Creative Business and Sustainability Journal (CBSJ)
Vol.47 No.1 January – June 2025, pp.88-108

Nowcasting the Condominium Price Index Using Google Search Data

Pornpun Ruankong*

Abstract

As land prices continue to escalate, the demand for condominiums has correspondingly increased. However, if this increasing demand is driven by speculation, condominium prices may not accurately reflect actual market demand. Analyzing trends in the Condominium Price Index is crucial for entrepreneurs, investors, and the public to make informed decisions. Additionally, understanding these trends helps clarify the relationship between the real estate cycle and the business cycle, both of which serve as indicators of economic downturns and recoveries.

This study investigates the potential of Google Trends as a leading indicator for the Condominium Price Index by employing a nowcasting model. Unlike previous research, this study adopts mixed-data sampling (MIDAS) techniques to incorporate data with varying frequencies. The empirical findings indicate that integrating macroeconomic variables and Google Trends data into autoregressive (AR) models enhances their explanatory power. Furthermore, the Augmented Distributed Lag MIDAS (ADL-MIDAS) model demonstrates superior forecasting performance, particularly in atypical market conditions.

Keywords: Condo Price Index; Nowcasting; Google Trends; Mixed Frequency Data

Received: March 28, 2025 | **Revised:** June 11, 2025 | **Accepted:** June 24, 2025

* Department of Economics, Faculty of Humanities and Social Sciences, Burapha University, Thailand, Email: pornpun.ru@buu.ac.th. ORCID: 0009-0002-3334-5577.

Creative Business and Sustainability Journal (CBSJ)
Vol.47 No.1 January – June 2025, pp.88-108

Introduction

The Internet has become an integral part of daily life for most individuals, fundamentally transforming the way people communicate. It is now utilized across nearly all activities, with its role becoming particularly prominent during the COVID-19 pandemic. For instance, the Internet facilitates teleconferencing, online learning, e-commerce, and various financial transactions. Additionally, individuals frequently browse the Internet for leisure and seek information through popular search engines, such as Google, which is the most widely used search engine in Thailand[1]. As a result, an extensive amount of data is generated and stored in the form of Google Trends, which has broad applications in various economic fields. This study, however, focuses specifically on leveraging Google Trends data for nowcasting of the Condominium Price Index in Bangkok and its vicinity in Thailand.

Housing is one of the four fundamental necessities of life. In the past, owning and residing in a condominium was less prevalent due to its limited usable space and the availability of various alternative housing options. When comparing the cost of homeownership in terms of living space and landownership, a detached house was often considered a more viable alternative. However, with increasing urban population density and longer daily commuting times, there has been a notable shift in housing preferences. This shift is a worldwide phenomenon, particularly evident in high-density urban areas, where the younger generation is increasingly inclined to purchase condominiums located near their workplaces (Rosen & Walks, 2013).

The key advantages of condominiums include their prime locations with convenient access to public transportation, comprehensive amenities, enhanced security, and greater privacy. This trend became even more pronounced during the COVID-19 pandemic, as many organizations adopted remote work policies. Living in a shared household with family members presented challenges, including limited personal space and an increased risk of infection within the household. Consequently, the demand for condominiums has risen significantly.

It is essential, however, to distinguish between two types of condominium demand: real demand and artificial demand. Real demand refers to purchases made for actual residential use. In this case, higher demand naturally drives up prices according to market mechanisms. Artificial demand, on the other hand, arises from investors purchasing condominiums for resale or rental purposes. This speculative activity can lead to price distortions, where rising prices do not accurately reflect genuine demand. Such conditions may result in the bubble bursting, as observed in the 1997 financial crisis. Research indicates that the business cycle and real estate price cycle are closely interconnected (Wu & Brynjolfsson, 2015). A real estate bubble is often followed by an economic downturn, whereas a recovering real estate market signals economic expansion. Therefore, analyzing real estate price cycles provides valuable insights into broader economic situations.

To effectively track the signals and trends of real estate price cycles, this study employs Google search data. Google Trends serves as a suitable leading indicator for forecasting models, as individuals typically conduct online searches before making purchase or investment decisions. These search patterns, in turn, reflect shifts in real estate supply and demand, ultimately influencing market prices. Furthermore, earlier models for forecasting house price indices have predominantly relied on low-frequency macroeconomic variables, which limit the

[1] StatCounter. (2021). Search engine market share. Retrieved April 2021, from https://www.statcounter.com

timeliness of predictions. Incorporating Google Trends data, which is available at a higher frequency, enhances the predictability of short-term market movements, thereby improving decision-making in the real estate sector.

Although the integration of Google Trends variables enhances real-time data availability, a key challenge in forecasting lies in the varying frequencies of the data. To address this issue, this study adopts the Augmented Distributed Lag Mixed-Data Sampling (ADL-MIDAS) model proposed by Andreou et al. (2013). The MIDAS framework enables efficient handling of mixed-frequency data without compromising estimation accuracy through multiple-step procedures. Additionally, it allows for the extraction of valuable insights embedded in high-frequency data.

This study aims to evaluate the effectiveness of Google Trends variables as leading indicators for the Condominium Price Index. Moreover, it aims to enhance the efficiency and accuracy of short-term forecasting models that utilize mixed-frequency data. A comprehensive understanding of market dynamics is crucial for investors, policymakers, and financial institutions to make informed decisions. Additionally, this research has practical implications for various fields, including risk management, asset pricing, and investment analysis.

The remainder of this paper is structured as follows: Section 2 presents a review of relevant literature. Section 3 describes the data and model specifications. Section 4 reports empirical findings. Finally, Section 5 concludes the study.

Literature Review

Numerous studies have examined the macroeconomic factors influencing house prices. Glindro et al. (2018) investigated the fundamental determinants of house prices across Asia-Pacific economies. They categorized macroeconomic and institutional variables into four groups: demand-side factors (e.g., real GDP, population, real mortgage rates, and mortgage credit-to-GDP ratio), supply-side factors (e.g., land supply index and real construction costs), asset-related factors (e.g., equity prices and exchange rates), and institutional factors (proxied by a principal component of four indices: business freedom, corruption, financial sector development, and property rights). Similarly, Deghi et al. (2020) identified key macro-financial variables, namely, the Financial Conditions Index, real GDP growth, the credit-to-GDP ratio, and the price-to-GDP per capita ratio, as significant during economic crises.

With the rapid expansion of the Internet of Things (IoT) and technological advancements, vast amounts of real-time data are now available. Google Trends, which has been archiving search data since 2004, is a prominent source increasingly used in nowcasting and forecasting models due to its timeliness and broad coverage of consumer intentions.

A seminal study by Choi and Varian (2009) utilized search query data to forecast economic indicators, including retail sales, automobile sales, tourism, and real estate transactions, using basic autoregressive models. They found that incorporating Google Trends variables significantly reduced forecasting errors. Building on this, Nakavachara and Lekfuangfu (2018) assessed the predictive power of Google Trends in the Thai context across three major sectors: the labor market, the real sector, and the financial sector. Their results supported the importance of Google Trends in economic forecasting models and aligned with the findings of Choi and Varian (2009).

Subsequent research has focused on identifying additional determinants that influence real estate markets. Wu and Brynjolfsson (2015) were among the first to utilize Google search data to predict house price trends and sales volume across U.S. states. Their study revealed a strong correlation between the House Price Index (HPI), the Housing Search Index (HSI), and real estate sales. Specifically, a 1% increase in the search term index within the real estate category was associated with an increase of approximately 67,220 house sales in the following quarter. Additionally, the inclusion of HSI variables in the model reduced the mean absolute error (MAE) in out-of-sample forecasts compared to models without HSI. In a related study, Dietzel et al. (2014) employed Google Trends data to forecast commercial real estate market dynamics using Vector Autoregression (VAR) models. They further conducted a Granger causality test to examine the causal relationships among variables, concluding that Google Trends enhances forecasting accuracy and serves as a leading indicator for real estate price indices.

Further advancements in this field were made by Askitas (2016), who introduced the BUSE Index, defined as the ratio of “buy” to “sell” search queries in the real estate category. Their study demonstrated that the BUSE Index inversely correlates with real estate price indices and can be utilized for nowcasting real estate price fluctuations. Similarly, Oust and Eidjord (2020) employed the Error Correction Model (ECM) to assess the predictive capacity of Google Trends as a leading indicator of real estate price indices and speculative bubbles in various U.S. states. Their findings suggested that specific search terms, such as “real estate agent,” were helpful in predicting long-term house price trends. In contrast, searches for “housing bubble” indicated the presence of a speculative bubble.

However, some studies have presented conflicting results. For example, Limnios and You (2018) applied a linear pricing model incorporating Google Trends variables to forecast real estate inflation but found no improvement in predictive accuracy. Their findings stand in contrast to earlier studies such as Choi and Varian (2009), Wu and Brynjolfsson (2015), and Dietzel et al. (2014), which reported that search query data significantly enhanced forecast performance. This discrepancy suggests that the effectiveness of Google Trends may be highly sensitive to model specification, keyword selection, and contextual factors such as market structure or consumer behavior. Linear models may struggle to capture the nonlinear and dynamic nature of search behavior, whereas more flexible approaches, such as VAR or MIDAS, are better suited to integrating high-frequency, sentiment-driven data. Additionally, variations in Internet penetration, user intent behind searches, and regional differences in online behavior may limit the generalizability of results across settings. These conflicting findings necessitate more nuanced modeling strategies and robust validation to determine the conditions under which search data can reliably improve real estate forecasts.

The widespread adoption of Internet search data for economic forecasting is primarily driven by its real-time availability and its ability to capture consumer demand without the need for additional data collection. Conceptually, Google Trends can serve as a proxy for latent demand or sentiment signals in real estate markets, as search frequency may reflect buyers’ and sellers’ intentions ahead of actual market transactions and price adjustments. This high-frequency data also enables real-time monitoring of shifts in market expectations, helping to bridge the timing gap between behavioral changes and their eventual impact on price indices. However, while such data offers substantial advantages, researchers must exercise caution in selecting appropriate methodologies and ensuring rigorous data management to avoid misinterpretation.

Given the challenges associated with mixed-frequency data, this study employs the Mixed-Data Sampling (MIDAS) approach pioneered by Ghysels, Santa-Clara, and Valkanov

(2004). MIDAS is particularly well-suited for integrating low-frequency macroeconomic variables (e.g., monthly, quarterly, or annual data) with high-frequency datasets (e.g., daily or weekly observations). Previous studies have demonstrated that MIDAS models frequently outperform traditional forecasting models. For instance, Asimakopoulos et al. (2013) applied MIDAS models to forecast annual fiscal outcomes, concluding that MIDAS-based forecasts outperformed alternative approaches. Similarly, Asgharian et al. (2013) incorporated principal components to aggregate information from multiple macroeconomic variables within a single equation, demonstrating that the GARCH-MIDAS model exhibited superior forecasting performance when integrating low-frequency data. Conrad and Loch (2015) further validated these findings by utilizing the MIDAS framework to analyze lead-lag relationships between macroeconomic variables and long-term volatility in the U.S. stock market. Their results supported the effectiveness of MIDAS in capturing economic trends.

Despite the extensive use of MIDAS models in global research, applications in the Thai context remain limited. Kingnetr et al. (2018) applied various MIDAS model specifications to forecast Thailand's quarterly GDP growth using financial variables. Their study demonstrated that MIDAS models outperformed traditional time-aggregated models when appropriate weighting schemes were applied. More recently, Wichitaksorn (2022) employed the MIDAS model to forecast key macroeconomic indicators in Thailand, such as GDP growth and inflation rates. Their findings revealed that incorporating multiple data frequencies significantly improved forecast accuracy, particularly during the COVID-19 pandemic.

While global research in this domain continues to expand, the body of literature in Thailand remains underdeveloped. Given the increasing relevance of Internet search data in economic forecasting, further exploration of its application in real estate market predictions, particularly using the MIDAS technique, remains a valuable research endeavor.

Research Methodology

Data

This study employs secondary data spanning from March 2008 to June 2021, subject to data availability. The Condominium Price Index used in this research is collected from the Residential Property Price Index and the Land Price Index for condominium-type properties, both obtained from the Bank of Thailand. This monthly index is calculated based on mortgage loan data from 17 commercial banks operating in Bangkok and its vicinities, including Bangkok, Samut Prakan, Nonthaburi, Pathum Thani, Nakhon Pathom, and Samut Sakhon.

To incorporate Google Trends data related to condominium search queries, a list of potential keywords was compiled to reflect search terms commonly used by individuals interested in purchasing condominiums. Inspired by the methodology of Venkataraman et al. (2018), Figure 1 presents examples of these keywords[2], including fundamental terms such as "คอนโด" (condo) and other related terms. A comparative analysis revealed that the term "คอนโด" (condo), represented by the blue line, exhibited the highest search frequency among all examined queries. The term "อาคารชุด" (condominium), while technically accurate, is not widely used in Thailand.

Furthermore, search terms associated with transactional activities, including "ซื้อ" (buy), "ขาย" (sale), and "เช่า" (rent), were analyzed, as illustrated in Figure 2. The findings indicate that

[2] As of July 2021.

the terms "ขาย" (sale) and "เช่า" (rent) had higher search volumes than "ซื้อ" (buy). Consequently, these terms were combined using the "+" operator[3], forming composite queries such as "คอนโด+ขาย" (condo+sale) and "คอนโด+เช่า" (condo+rent). Additionally, Google's category-based search data were incorporated into the analysis. To ensure the relevance of these variables, correlation coefficients between the Google Trends series and the Condominium Price Index were calculated (see Appendix B for details). The two Google Trends variables with the highest correlations, namely, the "Real Estate" category (correlation = 0.6644) and the search term "คอนโด+ขาย" (condo+sale) (correlation = -0.4589), were selected for inclusion in the model. The relationship between the Condominium Price Index and these selected Google Trends data is then plotted in Figure 3.

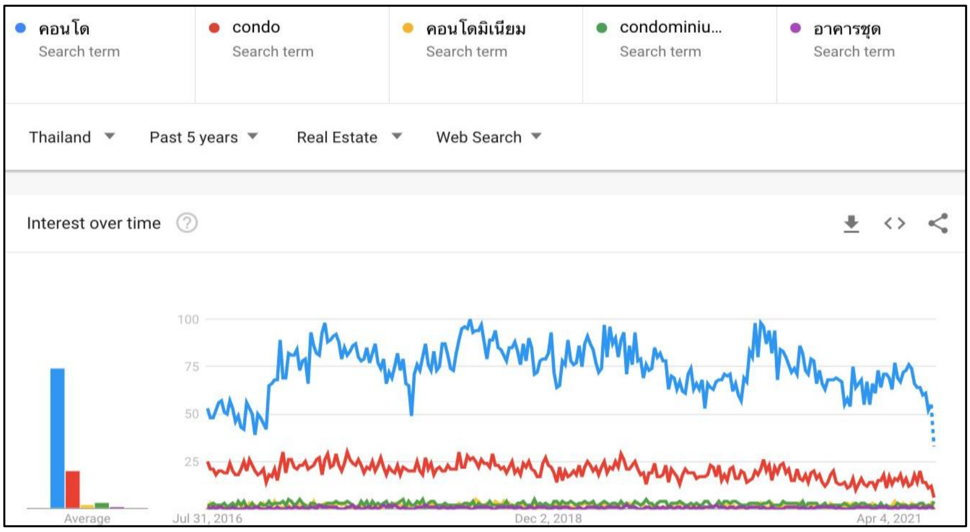


Figure 1: Comparison of Common Condominium-Related Search Terms

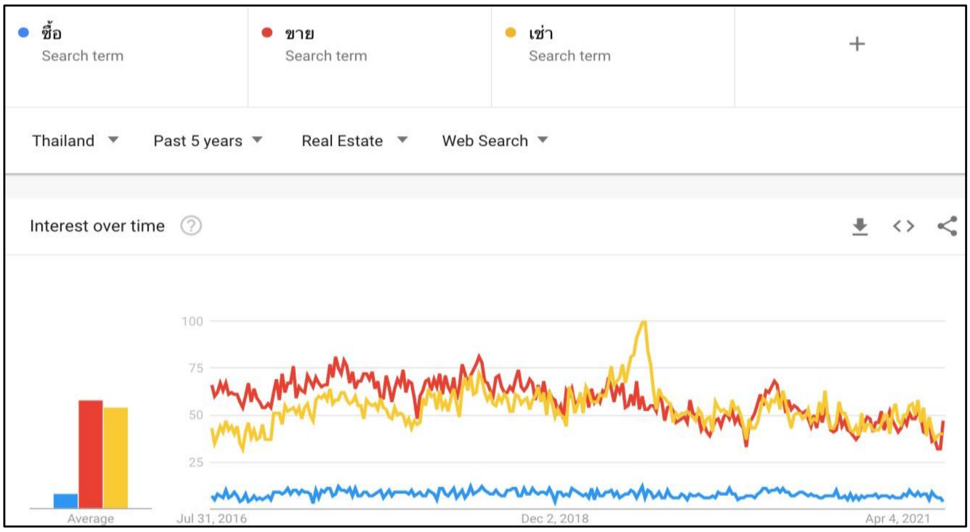
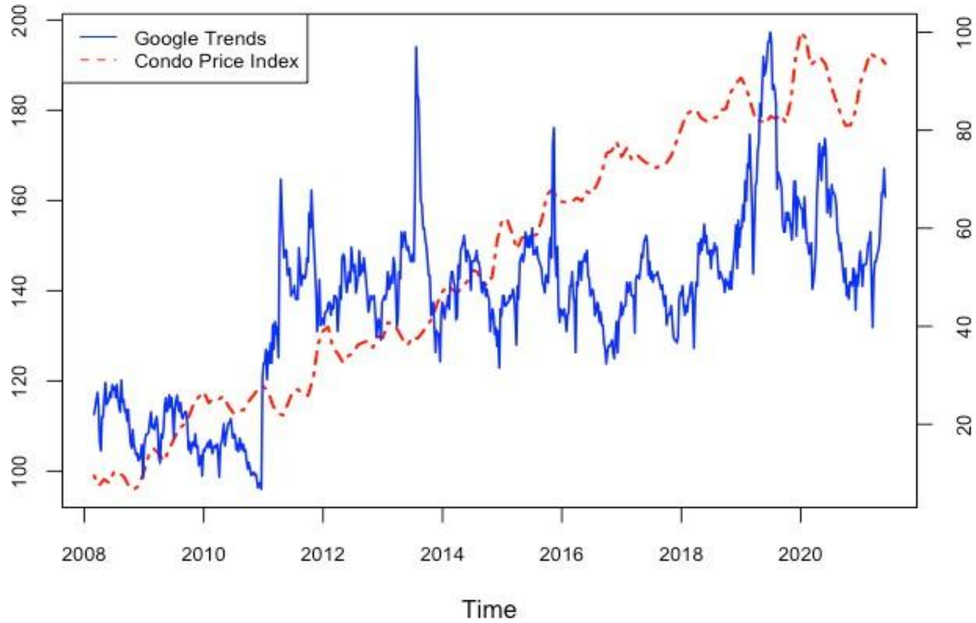
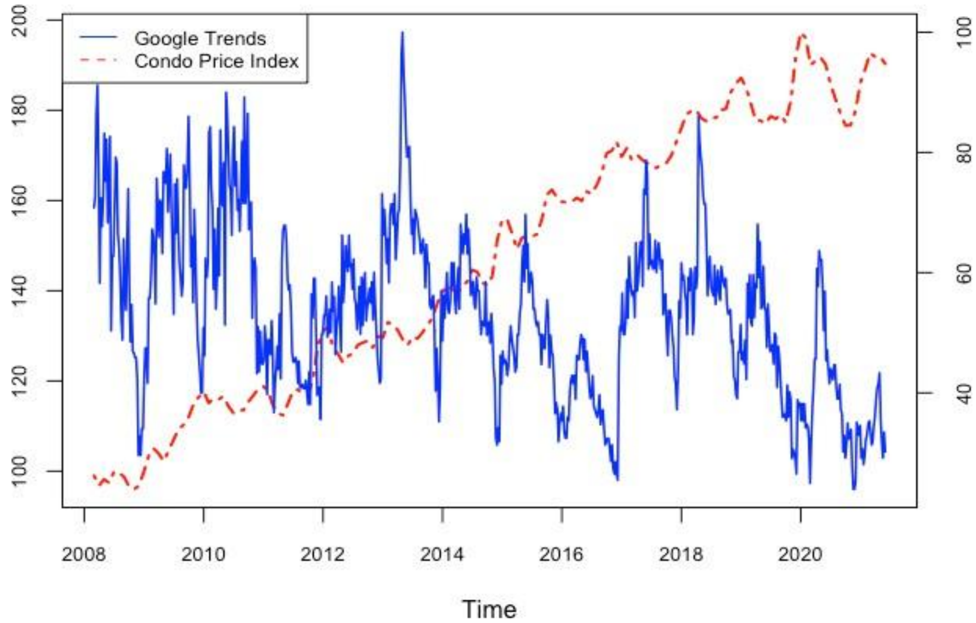


Figure 2: Comparison of Activity-Related Search Terms

[3] In search syntax: "+" means or, "–" excludes a word, a space implies and, and quotes ("") indicate a phrase match. (Stephens-Davidowitz & Varian, 2014).



(a) Google Trends: “Real Estate” category



(b) Google Trends: “คอนโด+ขาย” (condo+sale)

Figure 3: Relationship between Condo Price Index and Google Trends Data

Regarding macroeconomic variables, the study incorporates the Land Price Index, given its strong correlation with the Condominium Price Index and its extensive use in prior research. These macroeconomic indicators were obtained from the Bank of Thailand. A summary of the collected data is provided in Appendix C, while Figure 4 illustrates the relationship between the Condominium Price Index and the selected macroeconomic variables.

Since the dataset spans more than five years (March 2008–June 2021), Google Trends provides data at a monthly frequency for the entire period. To obtain weekly-level data, the

Creative Business and Sustainability Journal (CBSJ)
Vol.47 No.1 January – June 2025, pp.88-108

sample period was divided into three sub-periods: March 2008–December 2012, January 2013–December 2016, and January 2017–June 2021. However, due to Google Trends' normalization process, index values may vary across different timeframes. To address this issue, the index values were adjusted by weighting them with the monthly index derived from the full sample. An example of this calculation is presented in Appendix D.

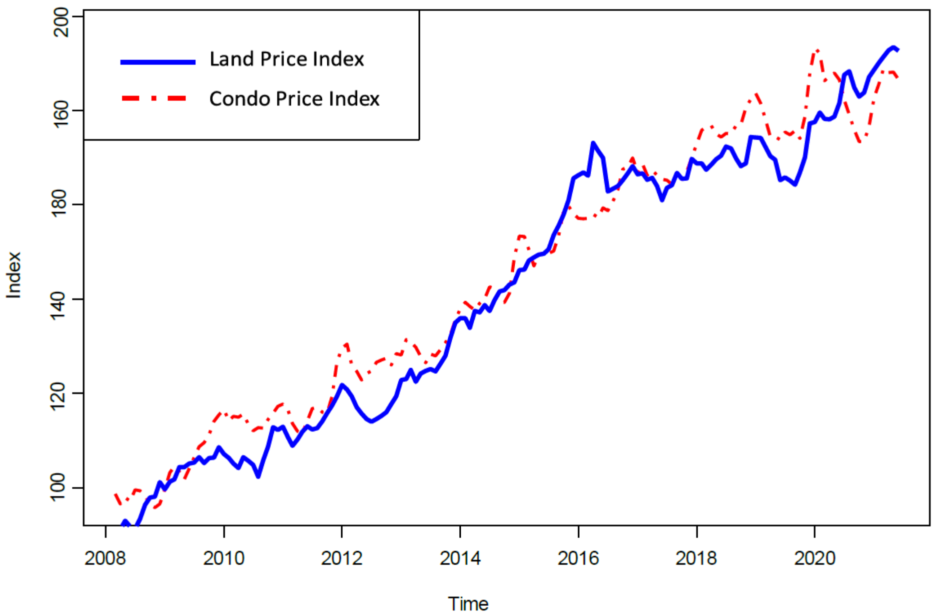


Figure 4: Relationship Between Condo Price Index and Macroeconomic Variables

Autoregressive (AR) model

Consider a simple autoregressive model, or AR(1), for seasonally adjusted data, which can be expressed as follows:

$$y_t = \mu_0 + \alpha_1 y_{t-1} + \varepsilon_t \tag{1}$$

where y_t represents the Condominium Price Index at time t, for t = 1, 2, …, T; y_{t-1} denotes the lagged Condominium Price Index at time t-1 (one month prior). The parameter μ_0 captures the constant term or baseline level of the index. The coefficient α_1 measures the degree of persistence in condominium prices by capturing the influence of the previous month's index, and ε_t is the error term.

According to the study by Choi and Varian (2009), incorporating Google Trends data into the model significantly improved predictive performance. Building on this approach, the following model is estimated and compared against the baseline model in Equation (1) to evaluate the potential of Google Trends variables as leading indicators for the Condominium Price Index:

$$y_t = \mu_0 + \alpha_1 y_{t-1} + \beta_1 x_t + \varepsilon_t \tag{2}$$

where x_t represents either Google Trends data or macroeconomic variables at time t. The coefficient β_1 quantifies the effect of the explanatory variable x_t on the Condominium Price Index at time t.

According to Equation (2), if x_t is a Google Trends variable, applying this data to the model will result in some data loss. This is because the frequencies of y_t and x_t data must be equal. Specifically, the Condominium Price Index (y_t) is available at a monthly frequency, whereas Google Trends data (x_t) is recorded at a weekly frequency. To address this mismatch, Choi and Varian (2009) suggested selecting either the value of x_t from the first week of each month or computing the average value over the first two weeks of each month.

To further refine the analysis, this study employs the Mixed Data Sampling (MIDAS) technique, initially introduced by Ghysels et al. (2004), which has been widely adopted for analyzing datasets with mixed frequencies. This approach enables the effective integration of high-frequency Google Trends data with lower-frequency macroeconomic indicators, thereby improving estimation accuracy and predictive capability.

Augmented Distributed Lag MIDAS (ADL-MIDAS) Model

This study employs one of the Mixed-Data Sampling (MIDAS)[4] specifications, commonly referred to as the ADL-MIDAS model, as introduced by Andreou et al. (2013). This approach enables the direct projection of a monthly Condominium Price Index onto weekly Google Trends data. To illustrate the underlying concept, consider a simple Augmented Distributed Lag Model—ADL(1,1)—for forecasting y_{t+1}, which depends on both the dependent variable y and the explanatory variable x, as expressed in the following equation:

$$y_{t+1} = \mu_0 + \alpha_1 y_t + \beta_1 x_t + \varepsilon_{t+1} \quad (3)$$

This equation can be generalized into the ADL (p,q) model, as discussed by Stock and Watson (2003):

$$y_{t+1} = \mu_0 + \alpha_1 B(L;\, \theta) y_t + \beta_1 B(L;\, \theta) x_t + \varepsilon_{t+1} \quad (4)$$

where the polynomial $B(L;\, \theta)$ represents a lag polynomial defined as $B(L;\, \theta) = \sum_{k=0}^{K} B(k;\, \theta) L^k$ where L denotes the lag operator such that $L^k x_t = x_{t-k}$. The equation above can be extended for h-step-ahead forecasting, yielding

$$y^h_{t+h} = \mu_0 + \alpha_1 B(L;\, \theta) y_t + \beta_1 B(L;\, \theta) x_t + \varepsilon^h_{t+h} \quad (5)$$

However, a challenge arises when the dependent variable y_t and the explanatory variable x_t are sampled at different frequencies. Specifically, while y_t is sampled at a fixed frequency (e.g., monthly), the explanatory variable x_t is observed m times within the same period. For example, suppose y_t is recorded monthly and x_t is collected weekly; then x_t can be denoted as $x_t^{(m)}$, where m equals 4. This discrepancy can lead to a well-known issue known as "parameter proliferation," where the number of parameters in the polynomial $B(L;\theta)$, which was assumed to be finite, becomes excessively large. For example, we use daily data spanning four months, assuming 22 trading days per month, which yields a total of K = 4 × 22 = 88 estimated parameters. This large number of parameters increases model complexity and computational inefficiency. Therefore, to mitigate the parameter proliferation issue, particularly in cases with high-frequency data or large datasets, more sophisticated modeling techniques are required.

To address this challenge, this study adopts the ADL-MIDAS approach, as proposed by Andreou et al. (2013) and further developed by Baumeister and Guérin (2021). By integrating the MIDAS framework with the traditional ADL model, the resulting ADL-MIDAS model can be expressed as follows:

[4] For general MIDAS models, see Ghysels et al. (2004) and Ghysels et al. (2007).

$$y_t^h = \mu_0^h + a^h y_{t-d}^h + \beta_1^h B\left(L^{\frac{1}{m}}; \theta^h\right) x_{t-h}^{(m)} + \varepsilon_t^{(h)} \quad (6)$$

where d = 1 for h = 0, $\frac{1}{m}$, $\frac{2}{m}$, …. , $\frac{m-1}{m}$ and d = h for h is an integer.

By applying a specialized filtering technique within the MIDAS framework, various alternatives for the weighting scheme $B(k; \theta)$ have been proposed, as discussed by Ghysels et al. (2007). Among these, the most widely utilized specifications are the Exponential Almon Lag and the Beta Lag, both of which are described in detail in Appendix E.

Evaluation of Forecasting Accuracy

To assess the forecasting performance of each model, we estimated the following specifications:

(1) The simple AR model
(2) The AR model with Google Trends variables
(3) The AR model with macro variables
(4) The ADL-MIDAS model with Google Trends variables

Subsequently, we conducted an out-of-sample forecasting evaluation using the AR model as the benchmark. To assess the models' forecasting performance, we computed both the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) for two forecast horizons: one-month-ahead (h = 1) and one-quarter-ahead (h = 3), of which the latter captures medium-term forecasting capabilities. To further evaluate the models' robustness under extreme conditions, MAE and RMSE were also separately calculated for the COVID-19 period. Additionally, formal statistical testing using the Diebold-Mariano (DM) test was conducted to determine whether the observed differences in forecasting performance across models were statistically significant.

Research Findings

This study examines the potential of Google Trends variables, which capture public interest in condominiums within the real estate market, as leading indicators of the Condominium Price Index. To achieve this objective, data from March 2008 to June 2021 were collected and analyzed using multiple forecasting models.

To determine the most appropriate keywords for Google Trends variables, several search terms related to condominiums were tested. Correlation analysis was then conducted to assess the relationship between each Google Trends variable and the Condominium Price Index. As presented in Appendix B, the "Real Estate" category, hereafter referred to as RE**,** exhibited the highest correlation with the Condominium Price Index (0.6644), indicating a strong positive relationship in which increased search interest in real estate corresponds with a rise in condominium prices. Conversely, the search term "คอนโด+ขาย" (condo+sale), hereinafter referred to as CS, exhibited the second-highest negative correlation with the Condominium Price Index at -0.4589, indicating an inverse relationship between selling-related search activity and price levels. This pattern can be attributed to both economic and behavioral factors. Economically, increased search interest may reflect a rise in potential supply, often driven by price decline expectations or financial stress, which can depress prices when not matched by demand. Behaviorally, such activity may signal loss aversion, herd behavior, and present bias, prompting individuals to sell preemptively out of fear or a preference for immediate liquidity. Thus, the search term "คอนโด+ขาย" may serve as a proxy for negative market sentiment and anticipatory selling pressure, contributing to downward movements in condominium prices. For monthly macroeconomic variables, the Land Price Index was selected for inclusion in the model due to its strong correlation with the Condominium Price Index (0.9817).

To ensure the validity of the time series data, all variables were first log-transformed to stabilize variance. Subsequently, the Augmented Dickey-Fuller (ADF) test, a standard unit root test, was conducted to assess stationarity. The results indicated that the log-transformed Condominium Price Index, Land Price Index, and Google Trends variables (RE and CS) were non-stationary in levels. However, their first differences were found to be stationary at the 1% significance level, confirming their appropriateness for use in further modeling and analysis.

To evaluate the predictive performance of various model specifications[5], the following models were considered:

Model 1: The AR model as a benchmark model

$$\Delta CoPI_t = \mu_0 + \alpha_1 \Delta CoPI_{t-1} + \alpha_2 \Delta CoPI_{t-3} + \varepsilon_t$$

Model 2: The AR model with Google Trends variables

$$\Delta CoPI_t = \mu_0 + \alpha_1 \Delta CoPI_{t-1} + \alpha_2 \Delta CoPI_{t-3} + \beta_1 \Delta RE_t + \varepsilon_t$$

and

$$\Delta CoPI_t = \mu_0 + \alpha_1 \Delta CoPI_{t-1} + \alpha_2 \Delta CoPI_{t-3} + \beta_1 \Delta CS_t + \varepsilon_t$$

Model 3: The AR model with macro variables

$$\Delta CoPI_t = \mu_0 + \alpha_1 \Delta CoPI_{t-1} + \alpha_2 \Delta CoPI_{t-3} + \beta_1 \Delta LPI_t + \varepsilon_t$$

Model 4: The ADL-MIDAS model with Google Trends variables

$$\Delta CoPI_t = \mu_0 + \alpha_1 \Delta CoPI_{t-1} + \alpha_2 \Delta CoPI_{t-3} + \beta \sum_{k=1}^{4} B(k;\theta) \Delta RE_t^{(k)} + \varepsilon_t$$

and

$$\Delta CoPI_t = \mu_0 + \alpha_1 \Delta CoPI_{t-1} + \alpha_2 \Delta CoPI_{t-3} + \beta \sum_{k=1}^{4} B(k;\theta) \Delta CS_t^{(k)} + \varepsilon_t$$

Where $CoPI_t$ is the natural logarithm of the Condominium Price Index at time t, t = 1, 2, …, T,
RE_t is the natural logarithm of Google Trends variables for the “Real Estate” category,
CS_t is the natural logarithm of Google Trends variables for “คอนโด+ขาย” (condo + sale),
LPI_t is the natural logarithm of the Land Price Index at time t, t =1,2, …, T, and
$B(k; \theta)$ is the beta lag.

Table 1 presents the regression results for these models. Models 1–3 are based on an autoregressive (AR) framework, which assumes that past condominium prices influence current prices. The results reveal a statistically significant positive relationship between the Condominium Price Index (CoPI) and its one-period lagged value, whereas the three-period lagged value displays a negative and significant effect.

[5] Several variations of the AR models were estimated, and the best-fitted model with lags of 1 and 3 was selected.

Creative Business and Sustainability Journal (CBSJ)
Vol.47 No.1 January – June 2025, pp.88-108

Table 1: Regression Analysis of CoPI on the Condominium-Related Search Index and Macro Variables

	(1) AR	(2) GG		(3) Macro	(4) ADL-MIDAS	
		RE	CS		RE	CS
$\Delta CoPI_{t-1}$	0.4230***	0.4276***	0.4409***	0.3927***	0.4037***	0.4105***
	(0.0644)	(0.0642)	(0.0636)	(0.0651)	(0.0712)	(0.0710)
$\Delta CoPI_{t-3}$	-0.4385***	-0.4364***	-0.4184***	-0.4334***	-0.4417***	-0.4529***
	(0.0638)	(0.0636)	(0.0632)	(0.0631)	(0.0715)	(0.0706)
ΔRE_t		0.0070			-0.0138	
		(0.0049)			(0.0156)	
CS_t			-0.0142**			-0.0188
			(0.0056)			(0.0194)
ΔLPI_t				0.1583**		
				(0.0729)		
Constant	0.0043 ***	0.0044***	0.0041***	0.0037***	0.0040***	0.0041***
	(0.0011)	(0.0011)	(0.0156)	(0.0011)	(0.0012)	(0.0012)
θ_1					1.1997	1.2626
θ_2					24.2670	24.3230
Adjusted R^2	0.3578	0.3623	0.3803	0.3730	0.3705	0.3690

Notes: *p<0.1, **p<0.05, ***p<0.01; robust standard errors are shown in parentheses. B(k; θ) is specified using the Beta lag structure.

To assess the influence of online search behavior on CoPI, Google search data encompassing real estate-related queries were incorporated into the analysis. Specifically, search interest from the "Real Estate" (RE) category and the "คอนโด+ขาย" (CS; condo+sale) keyword under the same category were included in Model 2. The results indicate a statistically significant negative relationship between CoPI and the CS search index. Specifically, a 1% increase in the Google search index for "คอนโด+ขาย" (condo+sale) was associated with a 0.0142% decrease in the Condominium Price Index. The finding suggests that heightened search activity may reflect growing interest in selling rather than buying, potentially signaling excess supply or weaker demand, which in turn could exert downward pressure on prices (Wu & Brynjolfsson, 2015). In contrast, the coefficient for the "Real Estate" (RE) search index was positive but not statistically significant. Moreover, Model 2 exhibits a slightly higher adjusted R^2 compared to the benchmark AR model. This improvement suggests that incorporating online search data may enhance the model's explanatory power, indicating the potential of Google Trends as a leading indicator of housing market activity.

In Model 3, the Google Trends variables were replaced with the Land Price Index (LPI), a key macroeconomic indicator. The results indicate a positive relationship between CoPI and LPI, further supporting the relevance of macroeconomic conditions in explaining condominium price dynamics. Additionally, the model showed a slight improvement in adjusted R^2 compared to the benchmark.

Model 4 applied the ADL-MIDAS approach to address the issue of mixed-frequency data, combining Google search data (weekly) and the Condominium Price Index (monthly) within a single framework. The results differ from those of Model 2 in that none of the search index variables exhibited statistically significant effects on the Condominium Price Index (CoPI). Furthermore, the adjusted R^2 value did not show substantial improvement over Model 2, suggesting that the traditional AR model remains a viable and parsimonious approach for forecasting condominium prices under normal market conditions.

To evaluate the forecasting accuracy of each model, a one-month-ahead forecast ($h = 1$) was conducted, and the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) were calculated. The results, presented in Table 2, indicate that the AR model augmented with Google Trends variables (RE and CS) and the AR model incorporating macroeconomic variables (Model 2–3) yielded MAE and RMSE values comparable to those of the baseline model (Model 1). This finding is consistent with the results of the Diebold-Mariano (DM) test, which produced p-values greater than 0.1 (0.3131, 0.9894, and 0.9747, respectively), suggesting that the null hypothesis cannot be rejected. In other words, the forecast accuracy of these models is not statistically different from the baseline.

In contrast, the ADL-MIDAS model (Model 4) demonstrated the highest predictive accuracy, as evidenced by the lowest MAE and RMSE values. Moreover, the DM test results support a statistically significant improvement over the baseline, indicating that the ADL-MIDAS model significantly outperforms the other models in terms of forecasting performance.

However, during the COVID-19 period, both the ADL-MIDAS model and the AR model with Google Trends (CS) showed statistically significant differences from the baseline. In contrast to the superior performance of the ADL-MIDAS model, the AR model with Google Trends (CS) exhibited significantly lower forecast accuracy (as indicated by the DM test's p-value of 0.0099), highlighting its reduced effectiveness under conditions of heightened market uncertainty.

Table 2: Out-of-Sample Forecast Performance

Model	one-month-ahead forecast (h=1)		one-quarter-ahead forecast (h=3)	
	MAE	**RMSE**	**MAE**	**RMSE**
Panel A: Full Sample (2008–2021)				
1) AR (Baseline Model)	0.0108	0.0133	0.0108	0.0134
2) GG with RE	0.0109	0.0135	0.0109	0.0135
GG with CS	0.0107	0.0133	0.0108	0.0134
3) Macro	0.0107	0.0133	0.0108	0.0135
4) ADL-MIDAS with RE	0.0028	0.0028	0.0051	0.0051
ADL-MIDAS with CS	0.0036	0.0036	0.0056	0.0056
Panel B: COVID Period (2020Q1–2021Q2)				
1) AR (Baseline Model)	0.0120	0.0149	0.0121	0.0151
2) GG with RE	0.0121	0.0150	0.0122	0.0152
GG with CS	0.0134	0.0160	0.0136	0.0162
3) Macro	0.0119	0.0152	0.0120	0.0154
4) ADL-MIDAS with RE	0.0028	0.0028	0.0051	0.0051
ADL-MIDAS with CS	0.0036	0.0036	0.0056	0.0056

The results for the one-quarter-ahead forecast (h = 3) reinforce the earlier findings from the one-month-ahead horizon. These results suggest that the ADL-MIDAS approach consistently enhances forecasting performance across different forecasting horizons and market conditions, including periods of heightened market uncertainty, such as the COVID-19 pandemic.

Conclusion

This study employed a nowcasting model to examine the relationship between the Condominium Price Index (CoPI) and Google search data. In addition to the widely recognized Autoregressive (AR) model utilized in previous research, the Augmented Distributed Lag Mixed-Data Sampling (ADL-MIDAS) model was applied to address data sampled at varying frequencies. Furthermore, macroeconomic variables were incorporated into the analysis to facilitate a comparison of model performance.

The key findings of this study can be summarized as follows. First, a significant negative correlation was observed between CoPI and Google Trends (CS), suggesting that increased search activity may reflect intentions to sell, which could signal potential downward pressure on prices.

Second, in terms of in-sample explanatory power, both the AR model augmented with Google Trends data and the model incorporating macroeconomic variables demonstrated slightly higher adjusted R^2 values compared to the benchmark AR model. The results suggest that online search activity and macroeconomic factors may provide further insight into condominium price dynamics. Although the improvements were modest, Google Trends data provided the advantage of higher frequency and timeliness, enabling rapid forecasting without delays associated with the release of official macroeconomic data.

Moreover, the ADL-MIDAS model did not exhibit significant superiority over the AR model incorporating either Google Trends or macroeconomic variables. Thus, in normal market conditions, the AR model remained sufficient even when variables were sampled at different frequencies. However, regarding out-of-sample forecasting performance, the ADL-MIDAS model consistently outperformed all other models across both short-term and medium-term horizons. These results highlight the benefits of the ADL-MIDAS framework in utilizing high-frequency data to enhance predictive accuracy.

Lastly, during the COVID-19 period, the ADL-MIDAS model maintained its superior forecasting performance, while the AR model with the Google Trends (CS) variable significantly underperformed the baseline. This suggests that the reliability of online search data may diminish during periods of market uncertainty unless appropriately modeled using a mixed-frequency approach.

Overall, the findings point out the advantages of Google Trends as a leading indicator in the real estate market. One of the primary advantages of utilizing Google Trends data is the absence of data collection costs, as well as the ability to obtain real-time insights without the need for extensive field surveys and interviews. Additionally, the timeliness and authenticity of user-generated search data make it a valuable tool for economic forecasting. When integrated with short-term forecasting models designed for mixed-frequency data, Google Trends can serve as an early warning signal for economic fluctuations. This capability is particularly beneficial for policymakers, including those responsible for monetary policy and financial regulation, as it enables real-time monitoring of market sentiment, detection of speculative bubbles, and the implementation of macroprudential measures to ensure financial stability. For example, the Bank of Thailand can utilize such models to formulate data-driven interventions, adjust credit conditions, or impose targeted regulations in response to overheating in specific property segments. Moreover, entrepreneurs, investors, and the public can apply these insights to make informed decisions regarding property transactions and investments.

Despite its contributions, this study has certain limitations. It did not conduct an in-depth analysis of real estate market trends during irregular economic disruptions. For instance, adverse shocks such as the outbreak of the Coronavirus Disease 2019 (COVID-19) may significantly impact consumption and investment behavior. Future research could extend this work by applying the nowcasting model to other types of real estate, such as detached houses or townhomes, and by comparing model performance across different regions in Thailand or other emerging markets. Additionally, incorporating alternative high-frequency data sources, such as social media sentiment or real-time transaction volumes, could further enhance the robustness of forecasting models. Such enhancements would also allow for the exploration of real-time policy applications, including early warning systems for housing bubbles or tools for macroprudential regulation.

Acknowledgments

The author gratefully acknowledges the financial support from the Puey Ungphakorn Institute for Economic Research, Thailand.

References

Alper, C. E., Fendoglu, S., & Saltoglu, B. (2008). Forecasting stock market volatilities using MIDAS regressions: An application to the emerging markets. (*MPRA Paper No. 7460*). University Library of Munich, Germany. https://mpra.ub.uni-muenchen.de/7460/

Andreou, E., Ghysels, E., & Kourtellos, A. (2013). Should macroeconomic forecasters use daily financial data and how?. *Journal of Business & Economic Statistics, 31*(2), 240–251.

Asgharian, H., Hou, A. J., & Javed, F. (2013). The importance of the macroeconomic variables in forecasting stock return variance: A GARCH-MIDAS approach. *Journal of Forecasting, 32*(7), 600–612.

Asimakopoulos, S., Paredes, J., & Warmedinger, T. (2013). *Forecasting fiscal time series using mixed frequency data*. ECB Working Paper.

Askitas, N. (2016). Trend-spotting in the housing market. *Cityscape, 18*(2), 165–178.

Baumeister, C., & Guérin, P. (2021). A comparison of monthly global indicators for forecasting growth. *International Journal of Forecasting, 37*(3), 1276–1295.

Choi, H., & Varian, H. (2009). Predicting initial claims for unemployment benefits. *Google Inc, 1*(2009), 1–5.

Conrad, C., & Loch, K. (2015). Anticipating long-term stock market volatility. *Journal of Applied Econometrics, 30*(7), 1090–1114.

Deghi, A., Katagiri, M., Shahid, S., & Valckx, N. (2020). *Predicting downside risks to house prices and macro-financial stability*. International Monetary Fund.

Dietzel, M. A., Braun, N., & Schäfers, W. (2014). Sentiment-based commercial real estate forecasting with Google search volume data. *Journal of Property Investment & Finance, 32*(6), 540–569.

Engle, R. F., Ghysels, E., & Sohn, B. (2013). Stock market volatility and macroeconomic fundamentals. *Review of Economics and Statistics, 95*(3), 776–797.

Ghysels, E., Sinko, A., & Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews, 26*(1), 53–90.

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2004). The MIDAS touch: Mixed data sampling regression models.

Glindro, E. T., Subhanij, T., Szeto, J., & Zhu, H. (2018). Determinants of house prices in nine Asia-Pacific economies. *International Journal of Central Banking, 26*, 1–28.

Kingnetr, N., Tungtrakul, T., & Sriboonchitta, S. (2017, December). Does forecasting benefit from mixed-frequency data sampling model: The evidence from forecasting GDP growth using financial factor in Thailand. In *International Conference of the Thailand Econometrics Society* (pp. 430–442). Cham: Springer International Publishing.

Limnios, C., & You, H. (2018). Can Google Trends actually improve housing market forecasts?. *Available at SSRN 2886705.*

Lindblad, A. (2017). *Exploring the time-variation in long-term stock market volatility forecasts*. Unpublished paper, University of Helsinki.

Nakavachara, V., & Lekfuangfu, N. W. (2018). Predicting the present revisited: The case of Thailand. *Thailand and The World Economy, 36*(3), 23–46.

Oust, A., & Eidjord, O. M. (2020). Can Google search data be used as a housing bubble indicator?. *International Real Estate Review*, *23*(2).

Rosen, G., & Walks, A. (2013). Rising cities: Condominium development and the private transformation of the metropolis. *Geoforum*, *49*, 160-172.

Stephens-Davidowitz, S., & Varian, H. (2014). A hands-on guide to Google data. *Further details on the construction can be found on the Google Trends page*.

Stock, J. H., & Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature, 41*(3), 788–829.

Venkataraman, M., Panchapagesan, V., & Jalan, E. (2018). Does Internet search intensity predict house prices in emerging markets? A case of India. *Property Management*, *36*(1), 103-118.

Wichitaksorn, N. (2022). Analyzing and forecasting Thai macroeconomic data using mixed-frequency approach. *Journal of Asian Economics*, *78*, 101421.

Wu, L., & Brynjolfsson, E. (2015). The future of prediction: How Google searches foreshadow housing prices and sales. In *Economic Analysis of the Digital Economy* (pp. 89–118). University of Chicago Press.

Appendix

Google Trends Usage

Google Trends does not provide absolute search counts but rather a normalized index (0–100) based on the relative popularity of a search term over time and location. For example, entering the term "ภาษี" (tax) for Thailand over the past five years shows a seasonal peak between February and March, as illustrated in Figure A1, which aligns with tax filing periods. In 2020 and 2021, the search volume shifted to later months due to deadline extensions during the COVID-19 pandemic, illustrating how search data reflects real-world events.

Researchers can adjust the timeframe, region, and category (e.g., "Shopping"), with Google returning monthly data for periods exceeding five years, weekly for shorter periods, and daily for durations of less than nine months. Comparative searches are also possible, such as analyzing "ฝุ่น" (dust) and "หน้ากากอนามัย" (mask), which showed correlated spikes during the PM2.5 air pollution period (Figure A2). While data may vary slightly due to real-time sampling, these fluctuations are minor. Careful keyword selection remains essential for reliable analysis.

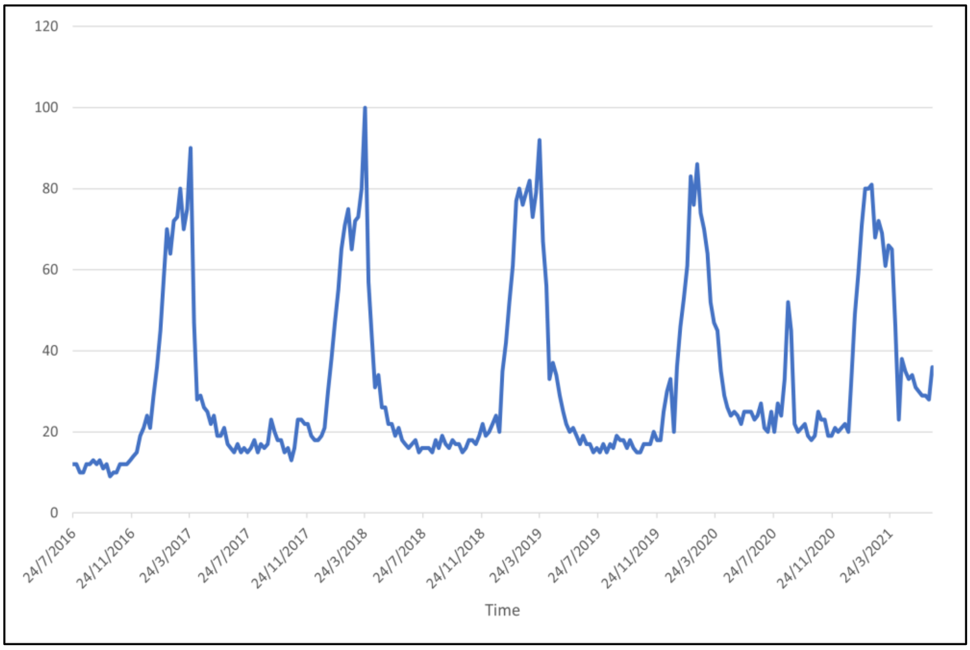


Figure A1: Google Trends for the Search Term "ภาษี" (tax)

Creative Business and Sustainability Journal (CBSJ)
Vol.47 No.1 January – June 2025, pp.88-108

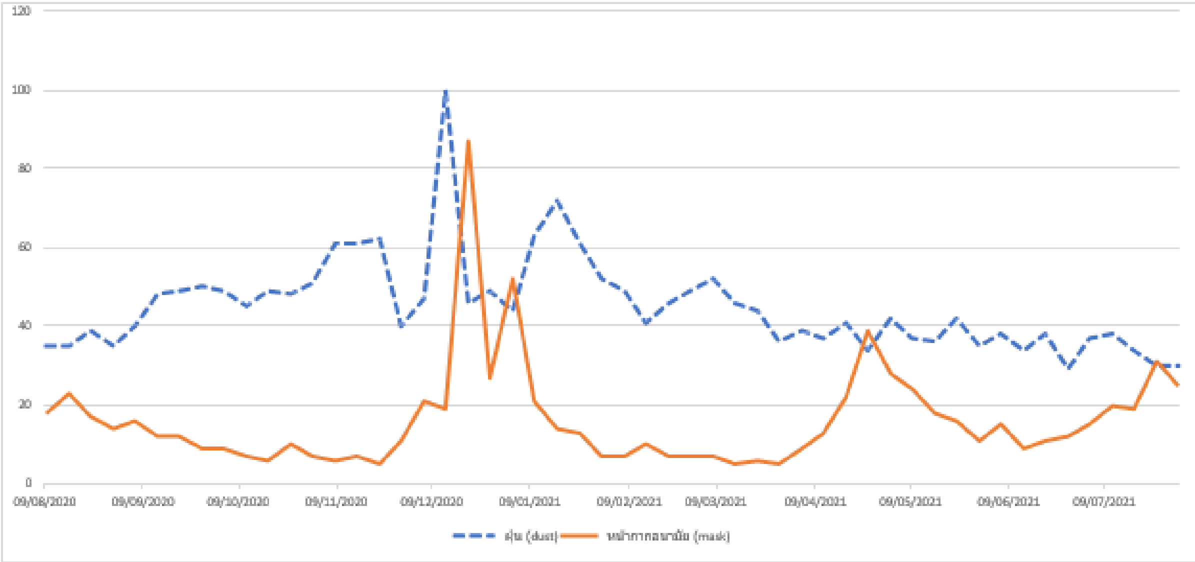


Figure A2: Comparison of Two Keywords, “dust” and “mask”

Correlations

Table A1: Correlations between Search Terms and Macroeconomic Variables with the Condo Price Index

	Correlations with Condo Price Index
Potential search terms	
Category “Real Estate”	0.6644
“คอนโด”(condo)	0.2952
“คอนโด+ขาย”(condo+sale)	-0.4589
“คอนโด+เช่า”(condo+rent)	-0.1121
Macro variables	
MRR rate	-0.3033
Population	0.9805
Land Price Index	0.9817
Construction Materials Price Index	0.0528
SET Index	0.8175
Exchange rates	-0.0237
Consumer Confidence Index	0.8580
Unemployment rate	0.2266

Note: All single search terms are obtained under the category “Real Estate.” The English translation is in parentheses.

Data summary

Table A2: Summary Statistics

Variables	Description	No.obs.	Frequency	Period	Mean	STD	Min	Max	Source
CoPI	Condo Price Index	160	monthly	Mar 08 – June 21	146.61	29.89	96.00	197.30	Bank of Thailand
RE	Category “Real Estate”	696	weekly	Mar 08 – June 21	45.02	17.54	6.75	100.00	Google Trends website
CS	“คอนโด+ขาย” (condo+sale)	696	weekly	Mar 08 – June 21	53.53	13.93	23.97	100.00	Google Trends website
LPI	Land Price Index	160	monthly	Mar 08 – June 21	143.60	31.32	90.4	197.4	Bank of Thailand

Adjusted Google Trends data

Table A3: Example Calculation of Adjusted Google Trends Data

Month	Weekly data		Monthly data	Adjusted data
Jan	5/1/2020	82	81	82×81/100 = 66.42
	12/1/2020	81	81	81×81/100 = 65.61
	19/1/2020	80	81	80×81/100 = 64.80
	26/1/2020	79	81	79×81/100 = 63.99
Feb	2/2/2020	80	80	80×80/100 = 64.00
	9/2/2020	75	80	75×80/100 = 60.00
	16/2/2020	83	80	83×80/100 = 66.40
	23/2/2020	77	80	77×80/100 = 61.60
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
Nov	1/11/2020	67	69	67×69/100 = 46.23
	8/11/2020	70	69	70×69/100 = 48.30
	15/11/2020	66	69	66×69/100 = 45.54
	22/11/2020	71	69	71×69/100 = 48.99
	29/11/2020	71	69	71×69/100 = 48.99
Dec	6/12/2020	68	71	68×71/100 = 48.28
	13/12/2020	73	71	73×71/100 = 51.83
	20/12/2020	72	71	72×71/100 = 51.12
	27/12/2020	61	71	61×71/100 = 43.31

Creative Business and Sustainability Journal (CBSJ)
Vol.47 No.1 January – June 2025, pp.88-108

Weighting Scheme

The exponential almon lag is defined as

$$B(k;\theta) = \frac{e^{\theta_1 k+\cdots+\theta_Q k^Q}}{\sum_{k=1}^{K} e^{\theta_1 k+\cdots+\theta_Q k^Q}}$$

where K is the number of lags in the MIDAS polynomial. If we have only two parameters (θ_1, θ_2), then this equation simply becomes

$$B(k;\theta) = \frac{e^{\theta_1 k+\theta_2 k^2}}{\sum_{k=1}^{K} e^{\theta_1 k+\theta_2 k^2}}$$

Note that for the special case when $\theta_1 = \theta_2 = 0$, the $B(k;\theta)$ is a standard equal weighting function. Another specification, "Beta Lag," can be written as

$$B(k;\theta_1,\theta_2) = \frac{f(\frac{k}{K},\theta_1,\theta_2)}{\sum_{k=1}^{K} f(\frac{k}{K},\theta_1,\theta_2)}$$

where

$$f(x,a,b) \ = \frac{x^{a-1}(1-x)^{b-1}\Gamma(\mathrm{a}+\mathrm{b})}{\Gamma(\mathrm{a})\Gamma(\mathrm{b})}$$

$$\Gamma(\mathrm{a}) = \int_0^\infty e^{-x} x^{a-1} dx$$

Similarly, when $\theta_1 = \theta_2 = 0$, we obtain the flat weights. As we can see from the equation, this functional form requires only two parameters.

Both specifications are nonnegative functions of θ for all k, which sums up to one. Engle et al. (2013) claim that their findings are robust regardless of the type of weighting schemes. Several studies, including those by Alper et al. (2008), Lindblad (2017), Conrad and Loch (2015), and Asgharian et al. (2013), have employed the beta weighting scheme due to its ability to generate various shapes of weighting functions using only two parameters.