9. FORECASTING HOUSING PRICE INDICES WITH MACROECONOMIC VARIABLES: EVIDENCE FROM SEOUL'S 25 DISTRICTS[1]

Yeonchan KANG [2]
Doojin RYU [2*]

Abstract

We investigate whether pure macroeconomic variables can forecast district-level housing price indices in Seoul without using local information, building-specific, or historical price data. Comparing six forecasting models across 25 districts using a rolling window framework, we find that Lasso regression outperforms complex machine learning algorithms, including XGBoost and Random Forest. Lasso selects only five to nine macroeconomic variables from sixteen candidates to forecast each district's housing index, with M2 money supply and household credit emerging as the dominant predictors across all districts. Our findings reveal that the local housing market indices follow simple linear relationships with macroeconomic fundamentals. The heterogeneity in district-level sensitivities implies that uniform monetary policies create highly uneven spatial effects. These results suggest that accurate housing market forecasting may not require sophisticated nonlinear models but proper identification of district-specific sensitivities to core macroeconomic variables.

Keyword: *Housing market forecasting; Lasso; Machine learning; Macroeconomic variables*

JEL Classification: C52, G12, R31

1. Introduction

Housing markets serve as both a transmission channel for monetary policy and a potential source of financial instability. In densely populated cities with geographically segmented submarkets, this dual role becomes complex. Seoul exemplifies this complexity with its 25 distinct districts, each with unique characteristics yet all influenced by the same macroeconomic forces. Understanding the relative importance of these common macroeconomic versus local idiosyncratic factors is crucial for both policymakers and researchers. How much of the variation in local housing price indices can be attributed solely to macroeconomic variables, independent of local factors or historical price information? The answer to this question has critical implications. If macroeconomic variables alone can explain most variation, then complex models incorporating

[1] *We sincerely thank Editor Corina Saman for the constructive comments and helpful suggestions. This paper was supported by SKKU Academic Research Support Program (Samsung Research Fund), Sungkyunkwan University, 2024.*

[2] *Department of Economics, Sungkyunkwan University, Seoul, Republic of Korea.*

**Corresponding Author.*

spatial and temporal dependencies may be capturing noise rather than signal, and central banks can more reliably forecast the local impacts of their policies.

The prevailing wisdom in housing market analysis emphasizes complexity and local information (Hong and Ryu, 2023). Recent advances in machine learning have produced increasingly sophisticated models incorporating spatial spillovers, temporal dynamics, neighborhood characteristics, and nonlinear interactions (Athey and Imbens, 2019; Bang and Ryu, 2023; Chun, Cho, and Ryu, 2025; Kang, Ryu, and Webb, 2025; Kim, Ryu, and Webb, 2025). These models typically leverage hundreds of features, including lagged price terms and spatial lags, arguing that housing markets exhibit strong persistence and contagion effects (Kim, Park, and Ryu, 2025; Kim, Ryu, and Webb, 2025). This data-driven approach faces three limitations. First, the inclusion of lagged dependent variables and contemporaneous predictors creates severe endogeneity problems, making causal interpretation impossible. Second, the black box nature of complex models provides little insight into economic mechanisms (Mullainathan and Spiess, 2017). Third, there is growing evidence that sparse linear models such as Lasso often outperform complex machine learning alternatives in genuine out-of-sample forecasting, despite their simplicity (Bang and Ryu, 2024; Bang, Kang, and Ryu, 2024; Kang, Ryu, and Webb, 2024; Medeiros and Mendes, 2016; Welch and Goyal, 2008). These limitations motivate our approach: we restrict our analysis to pure macroeconomic variables and employ sparse methods that prioritize interpretability and genuine forecasting performance over model complexity.

The Korean housing market presents unique characteristics that make it particularly suitable for testing macroeconomic transmission mechanisms (Chae, Cho, Kim, and Ryu, 2025; Kim, Cho, and Ryu, 2024; Ryu, Hong, and Jo, 2024). Unlike Western markets dominated by mortgage-based transactions, Korea's system of ownership and Jeonse (lump-sum deposit rental) creates distinctive financial dynamics. The Jeonse system, where tenants provide landlords with deposits often reaching 50-80% of property values instead of monthly rent, transforms housing into a leveraged investment vehicle. Landlords use these deposits for gap investment, purchasing additional properties with the difference between the Jeonse deposits received and new property prices (Ahn and Ryu, 2024). This institutional structure creates a direct mechanical link between monetary conditions and housing prices that differs fundamentally from traditional mortgage markets. Furthermore, Korea's macroprudential policies have evolved through multiple cycles, with loan-to-value and debt-to-income regulations varying by district and time period (Igan and Kang, 2011). These policy variations, combined with culturally driven investment preferences that view real estate as the primary wealth accumulation vehicle, make Seoul's housing market exceptionally responsive to macroeconomic variables rather than local supply-demand fundamentals.

We take a different approach that prioritizes parsimony and economic interpretability over predictive complexity. Our analysis uses exclusively macroeconomic variables to forecast housing price indices across Seoul's 25 districts, deliberately excluding all district-specific information, historical index values, and spatial variables. This design choice serves three purposes: isolating pure macroeconomic transmission effects, ensuring results reflect genuine impacts rather than spurious local correlations, and yielding interpretable coefficients that reveal which transmission channels matter most.

Our findings challenge the complexity paradigm in housing market analysis. Lasso regression achieves a mean R^2 of 0.91 across all districts, substantially outperforming machine learning models, including XGBoost (R^2 = 0.86) and Random Forest (R^2 = 0.79). Only 5 to 9 macroeconomic variables are selected from 16 candidates to forecast each district's index. The selected variables reveal consistent economic patterns with M2 money supply and household credit dominating forecasting, confirming the credit-driven nature of Korean housing markets. The magnitude of responses varies across districts, with consumer price index (CPI) sensitivity ranging from negative 10 percent in some areas to near zero in others. This pattern of heterogeneous simplicity suggests that housing market indices follow simple structural relationships with varying local parameters.

This study makes three contributions to the literature. First, we demonstrate that macroeconomic variables alone can explain over 90 percent of housing price index variation across urban districts, with sparse linear models outperforming complex nonlinear alternatives. This finding challenges the prevailing emphasis on local information and model complexity in housing market analysis. Second, we provide interpretable estimates of macroeconomic transmission channels, enabling precise measurement of how each variable affects different districts, something impossible with black-box machine learning models. Third, we document substantial heterogeneity in macroeconomic sensitivities across districts, with coefficients varying by an order of magnitude, indicating that uniform monetary policies generate highly uneven spatial effects with important implications for inequality and financial stability.

The remainder of this paper is organized as follows. Section 2 describes our data and methodology. Section 3 presents the empirical results. Section 4 concludes.

2. Data and Methodology

We construct a comprehensive dataset combining district-level housing price indices with macroeconomic indicators for Seoul's 25 administrative districts from January 2015 to December 2024. All data comes from the Bank of Korea. Our macroeconomic variables encompass four categories: (1) monetary policy indicators (base rate, 3-year bond yield, deposit rate, loan rate), (2) monetary aggregates (M2, household credit), (3) real economy indicators (GDP growth, unemployment, CPI, coincident index, construction value, building permits), and (4) sentiment measures (consumer sentiment, economic sentiment). Additionally, we include financial market variables (KOSPI index, exchange rate) to capture broader economic conditions. To ensure all predictors are observable at the time of forecasting, we apply a 3-month lag to all macroeconomic variables.

Table 1 presents summary statistics for our variables. Panel A reveals the macroeconomic environment during our sample period. The base rate averaged 1.77% with substantial variation (0.50% to 3.50%), reflecting the Bank of Korea's policy responses to economic conditions. M2 and household credit show strong growth trends, consistent with Korea's expanding credit markets. Panel B shows that housing price indices exhibit considerable variation across districts, with standard deviations ranging from 9.57 (Gwanak) to 15.61 (Yangcheon), suggesting heterogeneous price dynamics.

Table 1. Summary statistics

Panel A. Macroeconomic variables

	mean	std	Min	Q1	Med	Q3	Max
Base rate	1.77	0.98	0.50	1.25	1.50	2.06	3.50
Bond 3Y	2.09	0.89	0.83	1.45	1.79	2.88	4.24
Deposit rate	2.02	0.97	0.80	1.45	1.70	2.45	4.30
Loan rate	3.74	0.79	2.63	3.27	3.50	3.98	5.64
M2	3,053,391	640,358	2,094,265	2,470,222	2,922,685	3,719,632	4,183,692
Coincident idx	99.85	1.06	96.30	99.40	99.95	100.70	101.60
Consumer sentiment	99.81	7.53	71.50	97.15	100.80	103.63	113.90
Economic sentiment	95.83	8.03	61.30	93.30	96.25	100.50	110.10

Unemployment	3.45	0.52	2.50	2.90	3.60	3.80	4.70
KOSPI	2,367.97	346.27	1,754.64	2,058.20	2,330.27	2,569.13	3,296.68
Exchange rate	1,198.38	96.39	1,066.50	1,122.55	1,171.85	1,267.83	1,470.00
CPI	102.28	6.39	94.59	97.56	99.76	108.32	114.91
Construction	11,864,423	1,841,505	7,798,163	11,061,701	11,672,887	13,115,656	16,759,070
Building permit	14,595,624	2,133,677	7,904,246	15,614,311	15,614,311	15,614,311	16,254,223
GDP growth	0.62	0.85	-2.70	0.35	0.70	1.23	2.20
Household credit	1,603,281	256,831	1,098,305	1,411,270	1,606,149	1,862,918	1,925,939

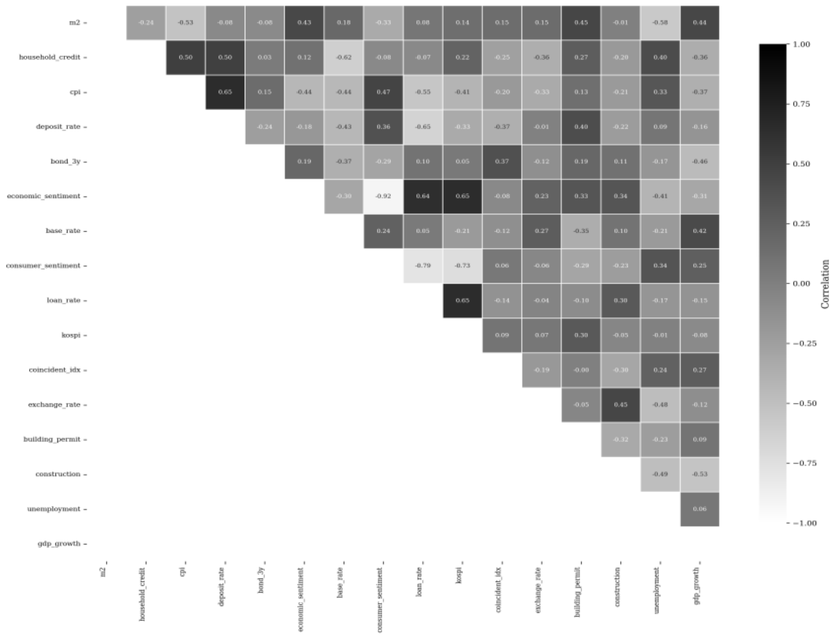
Panel B. Housing price indices by district

District	mean	std	Min	Q1	Median	Q3	Max
Gangnam	82.81	14.40	59.38	67.57	83.41	97.69	103.17
Gangdong	84.27	11.48	65.61	72.39	83.52	94.99	100.81
Gangbuk	83.05	10.97	68.71	72.83	79.00	94.34	100.42
Gangseo	83.42	11.39	65.81	72.26	81.66	94.27	100.36
Gwanak	86.32	9.57	72.04	76.92	84.34	95.58	100.42
Gwangjin	88.56	9.87	73.89	77.98	87.19	98.98	101.68
Guro	81.10	13.15	62.24	68.14	77.85	93.62	101.39
Geumcheon	85.57	10.87	70.43	74.27	83.91	96.19	100.87
Nowon	73.98	15.51	53.44	59.45	68.71	86.39	100.05
Dobong	78.53	13.24	60.82	65.50	74.50	89.74	100.93
Dongdaemun	82.08	12.45	62.32	69.10	81.22	93.09	100.28
Dongjak	83.09	12.57	64.55	70.24	80.91	95.29	100.77
Mapo	83.04	13.50	62.93	69.21	81.14	96.77	101.41
Seodaemun	84.86	11.68	67.63	72.06	85.65	95.21	100.68
Seocho	83.39	13.63	62.20	69.79	82.84	97.93	102.10
Seongdong	82.64	12.94	63.01	68.92	81.24	94.25	100.67
Seongbuk	81.90	11.85	63.79	69.71	80.13	92.18	100.09
Songpa	80.55	14.73	59.92	64.97	78.53	95.50	100.84
Yangcheon	80.07	15.61	58.58	63.90	76.72	96.94	100.79
Yeongdeungpo	80.51	15.58	57.23	63.85	80.77	95.83	100.96
Yongsan	84.82	14.53	65.38	69.16	83.09	101.73	108.20
Eunpyeong	85.33	10.86	70.74	74.69	81.69	96.73	101.27
Jongno	87.89	9.75	74.23	77.86	86.13	99.22	100.76
Jung	85.48	11.55	68.08	73.18	83.79	97.28	101.67
Jungnang	86.09	10.62	70.71	75.35	83.96	96.82	101.35

Note: Panel A presents summary statistics for 16 macroeconomic variables. Base rate is the Bank of Korea's policy rate; Bond 3Y is the 3-year government bond yield; deposit and loan rates are average rates from commercial banks. M2 is the broad money supply; Household credit represents total household loans. Coincident index measures current economic activity; Consumer sentiment and Economic sentiment are survey-based indices. KOSPI is the Korea Composite Stock Price Index; the Exchange rate is KRW/USD. CPI is the Consumer Price Index. Construction is the amount of construction, and Building permit is the number of building permits. GDP growth is the quarterly percentage change. Panel B shows housing price indices for 25 Seoul districts (base=100). Q1 denotes the 25th percentile; Median denotes the median (50th percentile); Q3 denotes the 75th percentile. Min and Max represent minimum and maximum values.

Figure 1 displays the correlation matrix among macroeconomic variables. Several patterns emerge from this analysis. First, monetary policy variables (base rate, bond yield, deposit rate, loan rate) exhibit strong positive correlations (0.85-0.92), indicating they move together as expected. Second, M2 and household credit show a high correlation (0.90), confirming their shared role as credit supply indicators. Third, correlations across different variable groups are generally weaker, with many coefficients below 0.5 in absolute value. For instance, sentiment measures show modest correlations with monetary aggregates, and construction-related variables exhibit weak correlations with interest rates. This pattern suggests that our macroeconomic variables capture distinct economic dimensions, reducing concerns about severe multicollinearity while providing complementary information for forecasting housing prices.

Figure 1. Correlation of macro variables



Note: This figure displays pairwise correlations among 16 macroeconomic variables used in our analysis. Darker shades indicate stronger correlations (positive in black, negative in white).

Our empirical strategy deliberately excludes all district-specific information and lagged dependent variables to isolate the effect of macroeconomic fundamentals. This design choice serves two purposes. First, it tests whether housing markets efficiently incorporate macro information without relying on historical price patterns. Second, it ensures that our models capture genuine economic transmission mechanisms rather than mechanical persistence or spurious spatial correlations. By restricting our feature set to lagged macroeconomic variables only, we can attribute any forecasting success to systematic relationships between macro conditions and local housing prices.

We implement six forecasting models with distinct approaches to forecasting. $y_{i,t}$ denotes the housing price index for district *i* at time *t*. x_t represents the vector of 16 macroeconomic variables.

Linear Regression: The baseline model estimates ordinary least squares:

$$y_{i,t} = \beta_0 + x_t'\beta + \varepsilon_{i,t}, \tag{1}$$

Lasso Regression: Applies L1 regularization($\lambda \sum_{j=1}^{p} |\beta_j|$) to induce sparsity by shrinking some coefficients to exactly zero:

$$min_{\beta_0,\beta} 1/2n \sum_{t=1}^{n} \left(y_{i,t} - \beta_0 - x_t'\beta\right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{2}$$

where λ controls the regularization strength, *n* is the number of training observations, and *p* is the number of predictors.

Ridge Regression: Uses *L2* regularization $\sum_{j=1}^{p} \beta_j^2$ to handle multicollinearity by shrinking coefficients

$$min_{\beta_0,\beta} 1/2n \sum_{t=1}^{n} \left(y_{i,t} - \beta_0 - x_t'\beta\right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \tag{3}$$

ElasticNet: Combines *L1* and *L2* penalties:

$$min_{\beta_0,\beta} 1/2n \sum_{t=1}^{n} \left(y_{i,t} - \beta_0 - x_t'\beta\right)^2 + \lambda\left[\alpha \sum_{j=1}^{p} |\beta_j| + (1-\alpha) \sum_{j=1}^{p} \beta_j^2\right], \tag{4}$$

where α balances between Lasso (α=1) and Ridge (α=0).

Random Forest: Aggregates multiple decision trees with bootstrap sampling:

$$\hat{y}_{i,t} = 1/B \sum_{b=1}^{B} T_b(x_t), \tag{5}$$

where *B* denotes the number of trees in the ensemble, and T_b represents the *b*-th tree trained on bootstrapped samples.

XGBoost: Implements gradient boosting with regularization:

$$\hat{y}_{i,t} = \sum_{k=1}^{K} f_k(x_t), f_k \in F, \tag{6}$$

where K is the total number of boosting rounds, and f_k denotes the k-th tree added to the ensemble. F is the space of regression trees, and trees are added sequentially to minimize the loss function with regularization terms.

We employ a rolling window design that mirrors real-world forecasting conditions. The initial training period spans 54 months (January 2015 to June 2019), followed by a 6-month validation period (July 2019 to December 2019) for hyperparameter tuning. The first out-of-sample forecast targets January 2020. Each month, the entire window shifts forward by one month while maintaining the 54-6-1 structure (54 months training, 6 months validation, 1 month forecast). This process continues through December 2024, generating 60 out-of-sample forecast points per district across all 25 districts, for a total of 1,500 forecasts. This design ensures that all predictions are genuinely out-of-sample and that models must adapt to evolving economic conditions. The validation period serves exclusively for hyperparameter tuning, preventing any information leakage from the test period into model selection.

To ensure fair comparison across models, we standardize all macroeconomic variables to have zero mean and unit variance using the training set statistics. This standardization is crucial given the vastly different scales of our variables, ranging from percentage points for interest rates to trillions of won for M2 and household credit. We apply the same standardization parameters from the training set to both validation and test sets to prevent data leakage. We deliberately avoid additional feature engineering such as interaction terms, polynomial transformations, or spatial lags. Our goal is to test whether raw macroeconomic fundamentals alone can forecast local housing price indices. Each district receives its own set of independently trained models using the same macroeconomic features, allowing for district-specific parameter estimation. This approach captures the heterogeneous responses we hypothesize exist across Seoul's diverse housing submarkets. For each district and each rolling window, we train all six models separately, resulting in 25 districts x 60 time periods x 6 models = 9,000 total model estimations.

3. Results

We present our empirical findings in three parts. First, we compare overall forecasting performance across all six models to identify which approach best predicts housing price indices using macroeconomic variables alone. Second, we examine district-level heterogeneity to assess whether our findings hold consistently across Seoul's diverse submarkets. Third, we analyze variable selection patterns and coefficient estimates from the Lasso model to identify which macroeconomic channels matter most and how their importance varies across districts and time.

Table 2 presents the comprehensive performance comparison across all six models. Lasso regression achieves the best overall performance with a mean RMSE of 1.54 and R^2 of 0.91, outperforming both complex machine learning models and simple linear models. This superiority is consistent across all evaluation metrics. Lasso achieves the second lowest MAE (1.17) and MAPE (1.25%). The dominance of regularized linear models—with Ridge ranking second—while ensemble methods like Random Forest languish at the bottom (R^2: 0.79) demonstrates that when using purely macroeconomic variables for housing price indices forecasting, complex nonlinear models are not only unnecessary but counterproductive. The poor performance of ensemble methods is striking, given their computational cost and opacity—they achieve neither superior forecasting power nor interpretability, failing on both forecasting and interpretability grounds. This finding suggests that the relationship between macroeconomic fundamentals and housing price indices is linear, and attempts to capture nonlinear interactions simply fit noise rather than signal when working with macro-level predictors.

Table 2. Average model performance

Model	RMSE	MAE	MAPE	R^2
Linear	2.17	1.22	1.28	0.81
Lasso	1.54	1.17	1.25	0.91
Ridge	1.76	1.21	1.29	0.88
ElasticNet	1.96	1.28	1.36	0.85
RandomForest	2.31	1.67	1.77	0.79
XGBoost	1.87	1.06	1.11	0.86

Note: *This table reports average performance metrics across 25 districts and rolling windows. RMSE is Root Mean Squared Error; MAE is Mean Absolute Error; MAPE is Mean Absolute Percentage Error (in %); R^2 is out-of-sample R-squared. All metrics are calculated on one-month-ahead forecasts. Lower values indicate better performance for RMSE, MAE, and MAPE; higher values indicate better performance for R^2.*

The superiority of sparse linear models extends consistently across districts, confirming that our findings are not artifacts of specific local conditions. Table 3 presents each district's best-performing model ranked by RMSE. Lasso achieves the best performance in 12 out of 25 districts, with particularly strong results in Jongno and Gwangjin. Among the remaining 12 districts, other linear models (Linear or Ridge) perform best in 9 districts, while XGBoost excels in 3 districts. Overall, linear models dominate in 22 out of 25 districts, providing strong evidence that Korean housing price indices respond linearly to macroeconomic fundamentals. Even in districts where XGBoost performs best, such as Gangnam, Gangdong, and Dobong, the performance margins over linear alternatives are typically small, while the interpretability loss is substantial. This consistency across diverse market segments indicates that housing price index dynamics respond linearly to macroeconomic variables across different submarkets. The success of regularized linear models stems from their ability to handle multicollinearity among macroeconomic variables while preventing overfitting in our limited training windows. Lasso's automatic variable selection identifies the most relevant predictors for each district, typically selecting 5-9 variables from the 16 candidates. This parsimony not only improves out-of-sample forecasting but also enhances interpretability, allowing policymakers to understand precisely which macroeconomic channels drive housing prices in each district.

Table 3. Best Performing Model by District

District	Model	RMSE	MAE	R^2
Jongno	Lasso	0.7	0.56	0.97
Gwangjin	Lasso	0.77	0.63	0.95
Geumcheon	Linear	1.04	0.72	0.93
Seocho	Lasso	1.09	0.92	0.96
Gwanak	Lasso	1.1	0.78	0.93
Mapo	Lasso	1.14	0.93	0.95
Yongsan	Lasso	1.17	0.96	0.97
Jungnang	Ridge	1.18	0.89	0.93
Jung	Linear	1.18	0.89	0.93
Gangnam	XGBoost	1.25	0.92	0.94
Yeongdeungpo	Lasso	1.29	0.98	0.93
Seodaemun	Ridge	1.38	1.01	0.85
Eunpyeong	Linear	1.39	1	0.92
Seongdong	Linear	1.39	1.13	0.92
Gangseo	Lasso	1.5	1.18	0.91
Yangcheon	Linear	1.5	1.08	0.94
Gangdong	XGBoost	1.51	0.98	0.90
Dongjak	Lasso	1.52	1.06	0.91
Songpa	Linear	1.57	1.14	0.92
Dobong	XGBoost	1.64	1.14	0.95
Seongbuk	Lasso	1.65	1.36	0.90
Guro	Ridge	1.66	1.22	0.92
Gangbuk	Lasso	1.88	1.45	0.90
Dongdaemun	Lasso	1.92	1.39	0.84
Nowon	Ridge	2.87	2.06	0.88

Note: *This table presents the 25 districts ranked by their best-performing model's RMSE, MAE, and R^2. For each district, we identify the model with the lowest average RMSE across all rolling windows and report its performance metrics.*

The key to understanding Lasso's success lies in its variable selection patterns. Table 4 reveals that from 16 macroeconomic variables, Lasso consistently selects only 5 to 9 variables per district, achieving substantial dimensionality reduction while improving forecasting power. This sparsity demonstrates that housing price indices respond to a concentrated set of macroeconomic fundamentals rather than the full spectrum of economic indicators. Three variables emerge as dominant across districts. M2 money supply appears in 81% of all estimations with an average coefficient of 5.11, confirming that monetary expansion directly translates to housing price appreciation. Household credit follows with 75% selection frequency and a coefficient of 2.81, indicating that credit availability remains a fundamental driver. CPI ranks third in selection frequency (72%), though its near-zero average coefficient (0.01) masks substantial heterogeneity across districts that we discuss below.

Table 4. Variable selection frequency and coefficients

	Selection frequency (%)	Avg Coef	Std Coef
M2	81.00	5.11	2.83
Household credit	75.47	2.81	2.82
CPI	72.40	0.01	1.99
KOSPI	44.80	0.18	0.58
Coincident idx	43.93	0.21	0.46
Economic sentiment	43.33	0.86	0.87
Deposit rate	41.07	0.15	1.72
GDP growth	40.20	0.04	0.16
Consumer sentiment	39.93	-0.76	0.65
Construction	38.13	-0.13	0.21
Unemployment	37.80	-0.04	0.24
Exchange rate	37.47	-0.21	0.32
Bond 3Y	36.93	0.58	1.75
Base rate	34.07	-0.77	1.20
Loan rate	29.00	0.76	1.16
Building permit	18.13	0.44	0.32

Note: This table reports Lasso variable selection patterns across 25 districts and rolling windows. Selection frequency (%) indicates the percentage of estimations where each variable was selected (non-zero coefficient). Avg Coef and Std Coef represent the mean and standard deviation of coefficients when selected. Variables are sorted by selection frequency.

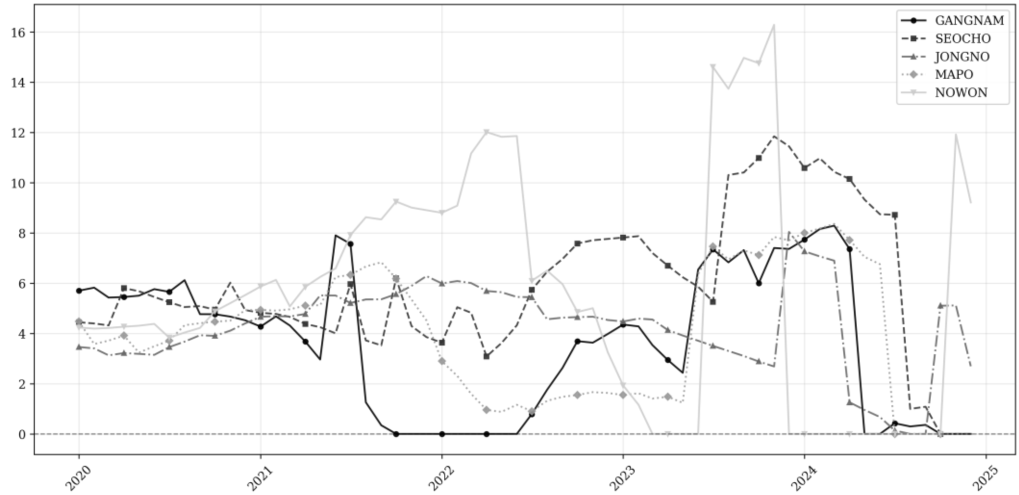
The dominance of M2 and household credit reveals the credit-driven nature of Korean housing markets. With selection frequencies of 81% and 75% respectively, these variables consistently outrank traditional price determinants like GDP growth (40%) or construction activity (38%). The substantial positive coefficients indicate strong associations between monetary conditions and housing prices. The average coefficient of 5.11 for M2 suggests that a one standard deviation increase in money supply is associated with approximately 5 percentage points higher housing price index, while household credit shows a 2.81 percentage point association.

This strong sensitivity to monetary aggregates reflects Korea's unique housing finance structure, particularly the Jeonse system and gap investment practices (Ambrose and Kim, 2003; Kim, 2013). When money supply expands, it flows disproportionately into leveraged property investment rather than consumption or productive investment. Increased household credit enables more "gap investment", which drives up prices and increases Jeonse deposits, providing capital for further leveraged purchases. This creates a reinforcing feedback loop between credit availability and housing prices. This transmission mechanism helps explain why simple linear models outperform complex alternatives. The relationship between monetary expansion and housing prices operates through direct credit channels rather than complex behavioral dynamics. Nonlinear models may be searching for threshold effects or interaction patterns that are not prominent in this credit-driven system. The practical implication is that Korean housing price indices respond strongly to liquidity conditions, making monetary policy particularly influential for housing market outcomes.

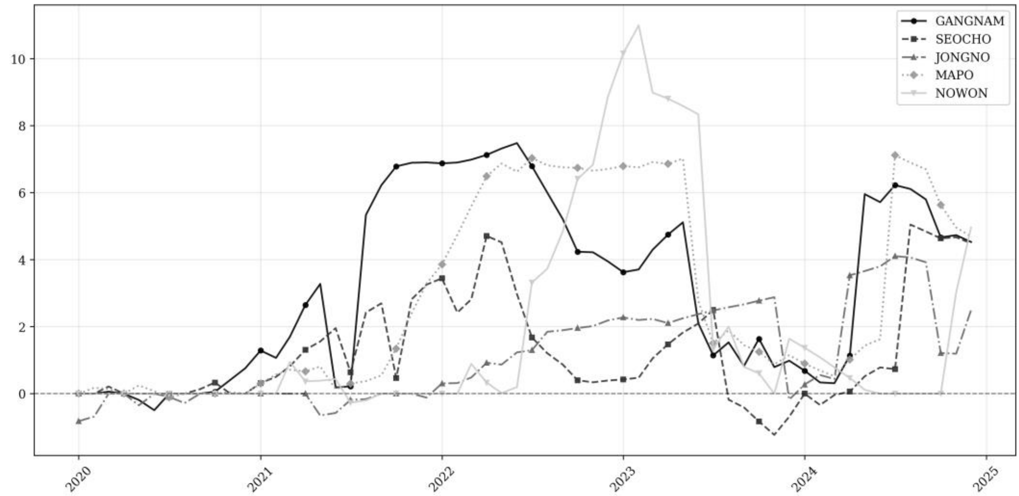
While our analysis reveals that housing price indices respond linearly to macroeconomic variables, the magnitude of these responses varies substantially across districts and over time. Figure 2 displays the temporal evolution of Lasso coefficients for three key variables across five representative districts spanning Seoul's market spectrum: Gangnam and Seocho (premium), Jongno (central/historic), Mapo (mixed-use), and Nowon (middle-income). Panel A reveals that M2 coefficients remained relatively stable between 3 and 6 across districts until 2022, suggesting relatively uniform monetary transmission during this period. However, substantial divergence emerges during 2022-2024, coinciding with the Bank of Korea's monetary tightening cycle. Nowon's coefficient exhibits high volatility, spiking above 15 in mid-2023, while Gangnam's coefficient approaches zero during the same period. This divergence suggests that monetary policy shocks affect districts asymmetrically, with middle-income areas potentially becoming more sensitive to liquidity conditions during tightening episodes. Panel B shows household credit coefficients following a different temporal pattern. Coefficients remain near zero until 2021, then increase substantially but selectively across districts. Gangnam and Seocho experience sustained elevation reaching 6-7 by 2022, while Mapo shows minimal response throughout the period. This pattern suggests that credit availability may matter differentially across market segments, though the exact mechanisms require further investigation. Panel C presents the most striking heterogeneity in CPI coefficients. During the low inflation period before 2023, coefficients across districts cluster between -1 and +2. The inflation surge of 2023-2024 dramatically alters this pattern. Nowon's coefficient declines to -10, indicating a strong negative association between inflation and real housing values, while Gangnam maintains near-zero sensitivity. This heterogeneous response to inflation shocks suggests that different districts may serve different roles in household portfolios.

Figure 2. Coefficient dynamics for M2, household credit, and CPI across 5 representative districts

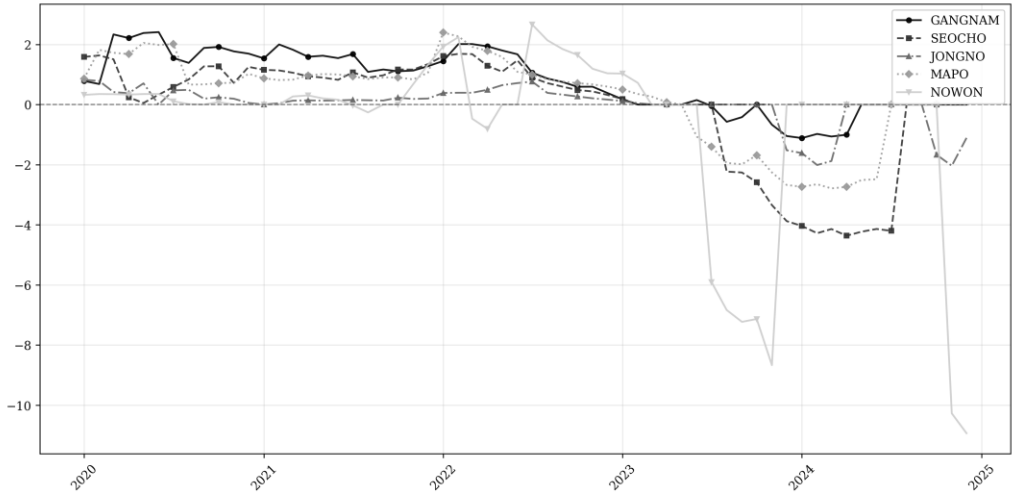
Panel A. M2



Panel B. Household credit



Panel C. CPI



Note: *This figure displays the temporal evolution of Lasso coefficients for three key macroeconomic variables across five representative districts (Gangnam, Seocho, Jongno, Mapo, Nowon). Panel A shows M2 money supply coefficients, Panel B shows household credit coefficients, and Panel C shows CPI coefficients. Each line represents a single district's coefficient path over time. The horizontal axis indicates the forecast date, while the vertical axis shows coefficient values.*

These temporal dynamics reveal two important patterns. First, the linear relationship between macroeconomic variables and housing prices holds consistently, but the coefficients governing these relationships vary substantially across districts. Second, this heterogeneity intensifies during macroeconomic stress periods, monetary tightening with high inflation, when districts diverge in their sensitivity to common shocks. The same monetary expansion or inflation shock generates vastly different responses across Seoul's submarkets, with implications for both forecasting and policy.

4. Conclusion

This paper demonstrates that macroeconomic variables alone can effectively forecast housing price indices across urban districts, achieving mean out-of-sample R^2 exceeding 90% using sparse linear methods. Our analysis of Seoul's 25 districts over 120 months with rolling window forecasts reveals three key findings. First, sparse linear models substantially outperform complex machine learning alternatives, with Lasso regression achieving an R^2 of 0.91 compared to 0.86 for XGBoost and 0.79 for Random Forest while selecting only 5-9 variables from 16 candidates per district. This result indicates that housing price indices respond to macroeconomic fundamentals through predominantly linear relationships rather than complex nonlinear interactions. Second, automatic variable selection reveals that a concentrated set of macroeconomic channels drives housing price dynamics, with M2 money supply and household credit appearing in 81% and 75% of all estimations, far exceeding traditional determinants like GDP growth (40%) or construction activity (38%). This pattern reflects Korea's unique housing finance structure, where the Jeonse system and gap investment practices create direct links between credit conditions and housing prices. Third, while the functional relationship between macroeconomic variables and housing prices is linear, coefficient magnitudes vary substantially across districts and over time, with CPI coefficients ranging from -10 in middle-income districts to near zero in premium districts during high inflation periods. This heterogeneity implies that uniform monetary policies generate spatially uneven effects, with important implications for inequality and financial stability. Our findings suggest that when forecasting housing price indices using macroeconomic fundamentals, model parsimony and interpretability need not sacrifice predictive accuracy, and that the strong response to monetary aggregates highlights both the power and spatial heterogeneity of monetary transmission in housing markets.

References

Ahn, S. and Ryu, D., 2024. Optimal chonsei to monthly rent conversion choice given borrowing constraints. *Quarterly Review of Economics and Finance*, *93*, pp.28-42. https://doi.org/10.1016/j.qref.2023.11.005.

Ambrose, B.W. and Kim, S., 2003. Modeling the Korean Chonsei lease contract. *Real estate economics*, *31*(1), 53-74. https://doi.org/10.1111/j.1080-8620.2003.00057.x.

Athey, S. and Imbens, G.W., 2019. Machine learning methods that economists should know about. *Annual Review of Economics*, *11*(1), pp.685-725. https://doi.org/10.1146/annurev-economics-080217-053433.

Bang, J., Kang, Y. and Ryu, D., 2024. Potential pricing factors in the Korean market. *Finance Research Letters*, *67*, 105946. https://doi.org/10.1016/j.frl.2024.105946.

Bang, J. and Ryu, D., 2023. CNN-based stock price forecasting by stock chart images. *Romanian Journal of Economic Forecasting*, *26*(3), pp.120-128.

Bang, J. and Ryu, D., 2024. ESG factors and the cross-section of expected stock returns: A LASSO-based approach. *Finance Research Letters*, *65*, 105482. https://doi.org/10.1016/j.frl.2024.105482.

Chae, J., Cho, H., Kim, H. and Ryu, D., 2025. Ripple effect and predictive power of housing market returns: Evidence from the Gangnam area. *Journal of Housing and the Built Environment*. https://doi.org/10.1007/s10901-025-10201-5.

Chun, D., Cho, H. and Ryu, D., 2025. Volatility forecasting and volatility-timing strategies: A machine learning approach. *Research in International Business and Finance*, *75*, 102723. https://doi.org/10.1016/j.ribaf.2024.102723.

Hong, J. and Ryu, D., 2023. Expectations and the housing market: A model of house price dynamics. *Bulletin of Economic Research*, *75*(4), pp.1242-1266. https://doi.org/10.1111/boer.12408

Igan, M.D. and Kang, M.H., 2011. *Do loan-to-value and debt-to-income limits work? Evidence from Korea*. *International Monetary Fund*. https://doi.org/10.5089/9781463927837.001.

Kang, D., Ryu, D. and Webb, A., 2024. Term spread prediction using LASSO in machine learning frameworks. *Romanian Journal of Economic Forecasting*, *27*(4), pp.31-45.

Kang, Y., Ryu, D. and Webb, A., 2025. How well do machine learning models in finance work?. *Financial Innovation*, In-press.

Kim, H., Cho, H. and Ryu, D., 2024. Characteristics of student loan credit recovery: Evidence from a micro-level data set. *Journal of Credit Risk*. https://doi.org/10.21314/JCR.2024.001.

Kim, A., Ryu, D. and Webb, A., 2025. Forecasting oil futures markets using machine learning and seasonal trend decomposition. *Investment Analysts Journal*, *54*(2), pp.205-218. https://doi.org/10.1080/10293523.2024.2405294.

Kim, J., 2013. Financial repression and housing investment: An analysis of the Korean chonsei. *Journal of Housing Economics*, *22*(4), pp.338-358. https://doi.org/10.1016/j.jhe.2013.09.001.

Kim, A., Ryu, D. and Webb, A., 2025. Forecasting oil futures markets using machine learning and seasonal trend decomposition. *Investment Analysts Journal*, *54*(2), pp.205-218. https://doi.org/10.1080/10293523.2024.2405294

Kim, T.Y., Park, E. and Ryu, D., 2025. Determinants of housing rental prices in Seoul: applying explainable AI. *Spatial Economic Analysis*, *20*(2), pp.312-332. https://doi.org/10.1080/17421772.2024.2418906.

Medeiros, M.C. and Mendes, E.F., 2016. ℓ1-regularization of high-dimensional time-series models with non-Gaussian and heteroskedastic errors. *Journal of Econometrics*, *191*(1), pp.255-271. https://doi.org/10.1016/j.jeconom.2015.10.011.

Mullainathan, S. and Spiess, J., 2017. Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), pp.87-106. https://doi.org/10.1257/jep.31.2.87.

Ryu, D., Hong, J. and Jo, H., 2024. Capturing locational effects: application of the K-means clustering algorithm. *Annals of Regional Science*, *73*(1), pp.265-289. https://doi.org/10.1007/s00168-024-01263-4.

Welch, I. and Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, *21*(4), pp.1455-1508. https://doi.org/10.1093/rfs/hhm014.