investor sentiment, influencing decision-making processes in the real estate market. Numerous studies have explored the correlation between news sentiment and asset prices, demonstrating that positive or negative media coverage can impact market dynamics.

For instance, Tetlock (2007) found that high levels of pessimistic news coverage are associated with declining stock prices, suggesting a similar potential effect in real estate markets. In the housing market, several studies have shown that news sentiment can serve as a leading indicator of price movements. Garcia (2013) demonstrated that changes in news sentiment about housing market conditions could predict subsequent changes in home prices.

In recent years, researchers have leveraged advanced natural language processing techniques to quantify news sentiment more accurately. These techniques involve analyzing the tone, frequency, and context of words in news articles to construct sentiment indices. For example, Soo (2018) employed sentiment analysis of real estate-related news to predict U.S. housing market trends, finding that sentiment indices significantly forecast future home price changes. The study highlighted that media sentiment could capture a broader range of market drivers than traditional economic indicators, such as consumer confidence or interest rates.

Damianov, Wang, and Yan (2020) examined the predictive power of household sentiment derived from Google searches on house prices and foreclosure rates. They found that the intensity of searches for "mortgage assistance" and "foreclosure help" was negatively associated with house prices but correlated with lower future foreclosure rates. They highlight the importance of sentiment analysis and its implications for real estate market predictions.

Nti et al. (2020) examined the predictability of stock market movements using sentiment analysis derived from web news, Twitter, forums, and Google Trends. By employing artificial neural networks (ANN), the study achieved improved prediction accuracy, indicating the significance of diverse data sources and AI in forecasting market trends. Though focused on the stock market, the findings are applicable to real estate market predictions, where investor sentiment and online search trends play a crucial role.

Fan and Chen (2022) proposed a model for predicting real estate stock prices based on investor sentiment, using the stock bar comment scores and the Baidu search index to construct a composite sentiment index. They highlighted the significance of sentiment analysis, particularly in the context of increasing risk aversion among investors in the real estate development business.

1.4 Bubble Indicators in Real Estate

The concept of bubbles in real estate refers to a situation where property prices significantly exceed their intrinsic values, often due to speculative activities. This issue is extensively analyzed in the works of Case and Shiller (2003), Glaeser (2013), Glaeser and Nathanson (2017), Granziera and Kozicki (2015), and Pavlidis et al. (2016), as well as in books "Animal Spirits" by Akerlof and

Shiller (2010) and "This Time It’s Different: Eight Centuries of Financial Folly" by Reinhart and Rogoff (2009). Identifying bubbles is critical for predicting potential market corrections and protecting the economy from severe downturns.

Real estate markets are susceptible to speculative bubbles, characterized by rapid increases in property prices driven by exuberant market behavior rather than fundamental values. Identifying and understanding these bubbles are crucial for forecasting potential market corrections and mitigating economic impacts. Oust and Hrafnkelsson (2017) provided a descriptive definition of housing price bubbles, identifying significant price movements around peaks and troughs as indicators of speculative bubbles. Ardila, Sanadgol, and Sornette (2018) analyzed the performance of statistical tests for bubble detection in real estate markets, suggesting the effectiveness of binary indicators derived from super-exponential trends in forecasting housing bubble tipping points. Hagemann and Wohlmann (2019) developed an early warning system to identify speculative price bubbles in housing markets using a logit regression approach, emphasizing monetary developments as significant predictors. Asal (2019) applied multiple methods, including affordability indicators, asset-pricing approaches, and right-tailed unit root tests, to assess the presence of housing bubbles in Sweden, suggesting periodic deviations from fundamental values. Jovanović (2021) explored the use of quantitative ratios from the System of National Accounts in predicting price bubbles in the housing market, suggesting that excessive construction activity could lead to the formation of price bubbles.

This study focuses on two significant contributions to bubble detection methodologies: the PSY bubble indicators developed by Phillips, Shi, and Yu (2015) and the bubble statistics proposed by Whitehouse, Harvey, and Leybourne (2022).

Phillips, Shi, and Yu (2015) introduced a robust econometric framework for detecting and dating speculative bubbles in time series data. Their methodology, known as the PSY test, extends the traditional Augmented Dickey-Fuller (ADF) test by incorporating a recursive right-tailed unit root test, allowing for the identification of multiple bubble periods within a single time series.

The PSY methodology has been widely applied in various contexts, including stock markets, housing markets, and foreign exchange markets. Its ability to identify both the start and end points of bubble periods provides valuable insights for policymakers and market participants, enabling timely interventions and informed decision-making.

Whitehouse, Harvey, and Leybourne (2022) proposed a complementary approach to bubble detection, focusing on real-time monitoring procedures to identify explosive and collapsing regimes in financial time series. Their framework introduces two key statistics: the $A(e, m)$ statistic for detecting explosive behavior and the $S(e, m, n)$ statistic for identifying subsequent collapses.

1.5 The Role of Google Trends in Economic Forecasting

Google Trends, which provides data on the popularity of search queries over time, has emerged as a valuable tool in economic forecasting. Research has shown that Google Trends data can predict economic indicators such as unemployment rates, consumer confidence, and even real estate market interest. Its real-time nature offers a unique advantage in capturing the current economic sentiment and interest.

Dietzel, Braun, & Schäfers (2014) explored the impact of Internet search query data on commercial real estate forecasting models. By incorporating Google Trends data, the study demonstrated significant improvements in the accuracy of forecasting models for commercial real estate transactions and price indices. This suggests that the sentiment reflected in search behavior extends beyond residential real estate to commercial markets as well, underlining the robustness of online search data as a sentiment indicator.

Dietzel (2016) demonstrated that Google search volume data could effectively serve as a leading sentiment indicator for the housing market. By analyzing real-estate-related search volumes, Dietzel found a reliable model for predicting market turning points in the US housing market, emphasizing the potential for applying similar methodologies in the Canadian context.

Oust & Eidjord Ole Martin (2018) investigated the predictive power of Google search volume indices for identifying housing bubbles, finding keywords related to real estate as effective indicators. Bulczak (2021) further explored the utility of Google Trends data in the UK real estate market, demonstrating its efficacy as a supplementary data source for forecasting real estate market movements. This study supports the inclusion of Google Trends data in predictive models for real estate price movements.

1.6 Forecasting Models and Integration of Predictors

Forecasting real estate prices is a complex and multifaceted task that has been the subject of extensive academic research. Various modeling approaches have been developed to capture the intricate dynamics of real estate markets, each with its own strengths and limitations. Among these, time series models, structural models, and machine learning models have emerged as the most prominent types of forecasting techniques. On the other hand, recent studies like Li et al. (2020) have explored the integration of sentiment analysis, bubble indicators, and Google Trends data for a more holistic approach to real estate price prediction. This integrated method aims to capture not only the fundamental and economic factors but also the psychological and speculative elements affecting the market.

Time series models are widely used in real estate price prediction due to their ability to capture temporal dependencies in data. One common type of time series model is the Autoregressive Integrated Moving Average (ARIMA) model, which combines autoregressive (AR) and moving average (MA) components to model the underlying patterns in the time series data. For example, Tsai et al. (2010) utilized an ARIMA model to forecast housing prices in Taiwan, demonstrating the model's efficacy in capturing short-term price movements.

The ARIMAX model extends the capabilities of ARIMA by integrating external (exogenous) variables into the forecasting model. Vishwakarma's (2013) research into the Canadian real estate market exemplifies this by including macroeconomic variables such as GDP, CPI, interest rate differentials, and the exchange rate of the Canadian dollar against the US dollar.

Structural models incorporate economic theories and principles to establish relationships between various economic variables and real estate prices. These models often include supply and demand factors, interest rates, and demographic changes. For instance, Meen (1999) developed a structural model to predict UK house prices by considering factors such as real income, housing stock, and mortgage rates. This approach provides insights into the causal relationships between economic variables and real estate prices.

With the advent of big data and advanced computing capabilities, machine learning models have gained popularity in forecasting real estate prices. These models, such as decision trees, neural networks, and support vector machines, can capture complex, non-linear relationships in the data. For example, Kok et al. (2017) employed a machine learning approach using random forests to predict housing prices in the U.S., highlighting the potential of these models to improve prediction accuracy.

In addition to these primary types, researchers have increasingly incorporated exogenous variables into traditional forecasting models to enhance their predictive power. Autoregressive models with exogenous variables (ARX) are a particular type of time series model that incorporates external factors—variables that are not directly related to past values of the target variable but may influence future values. By including these exogenous variables, researchers aim to account for external influences that might affect real estate prices.

For example, Shiller (1990) utilized an AR model augmented with macroeconomic indicators as exogenous variables to predict U.S. real estate prices, demonstrating that incorporating broader economic conditions can improve model accuracy. Another study by Case and Shiller (2003) included consumer sentiment as an exogenous variable in their AR model to capture psychological factors influencing housing market behavior. Similarly, Goodhart and Hofmann (2008) used an AR model with exogenous variables such as interest rates and credit conditions to forecast housing prices, highlighting the importance of financial variables in real estate market analysis.

This thesis adopts a similar approach, employing autoregressive models with exogenous variables to forecast real estate prices in Canada. Specifically, instead of macroeconomic variables, we explore the predictive power of psychological and behavioral factors. We investigate three distinct exogenous variables—news sentiment, Google Trends, and bubble indicators—by incorporating each variable separately into the AR model. The inclusion of these variables aims to capture external factors that may not be directly observable in past real estate price data but significantly influence market dynamics. By isolating the impact of each exogenous variable, this research seeks to provide a nuanced understanding of their individual contributions to real estate price forecasting, offering valuable insights for investors, policymakers, and market analysts. While news sentiment provides insights into market mood and investor confidence, Google Trends captures public interest and potential demand, and bubble indicators highlight the presence of speculative activity or market overvaluation.

2. Theoretical Background

This section examines the crucial determinants of the housing market to develop a robust regression model. While macro-financial trends significantly influence homeownership access, emerging evidence indicates that psychological factors also play a vital role in housing price dynamics. Notably, irrational public speculation can contribute to the formation of a housing bubble.

From a financial perspective, a house can be considered an asset whose price reflects the present value of the stream of housing services (or rental income for non-occupying owners) it generates. In this framework, housing market factors, termed "fundamentals," are those that directly influence the net value of these housing services, such as mortgage rates, employment, and vacancy rates. However, most households make purchasing decisions based on naive and simplistic speculations about local housing trends rather than a thorough asset-pricing rationale.

Two surveys conducted in 1988 and 2003, each sampling 2000 American homebuyers, support the hypothesis that price changes are closely linked to public market anticipation, particularly during periods of pronounced optimism or pessimism (Case and Shiller, 2003). In both surveys, more than half of the respondents indicated that their purchasing decisions were influenced by the excitement conveyed through word-of-mouth. This underscores the strong relationship between price changes and public market anticipation, especially in times of evident optimism or pessimism (Case and Shiller, 2003).

Studies by Kindleberger and Aliber (2011) and Shiller (2008) highlight the dangers of assuming consumer rationality. The limited cognition of agents is especially evident in housing finance, where capital gains (or losses) significantly influence consumer sentiment. Glaeser (2013) documents the history of US real estate fluctuations, emphasizing that public predictions of regional real estate growth are fundamentally flawed. Market participants consistently form expectations based on isolated, regional stimuli, neglecting broader macroeconomic factors (Glaeser, 2013). Even professional investors can exhibit excessive optimism, causing stock prices to deviate from their fundamental values. The UBS/Gallup Investor Survey, which surveyed 1,000

U.S. investors, discovered that as investors become more optimistic about the stock market, there tends to be a corresponding increase in the price-to-dividend ratios of stocks. This finding challenges the theory proposed by Fama and French (1988), which suggests that rational return expectations should not be influenced by such investor sentiments. (Adam, Marcet, and Beutel, 2017).

In our analysis, we discern the effects of market sentiment, irrational economic behavior, and latent demand. To that end, we perform forecasting by including each of the following in the model separately: news articles, bubble indicators, and Google Trends. News articles represent the market sentiment toward future real estate prices in the media. Bubble indicators aim to capture public speculation, market optimism, and extrapolative expectations. Google Trends shows the volume of the public's online search activity trying to find a home to buy, expressing potential future demand.

2.1 News Sentiment

The application of sentiment indices in real estate markets is relatively novel but promising. Real estate markets are influenced by a variety of factors, including economic conditions, government policies, and market sentiment. Market sentiment refers to the overall attitude or mood of investors and market participants toward a particular market or asset. It reflects the collective emotions and outlooks that investors have, which can be either positive (bullish) or negative (bearish). Market sentiment is influenced by various factors, including economic indicators, news, events, and broader market trends.

Market sentiment can be based on both rational and irrational behavior. Sometimes, market sentiment is grounded in objective analysis and rational decision-making. For example, investors may react positively to strong economic data or company earnings, leading to optimism in the market. This type of sentiment aligns with fundamental analysis and informed expectations about future performance. At other times, market sentiment can be driven by emotions, biases, or herd behavior, leading to irrational decisions. This can result in market bubbles, where asset prices inflate beyond their intrinsic values, or panic selling, where prices plummet despite little change in underlying fundamentals. Irrational sentiment often leads to volatility and can cause markets to deviate from their fundamental values. Overall, market sentiment is a blend of rational assessments and psychological influences, making it a complex and dynamic aspect of financial markets.

Sentiment analysis, also known as opinion mining, involves extracting subjective information from textual data to gauge mood or opinion. In financial markets, sentiment analysis has been widely applied to predict stock prices, market volatility, and economic indicators. The underlying assumption is that public sentiment, as reflected in news media, social media, and other textual sources, can influence investor behavior and, consequently, market dynamics. News sentiment can capture market expectations, investor confidence, and public interest, which are not always reflected in traditional economic indicators. By incorporating market sentiment into forecasting models, we can potentially enhance the accuracy and timeliness of real estate price predictions.

Market sentiment, the overall attitude of investors towards a particular market, has been identified as a key factor influencing real estate prices. Various studies have employed sentiment analysis, leveraging news articles, social media posts, and expert opinions to gauge market mood (Turner et al., 2021). The impact of market sentiment on real estate prices has been well-documented, with positive sentiment often correlating with price increases and negative sentiment with price decreases (Mudinas et al., 2018).

Several studies have demonstrated the efficacy of sentiment indices in financial forecasting. For instance, Tetlock (2007) showed that high levels of pessimism in news articles predicted lower stock market returns. Similarly, Da, Engelberg, and Gao (2011) found that sentiment extracted from Google search volumes could predict future stock prices. Nti, Adekoya, and Weyori (2020) explored the predictability of future stock price movements using public sentiments from web news, financial tweets, Google trends, and forum discussions. They applied Artificial Neural Networks (ANN) to stock data from the Ghana Stock Exchange and found that combining multiple data sources improved prediction accuracy significantly. This study underscores the potential of using diverse internet-based data, including news analyzed by LLMs, in forecasting stock market behaviors.

The inclusion of sentiment indices in real estate price forecasting models offers several advantages. It provides a more comprehensive view of market dynamics by incorporating psychological and behavioral factors. This approach can enhance the predictive power of traditional models, leading to more accurate and timely forecasts. For investors, policymakers, and other stakeholders, these improved predictions can inform better decision-making and risk management strategies. The theoretical foundation for using sentiment indices derived from news as predictors in real estate price forecasting models is well-supported by empirical evidence and methodological advancements. By integrating sentiment analysis into predictive models, this thesis aims to enhance the accuracy and reliability of real estate price predictions, offering valuable insights into market sentiment and its impact on housing prices.

2.2 Bubble Indicators

Real estate markets are often subject to periods of speculative bubbles, where property prices deviate significantly from their fundamental values. These bubbles can lead to substantial economic disruptions when they burst. Accurately predicting real estate prices, therefore, requires robust methodologies that can identify and account for these speculative episodes. This thesis integrates advanced bubble detection techniques into the predictive modeling of real estate prices, leveraging the frameworks developed by Phillips, Shi, and Yu (2015), and Whitehouse, Harvey, and Leybourne (2022).

Phillips, Shi, and Yu (2015) introduced a seminal approach to detecting speculative bubbles in financial time series. Their method extends the conventional Augmented Dickey-Fuller (ADF) test by incorporating a recursive right-tailed unit root test, allowing for the identification of multiple bubble periods within a single dataset. This framework includes two primary tests: the Supremum ADF (SADF) test and the Backward SADF (BSADF) test.

Supremum ADF (SADF) Statistic: This test recursively applies the ADF test over an expanding window of the sample, capturing the most extreme test statistic over all windows. It is particularly effective in detecting the initial emergence of a bubble by identifying the first instance of explosive behavior.

Backward SADF (BSADF) Statistic: To enhance the robustness of bubble detection, the BSADF test applies a backward recursive approach. This involves calculating the ADF test over rolling windows, which allows for the detection of multiple bubble episodes within the data. The BSADF test is thus more flexible and can identify both the formation and the end of speculative bubbles.

To account for episodes of market irrational behavior, we develop covariates that measure the intensity of price surges and declines. Pavlidis et al. (2016) employ the Backward Supremum Augmented Dickey-Fuller Test (BSADF) to identify housing bubbles. The BSADF test statistic serves as an indicator of exuberance, designed to detect multiple episodes of explosive dynamics in a univariate time series. Utilizing data from the International House Price Database of the Federal Reserve Bank of Dallas (FRB Dallas), the authors derive annual country-specific BSADF test statistics for real house prices and price-to-income ratios. The FRB Dallas maintains these updated statistics as part of its housing database.

The BSADF statistic for house prices tracks the explosiveness of house prices, while the BSADF statistic for house-price-to-income ratios highlights periods when house price movements diverge from market fundamentals. During a bubble, prices may significantly deviate from income trends, driven by speculative behavior. By examining price-to-income explosiveness, the authors discovered that price increases appear unsustainable when they depart from fundamental trends such as income. They also found that episodes of exuberance for house prices and price-to-income ratios generally coincide, but periods of exuberance are shorter for price-to-fundamentals. Additionally, negative exuberance levels are observed when time-series exhibit descending explosive-root-like trajectories.

The PSY methodology has proven effective in various financial markets, including real estate, where it helps distinguish between fundamental-driven price movements and speculative bubbles. By integrating these bubble indicators as exogenous variables in forecasting models, we can improve the accuracy and reliability of real estate price predictions.

Whitehouse, Harvey, and Leybourne (2022) developed complementary statistics for real-time bubble detection and monitoring, focusing on both the formation and collapse of bubbles. Their framework includes two key statistics: the $A(e,m)$ statistic for detecting explosive growth and the $S(e,m,n)$ statistic for identifying subsequent collapses.

$\boldsymbol{A(e,m)}$ **Statistic**: This statistic detects upward trends in the first differences of a time series during an explosive regime. By calculating the statistic over rolling sub-samples and comparing it to a critical value derived from a training sample, the $A(e,m)$ statistic effectively identifies periods of rapid price increases that signify speculative bubbles.

$\boldsymbol{S(e,m,n)}$ **Statistic**: The $S(e,m,n)$ statistic is designed to detect the transition from an explosive regime to a stationary collapse. It assesses the product of the means of the first differences before and after the suspected transition point, providing a robust indicator of market corrections following a bubble.

These statistics offer a comprehensive approach to bubble monitoring, enabling the detection of both the onset and the bursting of bubbles in real-time. Integrating the $A(e,m)$and $S(e,m,n)$ statistics into predictive models can enhance the forecasting of real estate prices by accounting for the dynamic nature of market sentiment and speculative behavior.

The PSY bubble indicators by Phillips, Shi, and Yu (2015) and the bubble statistics by Whitehouse, Harvey, and Leybourne (2022) represent significant advancements in the detection and analysis of speculative bubbles in financial markets. Their integration provides a comprehensive framework for understanding bubble dynamics that can be used in real estate price forecasting models. This research investigates the role of incorporating robust bubble detection methodologies in the real estate predicting model, paving the way for more informed decision-making and effective market regulation.

2.3 Housing Search Activity

Google Trends is a web-based tool that provides insights into the relative popularity of search queries over time. By analyzing search volume data, Google Trends can reflect the collective interest and behavior of internet users, making it a valuable proxy for latent market demand (Moller et al., 2023). The data from Google Trends is normalized, allowing for meaningful comparisons across different search terms and regions.

The use of Google Trends data in economic forecasting is grounded in the idea that search behavior reflects real-time public interest and concerns, which can be leading indicators of market demand. Several studies have demonstrated the predictive power of Google Trends data in various domains: Preis, Moat, and Stanley (2013) showed that changes in Google search volumes could predict stock market movements, as increased search activity often precedes significant market events. Choi and Varian (2012) found that Google search data could improve the accuracy of unemployment rate forecasts by capturing real-time job search behavior. Vosen and Schmidt (2011) demonstrated that Google Trends data could serve as a proxy for consumer confidence, reflecting public demand and economic conditions. These studies underscore the potential of Google Trends as a real-time indicator that can enhance traditional forecasting models.

Real estate markets are influenced by various factors, including economic conditions, public sentiment, and market expectations. Google Trends data can capture these elements by reflecting the search behavior related to housing market activities. For instance, an increase in searches for terms like "buying a house," "mortgage rates," or "real estate market" can indicate rising interest in the housing market, which may precede actual market movements.

Several studies have explored the use of Google Trends data in real estate market analysis: Wu and Brynjolfsson (2015) examined the predictive power of online search activity for house prices and sales, finding that search volumes could serve as early indicators of market trends. Dietzel, Braun, and Schafers (2014) integrated Google Trends data into commercial real estate forecasting models, demonstrating significant improvements in prediction accuracy. Moller et al. (2023) leveraged Google Trends data to construct a housing search index (HSI) that predicted housing price movements, highlighting the growing importance of digital footprints in economic forecasting. These studies provide a solid empirical foundation for using Google Trends data in real estate price prediction.

Incorporating Google Trends data into real estate price forecasting involves several steps. Data Collection: Identify relevant search terms related to the real estate market and collect their search volume data from Google Trends. Data Preprocessing: Normalize the search volume data to account for seasonal variations and trends. This can involve calculating moving averages or differencing the data to remove seasonality. Index Construction: Construct a composite index, such as the Housing Search Index (HSI), by aggregating the search volumes of multiple relevant terms. This index serves as a proxy for latent demand in the housing market.

The process of forecasting real estate prices using Google Trends data involves identifying the most relevant search queries indicative of market interest. Prospective buyers and sellers utilize a diverse array of search terms, and not all of these terms are equally reflective of price movements. Therefore, it is imperative to establish a methodical approach for selecting the most informative and pertinent search queries.

To construct a robust search index, it is necessary to blend a sample of Google Trends search terms that accurately capture the underlying search activity related to the real estate market. This involves the creation of multiple search indices, each derived from different sets of search queries. Given that different search terms are used at varying times leading up to a real estate transaction, constructing distinct indices allows for the capture of temporal variations in search behavior.

In summary, while macro-financial factors influence house prices, public economic misconceptions can lead to prices deviating from their fundamental values. Akerlof and Shiller (2010) emphasize that financial and economic analysts have often overlooked the significant impact of human psychology on economic decisions. Therefore, we utilize a variety of psychological indices such as news sentiment, bubble indicators, and Google Trends to account for this concern.

3. Data

In this section, we describe Canada's House Price Index retrieved from Teranet–National Bank. We also detail news articles related to the Canadian real estate market outlook obtained from Lexis Uni. Moreover, we explain different options that we can use to find news related to our topic. Finally, we dive into the different types and various search terms available in Google Trend search data. All three data categories mentioned are in monthly time intervals.

3.1 Teranet–National Bank House Price Index

There are a variety of methods for calculating HPIs (also called Residential Property PriceIndices – RPPIs). Simple mean or median methods usually track the median price of a property sold from one period to the next. Hedonic regression methods estimate the extent to which the characteristics of the property (e.g., size, appearance) and the characteristics of the surrounding environment (e.g., neighborhood) affect the market price of the property. Appraisal methods use the sale price as the base price; a new value (and increase ratio) is calculated based on an appraisal price for the property. Repeat sales methods use information on properties that have sold more than once, comparing the sale price at two (or more) periods in time. Each of these methods is subject to uncertainties, including unpredictable transaction times or cycles, differences across geographic locations and structures, and unknowns related to depreciation and renovation.

Developers of HPIs rely on public and/or private data collection and data sets because house-by-house or owner-based, housing-sale price surveys over time are unreasonable in terms of cost, time, and/or accuracy. In Canada, HPIs are based on two different data sets. Land-registry data: Administrative and legal data collected, stored, and shared via government-managed systems. Non-registry data: Data collected, stored, and shared by a private entity. An example of non-registry data is the sales data provided by the Multiple Listings Service® (MLS®). MLS data provides information on properties sold by accredited members of the Canadian Real Estate Association. Though no dataset is perfect, studies show that the use of land-registry data provides a more complete and accurate picture of transactional data (Costello and Watkins, 2002).

The Teranet-National Bank HPI[1] is based on the methodology developed by Bailey, Muth, and Nourse (1963) and extended by Case and Shiller (1987), whose work has significantly contributed to understanding housing price dynamics. This methodology is also employed by other notable indices globally, such as the Standard & Poor's/Case-Shiller HPI in the US, the Federal Housing Finance Agency's index in the US, the UK Land Registry's index, and Residex in Australia.

The Teranet-National Bank House Price Index (HPI) is an independent measure of single-family home prices across Canada. The index is based on the repeat-sales method of price measurement. It is derived from sale prices recorded in land registry data over a designated period. To be included in the index calculations, a property must have been sold at least twice, forming "sales pairs" that indicate the linear change in property value. The repeat-sales method tracks the sale prices of the same properties over time. This method helps to isolate the price changes of individual homes and provides a more accurate reflection of the housing market's movements. To avoid distortions from atypical sales, properties affected by internal factors are excluded from the calculations such as non-arm's-length sales, changes in the type of property, data errors, or high turnover frequency.

The data for the Teranet-National Bank HPI comes directly from the property records of public land registries without any third-party involvement. The index includes over 30 years of historical information from key residential markets across different property types. It includes all complete transactions recorded in the land registry systems, including private and published Multiple Listing Service (MLS) sales.

The Teranet-National Bank HPI for Canada is available from June 1990. Its HPI, the Composite 11 Index, includes Vancouver, Victoria, Calgary, Edmonton, Winnipeg, Hamilton, Toronto, Ottawa-Gatineau, Montreal, Quebec City, and Halifax. We use the monthly HPI of

[1] https://housepriceindex.ca/index-history/

Canada, which is not seasonally adjusted and spans 20 years, from January 2004 to December 2023.

3.2 Nexis Uni News Articles

There are several news aggregator databases, such as Nexis Uni, Factiva, ProQuest, and Access World News. Nexis Uni is a comprehensive news aggregator database provided by LexisNexis. It is designed to support academic research by offering access to 17,000 sources, including newspapers, magazines, legal documents, and business publications. The platform indexes various types of records, such as text articles, scanned newspaper pages, photos, and more. The coverage includes international, national, regional, and local titles.

When compared to other news aggregator databases such as Factiva, U.S. Newsstream (ProQuest), and Access World News (NewsBank), Nexis Uni exhibits both strengths and limitations: Factiva offers the most extensive content coverage with over 30,000 indexed sources, but imposes stricter download limits. U.S. Newsstream (ProQuest) provides substantial current full-text coverage of top circulating newspapers and regional papers. Access World News (NewsBank) specialises in regional papers and offers multimedia content such as podcasts and videos. Nexis Uni allows for saving searches and downloading a significant number of articles, more than Factiva but less than U.S. Newsstream and Newspaper Source Plus (EBSCO). Factiva and U.S. Newsstream both offer APIs for large-scale data extraction, though these features may incur additional costs. Overall, while Nexis Uni may not have the highest number of sources or the most generous download limits, its comprehensive content coverage, advanced search functionalities, and strong user support make it a highly valuable tool for academic research

Upon executing a search on Nexis Uni, the service automatically decides whether to process it as plain language or as a Boolean search. Plain language searches allow us to use natural language, entering terms, questions, or key phrases relevant to the topic, much like how we would in a regular conversation (e.g. "Real estate perspective in Canada"). On the other hand, if we want to define the relationship between our search terms, we can use a terms and connectors search (Boolean search). Connectors we can use include *and, or,* and *not*, as well as proximity connectors such as *w/n*. For example, we could enter "Real estate w/5 perspective or prediction and Canada".

We decided to conduct a Boolean search using proximity connectors for two reasons. First, the plain language search doesn't restrict the articles to those that have the search terms and tries to find the relevant news. Then, it returns the news sorted by relevancy. We found that the news articles in results with the least relevancy are almost irrelevant. Second, this method results in a large number of news, which makes it hard to process and extract the sentiment.

We narrowed the results to news articles in the real estate industry and picked those whose content geography was Canada. We didn't have any restrictions on the publication location, though the news articles published outside Canada were about 1% of the results. The type of news source covered a variety of options, such as Newspapers, Newswires & Press Releases, Web-based Publications, Industry Trade Press, Magazines & Journals, News, Blogs, Newsletters, News Transcripts, Business Opportunities, Aggregate News Sources, Law Reviews & Journals, Market Research Reports.

We got 3,960 News articles from Lexis Uni from January 1, 2010, to October 31, 2023. There were 2,833 news articles with unique titles among them. The average and maximum number of tokens in news articles were 1,120 and 6,964 respectively. The search term we used is "(real estate or house price or home price) pre/5 (forecast or outlook or predict or anticipate or perspective)". This means that the news article title or body should have one of "real estate", "house price", or "home price", and in less than 5 words after, it should have one of "forecast", "outlook", "predict", "anticipate", or "perspective". We also excluded the stock market reports that mainly discussed stock price changes in the previous days and stock information releases. It is worth noting that Nexis Uni is not sensitive to capitalization. The filter settings we used in this research are in the appendix.

3.3 Google Trends Search Data

Google Trends is a public web facility of Google Inc., based on the world's largest search engine, that shows how often a particular search term is entered across various regions of the world and over different time frames. The search data is normalized to make comparisons between terms and across regions more meaningful. The values are presented as percentages of the highest point of search activity during the selected period and region on a range of 0 to 100. This means that a value of 100 represents peak popularity, while a value of 50 means the term is half as popular. The tool breaks down search data by regions, subregions, cities, and metros, allowing for geographic comparison of interest levels. Search terms are categorized into topics to enhance the relevance and accuracy of trend analyses. Google Trends groups similar search terms and identifies related searches to provide a comprehensive view of interest in broader subjects. Google Trends data can be filtered by different types of searches, such as web searches, image searches, news searches, Google Shopping, and YouTube searches.

Google Trends is widely used in various fields of research, including economics, sociology, epidemiology, and market research. Google dominates the Canadian search engine market with a 92.4% market share as of April 2023 (Statista 2023). Its ability to provide real-time data on public interest makes it a valuable tool for understanding social phenomena, predicting economic trends, and analyzing consumer behavior. By leveraging Google Trends, researchers can obtain a dynamic and immediate view of how interest in specific topics evolves over time and across different regions, offering a powerful complement to traditional data sources.

While Google Trends provides powerful insights, it is important to acknowledge its limitations. The data represents a sample of Google searches and may not capture all search activity. It may also be subject to biases inherent in search behavior. Moreover, High search volumes do not necessarily indicate high demand; they reflect interest or curiosity, which may be driven by various factors, including news events, cultural phenomena, or significant occurrences. It is crucial to interpret the data in the context of the specific search terms and market conditions. In addition, according to D'Amuri and Marcucci (2017), Google Trends are based on a sample of queries that vary depending on the time and IP address used to download the data. This variation may cause an issue for reproducibility. In order to address sampling error, Moller et al. (2023) calculated the index for all Google Trends queries using an average of over 15 different days. The correlation across different samples was consistently above 0.99. Therefore, the results are robust and reliable for practical purposes.

Google Trends data can be refined by various parameters, including time, region, category, and search environment. Additionally, the data can be downloaded for specific search terms as well as predefined search topics. The available data spans from 2004 to the present. For this study, our sample period extends from January 2004 to December 2023, with data aggregated on a monthly basis. The geographical focus of our analysis is Canada, and we utilized the "All categories" filter to ensure comprehensive coverage. Furthermore, the search environment was specified as "Web Search."

Moller et al. (2023) developed a housing search index (HSI) extracted from online search activity on a limited set of keywords related to the house-buying process for the US market. To obtain a measure of housing demand, they initially used “buying a house” as the main search term and subsequently obtained a list of 22 related terms. They excluded three remaining related search terms either because they are unrelated to housing (“buying a car”) or because the search volume is low. They define low-volume series as those for which more than 10% of observations equal zero. Like Moller et al. (2023), we exclude low-volume series.

Out of their 24 related search terms, we use 10 whose volume in Canada is not low: “when buying a house,” “buying a home,” “buy a house,” “mortgage,” “how to buy a house,” “real estate,” “homes for sale,” “building a house,” “mortgage calculator,” and “houses for sale.”[2] Moreover, we add 9 new search terms obtained in Canada in the same way that are not in the mentioned search terms: "buying a house," " buying a condo," " buying a house in Canada," "closing costs," "house for sale," "land transfer tax," " mortgage calculator Canada," " mortgage rates," and " selling a house."

Moller et al. (2023) constructed an HSI that focuses on the buying side of the housing market through the chosen keywords. Accordingly, they interpret the search index as a proxy for latent demand. Wu and Brynjolfsson (2015) also consider the use of online search activity to predict house prices and sales. Instead of using specific keywords, they consider predefined search topics supplied by Google Trends, namely, “real estate agencies” and “real estate listings.” Google

[2] The 14 excluded terms are “help buying a house,” “buying a house cash,” “buying a new house,” “before buying a house,” “steps to buying a house,” “buying a house calculator,” “first time buying a house,” “buying a house process,” “house buying process,” “buying a house with bad credit,” “cost of buying a house,” “buying a house to rent,” “buying a house tips,” and “buying a foreclosure house.”

classifies search queries into topics using an undisclosed natural language classification engine (Choi and Varian 2012), and it is unclear how we should interpret these topics other than that they relate to the topic given by the name of the topic. Wu and Brynjolfsson (2015) find that these two search topics hold some predictive power for future house prices and also that prices are more difficult to predict than house sales.

We also use Wu and Brynjolfsson's (2015) predefined search topics and add 3 new search topics related to our target: "Mortgage loan," "Real Estate," and "Real Estate Broker". Moreover, we use 3 predefined real estate companies to construct an index for searching for brokers and online listings: "RE/MAX," "Royal LePage," and "Realtor.ca". We add their values to make a new index since the search for brokers and new listings can be substituted for each other.

In the Google Trends panel, the category filter allows us to narrow down the search to a specific area of interest. This helps in refining the data to show trends that are more relevant to a particular topic. The categories range from broad topics like "Business" or "Health" to more specific ones like "Real Estate" or "Fitness." By selecting a category, we can see how search interest varies within that particular context. We obtained all previous data by category set to "All categories". In order to get proxies for supply and demand in the real estate market, we add the search volume data for "buy" and "sell" to our data set while we set the category option to "Real Estate".

To make a comprehensive dataset, we also asked ChatGPT to give us the most searched terms for people who are trying to buy a home. 3 search terms of what it returned were new and not low volume: "buy house," "home buy," and "real estate agents."

Finally, our search volume data set consisted of 30 search volume series. It is worth mentioning that Google Trends data is not sensitive to capitalization. These search terms are all related to the home-buying process and, as such, should serve as a proxy for housing demand.

4. Methodology

In this section, we first demonstrate the auto-regressive model we use to forecast HPI. Second, we illustrate the process of extracting market sentiment from news articles and mapping it to a numeric index. Third, we detail the computation of the SADF and BSADF test statistics proposed by Phillips, Shi, and Yu (2015, PSY hereafter) to generate HPI exuberance levels. Fourth, we describe $A(e, m)$ and $S(e, m, n)$ statistics developed by Whitehouse, Harvey, and Leybourne (2022, WHL hereafter) to detect bubbles in the real estate market. Fifth, we describe the process of calculating the House Search Index (HSI) from Google Trends data. Finally, we compare different forecasting models by the forecast-encompassing test.

4.1 Forecasting Model

The forecasting model can be specified as an autoregressive model with exogenous variables (ARX), where the real estate price index is regressed on its lagged values and the exogenous variable lagged values. The reason is that we aim to improve the forecasting power of a simple AR model by adding the exogenous variable. This specification allows the model to leverage historical price data and predictor indices to predict future price movements. Our model also includes constant and seasonal dummies. We forecast the HPI for 1, 3, 6, and 12 months ahead.

$$(1) \qquad Y_{t+h} = c + \sum_{i=0}^{m-1} \alpha_i * Y_{t-i} + \sum_{i=0}^{n-1} \beta_i * X_{t-i} + \sum_{i=1}^{12} \gamma_i * d_{i,t+h} + \varepsilon_{t+h},$$

where Y_{t+h} is the forecasted HPI (or preprocessed HPI), Y_{t-i} represents lagged values of the HPI (or preprocessed HPI), X_{t-i} includes sentiment indices, $d_{i,t+h}$ are the month dummies, and ε_{t+h} is the error term. h is the forecast horizon, m is the number of HPI lags, and n is the number of exogenous variable lags. We pick the first 80% of the data to estimate the model and the last 20% for validation.

4.2 News Sentiment

News articles are a rich source of information that can reflect market sentiment. By analyzing the tone and content of news articles, it is possible to construct sentiment indices that capture the public's perception of the real estate market. These indices can then be used as predictors in forecasting models. The process of deriving sentiment indices from news involves several steps. Data Collection: Gathering a large corpus of news articles relevant to the real estate market. This can be achieved using news aggregator databases such as Nexis Uni, Factiva, or similar platforms. Text Preprocessing: Cleaning and preparing the text data for analysis. This includes removing stop words, punctuation, and irrelevant content, as well as tokenizing the text into meaningful units. Sentiment Scoring: Applying sentiment analysis techniques to assign sentiment scores to the text. This can be done using lexicon-based approaches, machine learning models, or deep learning techniques. The scores typically range from negative to positive, indicating the sentiment expressed in the text. Index Construction: Aggregating the sentiment scores over a specified time period to create a sentiment index. This index reflects the overall sentiment trend in news articles and can be used as an exogenous variable in forecasting models.

Various techniques can be employed to analyze sentiment in news articles. Lexicon-based approaches use predefined dictionaries of positive and negative words to score text, while machine learning models can be trained on labeled datasets to classify sentiment. More advanced methods, such as deep learning models, leverage neural networks to capture complex patterns in the text. Finally, we can use LLMs to extract the sentiment score from news text.

In this section, we detail the process of extracting a numeric sentiment score from news text. In this order, we used the 'gpt-3.5-0613' model of OpenAI to implement this procedure. We embed the news in a prompt and ask the model to return if the news is bullish or bearish.

3,960 news articles are downloaded from Nexis Uni in Word format. We first extract the news article's title, date, and body. Then, for each news article, we embed this information in the prompt template and send it to the model. We ask the model to return its answer in a JSON format so it is easy to interpret. Our prompt consists of roles 'system' and 'user.' Based on OpenAI guidelines for prompting, we clearly specify the steps the model should take, including expressing the news sentiments in up to five sentences and then returning a label as the news sentiment. We also determine the expected output format for the model. The prompt roles and their contents are in the appendix.

Some news articles usually present different viewpoints for short-term and long-term horizons. Their timing format to express short-term and long-term is usually the first half and second half of the year or this year and next year. To simplify the model's task and get robust results, we specify the short-term and long-term dates in the prompt rather than letting the model decide about timing short-term and long-term horizons. We specify these two dates based on the news's publication date and ask the model to extract the news opinion for each date. If the news is published in the first quarter, the short-term date is the first half of the publication year, and the long-term date is the second half of that year. If the news is published in the second and third quarters, the short-term date is the publication year, and the long-term date is the next year. Finally, if the news is published in the fourth quarter, the short-term date is the first half of the year after the publication year, and the long-term is the second half of that year.

Before feeding the news body to the model, we must ensure that the number of text tokens and the prompt does not exceed the model's token window. We subtract the number of template prompt tokens from the model's token limit to find the maximum feasible number of tokens for the news text. Then, we split the news text into small chunks, each chunk equal to a sentence. To find the most relevant sentences, we need to specify a topic and calculate the embedding vector for the topic and each sentence. Then, we calculate the distance of each sentence embedding vector to the topic embedding vector. The sentences with embedding vectors closer to the topic embedding vector are the more relevant sentences. In this regard, we choose 'future real estate price' as the topic. Then, we use the OpenAI embedding model 'text-embedding-ada-002' to calculate the embedding vectors for news sentences and the topic. We start with the most relevant sentences and pick sentences until reaching the specified maximum feasible number of tokens for the news text. Finally, we use these sentences as news text in our prompt.

To facilitate extracting the news sentiment, we ask the model to choose a predefined qualitative label as the news sentiment rather than asking to return a numeric sentiment score directly. Then, we map these labels to a sentiment score. To address hallucinations, we give the model the option to return that the news is irrelevant to the topic. The qualitative terms we provide the model to choose from are: 'strongly negative,' 'weakly negative,' 'neutral,' 'weakly positive,' 'strongly positive,' and 'irrelevant.' Then, we map the relevant terms to -2, -1, 0, 1, and 2, respectively, and drop the news with an 'irrelevant' label.

Finally, we take the monthly average of news sentiment scores and create a sentiment index to use as a predictor in the forecasting model. We calculate two series from each short-term and long-term sentiment index to account for the different impacts that different news articles may have on their audience. Some news articles are mirrored in different news outlets. This can either show that they are important or republishing the news can have more impact on the audience. The news article text size is another factor that can be a potential source of the heterogeneous impact of news. We use these two features to calculate two monthly weighted average sentiment indices. For each news article i and its sentiment S_i, we calculate new sentiment scores $S1_i$ and $S2_i$ as follows:

$$S1_i = \sqrt{F_i} \times S_i$$

$$AN_i = \frac{\left(\sqrt{N_i} - \sqrt{N_{min}}\right)}{\left(\sqrt{N_{max}} - \sqrt{N_{min}}\right)} \times 0.1 + 1$$

$$S2_i = AN_i \times S_i$$

where F_i is frequency of news i's publication, N_i is the number of news tokens, and AN_i scaled value of N_i. We add a constant 10 to AN_i to level up the weight of news with tokens near N_{min} from zero. Finally, we construct an index by taking the average of short-term and long-term sentiment scores.

4.3 The PSY Test

The Backward Supremum Augmented Dickey-Fuller (BSADF) test is an extension of the Supremum Augmented Dickey-Fuller (SADF) test, which is used to identify the presence of explosive behavior or bubbles in time series data. The BSADF test enhances the flexibility and robustness of the SADF test by using a backward recursive approach, allowing for the detection of multiple bubbles within a single time series. We outline the steps and formulas involved in calculating the BSADF test statistic proposed by Phillips, Shi & Yu (2015).

The Backward Supremum ADF (BSADF) and Supremum ADF (SADF) tests are based on the Augmented Dickey-Fuller (ADF) unit-root test, which is in turn based on the regression:

$$\Delta y_t = a_{r_1,r_2} + \beta_{r_1,r_2} y_{t-1} + \sum_{j=1}^{k} \psi_{r_1,r_2}^{j} \Delta y_{t-j} + \varepsilon_t.$$

In the context of this model, y_t is a univariate time series, k denotes the number of auto-regressive lags, and ε_t is an independently and identically distributed (iid), normally distributed error term with standard deviation σ_{r_1,r_2}. The interval $[r_1, r_2]$ (where r_1 and r_2 are within the range [0,1]) specifies the portion of the sample used to compute the ADF statistic. For a sample with periods ranging from 0 to T, the $ADF_{r_1=n/T}^{r_2=m/T}$ statistic is based on a subset of periods ranging from n to m, where n and m are within the range $\{0,\ldots,T\}$ and n is less than m. The ADF test statistic is defined as:

$$ADF_{r_1}^{r_2} = \frac{\hat{\beta}_{r_1,r_2}}{\sigma_{\hat{\beta}_{r_1,r_2}}}.$$

The SADF test, introduced by Phillips, Wu, and Yu (2011), is a right-tailed unit root test applied recursively over a sequence of forward expanding windows. The test statistic is the supremum value of the Augmented Dickey-Fuller (ADF) statistic over these windows. The SADF test is defined as:

$$SADF(r_0) = \sup_{r_2 \in [r_0,1]} ADF_0^{r_2},$$

where r_1 and r_2 represent the starting and ending points of the time window, respectively, and $ADF_{r_1}^{r_2}$ is the ADF statistic calculated for the window from r_1 to r_2.

The BSADF test extends the SADF test by incorporating a backward recursive mechanism. For each endpoint r_2, apply a backward recursive mechanism by calculating the ADF statistic over windows ending at r_2 and starting from various points before r_2. This yields a sequence of ADF statistics for different starting points. The BSADF statistic is the supremum value of the ADF statistics obtained from the backward recursive mechanism. It is defined as:

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

where the minimum window size is indicated by r_0. By designating r_2 as the current period (the t-th period corresponds to $r_2 = t/T$) and allowing the start of our estimating period r_1 to fluctuate between the beginning of our sample (0) and $r_2 - r_0$, this approach can generate real-time exuberance levels. Then, for HPI, we can recursively generate a number of BSADF statistics.

The BSADF test statistic is used to identify periods of explosive behavior or bubbles in the time series. If the BSADF statistic exceeds the critical value from the right-tailed unit root distribution, it indicates the presence of an explosive period within the time series. Multiple such periods can be identified by applying the BSADF test recursively over different sub-samples of the data.

The BSADF test provides a more flexible and powerful method for detecting bubbles in time series data compared to the traditional SADF test. By allowing for multiple starting points and applying a backward recursive mechanism, the BSADF test can identify periods of explosive behavior more effectively, making it a valuable tool in econometric analysis (PSY, 2015).

4.4 The WHL Test

The detection of financial bubbles and crashes is critical for mitigating economic damage. Whitehouse, Harvey, and Leybourne (2022) propose real-time monitoring procedures to detect such phenomena. This document focuses on the calculation of two key statistics: $A_{(e,m)}$ and $S_{(e,m,n)}$, which are integral to identifying explosive and stationary regimes in time series data.

The $A_{(e,m)}$ statistic is employed to detect an explosive regime within a time series. It is based on the Taylor series expansion of the first differences during the explosive regime and tests for the presence of an upward trend in these differences. Let k denote the window width over which the statistic is computed and e the last observation used. The $A_{(e,m)}$ statistic is given by:

$$A(e,k) = \frac{\sum_{t=e-k+1}^{e} (t - e + k) \Delta y_t}{\sqrt{\sum_{t=e-k+1}^{e} (t - e + k) \Delta {y_t}^2}},$$

where Δy_t represents the first differences of the time series y_t.

In practice, the $A_{(e,m)}$ statistic is computed over rolling sub-samples of length k within an initial training sample. The maximum of these training sample statistics forms the critical value $A_{max,train}$:

$$A_{max,train} = \max_{e \in [k+1,T]} A(e,k)\,,$$

where T is the length of the training sample. Monitoring for an explosive regime begins at time $t = T + k$, and detection occurs if:

$$A(\mathrm{t},k) > A_{max,train}$$

The $S(e,m,n)$ statistic is used to detect the transition from an explosive regime to a stationary collapse regime. It leverages the differing signs of the means of the first differences in the explosive and stationary processes. Specifically, it assesses the product of these means over two sub-samples around the suspected transition point.

Consider the model expressed in the first differences near the endpoint of the explosive regime:

$$\Delta y_t = \begin{cases} \beta_1 + \varepsilon_t \; for \; t \in [\tau_2 T - m + 1, \;\; \tau_2 T] \\ \beta_2 + \varepsilon_t \; for \; t \in [\tau_2 T + 1, \;\; \tau_2 T + n] \end{cases}$$

Here, β_1 and β_2 are the means of the first differences before and after the transition, respectively. The statistic is then given by:

$$S(e,m,n) = \frac{1}{m} \sum_{t=e-n-m+1}^{e-n} \Delta y_t \cdot \frac{1}{n} \sum_{t=e-n+1}^{e} \Delta y_t$$

This statistic is standardized to account for possible variance changes, resulting in:

$$S(e,m,n) = \frac{\left(\sum_{t=e-n-m+1}^{e-n} \Delta y_t\right) \cdot \left(\sum_{t=e-n+1}^{e} \Delta y_t\right)}{\left(\sum_{t=e-n-m+1}^{e-n} \Delta y_t^2\right)^{1/2} \cdot \left(\sum_{t=e-n+1}^{e} \Delta y_t^2\right)^{1/2}}$$

The critical value for this statistic, $S_{min,train}$, is the minimum of the training sample statistics:

$$S_{min,train} = \min_{e \in [T-m-n+1,T]} S\,(e, m, n)$$

Monitoring for a stationary collapse begins after the detection of an explosive regime. A collapse is detected if:

$$S(e, m, n) < S_{min,\text{train}}$$

Whitehouse et al. demonstrate the efficacy of these statistics through simulations and empirical application to the US housing market. The $A(e, m)$ statistic effectively detects bubbles, while the $S(e, m, n)$ statistic identifies subsequent crashes. The flexibility of these procedures, particularly the user-chosen parameters m and n, allows practitioners to balance between rapid detection and false positive rates.

The $A(e, m)$ and $S(e, m, n)$ statistics provide robust tools for real-time monitoring of bubbles and crashes in financial time series. Their implementation, leveraging training samples for calibration, ensures both sensitivity and specificity in detecting critical regime changes, thereby aiding policymakers in timely interventions. To calculate $A(e, m)$ and $S(e, m, n)$ indices, we need to determine m and n values. We use $m = 10$ and $n = 2$ as the authors suggest that they will be suitable for many senarios.

Based on Figures 7 and 8 in the Appendix, the PSY indicators (PSY_SADF and PSY_BSADF) generally show smoother trends with clear peaks, indicating periods of potential bubble activity. The PSY_BSADF line tends to be more volatile than the PSY_SADF, capturing more pronounced peaks, particularly noticeable around 2008 and post-2015. In contrast, the WHL indicators (WHL_A and WHL_S) exhibit much higher volatility and frequent oscillations. This suggests that the WHL methodology is more sensitive to detecting short-term fluctuations. The WHL_S indicator, in particular, shows extreme variability, frequently crossing above and below the threshold line.

Overall, while both sets of indicators aim to identify bubble periods, the PSY indicators provide a more stable and gradual representation, potentially capturing longer-term bubble dynamics. The WHL indicators, with their high volatility, seem to be more responsive to short-term market movements. The choice between these methods may depend on whether the focus is on capturing broad, sustained bubbles or detecting frequent, short-term deviations.

4.5 House Search Index (HSI)

In the following, we construct multiple versions of the House Search Index (HSI) and select the best one to use as an exogenous variable in the forecasting model. In this regard, we get 30 relevant search query data from Google Trends. They include those suggested by Moller et al. (2023) and what can be obtained by following the paper's approach for Canada. The data contains monthly observations from 2004-01 to 2023-11. We only use data until 2019-12 to preprocess the data and train the model.

Some of the related terms may have more noise in their measurement than others. To eliminate the noise and more precisely estimate latent demand, Moller et al. (2023) employ a targeted principal component analysis (PCA) method that ensures only the most relevant search indices are included in calculating the latent demand factor. Specifically, their approach follows Bai and Ng (2008) as they use the elastic net estimator of Zou and Hastie (2005) to select the 10 most relevant search indices and then apply principal component analysis to summarize the most important information from these indices into one common component. Moller et al. (2023) interpret this principal component as a summary measure for housing search and refer to it as the HSI.

Moller et al. (2023) admit their main goal is to produce a simple and easy-to-interpret index of housing search, which is why they used a simple targeted PCA approach. Moreover, they claim the predictive results that they reported are generally highly robust to using more advanced machine learning techniques. Like Moller et al. (2023), we adopt the targeted principal component analysis (PCA) method to calculate the HSI. While Moller et al. (2023) construct one HSI, we construct various potential HSIs from different lags and combinations of the Google Trend series and select the best HSI for forecasting purposes.

The literature has no consensus regarding whether Google Trends data should be characterized as stationary, trend stationary, or possessing a unit root. This determination can be highly dependent on the specific query being analyzed. Vozlyublennaia (2014), Choi and Varian (2012), Bijl et al. (2016), and D'Amuri and Marcucci (2017) did not perform any differencing or detrending of the series, indicating that the Google Trends data they used are stationary. In contrast, Da et al. (2015) analyzed the log-differences (growth rates) of their data. To account for the possibility that the individual Google Trends series could follow different trends, Moller et al. (2023) adopted a sequential testing strategy in the spirit of Ayat and Burridge (2000) and similar to Borup and Schutte (2022). They further removed seasonality by regressing each series on monthly dummy variables and studied the residuals from this regression.

To find the best HSI, we take the following steps: First, we preprocess Google Trends search volume data and the HPI, including removing seasonal and trend factors. Second, we compute several HSIs by extracting the first principal component of search indices volumes. Third, we find the best lags for each HSI in the forecasting model using information criteria BIC. Finally, we pick the best HSI based on the cross-validation MSE in the forecasting model over the training sample.

4.5.1 Preprocess HPI

We first convert the HPI series to its natural logarithm to study its proportional changes. It is well-known that house prices display strong seasonal variation, with high prices during spring and summer and low prices during fall and winter. Furthermore, we remove seasonality and trend by regressing each series on monthly dummy variables and a linear time trend factor. Then, we study the residuals from this regression.

$$(2) \quad Y_t = c + \alpha * t + \sum_{i=1}^{12} \beta_i * d_{it} + \varepsilon_t \longrightarrow Y_t^{new} = \varepsilon_t$$

4.5.2 Preprocess Google Trends data

We transform the search indices as follows: First, to comply with changes in the HPI, we remove seasonality and trend in the same way.

$$(3) \quad X_t = c + \alpha * t + \sum_{i=1}^{12} \beta_i * d_{it} + \varepsilon_t \longrightarrow X_t' = \varepsilon_t$$

While HPI varies over time, search volume data are very volatile. To address this issue, we smooth search volumes by converting them into their 12-month moving average.

$$(4) \quad X_t^{new} = 1/12 * \sum_{i=0}^{11} X'_{t-i}$$

4.5.3 Compute the House Search Index (HSI)

Computing HSIs consists of several steps and multiple layers. We constitute a pool of input data, naming it A_k in layer k. We initiate A_1 as a set of all search indices and their lags up to 3 months. This allows us to match search indices with different lags since a typical home buyer searches for different queries at different times before buying a home.

In layer k, we adopt the following procedure. First, for each forecast horizon h, we try to predict HPI by regressing HPI on each series X in A_k. Then, we compute the cross-validation MSE over the training sample. In this regard, we shuffle data and compute the 5-fold cross-validation MSE equation (5). Because lagged values are treated as distinct independent variables, shuffling the time series does not disrupt the lag relationship at each individual estimation point in the autoregressive model. We repeat this procedure 20 times and then take the average MSE and call it CV_{ih}.

$$(5) \quad Y_{t+h} = c + \beta * X_t + \varepsilon_{t+h}$$

Second, for each forecast horizon h, we assign a weight (W_{ihk}) to each series i in A_k based on its cross-validation MSE (CV_{ih}). The weights are different for each forecast horizon and are computed as follows.

$$(6) \quad W_{ihk} = \frac{\left(\frac{1}{CV_{ih}}\right)^4}{\sum_{i \in A_k} \left(\frac{1}{CV_{ih}}\right)^4}$$

Third, for each forecast horizon h, we pick 10 series from A_k using a weighted random draw (W_{ihk}). To avoid combining different lags of one search index to make a new HSI, we only consider the one with the lowest cross-validation MSE (CV_{ih}) from different lags of a search index. Then, we compute the HSI by calculating the first principal component of these 10 series. We repeat the selection and compute the new HSI 60 times for each forecast horizon. So, we would have 60 HSIs for each forecasting horizon and 240 HSIs at all.

4.5.4 Repeat Step 4.5.3

Finally, we check if the minimum CV_{ih} for i in A_k is lower than that for i in A_{k-1} for all forecast horizons. If so, we make A_{k+1} by adding 240 new HSIs to A_k and then start a layer $k+1$. Otherwise, we stop generating new HSIs and proceed to the next step. In layer $k = 1$, new HSIs are built only from the search indices, but in layers $k \geq 2$, new HSIs are built from the search indices and old HSIs that are constructed in previous layers.

After we finish creating HSIs, we define a set of best candidate predictors for each forecast horizon. In this regard, we constitute a set B_h by picking 240 series from A_k in the last layer with the lowest CV_{ih} for each forecast horizon h. From now on, we will only work with B_h sets, and we won't need A_k.

4.5.5 Find the Optimal Lag Structure

To select the best predictor to use in the forecasting model, we first need to specify the lags in the forecasting model (1). We find the best combination of HPI lags and HSI lags by Bayesian Information Criterion (BIC) for each candidate predictor in B_h. The lags are considered consecutively, and the maximum lag is 3.

4.5.6 Select the Best HSI

After specifying the lags of each predictor of B_h in equation (1), we find the best predictor for forecasting purposes based on cross-validation MSE in the forecasting model. So, like before, we shuffle data and compute the 5-fold cross-validation MSE 20 times, but this time in the forecasting model (1) rather than equation (5). These cross-validation MSEs are in the complete setup of the forecasting model with the lags of HPI and the predictor. Then, we take the average of 20 MSE and compute the mean MSE. Finally, we calculate the ultimate HSI for each horizon by computing the first PC of 10 HSI that have the lowest mean MSE in their full setup forecasting model.

Although each HSI is intended for a specific horizon, we consider all candidate HSIs and search indices in B_h for each forecasting horizon h. So, one of the search indices may be selected as the best predictor.

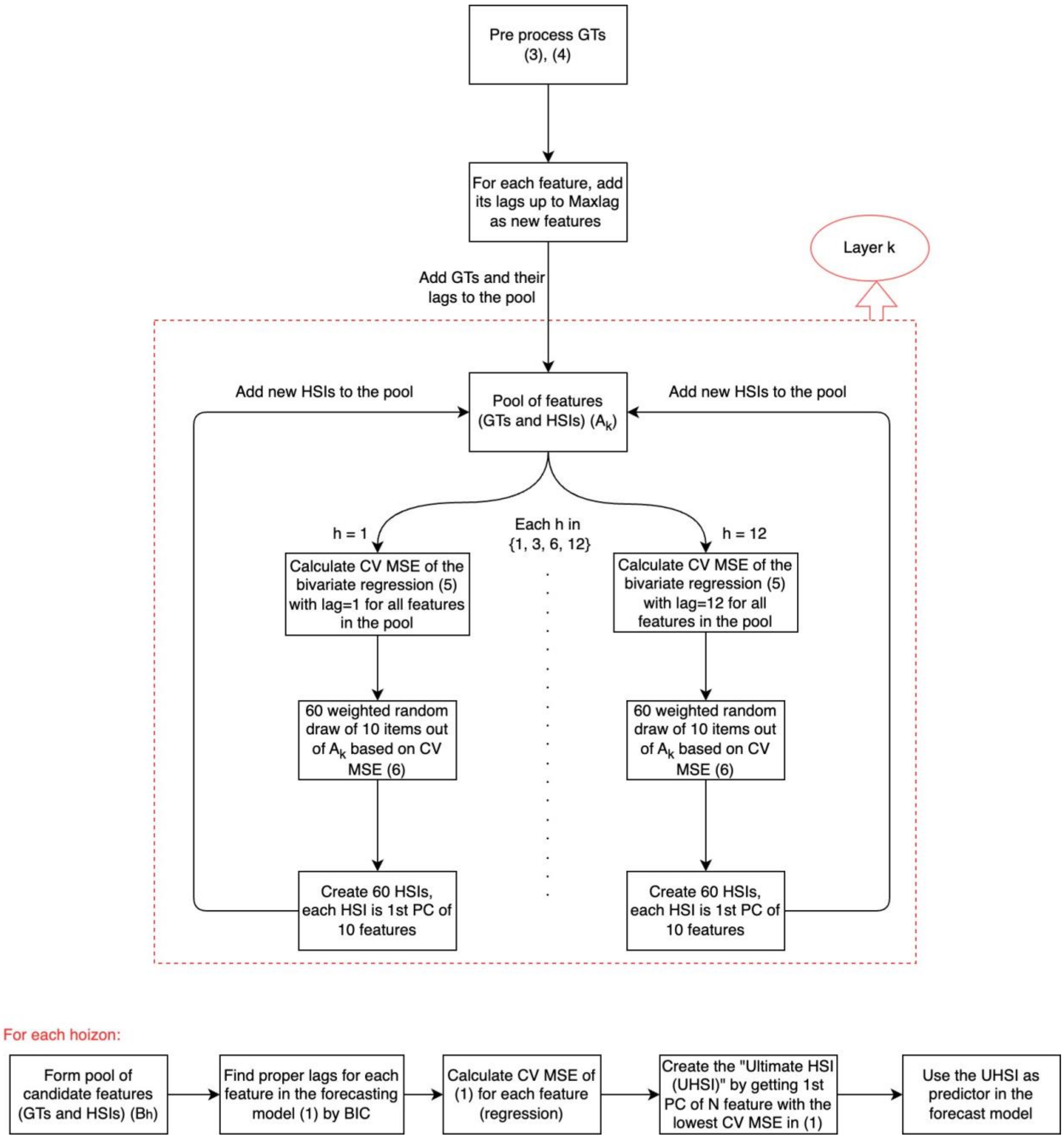

Figure 1: The process of creating UHSIs. This flowchart illustrates the process of generating Ultimate Housing Search Indices (UHSIs) and their integration into the forecasting model. The process begins with the preprocessing of Google Trends (GTs) data, followed by the addition of lagged features to the pool. HSIs are then created through weighted random draws from the pool of GTs and their lags, followed by bivariate regression across multiple forecast horizons (h = 1, 3, 6, 12). The HSIs are generated as the first principal component of the selected features. Layers are iteratively added to the procedure until the minimum cross-validated (CV) error of the HSIs in the most recent layer begins to increase, indicating the optimal depth of the model. Finally, the "Ultimate HSI" (UHSI) is derived from a set of best HSIs for each horizon h (B_h) and used as a predictor in the final forecasting model.

4.6 Statistical Tests

To assess the forecast accuracy and encompassing ability of competing models, we employ the Modified Diebold-Mariano (MDM) statistic developed by Harvey, Leybourne, and Newbold (1998). The MDM test, an extension of the original Diebold-Mariano (DM) test, is particularly suited for comparing the predictive accuracy of two competing forecasts, allowing for adjustments in finite sample distributions. This methodology ensures a robust comparison by considering the potential autocorrelation and heteroskedasticity in forecast errors.

The following steps outline the process: First, we generate out-of-sample forecasts from two competing models, M_A and M_B, for the variable of interest (e.g., housing prices). We ensure that the forecasts cover the same time horizon. The errors of the two models are calculated as e_t^A and e_t^B. Then, for the null hypothesis $H_{0,1}$ that model M_A and model M_B both have the same MAFE we calculate d_t as follows:

$$d_t = |e_t^A| - |e_t^B|$$

Similarly, for the null hypothesis $H_{0,2}$ that model M_A and model M_B both have the same MSFE we calculate d_t as follows:

$$d_t = (e_t^A)^2 - (e_t^B)^2$$

Finally, for the null hypothesis $H_{0,3}$ that model M_A forecast encompasses model M_B, we calculate d_t as follows:

$$d_t = (e_t^A - e_t^B)e_t^A$$

The test statistic MDM is then defined as:

$$\bar{\mathrm{d}} \equiv \tau^{-1} \cdot \sum_{t=T+1}^{T+\tau} \mathrm{d}_t$$

$$\sigma_{\mathrm{d}} \equiv \sqrt{\tau^{-1} \cdot \sum_{j=-(h-1)}^{h-1} \sum_{t=|j|+T+1}^{T+\tau} \mathrm{d}_t \cdot \mathrm{d}_{t-|j|}}$$

$$\mathrm{DM} \equiv \frac{\bar{\mathrm{d}}}{\sigma_{\mathrm{d}}}$$

$$\mathrm{MDM} \equiv \mathrm{DM} \cdot \sqrt{\frac{\tau + 1 - 2 \cdot h + h \cdot (h-1)/\tau}{\tau}}$$

where h is the forecast horizon, τ is the number of forecasts, T is the length of the estimation sample, σ_{d} is the standard deviation of d, and DM is the Diebold-Mariano statistic. MDM here follows $t(\tau - 1)$ distribution.

5. Results and Analysis

In this section, we evaluate the forecasting power of the selected predictors. We establish an autoregressive model without any exogenous variables as the base model. In Sections 5.1, 5.2, and 5.3, we assess the impact of incorporating each predictor into the base model. By comparing models that include different predictors to the base model separately, we are able to utilize the maximum available data length for each data category. For instance, the bubble indicators cover nearly twice the data range as the sentiment indices.

First, we predict the HPI using an autoregressive model with sentiment indices as the exogenous variables. Next, we apply the same model but substitute the sentiment indices with bubble test statistics as the exogenous variables. Finally, we use UHSIs derived from Google Trends for forecasting. These indices are based on the HSI introduced by Moller et al. (2023), who demonstrated that HSI effectively forecasts detrended and deseasonalized HPI. Accordingly, we apply the same preprocessing techniques to HPI in the third section.

In Section 5.4, we conduct a comparative analysis of the models incorporating different predictors by directly comparing them against one another. This approach involves designating certain predictors as the base model and evaluating whether the other predictors can outperform them. To ensure consistency, we use the same dependent variable, the same training period, the same validation sample, and the same lag selection method across all comparisons. Specifically, we utilize the detrended and deseasonalized HPI employed for the UHSIs.

Given that the validation sample and various segments of the training sample exhibit differing levels of volatility and trends, we select the same training sample for estimating the parameters of the forecasting model (Equation 1). For instance, the validation sample is characterized by higher volatility and a significant downtrend, whereas the training sample may vary between periods of stability and volatility. We hypothesize that a model estimated using the more volatile portion of the training sample will yield more accurate coefficients than one estimated using a stable period. Consequently, using different training samples for model estimation would introduce inconsistencies, making it an unfair comparison among the competing models.

We also report Mean Absolute Forecast Error (MAFE) and Mean Squared Forecast Error (MSFE) for all forecast models. Subsequently, we quantify the improvement in MAFE or MSFE achieved by the forecast model with predictors, expressed as a percentage relative to the base model. Additionally, we employ the Modified Diebold-Mariano (MDM) test to compare all forecasting models against the base model, testing various null hypotheses, and we report the corresponding p-values. The null hypotheses are as follows.

$H_{0,1}$: Both the model with external predictor and the base model have equal MAFE

$H_{0,2}$: Both the model with external predictor and the base model have equal MSFE

$H_{0,3}$: The model with external predictor forecast-encompasses the base model

5.1 News Sentiment Indices

In this section, we predict the natural logarithm of HPI by the forecast model (1) using sentiment indices. The news sentiment data spans from January 2010 to October 2023. We used the data until January 2021 to estimate the model and the rest for out-of-sample validation. First, for the base model without an exogenous variable, for each horizon, we select the lags of HPI on the right-hand side of the forecasting model using BIC scores. Then, we add the same lags of the predictor to constitute the forecast model with each predictor. In other words, we first constitute the base model with $n = 0$ to find m. Then, we set n equal to m for all forecasting models with exogenous predictors. We found that for $h \in \{1, 3, 6\}$, $m = 1$, and for $h = 12$, $m = 3$.

Forecast Horizon	Predictor	MAFE	MSFE	MAFE Improvement	MSFE Improvement	p-value		
						$H_{0,1}$	$H_{0,2}$	$H_{0,3}$
h=1	**None**	0.01127	0.00018	--	--	--	--	--
	STS	0.01063	0.00016	5.7%	11.3%	0.0059	0.0032	0.0045
	STS1	0.01055	0.00016	6.4%	12.7%	0.0064	0.0030	0.0043
	STS2	0.01063	0.00016	5.7%	11.4%	0.0059	0.0033	0.0046
	LTS	0.01116	0.00017	1.0%	2.0%	0.1052	0.0797	0.0901

Forecast Horizon	Predictor	MAFE	MSFE	MAFE Improvement	MSFE Improvement	p-value		
						$H_{0,1}$	$H_{0,2}$	$H_{0,3}$
	LTS1	0.01110	0.00017	1.5%	2.7%	0.0712	0.0588	0.0700
	LTS2	0.01116	0.00017	1.0%	2.0%	0.1043	0.0783	0.0885
	mixed	0.01086	0.00016	3.6%	7.3%	0.0024	0.0020	0.0022
h=3	**None**	0.02961	0.00128	--	--	--	--	--
	STS	0.02861	0.00119	3.4%	7.4%	0.2257	0.1509	0.1573
	STS1	0.02842	0.00117	4.0%	8.5%	0.2124	0.1542	0.1610
	STS2	0.02860	0.00119	3.4%	7.4%	0.2271	0.1511	0.1577
	LTS	0.02918	0.00124	1.5%	2.9%	0.2126	0.2156	0.2197
	LTS1	0.02905	0.00123	1.9%	3.6%	0.2095	0.2121	0.2172
	LTS2	0.02918	0.00124	1.5%	2.9%	0.2107	0.2146	0.2187
	mixed	0.02906	0.00123	1.9%	4.1%	0.1740	0.1485	0.1500
h=6	**None**	0.05142	0.00362	--	--	--	--	--
	STS	0.05103	0.00358	0.8%	1.2%	0.4083	0.4794	0.4826
	STS1	0.05098	0.00357	0.9%	1.3%	0.4139	0.4817	0.4854
	STS2	0.05101	0.00357	0.8%	1.2%	0.4082	0.4793	0.4826
	LTS	0.05081	0.00353	1.2%	2.5%	0.5447	0.5081	0.5166
	LTS1	0.05064	0.00350	1.5%	3.3%	0.5423	0.4958	0.5053
	LTS2	0.05080	0.00353	1.2%	2.6%	0.5433	0.5069	0.5154
	mixed	0.05129	0.00360	0.3%	0.5%	0.4299	0.4699	0.4709
h=12	**None**	0.07500	0.00791	--	--	--	--	--
	STS	0.07473	0.00785	0.4%	0.7%	0.8764	0.8411	0.8516
	STS1	0.07466	0.00782	0.5%	1.0%	0.8739	0.8149	0.8276
	STS2	0.0747	0.0078	0.4%	0.7%	0.8757	0.8418	0.8524
	LTS	0.07449	0.00772	0.7%	2.3%	0.8678	0.7775	0.7868

Forecast Horizon	Predictor	MAFE	MSFE	MAFE Improvement	MSFE Improvement	p-value		
						$H_{0,1}$	$H_{0,2}$	$H_{0,3}$
	LTS1	0.07465	0.0078	0.5%	1.9%	0.9089	0.8049	0.8153
	LTS2	0.07449	0.00772	0.7%	2.3%	0.8682	0.7774	0.7867
	mixed	0.07483	0.00787	0.2%	0.5%	0.9137	0.8722	0.8809

Table 1: Forecast by sentiment indices. This table reports MAFE and MSFE for predicting the logarithm of HPI in Canada using autoregressive models with seasonal dummies. The models evaluated include a base model with no external predictors and several models that incorporate different sentiment indices as external predictors. Additionally, the table shows the percentage improvement in forecast accuracy (MAFE and MSFE) of the models with sentiment indices relative to the base model. The p-values associated with the null hypotheses test whether the inclusion of each sentiment index significantly improves the model's forecasting performance compared to the base model. $\boldsymbol{H_{0,1}}$*: equal MAFE,* $\boldsymbol{H_{0,2}}$*: equal MSFE,* $\boldsymbol{H_{0,3}}$*: the predictor model encompasses the base model*

As we can see, short-term sentiment indices perform better than long-term sentiment indices in the short-term horizons ($h = 1, 3$). Similarly, long-term sentiment indices perform better than short-term sentiment indices in the long-term horizons ($h = 6, 12$). We can reject all null hypotheses for short-term sentiment indices in $h = 1$ at the 0.01 significance level. We can also reject all null hypotheses for long-term sentiment indices in $h = 1$ at the 0.10 significance level. However, we were unable to reject any null hypotheses of equal MAFE or MSFE, as well as forecast encompassing, across other forecast horizons.

The sentiment indices derived from news articles exhibited limited predictive power for long horizons for several potential reasons. Firstly, news authors often present scenarios for both upward and downward price movements, thereby avoiding definitive statements about future prices. Secondly, news articles frequently contain varied opinions pertaining to different cities and provinces. Lastly, as illustrated in Figure 3 in the appendix, when the volume of news articles containing sentiment is relatively low, the resulting sentiment index may become overly sensitive to the content of individual articles, potentially diminishing its reliability as an indicator of overall market sentiment. This sensitivity underscores the need for careful interpretation of sentiment indices during periods with sparse news coverage.

5.2 Bubble Indicators

We have the bubble indicators from November 2001 to November 2023. We used the data until June 2019 to estimate the model and the rest for forecast validation. Like before, for each horizon, we first select the lags of HPI on the right-hand side of the forecasting model using BIC scores for the base model without exogenous variables. Then, we add the same lags of the bubble indicators to constitute the forecast model with each indicator. In other words, we first find m and constitute the base model with $n = 0$. Then, we set n equal to m. We found that for $h \in \{1, 3\}$, $m = 3$ and for $h \in \{6, 12\}$, $m = 2$.

Forecast Horizon	Predictor	MAFE	MSFE	MAFE Improvement	MSFE Improvement	p-value		
						$H_{0,1}$	$H_{0,2}$	$H_{0,3}$
h=1	None	0.00815	9.89E-05	--	--	--	--	--
	SADF	0.00818	0.00010	-0.4%	-2.1%	0.8467	0.5789	0.3332
	BSADF	0.00822	9.86E-05	-0.8%	0.3%	0.6745	0.9075	0.7375
	WHL_A	0.00813	9.86E-05	0.2%	0.3%	0.8726	0.9064	0.8262
	WHL_S	0.00798	9.75E-05	2.1%	1.4%	0.3591	0.6724	0.9104
h=3	None	0.02203	0.00080	--	--	--	--	--
	SADF	0.02226	0.0008	-1.0%	0.3%	0.7723	0.9756	0.9484
	BSADF	0.02233	0.00081	-1.4%	-1.6%	0.6888	0.7972	0.7105
	WHL_A	0.02155	0.00077	2.2%	4.0%	0.6619	0.6094	0.7579
	WHL_S	0.02206	0.00080	-0.2%	-0.4%	0.9710	0.9468	0.8065
h=6	None	0.03854	0.00238	--	--	--	--	--
	SADF	0.03906	0.00242	-1.4%	-1.8%	0.7985	0.8841	0.8361
	BSADF	0.03902	0.00243	-1.2%	-2.3%	0.7505	0.7302	0.6880
	WHL_A	0.03849	0.00229	0.1%	3.8%	0.9851	0.7217	0.8233
	WHL_S	0.03782	0.00237	1.9%	0.3%	0.6731	0.9759	0.9795
h=12	None	0.05876	0.00522	--	--	--	--	--
	SADF	0.05903	0.0053	-0.5%	-1.3%	0.9755	0.9606	0.8615
	BSADF	0.05903	0.00523	-0.5%	-0.3%	0.9470	0.9826	0.9408
	WHL_A	0.05868	0.00526	0.1%	-0.9%	0.9681	0.8897	0.8725

Forecast Horizon	Predictor	MAFE	MSFE	MAFE Improvement	MSFE Improvement	p-value		
						$H_{0,1}$	$H_{0,2}$	$H_{0,3}$
	WHL_S	0.05891	0.00515	-0.3%	1.3%	0.9765	0.9413	0.9832

Table 2: Forecast by bubble indicators. This table reports MAFE and MSFE for predicting the logarithm of HPI in Canada using autoregressive models with seasonal dummies. The models evaluated include a base model with no external predictors and several models that incorporate different bubble indicators as external predictors. Additionally, the table shows the percentage improvement in forecast accuracy (MAFE and MSFE) of the models with bubble indicators relative to the base model. The p-values associated with the null hypotheses test whether the inclusion of each bubble indicator significantly improves the model's forecasting performance compared to the base model. $H_{0,1}$: equal MAFE, $H_{0,2}$: equal MSFE, $H_{0,3}$: the predictor model encompasses the base model.

As we can see, most bubble indicators do not show significant forecasting power. However, the table shows that the WHL statistics provide some slight improvement over the PSY statistics. WHL_A in $h = 3, 6$ and WHL_S in $h = 1, 6$ could slightly improve forecasting errors, but we couldn't reject any null hypotheses of equal MAFE or MSFE and forecast encompass.

5.3 Housing Search Index (HSI)

Google Trends data are available from January 2004, and our data spans from January 2004 to November 2023. We used data until the end of 2019 to estimate the model and from January 2020 to November 2023 to calculate forecasting measurements like MAFE and MSFE. To find the lags for the predictors on the right side of the forecasting model, we consider all combinations of HPI and exogenous variable consecutive lags and choose m and n based on Bayesian Information Criterion (BIC). We have one *UHSI* constructed for each forecast horizon, but we perform predictions using each for each forecast horizon. The following tables report the lags set selected for HPI and HSIs, MAFE, MSFE, MAFE improvement, MSFE improvement, and p-values for all the null hypotheses.

First, we start forecasting with a set named G1 which is a set of search queries relevant to the real estate market that a typical homebuyer uses while searching for a home, representing the market demand. Second, to check whether the results are reliable, we repeat the forecasting using a set of irrelevant keywords called G2 and examine how the new results are different. Finally, we investigate the HSI algorithm's performance in picking the most informative search data. In this regard, we combine all relevant and irrelevant keywords and use them to generate HSIs for forecasting. We call the last set G3.

5.3.1 Relevant Keywords

Forecast Horizon	Predictor	m	n	MAFE	MSFE	MAFE Improvement	MSFE Improvement	p-value		
								$H_{0,1}$	$H_{0,2}$	$H_{0,3}$
h=1	**None**	3	0	0.00850	0.00010	--	--	--	--	--
	UHSI_1	3	3	0.00825	0.00009	2.9%	12.7%	0.6684	0.2044	0.3378
	UHSI_3	3	1	0.00812	0.00009	4.5%	9.8%	0.0782	0.0162	0.0474
	UHSI_6	3	1	0.00813	0.00009	4.3%	9.3%	0.0814	0.0190	0.0517
	UHSI_12	3	0	0.00850	0.00010	--	--	--	--	--
h=3	**None**	3	0	0.02292	0.00082	--	--	--	--	--
	UHSI_1	3	3	0.02061	0.00063	10.1%	23.1%	0.4451	0.2999	0.5129
	UHSI_3	3	3	0.02088	0.00064	8.9%	21.4%	0.4952	0.3325	0.5867
	UHSI_6	3	1	0.02131	0.00070	7.0%	14.1%	0.2896	0.1956	0.2725
	UHSI_12	3	1	0.02089	0.00067	8.9%	17.5%	0.2977	0.1969	0.2878
h=6	**None**	3	0	0.03862	0.00240	--	--	--	--	--
	UHSI_1	3	1	0.03065	0.00158	20.7%	34.1%	0.3136	0.3540	0.4495
	UHSI_3	3	1	0.03040	0.00155	21.3%	35.5%	0.3156	0.3454	0.4458
	UHSI_6	3	1	0.03388	0.00180	12.3%	25.1%	0.4149	0.3334	0.4259
	UHSI_12	3	1	0.03114	0.00160	19.4%	33.2%	0.3231	0.3353	0.4257
h=12	**None**	3	0	0.05580	0.00488	--	--	--	--	--
	UHSI_1	3	1	0.03981	0.00206	28.7%	57.8%	0.5903	0.5231	0.6550
	UHSI_3	3	1	0.03982	0.00207	28.6%	57.6%	0.5879	0.5193	0.6546
	UHSI_6	3	1	0.04693	0.00311	15.9%	36.2%	0.6471	0.5287	0.6594
	UHSI_12	3	1	0.04121	0.00223	26.1%	54.2%	0.5914	0.5158	0.6426

Table 3: Forecast by UHSIs using relevant keywords. This table reports the optimal number of lags (m, n), MAFE, and MSFE for predicting the logarithm of HPI adjusted for seasonality and time trends in Canada using autoregressive models. The models evaluated include a base model with no external predictors and several models that incorporate different UHSIs driven from relevant keywords search data as external predictors. Additionally, the table shows the percentage improvement in forecast accuracy (MAFE and MSFE) of the models with UHSIs relative to the base model. The p-values associated with the null hypotheses test whether the inclusion of each UHSI significantly improves the model's forecasting performance compared to the base model. $H_{0,1}$*: equal MAFE,* $H_{0,2}$*: equal MSFE,* $H_{0,3}$*: the predictor model encompasses the base model.*

We employ the Bayesian Information Criterion (BIC) to determine the best lag structure. However, the selection process indicated that no lag was optimal for UHSI_12 in $h = 1$. In other words, we found that the optimal value of n in the forecast model is 0 for UHSI_12 in $h = 1$ and the forecast models became the base model. This outcome suggests that the lagged values of the dependent variable provide the most information for predicting the target variable, and including past values of the exogenous variable does not improve the model's performance according to the BIC. Thus, the model is specified without lagged terms for the exogenous variable, aligning with the criterion's emphasis on parsimony and minimizing overfitting. Consequently, we did not perform forecasting performance tests for these specific cases.

As demonstrated in the tables, HSIs provide greater predictive improvements over extended periods. While the MAFE and MSFE improvement increases by longer horizons, the p-values of null hypotheses of equal MAFE or MSFE increase too. This means that we can only reject some null hypotheses at the 0.05 and 0.1 significance levels in $h = 1$ and any null hypotheses in $h = 3, 6, 12$ cannot be rejected at the 0.15 significance level. The forecasting graphs are in the Appendix.

5.3.2 Irrelevant Keywords

The data we used so far to generate HSIs were search frequencies for a set of relevant keywords typically used by individuals seeking homes. The accuracy of the predictions was notably high, suggesting a strong correlation between search trends and real estate market movements. To validate the reliability of these results, a comparative analysis was conducted using predictions based on irrelevant keywords. This approach aimed to assess whether the predictive accuracy observed with the relevant keywords was genuinely indicative of market dynamics or merely coincidental. In this regard, we obtained 15 random keywords from ChatGPT which are: "Celestial," "Quasar," "Enigma," "Reverie," "Serendipity," "Nebula," "Ephemeral," "Labyrinth," "Zephyr," "Luminous," "Cascade," "Euphoria," "Paradox," "Halcyon," and "Axiom."

For the forecast horizon $h = 1$, the optimal value of n for all UHSIs is 0, resulting in the forecast model reducing to the base model. The results for other forecast horizons are provided in comparable tables below.

Forecast Horizon	Predictor	m	n	MAFE	MSFE	MAFE Improvement	MSFE Improvement	p-value		
								$H_{0,1}$	$H_{0,2}$	$H_{0,3}$
h=3	**None**	3	0	0.02292	0.00082	--	--	--	--	--
	UHSI_1	3	1	0.02270	0.00085	1.0%	-4.7%	0.8983	0.7457	0.5785
	UHSI_3	3	1	0.02274	0.00086	0.8%	-4.8%	0.9245	0.7270	0.4783
	UHSI_6	3	1	0.02296	0.00086	-0.2%	-5.4%	0.9858	0.7055	0.4058
	UHSI_12	3	1	0.02440	0.00093	-6.4%	-13.9%	0.6053	0.4394	0.2152
h=6	**None**	3	0	0.03862	0.00240	--	--	--	--	--
	UHSI_1	3	1	0.03815	0.00252	1.2%	-5.0%	0.9467	0.8840	0.6918
	UHSI_3	3	1	0.03593	0.00241	7.0%	-0.4%	0.7361	0.9891	0.6985
	UHSI_6	3	1	0.03557	0.00239	7.9%	0.2%	0.7327	0.9950	0.6218
	UHSI_12	3	1	0.04056	0.00278	-5.0%	-15.9%	0.8492	0.6896	0.3284
h=12	**None**	3	0	0.0558	0.0049	--	--	--	--	--
	UHSI_1	3	2	0.0614	0.0056	-10.0%	-14.4%	0.7902	0.8131	0.6418
	UHSI_3	3	2	0.0614	0.0054	-10.1%	-10.0%	0.8001	0.8663	0.6474
	UHSI_6	3	2	0.0611	0.0053	-9.6%	-7.6%	0.8347	0.9067	0.6354
	UHSI_12	3	1	0.0633	0.0051	-13.5%	-4.7%	0.8043	0.9492	0.3717

Table 4: Forecast by UHSIs using irrelevant keywords. This table reports the optimal number of lags (m, n), MAFE, and MSFE for predicting the logarithm of HPI adjusted for seasonality and time trends in Canada using autoregressive models. The models evaluated include a base model with no external predictors and several models that incorporate different UHSIs driven from irrelevant keywords search data as external predictors. Additionally, the table shows the percentage improvement in forecast accuracy (MAFE and MSFE) of the models with UHSIs relative to the base model. The p-values associated with the null hypotheses test whether the inclusion of each UHSI significantly improves the model's forecasting performance compared to the base model. $\boldsymbol{H_{0,1}}$*: equal MAFE,* $\boldsymbol{H_{0,2}}$*: equal MSFE,* $\boldsymbol{H_{0,3}}$*: the predictor model encompasses the base model.*

The comparison between the forecasts generated from relevant and irrelevant keywords revealed an intrinsic relationship between search behavior and market dynamics, rather than mere coincidence. This suggests that Google Trends data, when carefully selected to reflect pertinent a search topic, provides a meaningful and reliable indicator of real estate market movements. Consequently, the strong predictive performance observed in this study highlights the potential of leveraging online search behavior as a valuable tool in forecasting real estate prices, offering insights that are both practical and theoretically sound.

5.3.3 Combination of Relevant and Irrelevant Keywords

The previous analysis comparing relevant and irrelevant keywords demonstrated that the relevance of search terms is crucial for achieving accurate real estate price forecasts. The results highlighted a significant difference in predictive power, with relevant keywords leading to more accurate predictions, while irrelevant keywords showed no forecasting capability.

In this section, we extend the analysis by combining both relevant and irrelevant keywords to evaluate whether the Housing Sentiment Index (HSI) generation algorithm can effectively discern the predictive value of the keywords. Specifically, we aim to determine if the algorithm can identify and prioritize the most predictive keywords from the mixed dataset, thereby enhancing the overall forecasting accuracy. This assessment will provide insights into the robustness and sensitivity of the HSI algorithm in selecting impactful indicators for real estate market predictions.

Like the previous section, the lag selection process indicated that the optimal n is 0 for all UHSIs and the forecast models became the base model in $h = 1$. However, since the BIC scores for $n = 0$ and $n = 1$ were close, we report the results of $n = 1$ for $h = 1$ to be able to compare with the results of other keyword sets.

Forecast Horizon	Predictor	m	n	MAFE	MSFE	MAFE Improvement	MSFE Improvement	p-value		
								$H_{0,1}$	$H_{0,2}$	$H_{0,3}$
h=1	**None**	3	0	0.00850	0.00010	--	--	--	--	--
	UHSI_1	3	1	0.00813	0.00009	4.4%	11.0%	0.1470	0.0199	0.0804
	UHSI_3	3	1	0.00809	0.00009	4.8%	10.6%	0.0786	0.0154	0.0498
	UHSI_6	3	1	0.00819	0.00009	3.6%	7.5%	0.1065	0.0429	0.1002
	UHSI_12	3	1	0.00810	0.00009	4.6%	9.6%	0.1449	0.0555	0.1933
h=3	**None**	3	0	0.02292	0.00082	--	--	--	--	--
	UHSI_1	3	1	0.02105	0.00068	8.2%	16.7%	0.3007	0.1879	0.2657
	UHSI_3	3	1	0.02083	0.00067	9.1%	17.8%	0.2694	0.1746	0.2528
	UHSI_6	3	1	0.02133	0.00072	6.9%	12.2%	0.2661	0.2080	0.2859
	UHSI_12	3	1	0.02101	0.00069	8.3%	15.9%	0.3217	0.2103	0.3217
h=6	**None**	3	0	0.03862	0.00240	--	--	--	--	--

Forecast Horizon	Predictor	m	n	MAFE	MSFE	MAFE Improvement	MSFE Improvement	p-value		
								$H_{0,1}$	$H_{0,2}$	$H_{0,3}$
	UHSI_1	3	1	0.03211	0.00166	16.9%	30.7%	0.3296	0.3561	0.4199
	UHSI_3	3	1	0.03157	0.00162	18.3%	32.4%	0.3101	0.3291	0.3962
	UHSI_6	3	1	0.03360	0.00184	13.0%	23.3%	0.3693	0.2959	0.3876
	UHSI_12	3	1	0.03113	0.00163	19.4%	32.1%	0.3279	0.3179	0.4312
h=12	**None**	3	0	0.05580	0.00488	--	--	--	--	--
	UHSI_1	3	1	0.04246	0.00246	23.9%	49.6%	0.5799	0.5224	0.6024
	UHSI_3	3	1	0.04185	0.00243	25.0%	50.1%	0.5613	0.5157	0.5929
	UHSI_6	3	1	0.04711	0.00317	15.6%	35.0%	0.6484	0.5407	0.6749
	UHSI_12	3	1	0.04102	0.00218	26.5%	55.4%	0.6002	0.5196	0.6740

Table 5: Forecast by UHSIs using relevant and irrelevant keywords. This table reports the optimal number of lags (m, n), MAFE, and MSFE for predicting the logarithm of HPI adjusted for seasonality and time trends in Canada using autoregressive models. The models evaluated include a base model with no external predictors and several models that incorporate different UHSIs driven by the combination of relevant and irrelevant keywords search data as external predictors. Additionally, the table shows the percentage improvement in forecast accuracy (MAFE and MSFE) of the models with UHSIs relative to the base model. The p-values associated with the null hypotheses test whether the inclusion of each UHSI significantly improves the model's forecasting performance compared to the base model. $H_{0,1}$*: equal MAFE,* $H_{0,2}$*: equal MSFE,* $H_{0,3}$*: the predictor model encompasses the base model.*

The analysis involving a combination of relevant and irrelevant keywords yielded results that were nearly as accurate as those obtained using only relevant keywords. This finding suggests that the generating HSI algorithm is adept at identifying and prioritizing the most predictive keywords within the dataset. As evidenced by the tables, the HSIs still provide significant predictive improvements over extended periods, even when irrelevant keywords are included in the input data.

While the MAFE and MSFE improvements increase with longer forecast horizons, the p-values associated with the null hypotheses of equal MAFE or MSFE also rise. This indicates that the null hypotheses can only be rejected at the 0.05 and 0.1 significance levels for the one-month horizon ($h = 1$), whereas for longer horizons ($h = 3,6,12$), the null hypotheses cannot be rejected even at the 0.15 significance level. These results underscore the robustness of the algorithm in leveraging relevant keywords effectively, despite the presence of irrelevant data, to enhance the predictive accuracy of real estate price forecasts.

5.3.4 Visual Comparison

Figures 2 and 3 illustrate the Kernel Density Estimates (KDE) of the improvements in MAFE and MSFE achieved by HSIs in B_{12} set (similar graphs for other horizons are in the Appendix). The data is segmented into three groups: relevant keywords (G1), irrelevant keywords (G2), and a combination of relevant and irrelevant keywords (G3). Each figure shows the KDE for each group's forecasting results, with vertical lines representing the improvements associated with three series chosen from the respective groups. The analysis aims to evaluate the predictive power of relevant keywords, the lack of forecasting capability in irrelevant keywords, and the ability of the forecasting algorithm to discern and utilize relevant keywords when mixed with irrelevant ones.

Figure 2, which represents MAFE improvements, the KDE for relevant keywords (blue curve) demonstrates a clear positive improvement, indicating substantial forecasting accuracy. The irrelevant keywords (red curve), on the other hand, show a distribution centered around zero or negative improvements, highlighting their lack of predictive power. The combined set (orange curve) shows a distribution similar to that of the relevant keywords, suggesting that the forecasting algorithm successfully identifies and prioritizes the relevant keywords within the mixed dataset. The vertical lines are the improvements associated with UHSI_12 for each group. They further emphasize this pattern, with the UHSI_12 for the relevant keyword series showing significant positive improvement, while the UHSI_12 for the irrelevant keywords aligns with minimal or negative improvement. The combined set's line closely follows the relevant keywords' improvement, reinforcing the algorithm's efficacy in distinguishing relevant data.

Figure 3, depicting MSFE improvements, echoes the findings from the MAFE analysis. The KDE for relevant keywords again shows a pronounced positive shift, indicative of strong predictive accuracy. In contrast, the KDE for irrelevant keywords is centered around negative or negligible improvements, reaffirming their lack of forecasting utility. The combined set's KDE mirrors the relevant keywords' distribution, confirming that the algorithm effectively identifies and utilizes relevant keywords even when irrelevant ones are present. The UHSI_12 lines in this figure also illustrate the same trend, with the relevant keyword series showing significant improvement and the irrelevant keyword series showing minimal improvement. The combined set's vertical line aligns closely with that of the relevant keywords, demonstrating the algorithm's robustness in filtering out irrelevant information.

Overall, these figures provide compelling evidence that the relevance of search keywords plays a crucial role in enhancing the accuracy of real estate price forecasts. Additionally, the similarity in the distribution of improvements between the relevant and combined keyword sets indicates that the forecasting algorithm is proficient in detecting and leveraging relevant keywords from a mixed dataset. This capability is essential for practical applications, where the input data may not always be perfectly curated. The results validate the use of Google Trends data with appropriate keyword selection as a powerful tool for predicting real estate market dynamics.

In conclusion, the analysis highlights the effectiveness of using Google Trends data with carefully selected keywords for real estate price prediction. The significant predictive improvements achieved with relevant keywords, along with the algorithm's ability to filter and prioritize relevant information from a mixed dataset, demonstrate the practical utility and robustness of this approach. These findings contribute to the broader understanding of how online search behavior can be harnessed to gain insights into market trends and inform decision-making in the real estate sector.

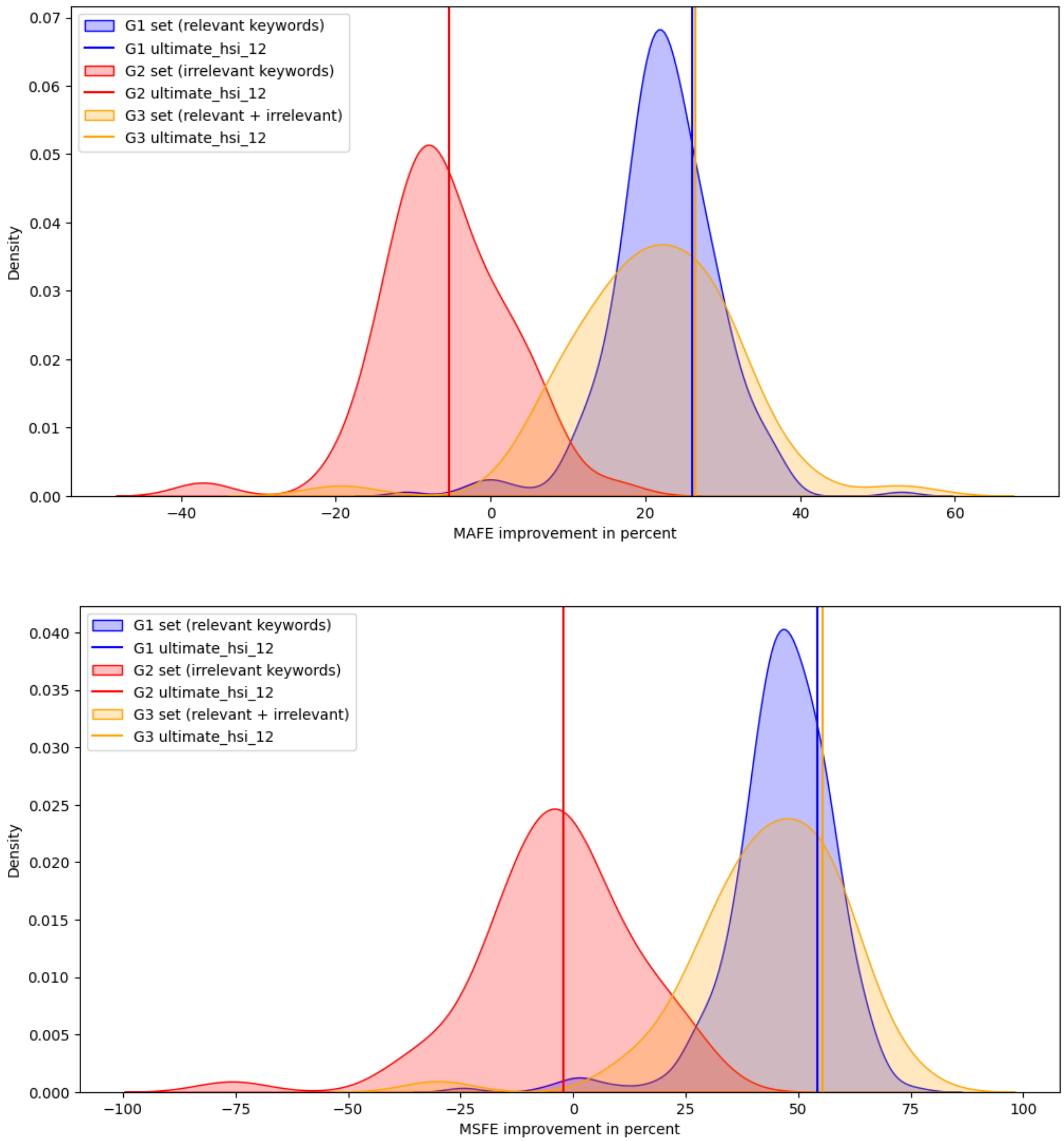

Figure 2: KDE distribution of MAFE and MSFE improvements at h=12. Kernel Density Estimation (KDE) plot illustrates the distribution of MAFE and MSFE improvement percentages relative to the base model at horizon h=12 for three groups of data: relevant keywords (blue), irrelevant keywords (red), and a combination of relevant and irrelevant keywords (orange). The figure shows the MAFE and MSFE improvement distributions for the HSIs in the B_{12} set, along with vertical lines representing the MAFE and MSFE improvement associated with the UHSI_12 for each group. The results indicate that the distributions for the relevant keywords group and the combination of relevant and irrelevant keywords group are similar, both showing distinct patterns compared to the irrelevant keywords group, which demonstrates a notably different distribution.

5.4 Comparison of All Predictors

In this section, we aim to compare the prediction power of three distinct groups of predictors: news sentiment indices, bubble indicators, and Google Trend UHSIs. In this regard, we try to predict the preprocessed HPI, as introduced in section 4.5.1, which is a detrended and deseasonalized version of the natural logarithm of HPI. To find the optimal lag structure in the forecasting model, we consider all combinations of HPI and exogenous variable consecutive lags and choose m and n based on the BIC. Due to differing time ranges across these groups, the analysis was constrained to the narrowest range, which corresponds to the news sentiment indices. We only include the predictors in the following tables whose optimal n value is not zero. In this regard, it is found that among news sentiment indices in all forecasting horizons, only the optimal n of STS1 in $h = 1$ has a nonzero value.

To evaluate the performance of these predictors, we compare the predictors in terms of MAFE and MSFE improvements against the base model. The MDM tests are conducted to check if the MAFE and MSFE of models with predictors are significantly different from those of the base model. We also use this test to see whether the null hypothesis of the models with predictors forecast encompass the base model can be rejected.

Forecast Horizon	Predictor	m	n	MAFE	MSFE	MAFE Improvement	MSFE Improvement	p-value		
								$H_{0,1}$	$H_{0,2}$	$H_{0,3}$
h=1	**None**	3	0	0.05580	0.00488	--	--	--	--	--
	UHSI_1	1	3	0.01051	0.00015	8.1%	14.3%	0.3449	0.1768	0.3752
	UHSI_3	1	2	0.00985	0.00015	13.9%	18.3%	0.0678	0.0793	0.7546
	UHSI_6	1	2	0.00965	0.00014	15.6%	21.5%	0.0400	0.0371	0.4810
	UHSI_12	1	3	0.00973	0.00014	14.9%	23.5%	0.0502	0.0134	0.4554
	SADF	1	1	0.01080	0.00016	5.6%	9.4%	0.0034	0.0039	0.0056
	BSADF	1	1	0.01057	0.00016	7.6%	11.4%	0.0047	0.0040	0.0080
	WHL_A	1	1	0.01066	0.00016	6.8%	13.6%	0.1131	0.0237	0.1167
	WHL_S	1	1	0.01112	0.00017	2.8%	7.9%	0.3259	0.1508	0.2671
	STS1	1	1	0.01072	0.00016	6.2%	12.8%	0.0055	0.0022	0.0032

Forecast Horizon	Predictor	m	n	MAFE	MSFE	MAFE Improvement	MSFE Improvement	p-value		
								$H_{0,1}$	$H_{0,2}$	$H_{0,3}$
h=3	**None**	2	0	0.02854	0.00108	--	--	--	--	--
	UHSI_1	1	3	0.02868	0.00111	-0.5%	-3.0%	0.9731	0.8913	0.5684
	UHSI_3	1	2	0.02718	0.00103	4.8%	4.1%	0.7094	0.8345	0.7194
	UHSI_6	1	2	0.02679	0.00100	6.1%	6.8%	0.6277	0.7174	0.8303
	UHSI_12	1	3	0.02740	0.00101	4.0%	6.0%	0.7105	0.7090	0.8923
	SADF	1	1	0.02912	0.00114	-2.0%	-5.8%	0.7783	0.6643	0.5701
	BSADF	1	1	0.02881	0.00113	-0.9%	-5.1%	0.8837	0.6221	0.4992
	WHL_A	1	1	0.02969	0.00116	-4.0%	-7.7%	0.6840	0.6495	0.5262
	WHL_S	1	1	0.03037	0.00121	-6.4%	-12.6%	0.3508	0.2682	0.2330
h=6	**None**	2	0	0.04953	0.00312	--	--	--	--	--
	UHSI_1	3	1	0.03977	0.00223	19.7%	28.7%	0.2709	0.3278	0.3727
	UHSI_3	1	2	0.04126	0.00241	16.7%	22.9%	0.4630	0.4653	0.6296
	UHSI_6	1	2	0.04158	0.00240	16.1%	23.1%	0.4690	0.4532	0.6031
	UHSI_12	1	2	0.04187	0.00241	15.5%	23.0%	0.4404	0.4162	0.5158
	SADF	1	1	0.04848	0.00295	2.1%	5.6%	0.8403	0.7957	0.8698
	BSADF	1	1	0.04811	0.00293	2.9%	6.2%	0.7486	0.6996	0.7604
	WHL_A	1	1	0.05287	0.00336	-6.7%	-7.5%	0.5528	0.7117	0.6217
	WHL_S	1	3	0.04934	0.00309	0.4%	1.0%	0.9681	0.9581	0.9518
h=12	**None**	3	0	0.07478	0.00677	--	--	--	--	--
	UHSI_1	3	1	0.05926	0.00438	20.8%	35.4%	0.5105	0.5210	0.5352
	UHSI_3	3	1	0.06169	0.00473	17.5%	30.2%	0.4891	0.5184	0.5343
	UHSI_6	3	2	0.05874	0.00431	21.5%	36.4%	0.5374	0.5658	0.5662
	UHSI_12	3	1	0.06059	0.00456	19.0%	32.7%	0.4764	0.5013	0.5148
	SADF	3	2	0.06199	0.00498	17.1%	26.4%	0.3945	0.4434	0.4703
	BSADF	3	1	0.06551	0.00523	12.4%	22.8%	0.4386	0.5029	0.5100

Table 6: Forecast using all predictors over a consistent period. This table reports the optimal number of lags (m, n), MAFE, and MSFE for predicting the logarithm of the HPI adjusted for seasonality and time trends in Canada using autoregressive models over the same training and validation sample. The models evaluated include a base model with no external predictors and several models that incorporate different sentiment indices, bubble indicators, and UHSIs as external predictors. For each Forecast horizon, the table includes only those predictors where the optimal number of lags (n) is greater than zero. Additionally, the table

shows the percentage improvement in forecast accuracy (MAFE and MSFE) of the models with predictors relative to the base model. The p-values associated with the null hypotheses test whether the inclusion of each predictor significantly improves the model's forecasting performance compared to the base model. $\boldsymbol{H_{0,1}}$*: equal MAFE,* $\boldsymbol{H_{0,2}}$*: equal MSFE,* $\boldsymbol{H_{0,3}}$*: the predictor model encompasses the base model.*

Among the three groups, the online search activity indices (UHSIs) demonstrated the greatest improvement in both MAFE and MSFE metrics. The bubble indicators, particularly SADF and BSADF statistics from the PSY test, also contributed to forecast improvements, albeit to a lesser extent than the UHSI indices. In contrast, the news sentiment indices did not show significant predictive power. Except for the STS1 index at a one-month horizon (h=1), which has a lag in the forecast model, other sentiment indices could not Justify the BIC to have lagged values in the forecast model. However, it is indicated in the previous sections that the models with sentiment lags can provide improvements in forecast accuracy. In line with the preceding sections, we can only reject null hypotheses at the 0.05 and 0.1 significance levels in h=1.

To test if the better forecasting performance of UHSIs compared to other predictors is statistically significant, we again employ the MDM test for the null hypotheses of equal MAFE and MSFE. Moreover, we utilize this test to determine whether the null hypothesis that the models incorporating UHSIs forecast encompass the models with other predictors can be rejected. We only test UHSIs against the predictors that have lagged values in the forecast model.

Forecast Horizon	Base Models	Validation Models									
		UHSI_1					UHSI_3				
		Improvement		p-value			Improvement		p-value		
		MAFE	MSFE	$H_{0,1}$	$H_{0,2}$	$H_{0,3}$	MAFE	MSFE	$H_{0,1}$	$H_{0,2}$	$H_{0,3}$
h=1	SADF	2.7%	5.5%	0.7601	0.6059	0.1416	8.8%	9.8%	0.2131	0.3262	0.7339
	BSADF	0.6%	3.8%	0.9486	0.7449	0.1300	6.8%	8.1%	0.2879	0.3617	0.7921
	WHL_A	1.4%	1.3%	0.8656	0.9365	0.0822	7.6%	5.8%	0.2059	0.5573	0.5866
	WHL_S	5.5%	7.2%	0.5139	0.5040	0.1616	11.4%	11.4%	0.0843	0.1885	0.9491
	STS1	2.0%	1.9%	0.8124	0.8661	0.0899	8.1%	6.4%	0.2396	0.5283	0.4558
h=3	SADF	1.5%	2.7%	0.9178	0.8985	0.7263	6.6%	9.4%	0.6095	0.6526	0.9860
	BSADF	0.5%	2.0%	0.9741	0.9231	0.7230	5.7%	8.8%	0.6343	0.6237	0.9330
	WHL_A	3.4%	4.4%	0.8341	0.8448	0.6961	8.5%	11.0%	0.4279	0.5500	0.8679
	WHL_S	5.6%	8.6%	0.6973	0.6759	0.8956	10.5%	14.9%	0.4010	0.4282	0.7623
h=6	SADF	18.0%	24.4%	0.3394	0.3820	0.4428	14.9%	18.3%	0.4995	0.5558	0.7529
	BSADF	17.3%	23.9%	0.3131	0.3880	0.4696	14.2%	17.7%	0.4536	0.4972	0.6573
	WHL_A	24.8%	33.6%	0.1838	0.2098	0.2566	22.0%	28.2%	0.2915	0.3037	0.3804
	WHL_S	19.4%	27.9%	0.2636	0.3334	0.3778	16.4%	22.1%	0.4534	0.4641	0.6146
h=12	SADF	4.4%	12.2%	0.8722	0.7911	0.9037	0.5%	5.1%	0.9839	0.8976	0.9694
	BSADF	9.5%	16.3%	0.7554	0.7601	0.8282	5.8%	9.6%	0.8075	0.8200	0.8985

Table 7: Comparison of UHSIs with other predictors. This table compares the forecasting power of UHSIs against other predictors for which the optimal number of lags (n) is greater than zero. It reports the improvements in MAFE and MSFE of models incorporating UHSIs relative to those with other predictors. The forecasting models are autoregressive, predicting the seasonal and trend-adjusted logarithm of the HPI in Canada using the same training and validation sample. The p-values associated with the null hypotheses test whether the forecasting accuracy of models with UHSIs significantly outperforms that of models with other predictors. $\boldsymbol{H_{0,1}}$*: equal MAFE,* $\boldsymbol{H_{0,2}}$*: equal MSFE,* $\boldsymbol{H_{0,3}}$*: the predictor model encompasses the base model.*

Forecast Horizon	Base Models	Validation Models									
		UHSI_6					UHSI_12				
		Improvement		p-value			Improvement		p-value		
		MAFE	MSFE	$H_{0,1}$	$H_{0,2}$	$H_{0,3}$	MAFE	MSFE	$H_{0,1}$	$H_{0,2}$	$H_{0,3}$
h=1	**SADF**	10.6%	13.5%	0.1263	0.1637	0.9218	9.9%	15.3%	0.1681	0.0660	0.9268
	BSADF	8.7%	11.9%	0.1678	0.1602	0.7977	7.9%	13.8%	0.2493	0.0713	0.9740
	WHL_A	9.5%	9.6%	0.1026	0.2870	0.9740	8.7%	11.5%	0.1719	0.1399	0.6705
	WHL_S	13.2%	15.1%	0.0451	0.0721	0.6345	12.5%	16.9%	0.0757	0.0383	0.9005
	STS1	10.0%	10.2%	0.1401	0.2908	0.7378	9.3%	12.1%	0.1771	0.1316	0.6073
h=3	**SADF**	8.0%	11.9%	0.5375	0.5548	0.7263	5.9%	11.2%	0.5914	0.4858	0.8147
	BSADF	7.0%	11.3%	0.5544	0.5128	0.7230	4.9%	10.6%	0.6323	0.4559	0.7309
	WHL_A	9.8%	13.5%	0.3361	0.4310	0.6961	7.7%	12.8%	0.4889	0.4191	0.7628
	WHL_S	11.8%	17.2%	0.3478	0.3471	0.8956	9.8%	16.6%	0.3810	0.3002	0.5617
h=6	**SADF**	14.2%	18.6%	0.5119	0.5342	0.7184	13.6%	18.4%	0.4394	0.4198	0.5265
	BSADF	13.6%	18.0%	0.4639	0.4775	0.6266	13.0%	17.8%	0.3871	0.3642	0.4407
	WHL_A	21.3%	28.5%	0.2843	0.2726	0.3193	20.8%	28.3%	0.2688	0.2203	0.2625
	WHL_S	15.7%	22.3%	0.4608	0.4566	0.5914	15.1%	22.2%	0.4180	0.3729	0.4635
h=12	**SADF**	5.2%	13.6%	0.8789	0.8300	0.9505	2.3%	8.5%	0.9224	0.8265	0.9503
	BSADF	10.3%	17.7%	0.7539	0.7796	0.8449	7.5%	12.8%	0.7625	0.7668	0.8319

Table 7 (Continued)

Analogous to the evaluation of UHSIs against the base model, the comparative analysis of UHSIs against other predictors reveals that the null hypotheses of equal MAFE and MSFE are predominantly rejectable only at the one-month forecast horizon ($h = 1$). However, the frequency of rejections is lower in this context compared to the base model analysis.

At $h = 1$, the null hypothesis of equal MSFE between UHSI_12 and all other predictors can be rejected at the 0.15 significance level. Similarly, the null hypothesis of equal MAFE between UHSI_6 and all other predictors, with the exception of the BSADF test statistic, is rejectable at the same significance threshold. In contrast, the WHL_S indicator exhibits the weakest statistical performance; the null hypotheses asserting equal MAFE and MSFE between

WHL_S and nearly all UHSIs are frequently rejectable. This pattern suggests that the rejection of null hypotheses is more prevalent for the WHL test statistics than for the PSY test indicators.

Furthermore, the findings indicate that UHSIs, excluding UHSI_1, demonstrate superior forecasting performance relative to other predictors. Moreover, the null hypotheses positing that UHSIs encompass the forecasting ability of other predictors cannot be rejected, except in the case of UHSI_1, for which the null hypotheses are rejectable against almost all predictors.

For forecast horizons beyond one month ($h > 1$), none of the null hypotheses could be rejected across all tests conducted. This outcome suggests that the significance of the predictive privilege of UHSIs against other predictors diminishes over extended forecasting periods.

6. Conclusion

This thesis employed psychological and behavioral data derived from news articles, bubble indicators, and Google Trends to predict real estate prices in the Canadian market. The main goal was to assess whether integrating these indicators could improve the predictive accuracy of existing real estate pricing models, while also comparing the relative performance of these predictors.

We tested two lag selection methods for news sentiment and bubble indicators, and the empirical results reveal that each approach yields different forecasting outcomes. When the lags of the predictors were aligned with the lags of the dependent variable, sentiment indices exhibited significant predictive improvements at the 0.01 and 0.1 significance levels for a one-month forecasting horizon. However, in this configuration, bubble indicators did not demonstrate any forecasting improvements. In contrast, when the lags of both the predictors and the dependent variable were selected based on the BIC, most sentiment indices, except for STS1, did not justify the inclusion of their lags in the model. Under this lag selection method, bubble indicators displayed significant forecast improvements at the one-month horizon.

Regarding the forecasts generated by the UHSIs derived from Google Trends, we identified optimal lags using the BIC. Our analysis showed that the UHSIs significantly improved forecast accuracy at the one-month horizon at the 0.05 and 0.1 significance levels, though the improvements for longer horizons were statistically insignificant. When comparing the performance of the various predictors, UHSIs consistently provided more accurate predictions than other predictors in certain cases at the one-month horizon.

This study also brings to light several challenges and limitations. Future research could focus on refining sentiment analysis techniques to extract more reliable sentiment scores from news sources and exploring innovative methodologies for integrating diverse data sources. Additionally, further studies could benefit from concentrating on the extraction of regional sentiment scores from news sources to enhance the granularity and relevance of the analysis. As illustrated in Figures 8 and 9 in the appendix, the forecast accuracies of UHSIs are nearly equivalent to the average accuracies of HSIs within B_h. Consequently, future research could aim to improve the procedure for extracting UHSI from HSIs in B_h, potentially leading to more accurate and robust forecasting models.

In conclusion, this thesis contributes to the growing body of literature on real estate price prediction by demonstrating the value of incorporating psychological and speculative elements into predictive models. By doing so, we can achieve more accurate and timely predictions, ultimately contributing to more informed decision-making in the real estate sector. This approach not only enriches academic research but also offers practical insights for improving the stability and efficiency of real estate markets.

Bibliography

Abidoye, Rotimi, Albert Chan, Funmilayo Abidoye, and Olalekan Oshodi (2019). Predicting property price index using artificial intelligence techniques: Evidence from Hong Kong, International Journal of Housing Markets and Analysis., 12(6), p. 1072-1092.

Adam, Klaus, Albert Marcet, and Johannes Beutel (2017). Stock Price Booms and Expected Capital Gains, American Economic Review, 107(8), p. 2352-2408.

Akerlof, George A. and Robert J. Shiller (2010). Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism, Princeton, New Jersey, Princeton University Press Princeton, p. 248.

Ardila, Diego, Dorsa Sanadgol, and Didier Sornette (2018). Out-of-sample forecasting of housing bubble tipping points, Quantitative Finance and Economics, 2(4), p. 904-930.

Ardila, Diego, Dorsa Sanadgol, Peter Cauwels, and Didier Sornette. (2016). Identification and critical time forecasting of real estate bubbles in the USA, Quantitative Finance, 17, p. 1-19.

Asal, Maher (2019), Is there a bubble in the Swedish housing market?, Journal of European Real Estate Research, 12(1), p. 32-61.

Bai, Jushan and Serena Ng (2008). Forecasting economic time series using targeted predictors, Journal of Econometrics, 146(2), p. 304-317.

Bailey, Martin J., Richard F. Muth, and Hugh O. Nourse (1963). A Regression Method for Real Estate Price Index Construction, Journal of the American Statistical Association, 58(304), p. 933–42.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis (2016). Measuring Economic Policy Uncertainty, The Quarterly Journal of Economics, 131(4), p. 1593–1636

Bulczak, Grzegorz M. (2021). Use of Google Trends to Predict the Real Estate Market: Evidence from the United Kingdom, International Real Estate Review, Global Social Science Institute, 24(4), p. 613-631.

Case, Karl E. and Robert J. Shiller (1987). Prices of Single Family Homes Since 1970: New Indexes for Four Cities, NBER Working Paper No. w2393

Case, Karl E. and Robert J. Shiller (2003). Is There a Bubble in the Housing Market?, Brookings Papers on Economic Activity, 68(2003-2), p. 299-342.

Choi, Hyunyoung, and Hal R. Varian (2009). Predicting the Present with Google Trends, Economic Record, 88, p. 2-9

Costello, Greg, and Craig Watkins (2002). Towards a System of Local House Price Indices, Housing Studies, 17(6), p. 857–873.

Cristescu, Marian Pompiliu, Raluca Andreea Nerisanu, Dumitru Alexandru Mara, and Simona-Vasilica Oprea. (2022). Using Market News Sentiment Analysis for Stock Market Prediction, Mathematics, 10(22), 4255.

Da, Zhi, Joseph Engelberg, and Pengjie Gao (2011). In Search of Attention, The Journal of Finance, 66, p. 1461-1499.

Damianov, Damian S., Xiangdong Wang, and Cheng Yan (2020). Google Search Queries, Foreclosures, and House Prices, The Journal of Real Estate Finance and Economics, 63, p. 177 - 209

Dietzel, Marian A., Nicole Braun, and Wolfgang Schäfers (2014). Sentiment-based commercial real estate forecasting with Google search volume data, Journal of Property Investment and Finance, 32(6), p. 540-569.

Dietzel, Marian A. (2016). Sentiment-based predictions of housing market turning points with Google trends, International Journal of Housing Markets and Analysis, 9, p. 108-136.

Fan, Xiaoyuan and Jiashuo Chen (2022). Stock Price Forecasting in Real Estate Industry Based on Investor Sentiment. Frontiers in Business, Economics and Management, 6, p. 54-59.

Garcia, Diego (2013). Sentiment during Recessions, The Journal of Finance, 68, p. 1267-1300.

Ghysels, Eric, Alberto Plazzi, Walter N. Torous, and Rossen Valkanov (2012). Forecasting Real Estate Prices, Handbook of Economic Forecasting, Elsevier, 2, p. 509-580.

Glaeser, Edward L. (2013). A Nation of Gamblers: Real Estate Speculation and American History, American Economic Review, 103(3), p. 1-42.

Glaeser, Edward L. and Charles G. Nathanson (2017). An Extrapolative Model of House Price Dynamics, Journal of Financial Economics, 126(1), p. 147-170.

Granziera, Eleonora and Sharon Kozicki (2015). House Price Dynamics: Fundamentals and Expectations, Journal of Economic Dynamics and Control, 60, p. 152-165.

Goodhart, Charles, and Boris Hofmann (2008). House prices, money, credit, and the macroeconomy, Oxford Review of Economic Policy, 24(1), p. 180–205.

Hagemann, Daniel and Monika Wohlmann (2019). An early warning system to identify house price bubbles, Journal of European Real Estate Research, 12, p. 291-310.

Kindelberger, Charles P., and Robert Z. Aliber (2011). Manias, Panics, and Crashes: A History of Financial Crises (6th ed.), London, Palgrave Macmillan.

Kok, Nils, Eija-Leena Koponen, Carmen A> Martínez-Barbosa (2017). Big Data in Real Estate? From Manual Appraisal to Automated Valuation, The Journal of Portfolio Management, 43, p. 202-211.

Li, Xiaodong, Pangjing Wu and Wenpeng Wang (2020). Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong, Inf. Process. Manag., 57, 102212.

Jovanović, Ivana (2021). Use of quantitative ratios from the system of national accounts in predicting price bubbles on the housing market, Ekonomske ideje i praksa, Faculty of Economics and Business, University of Belgrade, 43, p. 45-54.

Meen, Geoffrey (1999). Regional House Prices and the Ripple Effect: A New Interpretation, Housing Studies, 14(6), p. 733–53.

Moller, Stig V., Thomas Pedersen, Erik C. M. Schutte, Allan Timmermann (2023). Search and Predictability of Prices in the Housing Market, Management Science 70(1), p. 415-438.

Mudinas, Andrius, Dell Zhang, and Mark Levene (2018). Market Trend Prediction using Sentiment Analysis: Lessons Learned and Paths Forward, ArXiv, abs/1903.05440.

Nti, Isaac K., Adebayo F. Adekoya, Benjamin Weyori (2020). Predicting Stock Market Price Movement Using Sentiment Analysis: Evidence From Ghana. Applied Computer Systems, 25, p. 33-42.

Oust, Are, and Kjartan Hrafnkelsson (2017). What is a housing bubble, Economics Bulletin, 37, p. 806-836.

Oust, Are, Ole M. Eidjord (2020). Can Google Search Data be Used as a Housing Bubble Indicator?, International Real Estate Review, Global Social Science Institute, 23(2), p. 267-308.

Pavlidis, Efthymios, Alisa Yusupova, Ivan Paya, David Peel, Enrique Martínez-García, Adrienne Mack, et al. (2016). Episodes of Exuberance in Housing Markets: In Search of the Smoking Gun, The Journal of Real Estate Finance and Economics, 53(4), p. 419-449.

Phillips, Peter C. B., Shuping Shi, and Jun Yu (2015). Testing for Multiple Bubbles: Limit Theory of Real Time Detectors, International Economic Review, 56, p. 1079-1134.

Reinhart, Carmen M. and Kenneth S. Rogoff (2009). This Time Is Different: Eight Centuries of Financial Folly, Princeton, New Jersey, Princeton University Press, 512.

Shiller, Robert J. (1990). Speculative Prices and Popular Models, Journal of Economic Perspectives, 4(2), p. 55–65.

Shiller, Robert J. (2000). Review of Irrational Exuberance, The American Journal of Economics and Sociology, 59(3), 537–540.

Shiller, Robert, J. (2003). From Efficient Markets Theory to Behavioral Finance, Journal of Economic Perspectives, 17(1), p. 83–104.

Shiller, Robert J. (2008). Historic Turning Points in Real Estate, Eastern Economic Journal, 34(1), p. 1-13.

Soo, Cindy K. (2018). Quantifying Sentiment with News Media across Local Housing Markets, The Review of Financial Studies, 31(10), p. 3689–3719.

Statista (2023) Market share of leading search engines in Canada in April 2023. Accessed January 10, 2024, https://www.statista.com/statistics/475115/search-engine-market-share-canada/.

Tabales, Julia M. N., Jose M. C. Ocerin, and Francisco J. R. Carmona (2013). Artificial neural networks for predicting real estate prices. Revista de Métodos Cuantitativos para la Economía y la Empresa, 15, p. 29-44.

Tetlock, Paul C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market, The Journal of Finance, 62, p. 1139-1168.

Tsai, I.-Chun and Chien-Wen Peng (2011). Bubbles in the Taiwan housing market: The determinants and effects, Habitat International - HABITAT INT, 35, p. 379-390.

Turner, Zane, Kevin Labille, and Susan Gauch (2021). Lexicon-based sentiment analysis for stock movement prediction, Journal of Construction Materials, 2.

Vishwakarma, Vijay K. (2013). Forecasting Real Estate Business: Empirical Evidence from the Canadian Market. Global Journal of Business Research, 7(3), p. 1-14.

Whitehouse, Emily J., David I. Harvey, and Stephen J. Leybourne (2023). Real-Time Monitoring of Bubbles and Crashes,"Oxford Bulletin of Economics and Statistics, Department of Economics, University of Oxford, 85(3), p. 482-513.

Zhang, Jing, Shicheng Cui, Yan Xu, Qianmu Li and Tao Li (2018). A novel data-driven stock price trend prediction system, Expert Syst. Appl., 97, p. 60-69.

Zou, Hui and Trevor Hastie (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), p. 301-320.

Appendix

Nexis Uni Setting

Search terms:

(real estate or house price or home price) pre/5 (forecast or outlook or predict or anticipate or perspective)

Filters:

News - : News; Industry: Real Estate; Geography by Document: North America; Geography by Document: Canada; Exclusions: Exclude Stock Stories; Search Within Results: geography("canada"); Timeline: Jan 01, 2010 to Oct 31, 2023; Source Type: Newspapers, Newswires & Press Releases, Web-based Publications, Industry Trade Press, Magazines & Journals, News, Blogs, Newsletters, UNDEFINED, News Transcripts, Business Opportunities, Aggregate News Sources, Law Reviews & Journals, Market Research Reports;

Prompt to Extract Sentiment

```
prompt = """
A news article enclosed within triple backticks will be provided to you.
Your task is to extract the sentiment conveyed within this article toward the future prices of Canada's real estate.
You should analyze the text for {time1} and {time2}.

Perform the following actions:
1 - Extract and present up to ten sentences from the news that pertain directly to the future prices of Canada's real estate.
2 - Formulate the sentiment in at most five sentences for two horizons.
3 - Then represent the sentiments in a JSON object with "{time1}" and "{time2}" as the keys. Fill the JSON values with one of the following options: [strongly negative, weakly negative, neutral, weakly positive, strongly positive]. In the absence of the required sentiment in the text for each horizon, simply return 'N/A' for the JSON value.

Please follow the given format:
Related parts: <up to ten related parts>
Sentiment for {time1}: <up to five sentences>
Sentiment for {time2}: <up to five sentences>
Output JSON: <JSON object>

News article:
```{text}```
"""

message = [{"role": "system",
        "content": "You are a real estate analyst who extracts news sentiment
                about future national real estate prices in Canada."},
       {"role": "user", "content": prompt}]
```
```

News Articles Graphs

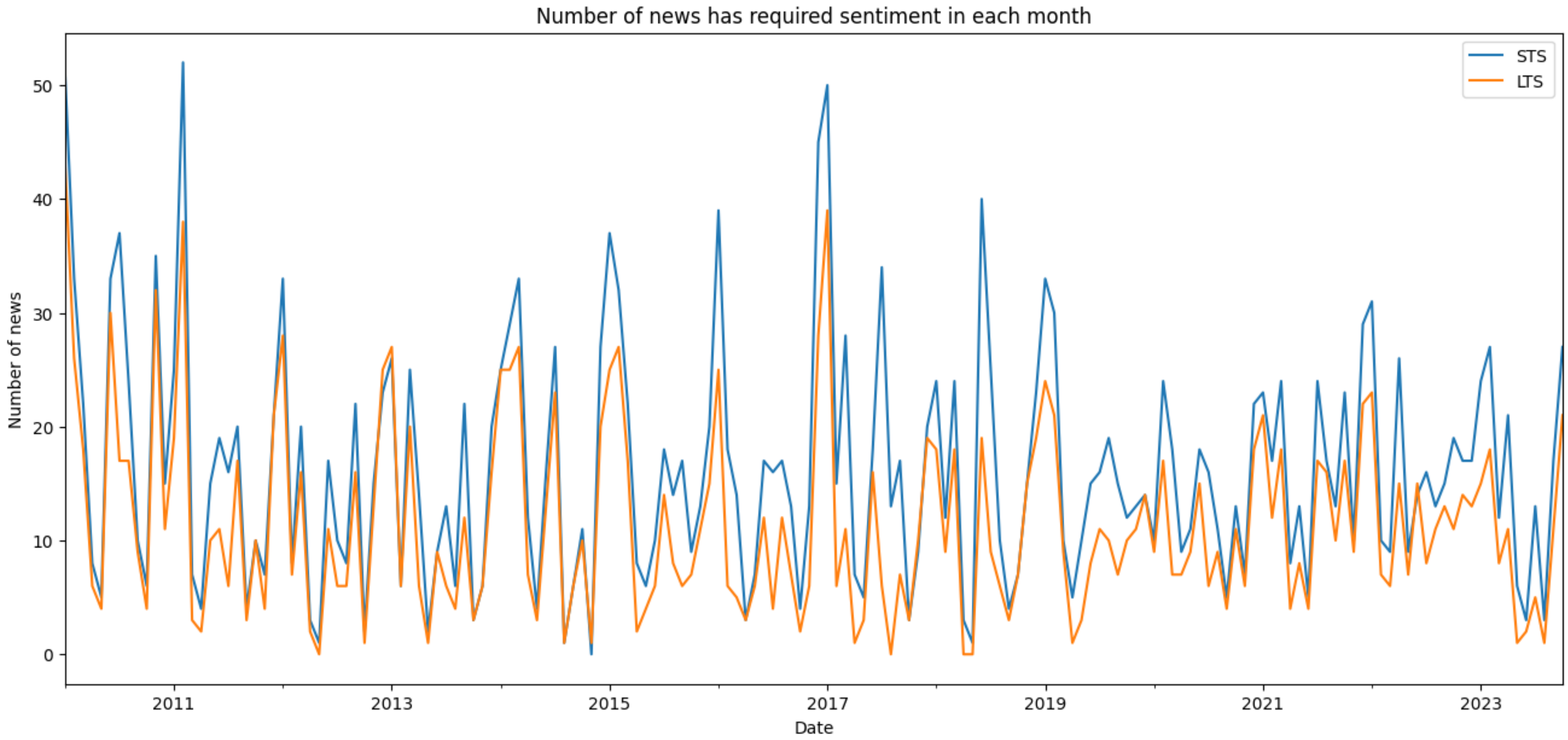

Figure 3: The number of news articles that have sentiment in each month. Time series plot depicting the monthly count of news articles containing short-term sentiment (STS, shown in blue) and long-term sentiment (LTS, shown in orange) from January 2010 to October 2023. The figure highlights fluctuations in the number of sentiment-driven articles over time, with visible variations in both STS and LTS across different periods. It is important to note that when the number of news articles containing sentiment is relatively low, the resulting sentiment index may become sensitive to individual articles, potentially making it a less reliable indicator of overall market sentiment.

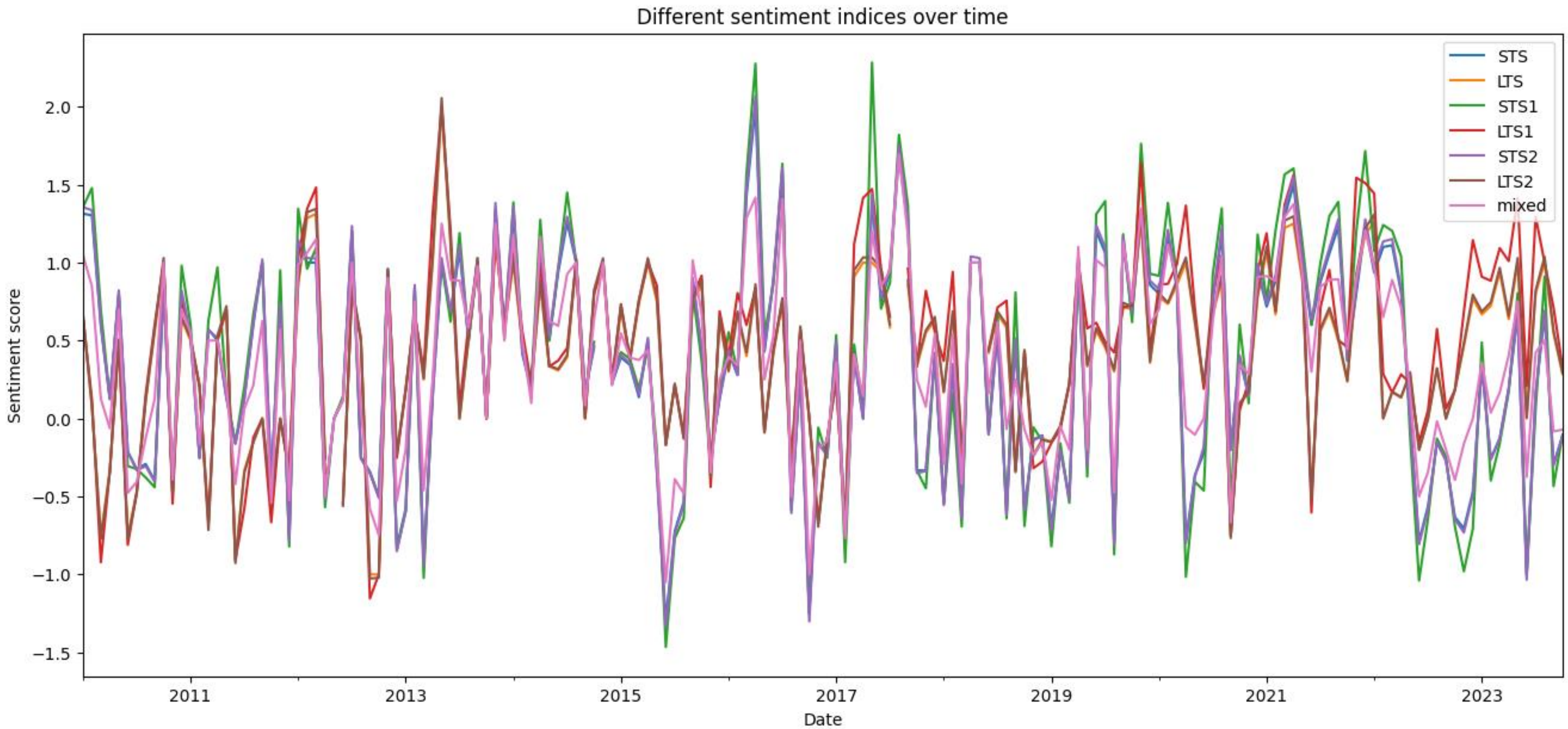

Figure 4: Time series of sentiment indices derived from news. This figure illustrates the time series of various sentiment indices derived from news articles, spanning from January 2010 to October 2023. The indices include different versions of short-term sentiments (STS, STS1, STS2), long-term sentiments (LTS, LTS1, LTS2), and a combined average of short-term and long-term sentiments (mixed). The plot captures the dynamic fluctuations in sentiment over time, with all indices generally following similar trends. However, a notable divergence among the sentiment indices appears toward the end of the period, suggesting increasing variation in sentiment as reflected by the different methodologies used to calculate these indices.

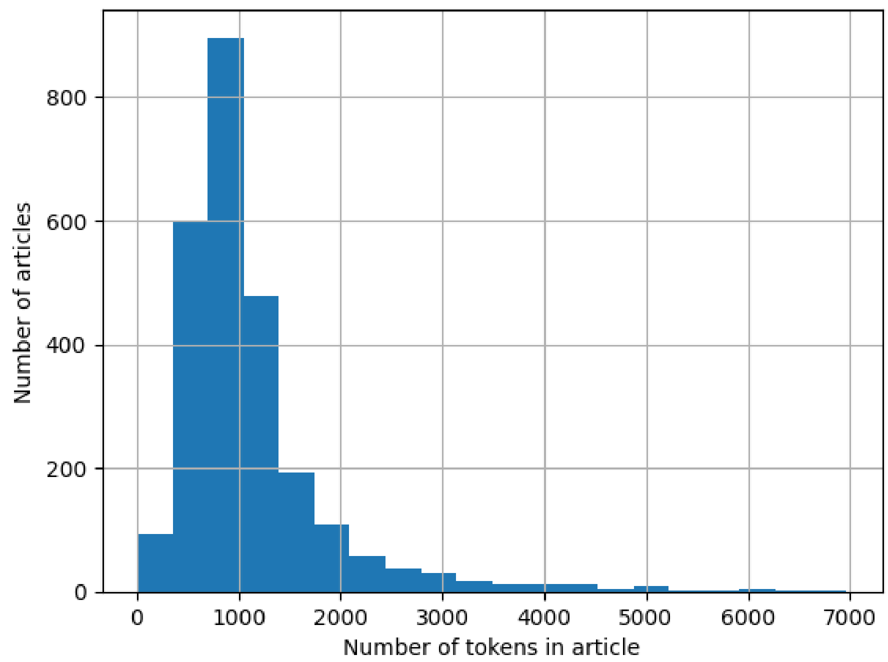

Figure 5: Distribution of news article lengths. This histogram displays the distribution of news article lengths, measured by the number of tokens, for 3,960 news articles related to Canadian real estate published between January 2010 and October 2023. The majority of the articles are relatively short, with token counts primarily ranging between 0 and 2,000, and a notable peak around 1,000 tokens. The distribution is skewed with a long tail to the right, indicating that while most articles are concise, a small number of articles contain significantly higher token counts, suggesting considerable variation in article lengths across the dataset.

Bubble Indicators Graphs

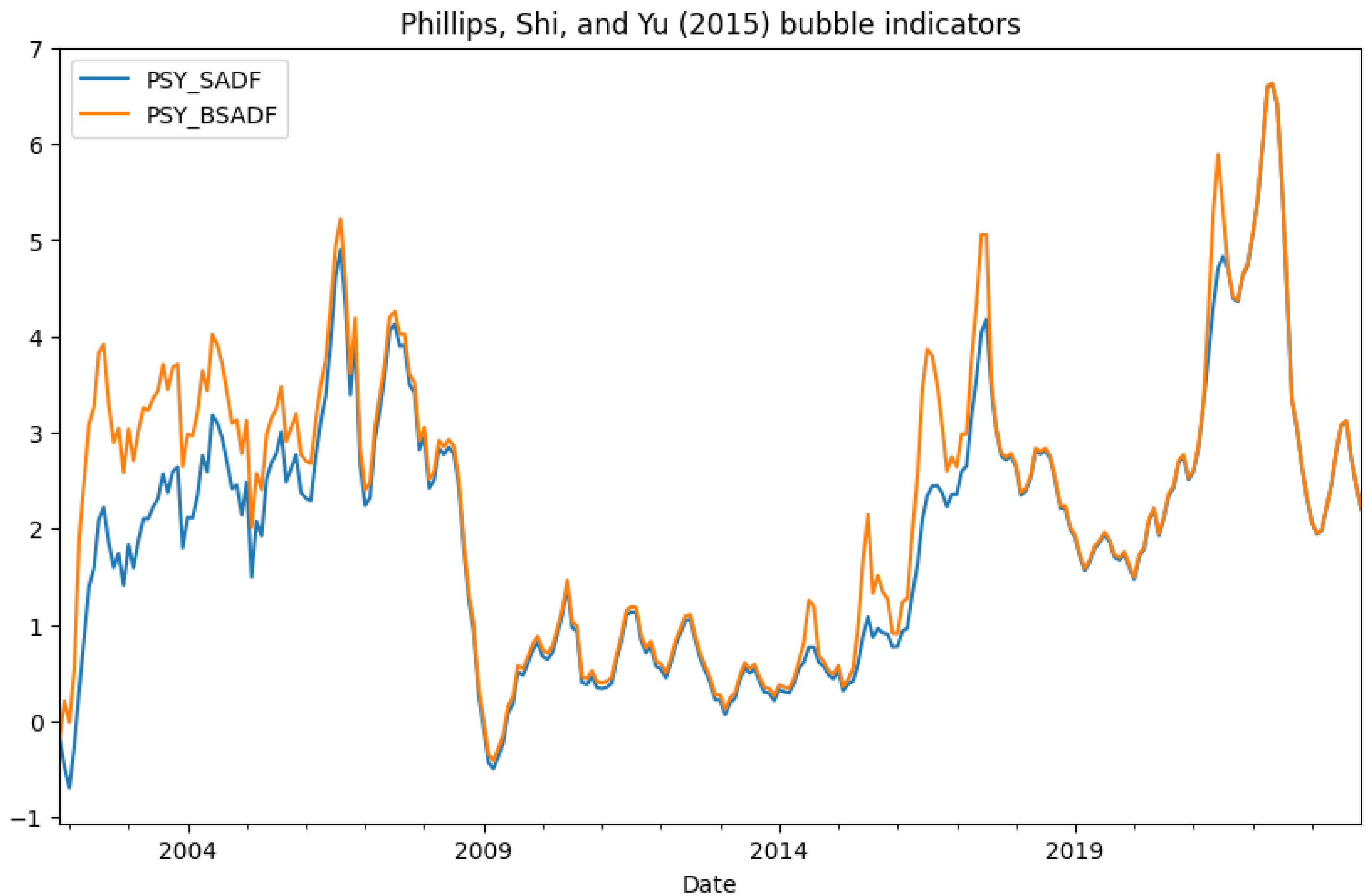

Figure 6: Phillips, Shi, and Yu (2015) bubble indicators. This time series plot displays the distribution of their two test statistics—PSY_SADF (used to detect a single bubble, shown in blue) and PSY_BSADF (used to detect multiple bubbles, shown in orange)—from November 2001 to November 2023. These bubble indicators are used to detect bubbles in the Canadian real estate market.

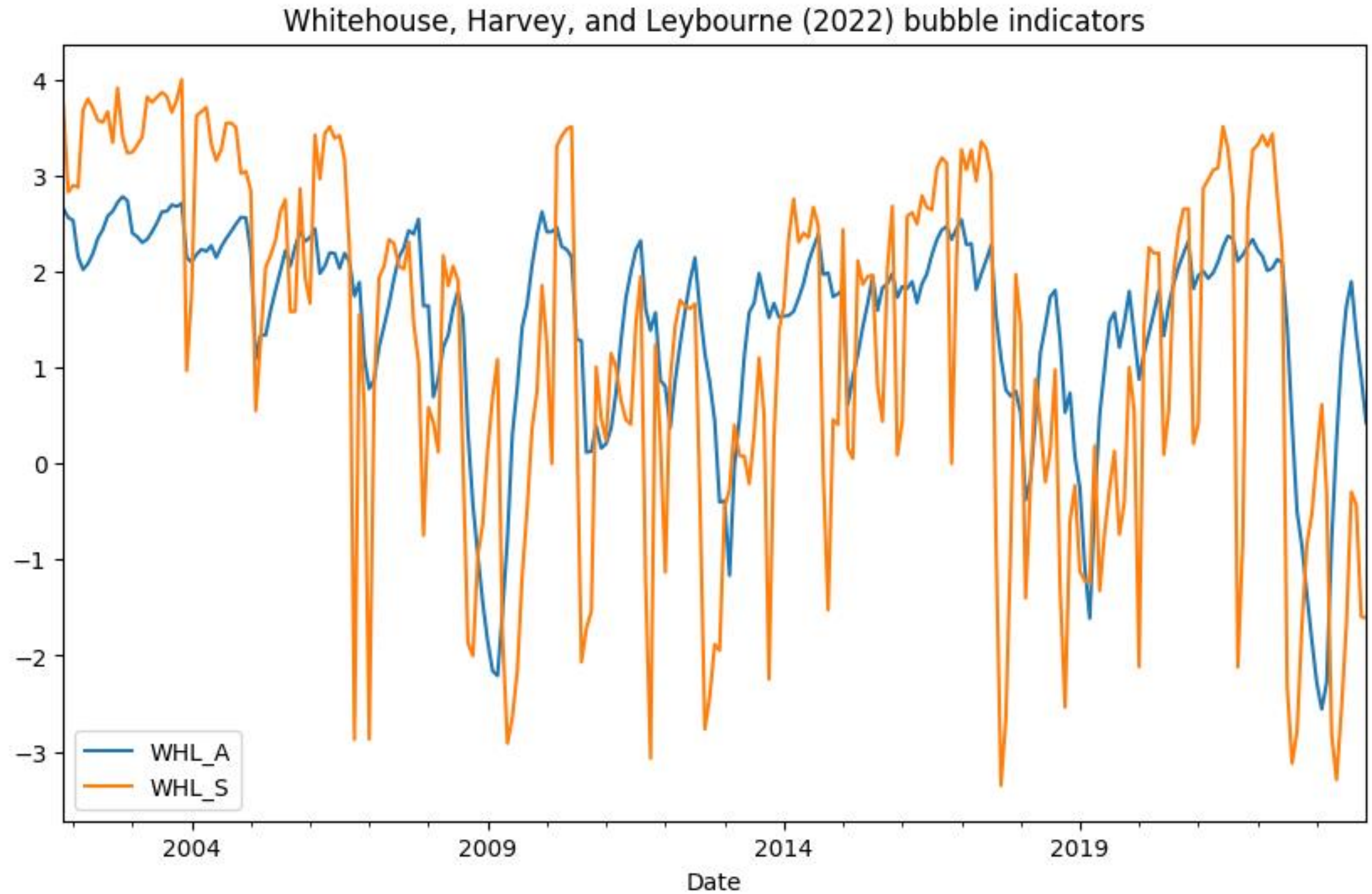

Figure 7: Whitehouse, Harvey, and Leybourne (2022) bubble indicators. This time series plot illustrates the distribution of two WHL test statistics—WHL_A and WHL_S—spanning from November 2001 to November 2023, specifically applied to the Canadian real estate market. The WHL_A statistic is designed to detect emerging bubbles. In contrast, the WHL_S statistic identifies collapsing bubbles. Compared to PSY indicators, WHL indicators exhibit greater volatility, reflecting their notable sensitivity to market fluctuations and more frequent detection of both emerging and collapsing bubbles.

Forecast Power of 10 HSIs Forming UHSI_1

Forecast Horizon	Predictor	m	n	MAFE	MSFE	MAFE improvement	MSFE improvement
h=1	**None**	3	0	0.0084975	0.00010089	0.0%	0.0%
	hsi_1_12_3	3	3	0.0077942	8.19E-05	8.3%	18.8%
	hsi_1_12_31	3	3	0.00809595	8.98E-05	4.7%	11.0%
	hsi_1_12_32	3	3	0.00761219	8.86E-05	10.4%	12.2%
	hsi_1_12_59	3	3	0.00760891	8.64E-05	10.5%	14.4%
	hsi_1_12_38	3	3	0.00731844	8.01E-05	13.9%	20.6%
	hsi_1_12_51	3	3	0.00789221	8.82E-05	7.1%	12.6%
	hsi_1_12_47	3	3	0.00720362	8.02E-05	15.2%	20.6%
	hsi_1_12_19	3	3	0.00760326	8.17E-05	10.5%	19.1%
	hsi_1_12_21	3	3	0.0080046	8.96E-05	5.8%	11.2%
	hsi_1_12_28	3	3	0.00794011	8.42E-05	6.6%	16.5%
h=3	**None**	3	0	0.02292311	0.00081504	0.0%	0.0%
	hsi_1_12_3	3	3	0.02091715	0.00063144	8.8%	22.5%
	hsi_1_12_31	3	3	0.02114254	0.00066815	7.8%	18.0%
	hsi_1_12_32	3	1	0.02155829	0.00071975	6.0%	11.7%
	hsi_1_12_59	3	3	0.01988789	0.00059969	13.2%	26.4%
	hsi_1_12_38	3	3	0.02038862	0.00063708	11.1%	21.8%
	hsi_1_12_51	3	1	0.02137553	0.00070674	6.8%	13.3%
	hsi_1_12_47	3	1	0.02093178	0.00066469	8.7%	18.4%
	hsi_1_12_19	3	1	0.02090397	0.00064868	8.8%	20.4%
	hsi_1_12_21	3	1	0.02178039	0.00074417	5.0%	8.7%
	hsi_1_12_28	3	3	0.02055814	0.00062185	10.3%	23.7%
h=6	**None**	3	0	0.0084975	0.00010089	0.0%	0.0%
	hsi_1_12_3	3	3	0.0077942	8.19E-05	8.3%	18.8%
	hsi_1_12_31	3	3	0.00809595	8.98E-05	4.7%	11.0%
	hsi_1_12_32	3	3	0.00761219	8.86E-05	10.4%	12.2%
	hsi_1_12_59	3	3	0.00760891	8.64E-05	10.5%	14.4%
	hsi_1_12_38	3	3	0.00731844	8.01E-05	13.9%	20.6%
	hsi_1_12_51	3	3	0.00789221	8.82E-05	7.1%	12.6%
	hsi_1_12_47	3	3	0.00720362	8.02E-05	15.2%	20.6%
	hsi_1_12_19	3	3	0.00760326	8.17E-05	10.5%	19.1%
	hsi_1_12_21	3	3	0.0080046	8.96E-05	5.8%	11.2%
	hsi_1_12_28	3	3	0.00794011	8.42E-05	6.6%	16.5%
h=12	**None**	3	0	0.0084975	0.00010089	0.0%	0.0%
	hsi_1_12_3	3	3	0.0077942	8.19E-05	8.3%	18.8%

Forecast Horizon	Predictor	m	n	MAFE	MSFE	MAFE improvement	MSFE improvement
	hsi_1_12_31	3	3	0.00809595	8.98E-05	4.7%	11.0%
	hsi_1_12_32	3	3	0.00761219	8.86E-05	10.4%	12.2%
	hsi_1_12_59	3	3	0.00760891	8.64E-05	10.5%	14.4%
	hsi_1_12_38	3	3	0.00731844	8.01E-05	13.9%	20.6%
	hsi_1_12_51	3	3	0.00789221	8.82E-05	7.1%	12.6%
	hsi_1_12_47	3	3	0.00720362	8.02E-05	15.2%	20.6%
	hsi_1_12_19	3	3	0.00760326	8.17E-05	10.5%	19.1%
	hsi_1_12_21	3	3	0.0080046	8.96E-05	5.8%	11.2%
	hsi_1_12_28	3	3	0.00794011	8.42E-05	6.6%	16.5%

Table 8: Forecast power of 10 components of UHSI_1. This table reports the optimal number of lags (m, n), MAFE, and MSFE for predicting the seasonally and trend-adjusted logarithm of the House HPI in Canada using autoregressive models applied to the same training and validation sample. The models evaluated include a base model with no external predictors and several models that incorporate 10 components of UHSI_1 as external predictors. Additionally, the table presents the percentage improvement in forecast accuracy (MAFE and MSFE) of the models with predictors relative to the base model.

MAFE Improvement of HSIs in B_h sets

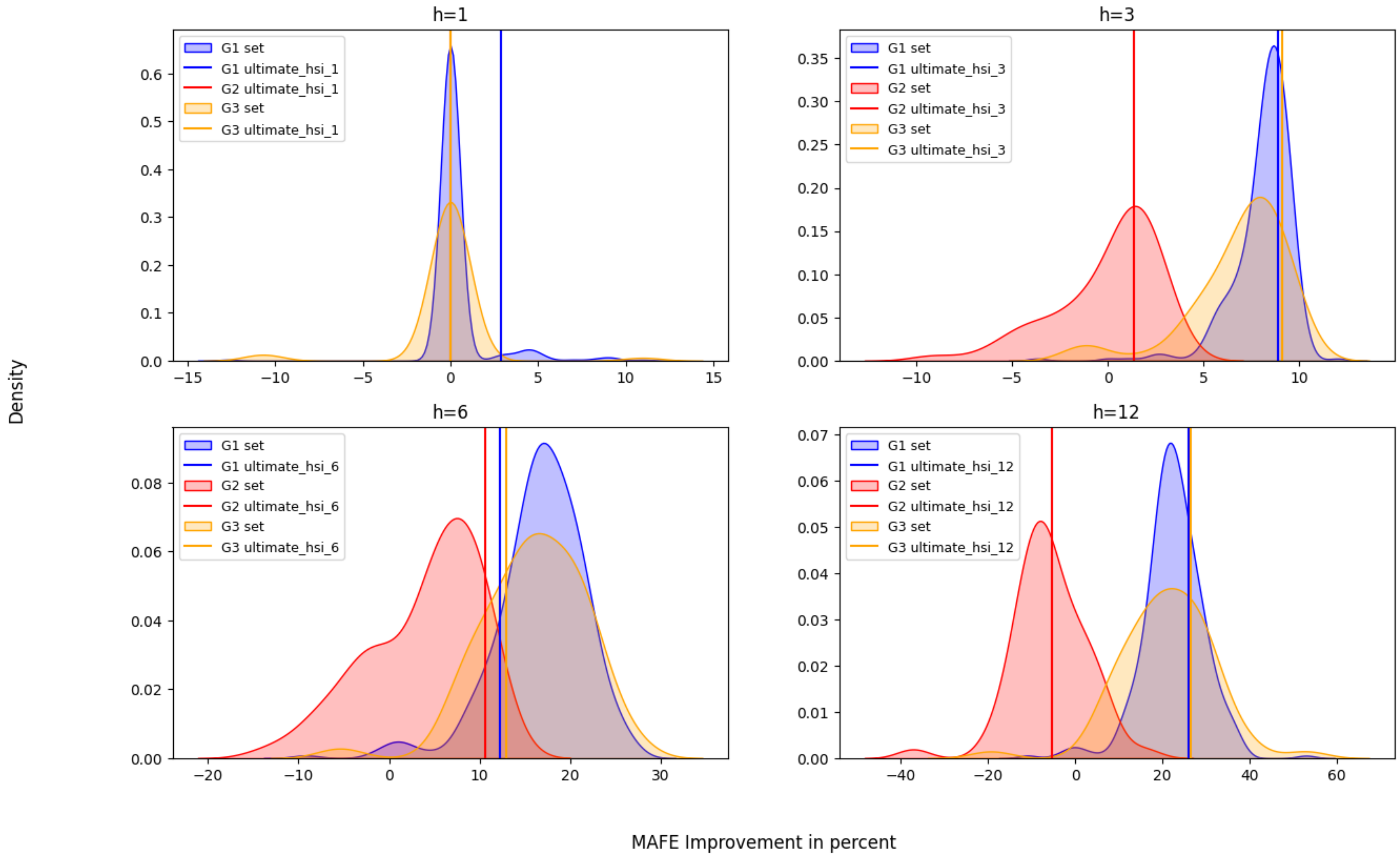

Figure 8: KDE distribution of MAFE improvements for all forecast horizons. This Kernel Density Estimation (KDE) plot illustrates the distribution of MAFE improvement percentages relative to the base model at all forecast horizons for three groups of data: relevant keywords (blue), irrelevant keywords (red), and a combination of relevant and irrelevant keywords (orange). The figure shows the MAFE improvement distributions for the HSIs in the B_h set, along with vertical lines representing the MAFE improvement associated with the UHSIs for each group. The results indicate that the distributions for the relevant keywords group and the combination of relevant and irrelevant keywords group are similar, both showing distinct patterns compared to the irrelevant keywords group, which demonstrates a notably different distribution.

MSFE Improvement of HSIs in B_h sets

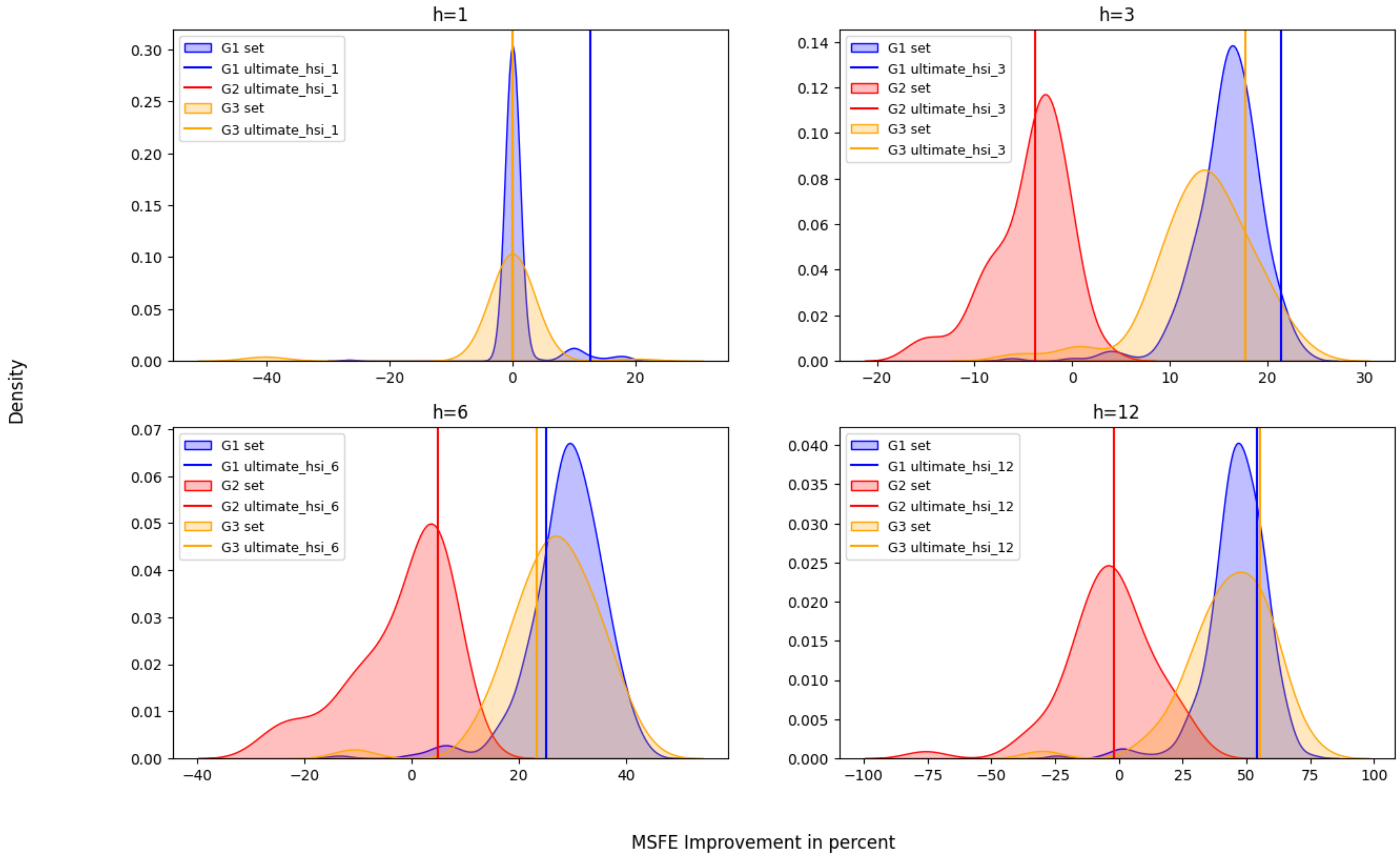

Figure 9: KDE distribution of MSFE improvements for all forecast horizons. This Kernel Density Estimation (KDE) plot illustrates the distribution of MSFE improvement percentages relative to the base model at all forecast horizons for three groups of data: relevant keywords (blue), irrelevant keywords (red), and a combination of relevant and irrelevant keywords (orange). The figure shows the MSFE improvement distributions for the HSIs in the B_h set, along with vertical lines representing the MSFE improvement associated with the UHSIs for each group. The results indicate that the distributions for the relevant keywords group and the combination of relevant and irrelevant keywords group are similar, both showing distinct patterns compared to the irrelevant keywords group, which demonstrates a notably different distribution.

Google Trends Preprocessing Graphs for h=12

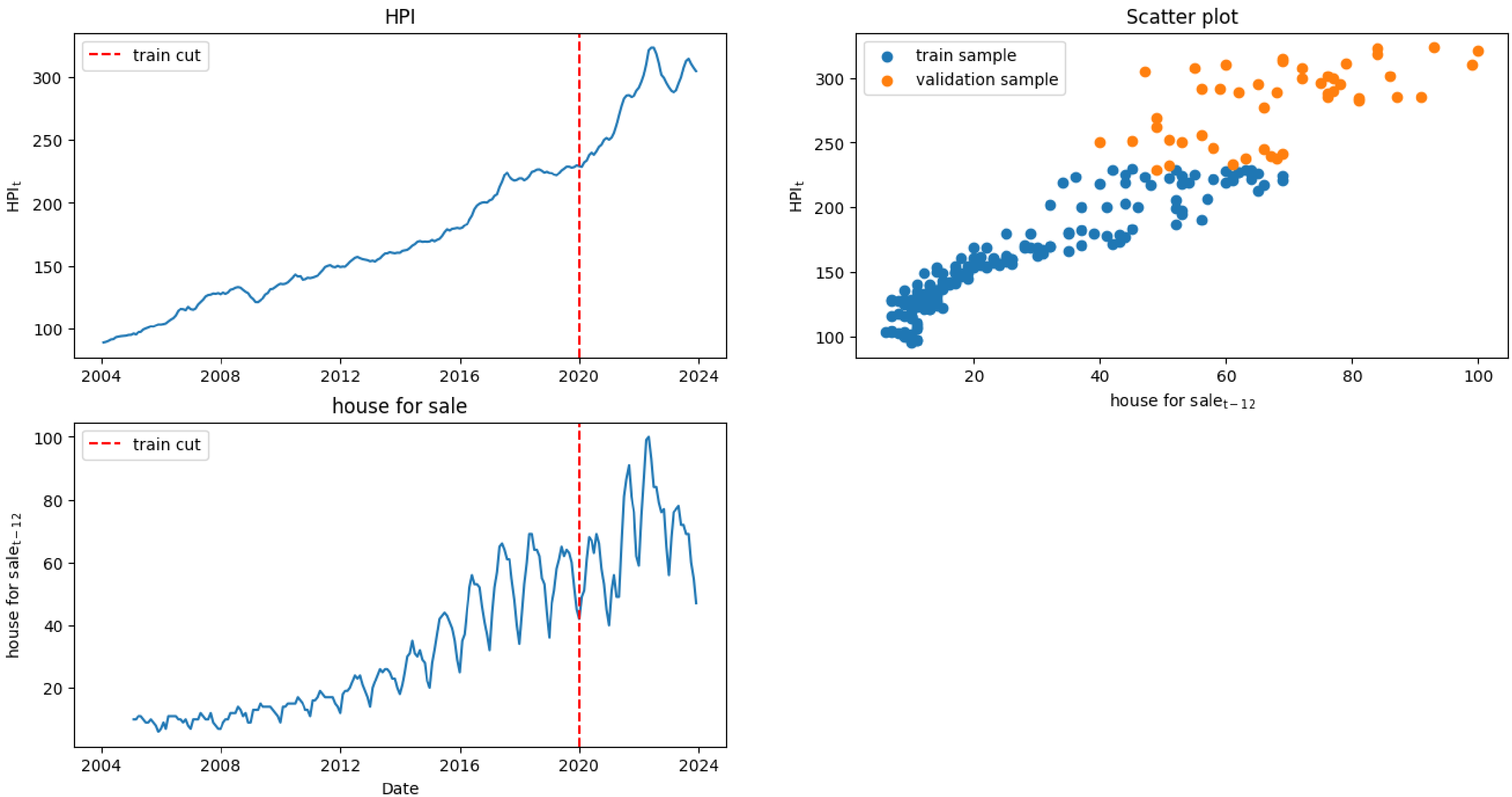

Figure 10: Google Trends preprocessing steps at h=12: Original data. This figure displays the original HPI and the "house for sale" query from Google Trends as an example of the raw time series data before any preprocessing, focusing on data from January 2004 to November 2023 in Canada. The left plots show the time series of HPI and the 12-month lagged "house for sale" query, with a red dashed line indicating the separation between the training and validation datasets. The 12-month lag is used because the analysis aims to investigate the forecasting procedure for a 12-month horizon (h=12). The scatter plot on the right illustrates the relationship between the HPI and the lagged "house for sale" query. Both the HPI and the query exhibit seasonality and a time trend, particularly in the query data.

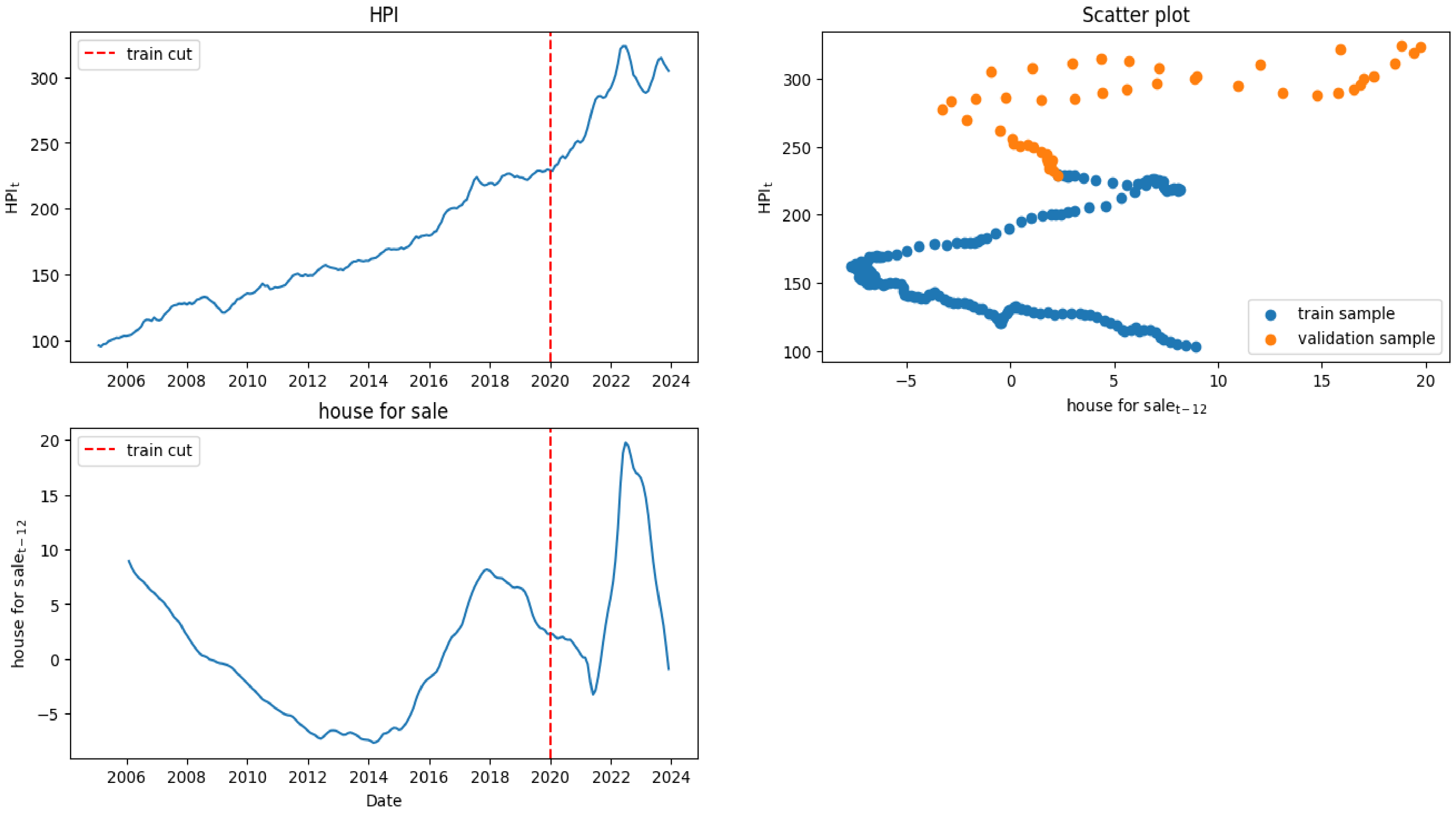

Figure 11: Google Trends preprocessing steps at h=12: After preprocessing "house for sale" query. This figure presents time series plots for Canadian data spanning from January 2004 to November 2023. The plots show the HPI data before preprocessing and the 12-month lag of the "house for sale" Google Trends query after preprocessing. The 12-month lag of the query is used to explore the forecasting procedure for a 12-month horizon (h=12). The red dashed line in the plots indicates the separation between training and validation datasets. While the HPI data remains unadjusted, the "house for sale" query has been preprocessed to account for seasonality and time trends.

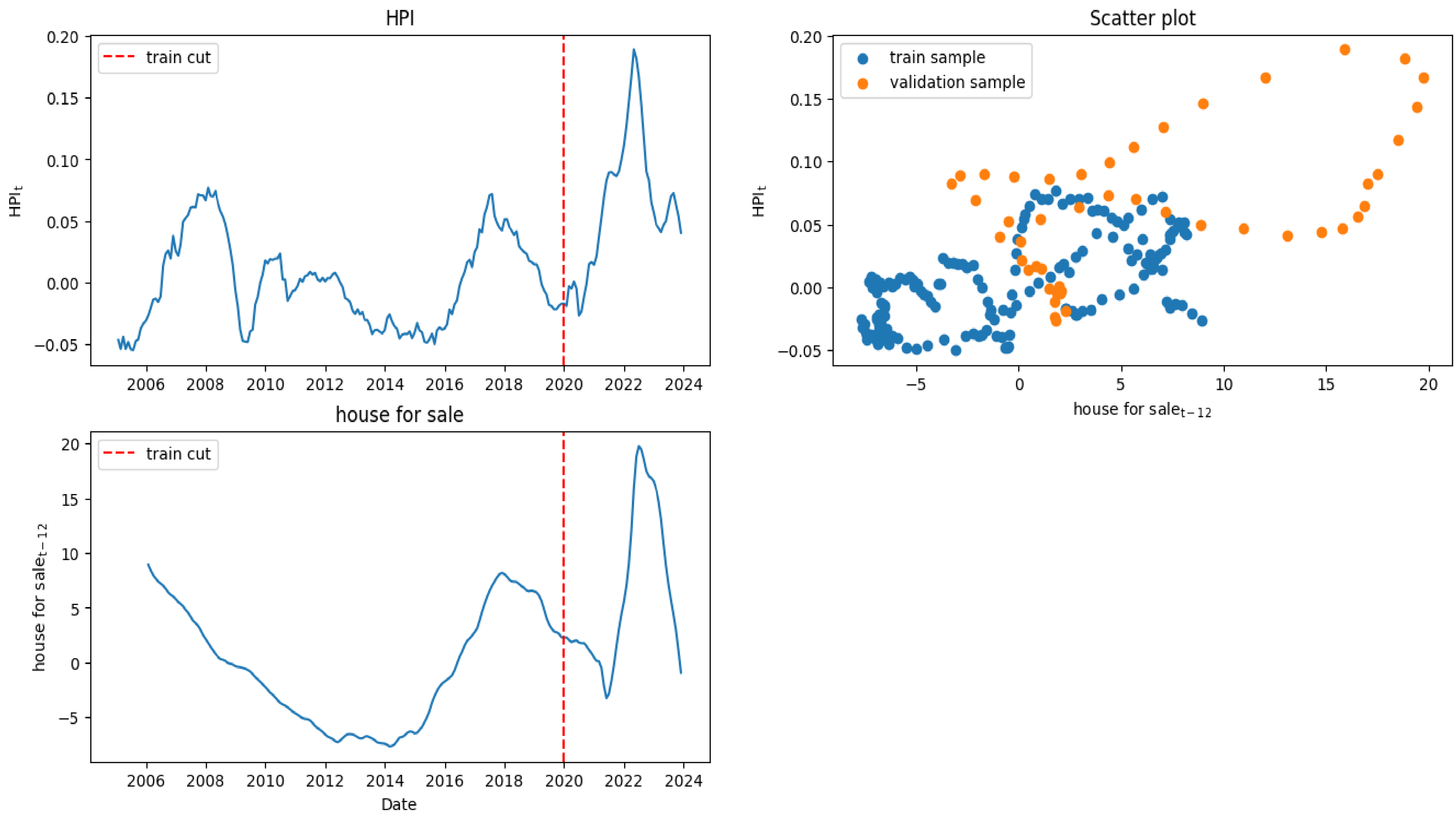

Figure 12: Google Trends preprocessing steps at h=12: After preprocessing HPI. This figure presents time series plots for Canadian data spanning from January 2004 to November 2023, showing the HPI and the 12-month lag of the "house for sale" Google Trends query after both series have been preprocessed. The preprocessing involved adjusting for seasonality and time trends in both the HPI and the query data. The 12-month lag of the query is used to explore the forecasting procedure for a 12-month horizon (h=12). The red dashed line in the plots indicates the separation between training and validation datasets. The scatter plot illustrates the correlation between the preprocessed HPI and the adjusted query, revealing a relationship between these variables after the preprocessing steps. This indicates that the "house for sale" query could be a valuable predictor for forecasting the HPI at a 12-month horizon.

Forecasting by HSIs Graphs

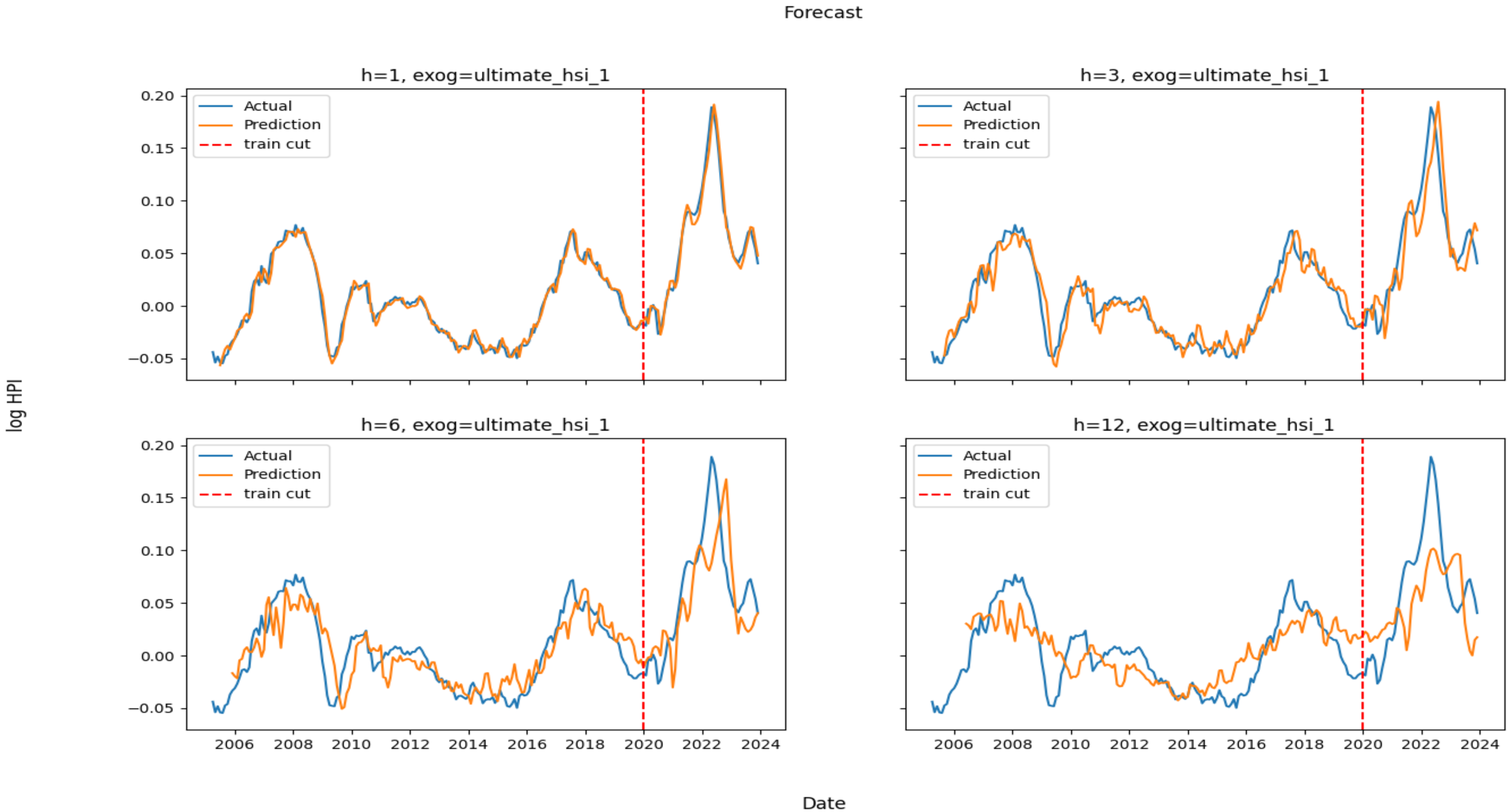

Figure 13: Forecast HPI using UHSI_1 for all horizons. This figure presents time series plots of forecasted versus actual log HPI values in Canada using the UHSI_1 as an external regressor across four forecast horizons ($h \in \{1, 3, 6, 12\}$). The data spans from January 2004 to November 2023. The red dashed line indicates the division between the training and validation datasets. As anticipated, the forecast errors increase with longer forecast horizons, with the predictive accuracy decreasing slightly from h=1 to h=12.

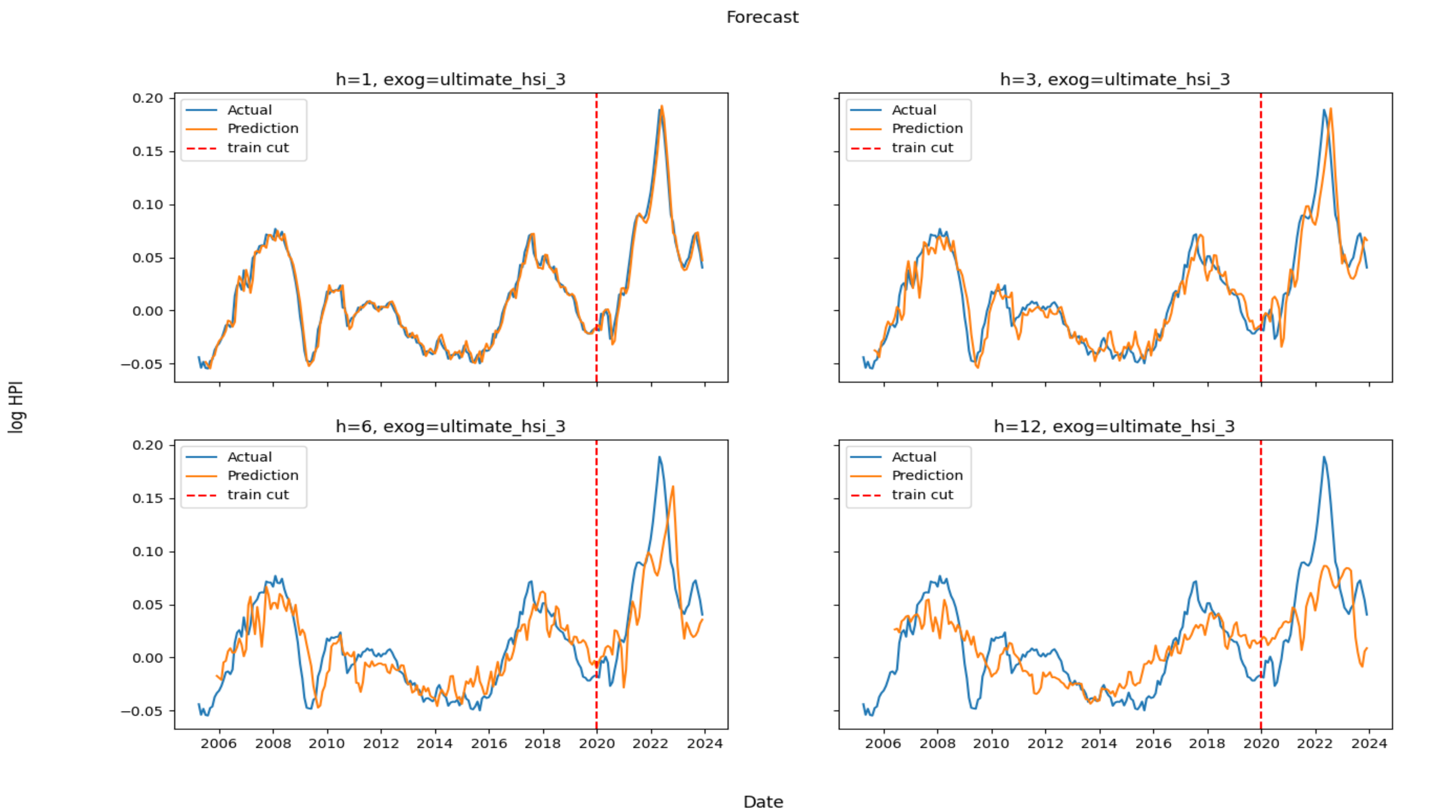

Figure 14: Forecast HPI using UHSI_3 for all horizons. This figure presents time series plots of forecasted versus actual log HPI values in Canada using the UHSI_3 as an external regressor across four forecast horizons ($h \in \{1, 3, 6, 12\}$). The data spans from January 2004 to November 2023. The red dashed line indicates the division between the training and validation datasets. As anticipated, the forecast errors increase with longer forecast horizons, with the predictive accuracy decreasing slightly from h=1 to h=12.

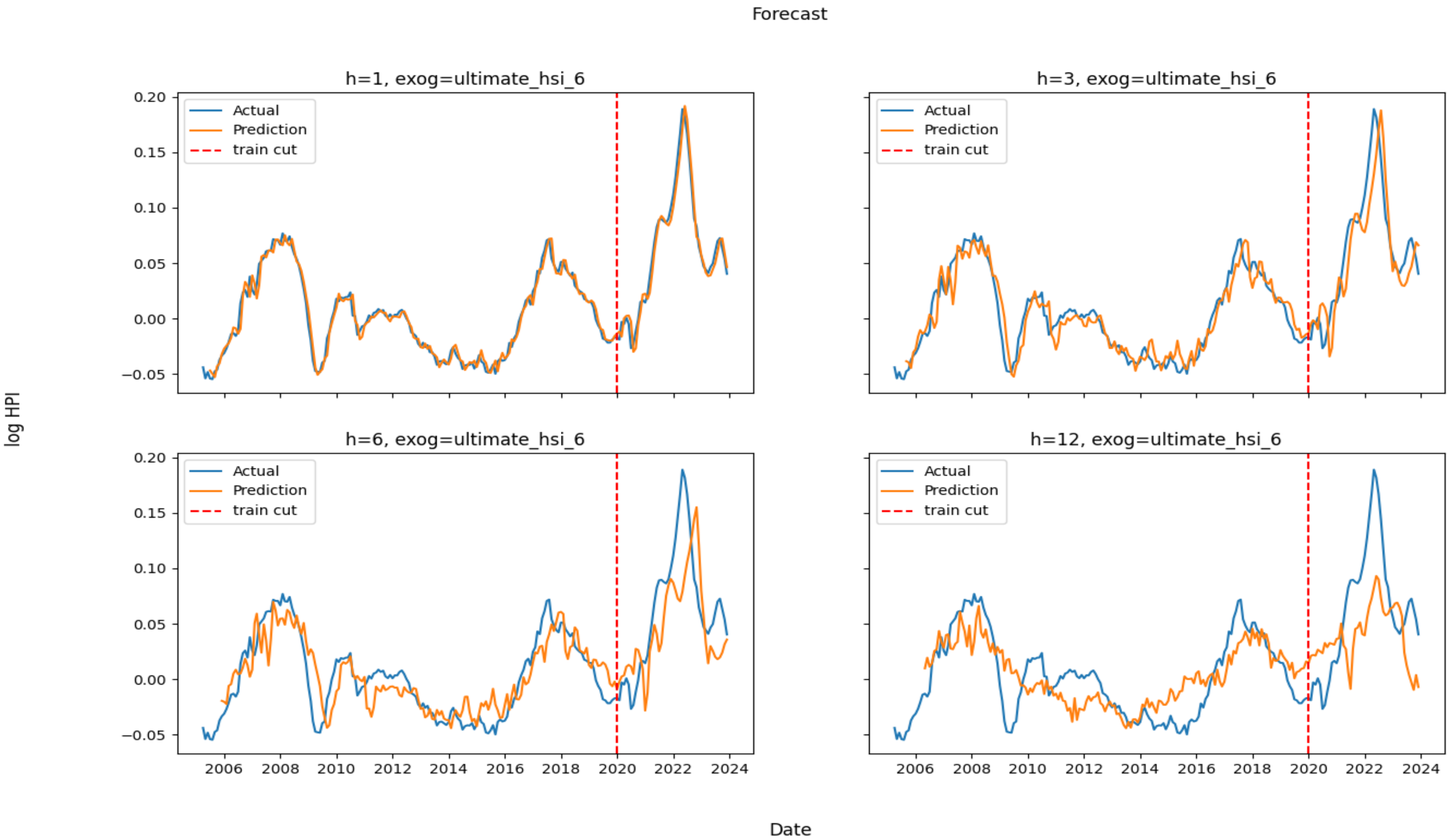

Figure 15: Forecast HPI using UHSI_6 for all horizons. This figure presents time series plots of forecasted versus actual log HPI values in Canada using the UHSI_6 as an external regressor across four forecast horizons ($h \in \{1, 3, 6, 12\}$). The data spans from January 2004 to November 2023. The red dashed line indicates the division between the training and validation datasets. As anticipated, the forecast errors increase with longer forecast horizons, with the predictive accuracy decreasing slightly from h=1 to h=12.

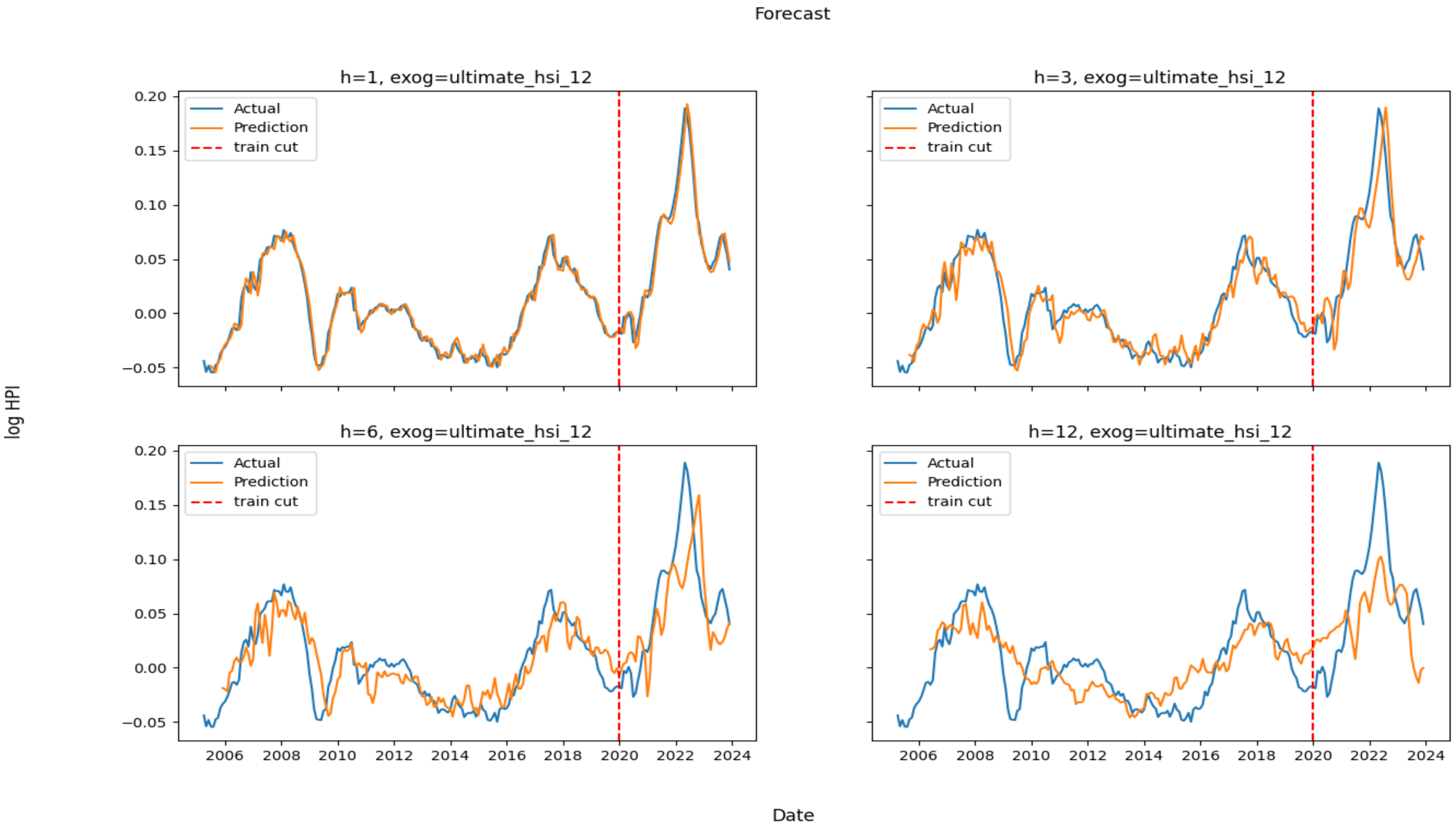

Figure 16: Forecast HPI using UHSI_12 for all horizons. This figure presents time series plots of forecasted versus actual log HPI values in Canada using the UHSI_12 as an external regressor across four forecast horizons ($h \in \{1, 3, 6, 12\}$). The data spans from January 2004 to November 2023. The red dashed line indicates the division between the training and validation datasets. As anticipated, the forecast errors increase with longer forecast horizons, with the predictive accuracy decreasing slightly from h=1 to h=12.