WORLD CONFERENCE ON TIMBER ENGINEERING 2025
BRISBANE, AUSTRALIA

Advancing Timber for the Future Built Environment

ASSESSING THE FACTORS BEHIND THE DEPLETION OF KOREAN TRADITIONAL WOODEN BUILDING 'HANOK' IN METROPOLIS SEOUL

Seongjun Koo[1], BongHee Jeon[2], Youngsoo Jeong[3]

ABSTRACT: This study investigates the causal factors behind the loss of hanok, traditional Korean houses, in Seoul amid rapid urban development. Seoul, a city celebrated for its blend of historical and modern landscapes, has experienced a marked decline in its hanok stock primarily due to urban development pressures. As of December 2023, only 8,983 hanok remained, further reduced to 8,849 by 2025—accounting for just 1.48% of the city's total building inventory. Despite their diminishing numbers, hanok clusters, especially in neighborhoods like Bukchon, continue to draw considerable attention and serve as emblematic symbols of the city's cultural heritage. Employing a causal inference framework, the study constructs a Directed Acyclic Graph (DAG) using the causal discovery algorithm to analyze the interplay of various urban and architectural factors affecting hanok survival. Based on the identification of policy-relevant causal paths and the estimation of their effects, it was confirmed that government financial support schemes effectively mitigate hanok loss. These insights might carry policy implications for the conservation of traditional architecture.

KEYWORDS: *Seoul, hanok, policy, machine learning, causal inference*

1 – INTRODUCTION

Since the founding of the Joseon Dynasty in the 14th century, Seoul has served as the capital of Korea and has undergone a transformation from a historical city to a modern metropolis through the process of modernization. As a result, both historical and contemporary elements coexist in the urban fabric. However, due to rapid development in the late 20th century, reinforced concrete buildings came to dominate the cityscape, while hanok, the traditional wooden houses of Korea, steadily declined in number. As of December 2023, only 8,983 hanok remained, reflecting the demolition of 2,793 structures over approximately ten years since 2014. By February 2025, an additional 134 hanok had been lost, leaving 8,849 still standing.

These 8,849 hanok account for only about 1.48% of all buildings in Seoul. Nevertheless, Bukchon, where hanok are densely clustered, has become a popular destination for both domestic and international tourists. Furthermore, in cities beyond Seoul—such as Jeonju and Gyeongju—areas with a similar concentration of hanok are increasingly emerging as nationally recognized attractions. This trend highlights the cultural and industrial potential of hanok. Although few in number, hanok constitute a significant component of Seoul's urban identity and contribute to architectural and landscape diversity. Recognizing this, the Seoul Metropolitan Government has implemented various policies, including providing subsidies and loans for the construction and repair of hanok since 2001. As of December 2023, a total of 1,202 hanok in Seoul's historic core have received such support. This type of direct financial assistance for the promotion of a specific architectural typology is exceptionally afforded to hanok alone, underscoring the city's growing commitment to their preservation. As such, hanok preservation is increasingly becoming a key agenda item.

This study aims to assess the risk of disappearance faced by hanok in Seoul, with a particular focus on the influence of various urban and architectural variables. To this end, the impact of each variable on hanok disappearance is evaluated using statistical and causal inference methodologies. Based on this approach, we aim to gain a deeper understanding of hanok loss in Seoul and to

[1] Seongjun Koo, Ph.D. Candidate, Department of Architecture and Architectural Engineering, Seoul National University, Seoul, Republic of Korea, https://orcid.org/0009-0009-0206-7958

[2] BongHee Jeon, Professor, Department of Architecture and Architectural Engineering, Seoul National University, Seoul, Republic of Korea, jeonpark@snu.ac.kr

[3] Youngsoo Jeong, M.A. Student, Department of Architecture and Architectural Engineering, Seoul National University, Seoul, Republic of Korea, https://orcid.org/0009-0009-7772-6460

https://doi.org/10.52202/080513-0231

identify more clearly which areas require intervention to prevent further disappearance.

2 - BACKGROUND

Since the 2000s, as preservation policies for hanok in Seoul have emerged as a significant public issue, several surveys have been conducted. These surveys report on various aspects such as the number, location, condition, and surrounding urban environment of hanok. However, few of them attempt to estimate the causes behind the depletion of hanok, and even when such attempts are made, they are mostly based on expert opinions derived from subjective interpretations of general spatial distributions. In other words, the causal dimensions of hanok depletion have rarely been examined with precision. This is a notable shortcoming, especially considering that these surveys were originally prompted by the rapid decline in hanok numbers.

The most notable study estimating the causes of hanok depletion was conducted in 2012. Adopting a typomorphological perspective, the study evaluated the risk of hanok disappearance in twelve selected areas densely populated with residential hanok [1]. However, the variables examined in this study were limited primarily to urban features such as roads and land plots. Additionally, the analysis relied on Pearson's r, which is applicable only when the variables follow a normal distribution—a methodological limitation. Moreover, since the study focused solely on specific areas with a high concentration of hanok, it is difficult to ascertain whether the identified patterns can be generalized to the broader urban landscape.

3 – PROJECT DESCRIPTION

The central question of this study is: Which hanok are at risk of depletion? The specific objective is to identify the causal pathways leading to hanok depletion and to estimate the causal effects along some of those paths. This represents an effort to move beyond the limitations of intuitive judgments derived from correlations, aiming instead to quantitatively uncover genuine causal relationships.

The subjects of this study are the 8,983 hanok that remained in Seoul as of December 2023. As noted earlier, 134 of these were depleted by February 2025, leaving 8,849 still in existence. The analysis includes both the depleted and surviving hanok, since understanding the causes of depletion requires examining not only the conditions of those that have disappeared but also the conditions under which others have persisted. The attributes considered in the analysis consist of both urban and architectural factors. Architectural attributes include hanok depletion, vacancy status, year of construction, building area, safety performance, conservation status, presence of building code violations, usage type, and participation in the registered hanok program. Urban attributes include designation within a redevelopment zone, street frontage, topographic shape, elevation, and the presence of adjacent hanok. Details of each attribute are discussed in Chapter 4.

The specific analytical process is grounded in the methodology of causal inference. The motivation for adopting causal inference in this study is as follows. In quantitative research, it is common to assume a particular model based on presumed functional relationships among variables, and to test this model using statistical methods. However, the assumptions underlying such models often stem from the subjective heuristics of domain experts. In contrast, this study involves the simultaneous consideration of numerous unfamiliar variables. Therefore, we place greater trust in the strengths of computational processes to guide the analysis.

Causal inference consists of two main stages: identification and estimation. The identification stage involves determining whether a causal effect can be computed from the given (i.e., observable) data and a specified causal model. A causal model refers to a structural representation of the causal relationships among variables, in which the connections and directions between causes and effects are explicitly defined and interpreted. Methods for expressing or constructing causal models include the Structural Equation Model (SEM), which represents causal relationships through systems of equations, and the Directed Acyclic Graph (DAG), which depicts these relationships graphically using vertices (representing variables) and directed edges (representing causal links). DAGs are particularly useful when dealing with complex systems that involve multiple interacting variables, especially when one wishes to explore underlying structures not directly apparent from descriptive statistics alone. Since this study aims to explore detailed and potentially high-dimensional relationships among a wide range of architectural and urban variables, DAGs are adopted as the primary framework for modeling. The DAG is derived using a causal discovery algorithm applied to the dataset, and from this data-driven graph, the causal paths to be estimated are systematically selected. This process completes the identification stage.

https://doi.org/10.52202/080513-0231

In the estimation stage, the actual causal effects identified in the previous step are estimated using the available data. Logistic regression is employed in this process to estimate the probability of hanok depletion, and since all variables have been categorized as categorical types, this method is particularly well suited for the analysis. Methodological details, model specifications, implementation procedures, and empirical results for each stage are presented and discussed in Chapter 5.

4 – EXPERIMENTAL SETUP AND THEORETICAL CONSIDERATIONS

This study aims to examine the interactions among various variables surrounding the depletion of hanok. To this end, the analysis proceeds in the following order: data preprocessing, derivation of the DAG based on the data, and the sequential execution of identification and estimation. This chapter focuses on the data preprocessing process and the derivation of the DAG from the dataset.

4.1 DATA PREPROCESSING

A causal model represents the data generating process and is, in essence, a hypothesis constructed from the data. In this study, the data consist of various architectural and urban characteristics pertaining to individual hanok. These facts serve as the foundation both for deriving the causal model and for calculating causal estimates. Depending on the context, such facts may be referred to as variables in general terms or as vertices within a DAG. The architectural and urban attributes adopted in this study are listed in Table 1.

Each attribute is described as follows. Depletion (abbr. DEMOL) refers to whether a hanok has been depleted or remains. This was determined by identifying buildings listed in the expired building register and those absent from the most recent GIS dataset updated by the Ministry of Land, Infrastructure and Transport of Korea. Vacancy (abbr. EMPTY) is based on field surveys conducted in 2023. Redevelopment zone (abbr. DEV) indicates whether the hanok is located within an officially designated redevelopment area as of 2023. Year built (abbr. YEAR) refers to the construction year of the hanok; however, for some hanok built in the early or mid-20th century, construction records could not be found, resulting in 755 missing entries (approximately 8.4% of the total). For analytical convenience, the construction year was recoded into categorical data with three intervals: built before 1969, between 1970 and 1999, and after 2000. Building area (abbr. AREA) was also encoded as categorical data, with four intervals: up to 37 m², 38–51 m², 52–73 m², and 74 m² or more. These categories were defined based on the quartiles of the data distribution. Safety performance (abbr. SF) indicates whether the primary structural components of the hanok show visible damage. Conservation status (abbr. ABC) refers to the extent to which the exposed timber structure of the hanok contributes to the streetscape, evaluated on a three-tier scale: A, B, and C. Street frontage (abbr. LOT_ROAD) classifies the relationship between the land parcel and the road into 12 types. Parcels that do not abut any road are classified as Landlocked. Parcels abutting a road are further categorized by the road's width: Narrow, Small, Medium, or Broad. Parcels located at a street corner and abutting more than one road are prefixed with Corner. Within the Narrow category, a distinction is made between roads accessible to vehicles and those that are not. Lot shape (abbr. LOT_SHP) is classified into Square, Rectangle (long-front), Rectangle (short-front), Trapezoid, Flag-shaped, and Irregular. Topography (abbr. LOT_HEIGHT) is categorized as Low, Flat, Slope, Steep, and High. Building code violation (abbr. VIOL) refers to confirmed cases of construction activities that violate building regulations, such as unauthorized extensions. Usage (abbr. USE) describes the functional use of the building and is classified as Residential, Commercial, Others, or Not identified. Adjacent hanok (abbr. ADJ) indicates whether another hanok is located within an 8-meter radius. Registered hanok (abbr. REGI) denotes whether the hanok has received financial support from the Seoul Metropolitan Government for new construction or repairs.

Table 1: Variables and Their Values

Variables in Abbr.	Type	Values
DEMOL	Architectural	O, X
EMPTY	Architectural	O, X
DEV	Urban	O, X
YEAR	Architectural	-1969, 1970–1999, 2000-
AREA	Architectural	-37m2, 38-51m2, 52-73m2, 74m2
SF	Architectural	Good, Bad
ABC	Architectural	A, B, C
LOT_ROAD	Urban	Landlocked, Narrow(Non-Traversable), Narrow(Traversable), Narrow Corner(Non-Traversable), Narrow Corner(Traversable), Small, Small Corner, Medium, Medium Corner, Broad Small Corner, Broad Narrow Corner, Broad
LOT_SHP	Urban	Square, Rectangle(long-front), Rectangle(short-front), Trapezoid, Flag-shaped, Irregular
LOT_HEIGHT	Urban	Low, Flat, Slope, Steep, High
VIOL	Architectural	O, X
USE	Architectural	Regidential, Commercial, Others, Not identified
ADJ	Urban	O, X
REGI	Architectural	O, X

4.2 ADOPTED ALGORITHM OUTLINE

In this study, the causal model is a Directed Acyclic Graph (DAG). A DAG is a model composed of vertices and the edges that connect them, and the relationships between these elements can be visually represented in the form of a diagram. As the name suggests, a DAG is directed and acyclic. The presence of direction means that each edge points in only one direction, thereby indicating causal or sequential precedence between vertices. Being acyclic means that there is no path that starts and ends at the same vertex (including self-loops), which ensures a hierarchical structure within the model and prevents the formation of infinite feedback loops.

We now examine how the given data are used to construct the causal model. To build the DAG-based causal model, we employed a causal discovery algorithm, specifically the Hill-Climb Search (HCS) method. The basic principle of HCS is as follows: the process begins with a graph containing only vertices and no edges. Then, edge additions, deletions, and direction switches are attempted randomly in order to improve the model. Model improvement is evaluated based on the Bayesian Information Criterion (BIC) score. This process is repeated iteratively, and once no further improvement can be made through additional iterations, the algorithm terminates and produces the final DAG (Fig. 1) [2].

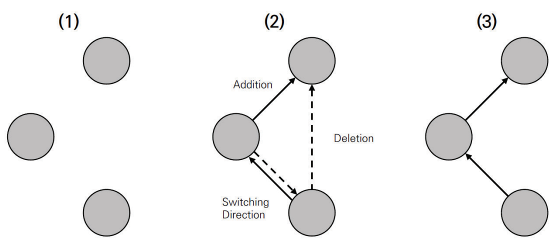

Figure 1: DAG Generating Process of HCS Algorithm

The Bayesian Information Criterion (BIC) score is a widely used evaluation metric that simultaneously considers both model fit and model complexity, offering a balance between accuracy and parsimony. Lower scores indicate better-fitting models. Model fit refers to how well the model aligns with the true underlying causal functional relationships assumed to generate the observed data; the better the fit, the more the score is reduced as a reward. Model complexity, on the other hand, typically refers to situations in which the model includes too many edges or parameters; as this complexity increases, the score is penalized accordingly to discourage overfitting.

$$\mathrm{BIC} = -2\ln(L) + k\ln(n) \tag{1}$$

In (1), L denotes the maximum likelihood of the causal model, and k refers to the number of parameters in the model, which, in the case of a DAG, corresponds to the number of edges. n represents the number of samples, that is, the total number of rows in the dataset used for learning.

Causal effect estimation is carried out using logistic regression analysis implemented through the Python library statsmodels. The statsmodels package offers a range of functionalities, including model fitting, prediction, and statistical testing [3]. In this study, all relevant variables are encoded as binary data for analysis, and the library is used to evaluate the causal effects exchanged among various variables presumed to be related to the depletion of hanok.

5 – RESULTS AND DISCUSSION

In this chapter, we examine notable causal paths derived from the DAG obtained through causal discovery, with particular attention given to the identification process for a core causal path of interest. This is followed by the estimation of causal effects along selected paths to quantitatively assess their influence.

5.1 CAUSAL PATHS

The concept of causality addressed in this study is grounded in the intuition that a cause brings about a change in some aspect of the outcome. In the DAG discussed below, the absence of an edge between two vertices does not necessarily imply that there is no functional relationship between them; rather, it simply indicates that no conditional dependence was identified given the data and the algorithm used. All interpretations of the derived DAG are made under the assumption that it can be trusted as a valid causal model. Let us now examine the structure of Fig. 2, where the vertices are connected by directed edges.

First, it is noteworthy that the only edge directed toward the primary variable of interest—hanok depletion (DEMOL)—originates from vacancy status (EMPTY). Of course, the fact that a hanok has become vacant does not imply a necessary condition for its depletion. Rather, this indicates that, during the DAG construction process, all other vertices were determined to be conditionally independent of hanok depletion (DEMOL), except for vacancy status (EMPTY). Other variables may still exert indirect effects along extended paths.

The DAG thus reveals that vacancy status (EMPTY) is the only direct path to hanok depletion (DEMOL). Furthermore, vacancy itself has only one incoming edge—

https://doi.org/10.52202/080513-0231

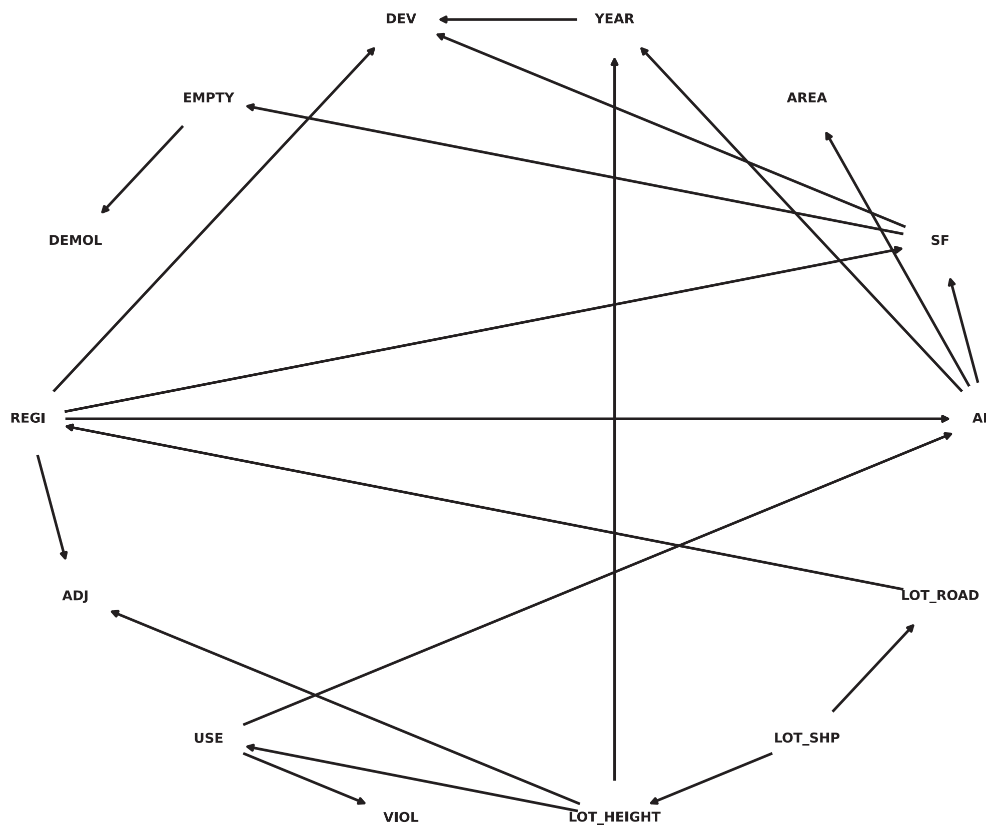

Figure 1. DAG Derived through Causal Discovery

from safety performance (SF). Therefore, the path SF → EMPTY → DEMOL can be considered the most important causal path. Notably, this causal chain is also intuitively plausible.

Next, attention should be drawn to the edge from conservation status (ABC) to safety performance (SF). The presence of this edge is justifiable from several perspectives. Statistically, hanok with better conservation status tend to exhibit better safety performance. Conceptually, the degree to which the timber structure is exposed often reflects the actual condition of the structural components; in other words, greater visibility of wooden members typically implies that those parts have not undergone significant deterioration.

The edge from usage type (USE) to conservation status (ABC) appears to be primarily driven by hanok classified as Other. These hanok tend to exhibit a high proportion of excellent conservation status, as they are typically used for special purposes such as Buddhist temples or exhibition spaces, where a strong visual identity as hanok must be maintained.

The vertex for participation in the registered hanok program (REGI) has five edges, four of which are directed toward other vertices. This suggests that registered hanok program (REGI) holds a central position in determining various other attributes. The vertices influenced by registered hanok program (REGI) include conservation status (ABC), safety performance (SF), redevelopment zone status (DEV), and presence of adjacent hanok (ADJ). While each of these connections invites interesting interpretations, the edges pointing to ABC and SF are particularly noteworthy. As mentioned earlier, safety performance (SF) is a crucial vertex on the path leading to hanok depletion (DEMOL), and it is influenced directly by only two vertices: registered hanok program (REGI) and conservation status (ABC). At the same time, registered hanok program (REGI) also influences safety performance (SF) indirectly through conservation status (ABC). Under these conditions, if one aims to estimate the direct causal effect of REGI on SF, it is necessary to control for ABC to properly adjust the estimation.

Street frontage (LOT_ROAD) is a vertex that naturally draws considerable attention. This is because, in general, parcels facing wide roads are considered to have greater development potential, whereas those facing narrow roads are seen as having more limited potential. This notion is intuitively plausible. However, the analysis shows that LOT_ROAD does not exert a direct causal influence on hanok depletion (DEMOL). This may be due to the nature of the dataset used in this study, which defines depletion based on a relatively short observation period of less than two years from December 2023. It is likely that hanok located on such high-development parcels had already been depleted in earlier years. In support of this interpretation, there exists an edge from LOT_ROAD to participation in the registered hanok program (REGI). This connection suggests that hanok on parcels facing narrower roads are more likely to be enrolled in the registration program. Such a pattern is particularly evident in areas like Bukchon, where hanok are densely concentrated.

5.2 IDENTIFICATION

As discussed above, we have examined the key causal paths. In preparation for estimation, these paths must first be properly identified, a process that requires prior knowledge. Specifically, to achieve valid identification of causal paths within a DAG, it is necessary to apply the concept of d-separation. There are three fundamental structures that underlie the causal relationships represented in a DAG: the chain, the fork, and the collider. A chain takes the form of X → Z → Y, where the causal effect is transmitted through an intermediate variable Z. In this structure, X and Y are indirectly connected via Z, and conditioning on Z blocks the causal influence of X on Y. A fork refers to a structure in which X → Y and X → Z occur simultaneously, meaning that X acts as a common cause of both Y and Z. Since both Y and Z are influenced by X, they appear to be statistically correlated, but this correlation disappears when X is conditioned on. Lastly, a collider has the structure X → Z ← Y. In this case, X and Y show no statistical association unless Z is conditioned on, in which case a spurious association between X and Y may emerge [4].

We now turn to the identification of causal paths centered on participation in the registered hanok program (REGI) and hanok depletion (DEMOL), which is particularly relevant when policy intervention is assumed. There are two causal paths leading from REGI to DEMOL. REGI has four outgoing edges directed toward redevelopment zone status (DEV), presence of adjacent hanok (ADJ), safety performance (SF), and conservation status (ABC). Among these, DEV and ADJ act as colliders, thus blocking their respective paths. As a result, information can flow only through the paths that go through SF and ABC. Therefore, the two active paths from REGI to DEMOL are: (1) REGI → SF → EMPTY → DEMOL and (2) REGI → ABC → SF → EMPTY → DEMOL. In this context, to estimate the effect of REGI on DEMOL, it is necessary to control for usage (USE), since USE explicitly influences one or more vertices along the paths from REGI to DEMOL.

5.3 ESTIMATION

Having identified the relevant causal path, the next step is to estimate the effect of REGI on DEMOL. This estimation will be carried out using logistic regression. The central question is: To what extent are hanok that have participated in the registration program less likely to be depleted compared to those that have not? The basic model is specified as follows:

$$\mathrm{logit}(P(Y=1)) = \beta_0 + \beta_1 X \quad (2)$$

At this point, the expression for P(Y=1) is given as follows.

$$P(Y=1) = \exp(\beta_0+\beta_1 X) / (1+\exp(\beta_0+\beta_1 X)) \quad (3)$$

Equation (2) expresses the log odds of Y = 1 as a linear combination of parameters. Here, β_0 represents the baseline level (i.e., the intercept), and β_1 indicates the change in the log odds resulting from a one-unit change in X, where X refers to participation in the registered hanok program (REGI) and serves as the treatment variable. Equation (3) applies the logistic function to convert this linear combination into a probability value bounded between 0 and 1.

To estimate the adjusted causal effect of participation in the registered hanok program (REGI) on hanok depletion (DEMOL), it is necessary to include a control term for building usage (USE) in the regression model. Among the four categories of USE, “Residential” is set as the reference category, and dummy variables for the remaining categories—“Commercial,” “Others,” and “Not identified”—are included in the model. This adjustment allows the potential confounding influence of the exogenous variable USE to be accounted for, enabling a more accurate evaluation of the causal effect of the registered hanok program (REGI).

As a result of estimating the adjusted causal effect of participation in the registered hanok program (REGI) on hanok depletion, the following findings were obtained. First, the coefficient of REGI in the regression model was approximately -1.3767, with a p-value of 0.007, indicating strong statistical significance. This result suggests that

https://doi.org/10.52202/080513-0231

hanok that participated in the registration program are at a lower risk of depletion compared to those that did not.

The effect of registered hanok program (REGI) can also be observed through the estimated conditional probabilities derived from the logistic regression model. When the usage type is Residential, the estimated probability of depletion is approximately 0.34% for hanok that participated in the program, compared to 1.35% for those that did not—a difference of about 1.01 percentage points. When the usage type is Commercial, the depletion probability is approximately 0.36% for participants and 1.40% for non-participants, showing a difference of about 1.04 percentage points. Results for the categories “Others” and “Not identified” are omitted due to small sample sizes. These results indicate that, regardless of usage type, participation in the registered hanok program is associated with a roughly 1 percentage point reduction in depletion probability, suggesting a tangible policy impact.

In this way, we were able to estimate the causal effect of program participation on hanok depletion. However, it is possible that unobserved confounding variables not accounted for in this study may have influenced the results. Future research should aim to construct more refined datasets and adopt more rigorous variable control techniques. Furthermore, assuming policy intervention, additional analysis of various causal paths may help reveal the mechanisms behind hanok preservation with greater precision.

6 – CONCLUSIONS

This study is grounded in a practical motivation: that understanding the actual causes of hanok depletion is essential for implementing effective policy interventions for their preservation. By approaching the depletion of hanok in Seoul from a causal perspective, the study aimed to identify the relevant contributing factors and assess the extent of their impact. To this end, a Directed Acyclic Graph (DAG) was constructed as a causal model based on a variety of architectural and urban variables, through which key causal paths influencing hanok depletion were identified. In particular, the effect of participation in the registered hanok program was examined using the identification and estimation steps of causal inference. The results show that hanok participating in the program are approximately 1 percentage point less likely to be depleted than those that are not, indicating that the program contributes meaningfully to reducing the risk of hanok loss.

This study carries several implications. First, it provides insights into the specific conditions under which hanok are situated in high-development-pressure urban areas. Moreover, although partial at this stage, the study represents a meaningful attempt to provide a quantitative foundation for systematic hanok preservation policies by analyzing a range of variables and clarifying both the processes and magnitudes of their influence on hanok depletion or preservation. Such an approach will require refinement in future research. Nonetheless, it has the potential to offer valuable insights when compared with similar cases in other cities around the world where historic wooden structures are disappearing under intense development pressure, as is the case in Seoul.

7 – ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2025-00515879).

This work was supported by the BK21 FOUR(Fostering Outstanding Universities for Research) Project in 2025. (No.4120200113771).

The Institute of Engineering Research at Seoul National University provided research facilities for this work.

8 – REFERENCES

[1] S. Baek. “A Typomorphological Interpretation of Depletion of Hanok in Hanok Residence.” PhD thesis. Seoul National University – Department of Civil & Environmental Engineering, 2012.

[2] A. Molak, *Causal Inference and Discovery in Python: Unlock the Secrets of Modern Causal Machine Learning with DoWhy, EconML, PyTorch and More*, 1st ed. Birmingham: Packt Publishing, 2023.

[3] Seabold, Skipper, and Josef Perktold. “statsmodels: Econometric and statistical modeling with python.” Proceedings of the 9th Python in Science Conference. 2010.

[4] J. Pearl, M. Glymour, and N. P. Jewell, *Causal Inference in Statistics: A Primer*, 1st ed. Chichester: John Wiley & Sons, 2016.